Effective Health Care Program

Volume 1

Registries for Evaluating Patient Outcomes: A User's Guide

Third Edition

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • **www.ahrq.gov**

The Effective Health Care Program of the Agency for Healthcare Research and Quality (AHRQ) conducts and supports research focused on the outcomes, effectiveness, comparative clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services. More information on the Effective Health Care Program can be found at www.effectivehealthcare.ahrq.gov. This report was produced under contract to AHRQ by the Outcome DEcIDE Center (Developing Evidence to Inform Decisions about Effectiveness) under Contract No. 290-2005-00351 TO7. The AHRQ Task Order Officer for this project was Elise Berliner, Ph.D.

Copyright Information:

Citation:

Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014. http://www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm.

Registries for Evaluating Patient Outcomes: A User's Guide
Third Edition

Volume 1

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2005-00351 TO7

Prepared by:

Outcome Sciences, Inc., A Quintiles Company
Cambridge, MA

Senior Editors

Richard E. Gliklich, M.D.
Nancy A. Dreyer, M.P.H., Ph.D.

Editor

Michelle B. Leavy, M.P.H.

AHRQ Publication No. 13(14)-EHC111
April 2014

Acknowledgements

The editors would like to acknowledge the efforts of the following individuals who contributed to the third edition of this document: Daniel Campion and Laura Khurana of Quintiles Outcome and Elise Berliner, Scott R. Smith, Sandy Cummings, Chris Heidenrich, and Frances Eisel of the Agency for Healthcare Research and Quality. We would especially like to thank Marion Torchia, who served as the managing editor for AHRQ for this User's Guide.

Preface

This project was performed under a contract from the Agency for Healthcare Research and Quality (AHRQ) in collaboration with the Centers for Medicare & Medicaid Services (CMS) through the Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) Network of AHRQ's Effective Health Care (EHC) Program. The purpose of the project was to update and expand *Registries for Evaluating Patient Outcomes: A User's Guide.* The *User's Guide* was first published in 2007 as a reference for establishing, maintaining, and evaluating the success of registries created to collect data about patient outcomes. The second edition, which provided updates to the existing topics and addressed four new topics, was published in 2010. The purpose of this revised and expanded third edition is to incorporate information on new methodological and technological advances into the existing chapters and to add 11 new chapters to address emerging topics in registry science.

Both the 2007 and 2010 versions and this third edition were created with support from a large group of stakeholders. Following award of the initial project on September 29, 2005, we created a draft outline for the document, which was posted for public comment on AHRQ's Effective Health Care Web site (http://www.effectivehealthcare.ahrq.gov) from January through March 2006. During that same period, we worked with AHRQ to create a process for selecting contributors and reviewers. We broadly solicited recommendations from a range of stakeholders, including government agencies, industry groups, medical professional societies, and other experts in the field; conducted a review of the pertinent literature; and contacted the initial list of contributors to confirm their interest and area of expertise and to seek further recommendations. Through that process and in collaboration with AHRQ and CMS, we arrived at a set of contributors and reviewers based on subject/content expertise, practical experience, and interest and availability, with balanced representation from key stakeholder groups for nearly all chapters. In addition, a request for submission of real-world case examples that could be used in the user's guide to illustrate issues and challenges in implementing registries was posted on the Effective Health Care Web site. The primary selection criteria for these examples concerned their utility in illustrating a practical challenge and its resolution.

An initial meeting of contributors was convened in February 2006. A second meeting including contributors and chapter reviewers was held in June 2006, following creation of an initial draft document and focused review by the reviewers. The collaborative efforts of contributors, reviewers, and editors resulted in a draft document that was posted for public comment on the Effective Health Care Web site in October and November 2006. In all, 39 contributors and 35 individual reviewers participated in the creation of the first document, which was released in April 2007 and has been published online and in print.

In August 2008, the user's guide update project was awarded. The project involved revising the existing chapters and case examples, creating new content to address four topics, and soliciting new case examples. From September to November 2008, we worked with AHRQ to select contributors and content reviewers for the new user's guide. We followed a process similar to that used in the creation of the original user's guide to arrive at a set of contributors and reviewers with subject matter expertise and a broad range of perspectives. The contributors drafted white papers on four topics: use of registries in product safety assessment, when to stop a registry, interfacing registries and electronic health records, and linking registry data. The white papers were reviewed and discussed at a meeting in April 2009. The papers were then posted for public comment in August and September 2009. After the papers were revised in response to public comments, the final papers were included in the expanded user's guide.

During the same timeframe, we contacted the authors and reviewers of the 2007 version of the user's guide. We asked authors and reviewers to update the existing chapters to address any new methodological, technological, or legal topics. The revised chapters were circulated for review and discussed at a meeting in July 2009. We also posted a new call for case examples on the Effective Health Care Web site in June 2009. The primary selection criteria for the new examples concerned their utility in illustrating issues and challenges related to the new topics addressed in the white papers. In addition, we contacted authors of the original case examples to obtain updated information on the registries. In all, 55 contributors and 49 individual reviewers participated in the creation of the second edition, which was published in September 2010.

The project to create the third edition of the user's guide was awarded in September 2010. The project involved revising the existing chapters and case examples, creating new content to address 11 topics, and soliciting new case examples. From October to December 2010, we followed a process similar to that used in the creation of the second edition to select contributors and reviewers with subject matter expertise and a broad range of perspectives. Beginning in January 2011, contributors drafted white papers on 11 new topics: registry transitions, analyzing linked datasets, patient identity management, informed consent for registries, protection of registry data, public-private partnerships, using patient-reported outcome measures in registries, rare disease registries, pregnancy registries, quality improvement registries, and medical device registries. The white papers were reviewed and discussed at a series of meetings held between July and October 2011. The papers were then posted for public comment in the spring and summer of 2012. After the papers were revised in response to public comments, the final papers were included in the expanded user's guide.

During the same timeframe, we contacted the authors and reviewers of the 2010 version of the user's guide. We asked them to update the existing chapters to address any new methodological, technological, or legal topics. The revised chapters were circulated for review and discussed at a meeting in July 2012. We also posted a new call for case examples on the Effective Health Care Web site in the spring of 2012. The primary selection criteria for the new examples concerned their utility in illustrating issues and challenges related to the new topics addressed in the white papers. In addition, we contacted authors of the original case examples to obtain updated information.

For all three editions, the contributors and reviewers participated as individuals and not necessarily as representatives of their organizations. We are grateful to all those who contributed to these documents, and who reviewed them and shared their comments.

To begin the discussion of registries, we would like to clarify some distinctions between registries and clinical trials. Although this subject is discussed further in Chapter 1, we offer here the following distinctions from a high-level perspective. A clinical trial is an experiment in which an active intervention intended to change a human subject's outcome is implemented, generally through a randomization procedure that takes decisionmaking away from the practitioner. The research protocol describes inclusion and exclusion criteria that are used to select the patients who will participate as human subjects, focusing the experiment on a homogeneous group. Human subjects and clinical researchers agree to adhere to a strict schedule of visits and to conduct protocol-specific tests and measurements.

In contrast, registries use an observational study design that does not specify treatments or require any therapies intended to change patient outcomes (except insofar as specific treatments or therapies may be inclusion criteria). IInclusion and exclusion criteria are kept to a minimum in an effort to study a broad range of patients in order to make the results more generalizable. Patients are typically observed as they present for care, and the data collected generally reflect whatever tests and measurements a provider customarily uses.

Patient registries represent a useful tool for a number of purposes. Their ideal use and their role in evidence development, design, operations, and evaluation resemble but differ from clinical trials in a number of substantive ways, and therefore they should not be evaluated with the same constructs. This user's guide presents what the contributors and reviewers consider good registry practices. Many registries today may not meet even the basic practices described. On the whole, registry science is in an active state of development. This third edition of the user's guide is an important step in developing the field.

This book is divided into two volumes and six sections. Volume 1 includes the first three sections: Creating Registries; Legal and Ethical Considerations for Registries; and Operating Registries. Volume 2 includes: Technical, Legal, and Analytic Considerations for Combining Registry Data with Other Data Sources; Special Applications in Registries; and Evaluating Registries.

The first three sections provide basic information on key areas of registry development and operations, highlighting the spectrum of practices in each of these areas and their potential strengths and weaknesses. Section I, "Creating Registries," contains six chapters. "Patient Registries" defines and characterizes types of registries, their purposes, and uses, and describes their place within the scope of this document. "Planning a Registry" focuses on the recommended steps in planning a registry, from determining if a registry is the right option to describing goals and objectives. "Registry Design" examines the specifics of designing a registry once the goals and objectives are known. "Data Elements for Registries" provides a scientific and practical approach to selecting data elements. "Use of Patient-Reported Outcomes in Registries" discusses the role that patient-reported outcome measures play in registries and addresses factors in selecting and using these types of measures. "Data Sources for Registries" describes how existing data sources (administrative, pharmacy, other registries, etc.) may be used to enhance the value of patient registries.

Section II, "Legal and Ethical Considerations for Registries," contains three chapters. "Principles of Registry Ethics, Data Ownership, and Privacy" reviews several key legal and ethical issues that should be considered in creating or operating a registry. "Informed Consent for Registries" discusses how the requirements of informed consent for patient registries differ from those for clinical trials and offers suggestions for creating informed consent documents that address the unique aspects of registries. "Protecting Data: Confidentiality and Legal Concerns of Providers, Manufacturers, and Health Plans" reviews the legal protections available for data about providers, manufacturers, and health plans contained in registries.

Section III, "Operating Registries," provides a practical guide to the day-to-day operational issues and decisions for producing and interpreting high-quality registries. "Recruiting and Retaining Participants in the Registry" describes strategies for recruiting and retaining providers and patients. "Data Collection and Quality Assurance" reviews key areas of data collection, cleaning, storing, and quality assurance for registries. "Adverse Event Detection, Processing, and Reporting" examines relevant practical and regulatory issues. "Analysis, Interpretation, and Reporting of Registry Data To Evaluate Outcomes" addresses key considerations in analyzing and interpreting registry data. "Modifying and Stopping Registries" discusses the process of modifying an existing registry as well as considerations for determining when to end a registry.

Section IV, "Technical, Legal, and Analytic Considerations for Combining Registry Data with Other Data Sources," reviews several issues related to the emerging trend of linking or integrating registry data with other data sources, such as electronic health records, administrative databases, or other registries.

"Interfacing Registries With Electronic Health Records" describes the current state of electronic health record (EHR) integration technology and maps out potential options for developing interfaces between registries and EHRs. "Linking Registry Data With Other Data Sources To Support New Studies" discusses the technical and legal issues surrounding the linkage of registry data with other data sources. "Managing Patient Identity Across Data Sources" reviews the options and strategies for linking patient information stored in multiple databases without the use of full personal identifiers. "Analysis of Linked Registry Data Sets" addresses issues that must be considered when analyzing combined or linked registry data, as well as issues related to using registry data to support secondary research studies.

Section V, "Special Applications in Patient Registries," highlights several specific types of patient registries that face unique challenges. "Use of Registries in Product Safety Assessment" describes the utility and challenges of designing a registry to assess safety. "Rare Disease Registries" discusses the increasing interest in using registries to study rare diseases and the related challenges in design, recruitment, retention, and analysis. "Pregnancy Registries" reviews the value of registries for understanding the effects of medication used during pregnancy and the challenges related to design, recruitment, analysis, and dissemination of results. "Quality Improvement Registries" examines the ability of registries to support efforts to improve quality of care through the use of specialized tools and reports. "Registries for Medical Devices" addresses the unique aspects of medical devices that must be considered in the development and analysis of a device-based registry. "Public-Private Partnerships" provides a review of public-private partnership models for supporting registries as well as a discussion of major considerations for planning and operating a registry using this type of model.

Interspersed throughout the first five sections of the user's guide are case examples. As discussed above, the choice of examples was limited to those submitted for consideration during the 2006, 2009, and 2012 public submission periods. Their purpose is solely to illustrate specific points in the text using real-world examples, regardless of whether the source of the example is within the scope of the user's guide as described in Chapter 1. Inclusion of a case example is not intended as an endorsement of the quality of the particular registry, nor do the case examples necessarily present registries that meet all the criteria described in Chapter 25 as essential elements of good practice. Rather, case examples are introduced to provide the reader with a richer description of the issue or question being addressed in the text. In some cases, we have no independent information on the registry other than what has been provided by the contributor.

Section VI is "Evaluating Registries." This final chapter on "Assessing Quality" summarizes key points from the earlier chapters in a manner that can be used to review the structure, data, or interpretations of patient registries. It describes good registry practice in terms of "essential elements" and "further indicators of quality." This information might be used by a person developing a registry, or by a reviewer or user of registry data or interpretations derived from registries.

Richard E. Gliklich
Nancy A. Dreyer
Senior Editors

Michelle B. Leavy
Editor

Contents Overview

Volume 1

Volume 2

Contents—Volume 1

Figures

Executive Summary

Defining Patient Registries

This User's Guide is intended to support the design, implementation, analysis, interpretation, and quality evaluation of registries created to increase understanding of patient outcomes. For the purposes of this guide, a patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes. A registry database is a file (or files) derived from the registry. Although registries can serve many purposes, this guide focuses on registries created for one or more of the following purposes: to describe the natural history of disease, to determine clinical effectiveness or cost-effectiveness of health care products and services, to measure or monitor safety and harm, and/or to measure quality of care.

Registries are classified according to how their populations are defined. For example, product registries include patients who have been exposed to biopharmaceutical products or medical devices. Health services registries consist of patients who have had a common procedure, clinical encounter, or hospitalization. Disease or condition registries are defined by patients having the same diagnosis, such as cystic fibrosis or heart failure.

Planning a Registry

There are several key steps in planning a patient registry, including articulating its purpose, determining whether it is an appropriate means of addressing the research question, identifying stakeholders, defining the scope and target population, assessing feasibility, and securing funding. The registry team and advisors should be selected based on their expertise and experience.

The plan for registry governance and oversight should clearly address such issues as overall direction and operations, scientific content, ethics, safety, data access, publications, and change management. It is also helpful to plan for the entire lifespan of a registry, including how and when the registry will end and any plans for transition at that time.

Registry Design

A patient registry should be designed with respect to its major purpose, with the understanding that different levels of rigor may be required for registries designed to address focused analytical questions to support decisionmaking, in contrast to registries intended primarily for descriptive purposes. The key points to consider in designing a registry include formulating a research question; choosing a study design; translating questions of clinical interest into measurable exposures and outcomes; choosing patients for study, including deciding whether a comparison group is needed; determining where data can be found; and deciding how many patients need to be studied and for how long. Once these key design issues have been settled, the registry design should be reviewed to evaluate potential sources of bias (systematic error); these should be addressed to the extent that is practical and achievable. The information value of a registry is enhanced by its ability to provide an assessment of the potential for bias and to quantify how this bias could affect the study results.

The specific research questions of interest will guide the registry's design, including the choice of exposures and outcomes to be studied and the definition of the target population (the population to which the findings are meant to apply). The registry population should be designed to approximate the characteristics of the target population as much as possible. The number of study subjects to be recruited and the length of observation (followup) should be planned in

accordance with the overall purpose of the registry. The desired study size (in terms of subjects or person-years of observation) is determined by specifying the magnitude of an expected, clinically meaningful effect or the desired precision of effect estimates. Study size determinants are also affected by practicality, cost, and whether the registry is intended to support regulatory decisionmaking. Depending on the purpose of the registry, internal, external, or historical comparison groups strengthen the understanding of whether the observed effects are indeed real and in fact different from what would have occurred under other circumstances.

Registry study designs often restrict eligibility for entry to individuals with certain characteristics (e.g., age) to ensure that the registry will have subgroups with sufficient numbers of patients for analysis. Or the registry may use some form of sampling—random selection, systematic sampling, or a haphazard, nonrandom approach—to achieve this end.

Data Elements

The selection of data elements requires balancing such factors as their importance for the integrity of the registry and for the analysis of primary outcomes, their reliability, their contribution to the overall burden for respondents, and the incremental costs associated with their collection. Selection begins with identifying relevant domains. Specific data elements are then selected with consideration for established clinical data standards, common data definitions, and whether patient identifiers will be used. It is important to determine which elements are absolutely necessary and which are desirable but not essential. In choosing measurement scales for the assessment of patient-reported outcomes, it is preferable to use scales that have been appropriately validated, when such tools exist. Once data elements have been selected, a data map should be created, and the data collection tools should be pilot tested. Testing allows assessment of respondent burden, the accuracy and completeness of questions, and potential areas of missing data. Inter-rater agreement for data collection instruments can also be assessed, especially in registries that rely on chart abstraction. Overall, the choice of data elements should be guided by parsimony, validity, and a focus on achieving the registry's purpose.

Use of Patient-Reported Outcomes in Registries

Patient-reported outcomes (PROs) are reports of health status taken directly from patients without interpretation by clinicians. PROs can provide useful information for registries designed for many purposes, including natural history of disease, quality improvement, effectiveness, and comparative effectiveness. Using PROs raises such questions as when and how often to collect the data, which method or combination of methods should be used (e.g., paper-based, electronic), and which instrument(s) should be used. Many validated instruments and measures are available, such as general assessment scales (e.g., health-related quality of life), disease-specific scales, symptom-specific scales, evaluations of functioning across a variety of domains (e.g., physical, social, emotional), and scales assessing satisfaction with care received. When selecting instruments or measures, it is important to define (1) the population of interest, (2) the outcomes of interest, (3) the intended users of the registry, and (4) the purpose of the registry. Defining these factors will help determine which PROs are useful and appropriate for the study. The instrument's validity, reliability, and ability to detect change should also be considered. Once PROs have been selected, the registry should focus on consistency, across patients and across sites, with respect to how the instruments are administered and how data are entered into the registry.

Data Sources

A single registry may integrate data from various sources. The form, structure, availability, and timeliness of the required data are important considerations. Data sources can be classified as primary or secondary. Primary data are collected by the registry for its direct purposes. Secondary data have been collected by a secondary source for purposes other than the registry, and may not be uniformly structured or validated with the same

rigor as the registry's primary data. Sufficient identifiers are necessary to guarantee an accurate match between data from secondary sources and registry patients. Furthermore, it is advisable to obtain a solid understanding of the original purpose of the secondary data, because the way those data were collected and verified or validated will help shape or limit their use in a registry. Common secondary sources of data linked to registries include medical records systems, institutional or organizational databases, administrative health insurance claims data, death and birth records, census databases, and related existing registry databases.

Ethics, Data Ownership, and Privacy

Critical ethical and legal considerations should guide the development and use of patient registries. The Common Rule is the uniform set of regulations on the ethical conduct of human subjects research, issued by the Federal agencies that fund such research. Institutions that conduct research agree to comply with the Common Rule for federally funded research, and may opt to apply that rule to all human subjects activities conducted within their facilities or by their employees and agents, regardless of the source of funding. The Privacy Rule, promulgated under the Health Insurance Portability and Accountability Act of 1996 (HIPAA), establishes Federal protections for the privacy of individually identifiable health information created and maintained by health plans, health care clearinghouses, and most health care providers (collectively, "covered entities"). The purpose of a registry, the type of entity that creates or maintains the registry, the types of entities that contribute data to the registry, and the extent to which registry data are individually identifiable affect how the regulatory requirements apply. Other important concerns include transparency of activities, oversight, and data ownership. This chapter of the User's Guide focuses solely on U.S. law. Health information is also legally protected in European and some other countries by distinctly different rules.

Informed Consent for Registries

The requirement of informed consent often raises different issues for patient registries versus clinical trials. For example, registries may be used for public health or quality improvement activities, which may not constitute "human subjects research." Also, registries may integrate data from multiple electronic sources (e.g., claims data, electronic health records) and may be linked to biobanks. Institutional review boards may approve waivers or alterations of informed consent (e.g., electronic consent, oral consent) for some registries, depending on the purpose and risk to participants. Established registries that undergo a change in scope (e.g., changes in data sharing policies, changes to the protocol, extension of the followup period) may need to ask patients to "re-consent." When planning informed consent procedures, registry developers should consider several factors, including documentation and format, consent revisions and re-consent, the applicability of regulatory requirements, withdrawal of participants from the study, and the physical and electronic security of patient data and biological specimens. In addition, registry developers may need to consider the individual authorization requirements of the HIPAA Privacy Rule, where applicable.

Confidentiality and Legal Concerns for Providers, Manufacturers, and Health Plans

As patient registries are increasingly recognized as a valuable data source, questions about privacy and the confidentiality of the data arise, particularly when data are desired for litigation or other judicial or administrative proceedings. In addition to patient data, registries often include private, confidential, and/or proprietary information about providers, manufacturers, and health plans. While significant attention has been paid to protecting the privacy of identifiable patient information, there is no single comprehensive Federal law governing

protection of registry data about providers, manufacturers, or health plans. Sources of protection for these data include the Patient Safety and Quality Improvement Act of 2005, the Health and Human Services Certificate of Confidentiality, the Agency for Healthcare Research and Quality Confidentiality Statute, the Privacy Act of 1974, the Federal Rules of Evidence and Civil Procedure, the Freedom of Information Act, Quality Improvement Organizations, the Federal Trade Secrets Act, and the Patient Protection and Affordable Care Act. Additional protections are available at the State level through safe harbor and peer review laws. Registry developers should consider this issue during the planning phase and clearly articulate the policies and procedures that the registry will follow in the case of a request for registry data (e.g., from litigation attorneys, regulatory authorities, the press, or members of the public).

Patient and Provider Recruitment and Management

Recruitment and retention of patients as registry participants, and of providers as registry sites, are essential to the success of a registry. Recruitment typically occurs at several levels, including facilities (hospitals, physicians' practices, and pharmacies), providers, and patients. The motivating factors for participation at each level and the factors necessary to achieve retention differ according to the registry. Factors that motivate participation include the perceived relevance, importance, or scientific credibility of the registry, as well as a favorable balance of any incentives for participation versus the risks and burdens thereof. Because patient and provider recruitment and retention can affect how well a registry represents the target population, well-planned strategies for enrollment and retention are critical. Goals for recruitment, retention, and followup should be explicitly laid out in the registry planning phase, and deviations during the conduct of the registry should be continuously evaluated for their risk of introducing bias.

Data Collection and Quality Assurance

The integrated system for collecting, cleaning, storing, monitoring, reviewing, and reporting on registry data determines the utility of those data for meeting the registry's goals. A broad range of data collection procedures and systems are available. Some are more suitable than others for particular purposes. Critical factors in the ultimate quality of the data include how data elements are structured and defined, how personnel are trained, and how data problems (e.g., missing, out-of range, or logically inconsistent values) are handled. Registries may also be required to conform to guidelines or to the standards of specific end users of the data (e.g., 21 Code of Federal Regulations, Part 11). Quality assurance aims to affirm that the data were, in fact, collected in accordance with established procedures and that they meet the requisite standards of quality to accomplish the registry's intended purposes and the intended use of the data.

Requirements for quality assurance should be defined during the registry's inception and creation. Because certain requirements may have significant cost implications, a risk-based approach to developing a quality assurance plan is recommended. It should be based on identifying the most important or likely sources of error or potential lapses in procedures that may affect the quality of the registry in the context of its intended purpose.

Adverse Event Detection, Processing, and Reporting

The U.S. Food and Drug Administration defines an adverse event (AE) as any untoward medical occurrence in a patient administered a pharmaceutical product, whether or not related to or considered to have a causal relationship with the treatment. AEs are categorized according to the seriousness and, for drugs, the expectedness of the event. Although AE reporting for all marketed products is dependent on the principle of "becoming aware," collection of AE data falls into

two categories: those events that are intentionally solicited (meaning data that are part of the uniform collection of information in the registry) and those that are unsolicited (meaning that the AE is volunteered or noted in an unsolicited manner). The determination of whether the registry should use a case report form to collect AEs should be based on the scientific importance of the information for evaluating the specified outcomes of interest. Regardless of whether or not AEs constitute a primary objective of the registry, it is important for any registry that has direct patient interaction to develop a plan for detecting, processing, and reporting AEs. If the registry receives sponsorship, in whole or in part, from a regulated industry (drugs or devices), the sponsor has mandated reporting requirements including stringent timelines, and the registry should establish the process for detecting and reporting AEs and should provide training to registry personnel on how to identify AEs and to whom they should be reported. Sponsors of registries designed specifically to meet requirements for surveillance of drug or device safety are encouraged to hold discussions with health authorities about the most appropriate process for reporting serious AEs.

Analysis, Interpretation, and Reporting of Registry Data

Analysis and interpretation of registry data begin with answering a series of core questions: Who was studied, and how were they chosen for study? How were the data collected, edited, and verified, and how were missing data handled? How were the analyses performed? Four populations are of interest in describing who was studied: the target population, the accessible population, the intended population, and the population actually studied (the "actual population"). The representativeness of the actual population to the target population is referred to as generalizability.

Analysis of registry outcomes first requires an analysis of recruitment and retention, of the completeness of data collection, and of data quality. Considerations include an evaluation of losses to followup; completeness for most, if not all, important covariates; and an understanding of how missing data were handled and reported. Analysis of a registry should provide information on the characteristics of the patient population, the exposures of interest, and the endpoints. Descriptive registry studies focus on describing frequency and patterns of various elements in a patient population, whereas analytical studies concentrate on associations between patients or treatment characteristics and health outcomes of interest. A statistical analysis plan describes the analytical plans and statistical techniques that will be used to evaluate the primary and secondary objectives specified in the study plan. Interpretation of registry data should be provided so that the conclusions can be understood in the appropriate context and any lessons from the registry can be applied to the target population and used to improve patient care and outcomes.

Modifying and Stopping Registries

Most, if not all, registries should undergo periodic critical evaluation by key stakeholders to ensure that the objectives are being met. When registry objectives are no longer being met or when clinical or other changes affect the registry (e.g., changes in treatment practices, the introduction of a new therapy), the registry may need to be adapted, or the registry may stop collecting new data. Many registries will undergo a modification or transition at some point in their lifecycle, and these changes will vary in scope and size. A major registry transition is a change in the registry's purpose, stakeholders, and/or technology platform that has a substantive impact on the ongoing conduct of the registry. Considerations for the transition of a registry are similar to those for starting a registry, but transitions can also present some unique challenges. It is important to select a leadership team that will carefully plan and implement the transition and consider the impacts of the planned changes (e.g., legal and ethical issues, technology, and data analysis). The transition team should also be prepared to handle unplanned or exigent circumstances that may arise during the transition and modify the project plan accordingly. Open, ongoing communication between the project team,

stakeholders, participants, and other resources is key to conducting a successful transition.

A registry may stop collecting new data because it has fulfilled its original purpose, is unable to fulfill its purpose, is no longer relevant, or is unable to maintain sufficient funding, staffing, or other support. If an open-ended registry is planned, reasonable goals should be set for data quality, study enrollment, and the amount of information needed to address specific endpoints of interest which will inform the decision if and when to end the registry.

Interfacing of Registries With Electronic Health Records

Achieving interoperability between electronic health records (EHRs) and registries will be increasingly important as adoption of EHRs and the use of patient registries for many purposes both grow significantly. Such interoperability should be based on open standards that enable any willing provider to interface with any applicable registry without requiring customization or permission from the EHR vendor. Interoperability for health information systems requires accurate and consistent data exchange and use of the information that has been exchanged. Syntactic interoperability (the ability to exchange data) and semantic interoperability (the ability to understand the exchanged data) are the core constructs of interoperability and must be present in order for EHRs and registries to share data successfully. Full interoperability is unlikely to be achieved for some time. The successive development, testing, and adoption of open standard building blocks (e.g., the Healthcare Information Technology Standards Panel's HITSP TP-50) is a pragmatic approach toward incrementally advancing interoperability while providing real benefits today. Care must be taken to ensure that integration efforts comply with legal and regulatory requirements for the protection of patient privacy and the security of individually identifiable health information.

Linking Registry Data With Other Data Sources to Support New Studies

Registry data may be linked to other data sources (e.g., administrative data sources, other registries) to examine questions that cannot be addressed using the registry data alone. Two equally weighted and important sets of questions must be addressed in the data linkage planning process: (1) What is a feasible technical approach to linking the data? (2) Is linkage legally feasible under the permissions, terms, and conditions that applied to the original compilations of each dataset? Many statistical techniques for linking records exist (e.g., deterministic matching, probabilistic matching); the choice of a technique should be guided by the types of data available. Linkage projects should include plans for managing common issues (e.g., records that exist in only one database and variations in units of measure). In addition, it is important to understand that linkage of de-identified data may result in accidental re-identification. Risks of re-identification vary depending on the variables used, and should be managed with guidance from legal and statistical experts to minimize risk and ensure compliance with the HIPAA Privacy Rule, the Common Rule, and other legal and regulatory requirements.

Managing Patient Identity Across Data Sources

As new technologies emerge to manage electronic health care data and create new opportunities for data linkage, patient identity management (PIM) strategies and standards grow increasingly important. If shared patient identifiers exist between two data sources, data can be linked using a unique patient identifier (UPI), such as a medical record number. The concept of a universal UPI has been the subject of debate for some time. Some view the UPI as a tool to reduce administrative workload and facilitate the exchange of electronic data, while others raise serious concerns about the privacy and protection of patient-identifiable information. To date, these concerns have halted efforts to implement universal UPIs in the United

States. As a result, common PIM practices in the United States include algorithms and other statistical methods to link and combine data when no shared patient identifiers are present. However, with no standardized PIM practices in place, methods can vary widely, making it difficult to ensure the accuracy and effectiveness of data linkage techniques.

Analysis of Linked Registry Data Sets

Retrospective database studies are studies that use data collected for a primary purpose other than research (e.g., administrative databases) or data collected for specific research objectives but used to support secondary studies focused on different objectives. These studies have yielded substantial information on the incidence, prevalence, and outcomes of many diseases and can be used to generate a rapid response to emerging research questions. However, these studies require special considerations related to conduct and interpretation because of the possibility of producing biased or invalid results. Challenges faced by retrospective database studies include inaccurate measurement of exposures, outcomes, and confounders and overweighting of results because of the large study population. To avoid these pitfalls, it is important to clearly define the study objective, patient population, and potential confounders and modifiers. Researchers must also understand the conditions under which the data were collected originally.

Use of Registries for Product Safety Assessment

Whether as part of a postmarketing requirement or out of a desire to supplement spontaneous reporting, prospective product and disease registries are also increasingly being considered as resources for examining unresolved safety issues and/or as tools for proactive risk assessment in the postapproval setting. Registries can be valuable tools for evaluating product safety, although they are only one of many approaches to safety assessments. When designing a registry for the purposes of safety, the size of the registry, the enrolled population, and the duration of followup are all critical characteristics to ensure validity of the inferences made based on the data collected. Consideration in the design phase must also be given to other recognized aspects of product use in the real world (e.g., the switching of therapies during followup, the use of multiple products in combination or in sequence, dose effects, delayed effects, and patient compliance).

Registries designed for safety assessment should also formulate a plan that ensures that appropriate information will reach the right stakeholders (through reporting either to the manufacturer or directly to the regulator) in a timely manner. Stakeholders include patients, clinicians, providers, product manufacturers and authorization holders, and payers such as private, State, and national insurers. Registries not designed specifically for safety assessment should, at a minimum, ensure that standard reporting mechanisms for AE information are described in the registry's standard operating procedures and are made clear to investigators.

Rare Disease Registries

A rare disease registry can be a valuable tool for increasing understanding of the disease and supporting the development of treatment protocols and therapies. Typical goals of a rare disease registry include generating knowledge around the natural history, evolution, risk, and outcomes of a specific disease; supporting research on genetic, molecular, and physiological basis of a disease; establishing a patient base for evaluating drug, medical devices, and orphan products; and facilitating connections among affected patients, families, and clinicians.

Stakeholders often play an important role in rare disease registries. Stakeholders may include patient advocacy groups, regulatory, funding, and public health agencies, clinicians, scientists, industry, payers, and individuals and families. Because of their limited patient population, rare disease registries face unique planning and design challenges. For example, little information may be available on the disease to guide development of a

research plan, and diagnostic criteria may be complex or evolving. Disease-specific patient-reported outcome measures may not be available. Long-term (even lifelong) followup may be needed. Due to these challenges, rare disease registries may need to adapt and change over time as knowledge increases or treatments become available. Retention of patients and providers can also be difficult over the duration of the registry, and registry developers should monitor followup rates over time to identify potential issues. Clear policies should be developed for governance, data access, and publications, particularly if multiple stakeholders are involved.

Pregnancy Registries

A pregnancy exposure registry is an observational prospective cohort of women receiving a biopharmaceutical product(s) of interest as part of their routine clinical care who are enrolled voluntarily during gestation, before outcomes can be known. Participants are followed until the end of pregnancy or longer to systematically collect information on specific pregnancy outcomes and evaluate their frequency relative to a scientifically valid reference population(s). While pregnancy registries are an efficient method for evaluating the effects of medications used during pregnancy, they present unique challenges related to patient recruitment and retention, the choice of reference or comparator groups, ways of mitigating bias, and generalizability of registry results. Analysis and interpretation of data from pregnancy registries also requires careful consideration. Because specific birth defects are rare events, pregnancy registries usually do not have sufficient sample size/power to evaluate increased risks for specific defects unless the relative risks are quite large. Most registries compare the overall proportion of all major defects combined in the exposed group to the overall proportion in the reference group.

Quality Improvement Registries

Quality improvement (QI) registries use systematic data collection and other QI tools to improve the quality of care on the local, regional, or national level. In a QI registry, patients are either exposed to a particular health service (e.g., a procedure registry), or they have a disease/condition that is tracked over time through multiple health care providers and services. Most of the steps for planning a QI registry are similar to the steps used for other types of registries, with two major differences. First, the identification of active, engaged participants, often called “champions,” is critical for the early success of the registry. Second, the registry must collect actionable information that can be used to modify behaviors, processes, or systems of care. Actionable information is typically presented to providers in the form of process of care or quality measures. The selection of these measures requires balancing the goals of the registry with the desire to meet other needs for providers. In the design phase, QI registries can use the process of care or quality measures to drive the selection of data elements. Because many data elements collected in QI registries are often collected for other purposes (e.g., claims, medical records), integration with other data sources may be important for encouraging participation. Motivations for participation often differ from other types of registries, and incentives for participation focus on QI (e.g., recognition programs, QI tools, and benchmarking reports). Reporting information is also an important component of QI registries. Registries may report blinded or unblinded data at the individual patient, provider, or institution level. Lastly, QI registries must be able to adapt to new evidence and improvements in care over time, and they may face questions from institutional review boards less familiar with these types of registries.

Registries for Medical Devices

Medical device registries are an increasingly important tool for capturing patients' experience with medical devices throughout the device lifecycle. Registries help bridge the gap between device performance in clinical trial settings and in routine practice. However, the unique features of medical devices require special consideration when developing a registry. Regulations and approval guidelines for medical devices differ greatly from those for drugs. Compared with drugs, device technologies tend to see more rapid change over shorter time periods, and device registries must adapt to these changes. The current lack of unique device identifiers is also challenging— although efforts are underway to create them. In many cases, multiple devices are used, and devices may be used in combination with a drug component, further complicating efforts to examine safety and effectiveness. In addition, providers may have different levels of experience with the device, which may affect patient outcomes (especially with implantable devices). Medical device registries should attempt to classify all parts of a device with as much identifying information as possible. Many registries collect information on provider training and experience as well. An emerging trend is the ability for medical devices to transmit data directly to an electronic health record or registry. This new technology may reduce the burden of data entry for registries and improve the timeliness of registry data.

Public-Private Partnerships

A public-private partnership (PPP) refers to any partnership in which one entity is a public agency (e.g., a government entity) and the other entity is a private organization. PPPs are increasingly used as a means to develop patient registries, in part because of a growing interest from governments and payers in using registry data to make decisions about approval, coverage, and public health needs. Many models for PPPs exist. For example, a PPP may involve a partnership with Federal agencies and academia, health agencies from several countries and industry, or professional associations and public payers. During the planning phase of a PPP, it is important to define clear, transparent plans for governance, with documented roles for each stakeholder. Formal policies for analyses, publications, and data sharing are also critical, as are plans for managing conflicts of interest. During the operational phase, PPPs should focus on consistent communication with stakeholders to maintain their interest. PPP registries are more likely to succeed if they have clear, agreed-upon goals; explicit roles and responsibilities for each stakeholder; strong leaders who are respected in the field; consistent data collection and analysis plans; and the flexibility to adapt to changing conditions.

Evaluating Registries

Although registries can provide useful information, there are levels of rigor that enhance validity and make the information from some registries more useful for guiding decisions. The term "quality" can be applied to registries to describe the confidence that the design, conduct, and analysis of the registry can be shown to protect against bias and errors in inference—that is, erroneous conclusions drawn from the registry. Although there are limitations to any assessment of quality, a quality component analysis is used both to evaluate high-level factors that may affect results and to differentiate between research quality (which pertains to the scientific process) and evidence quality (which pertains to the data/findings emanating from the research process). Quality components are classified as either "basic elements of good practice," which can be viewed as a checklist that should be considered for all patient registries, or as "potential enhancements to good practice," which may strengthen the value of the information in particular circumstances. The results of such an evaluation should be considered in the context of the disease area(s), the type of registry, and the purpose of the registry, and should also take into account feasibility and affordability.

Section I

Creating Registries

Chapter 1. Patient Registries

1. Introduction

The purpose of this document is to serve as a guide for the design and use of patient registries for scientific, clinical, and health policy purposes. Properly designed and executed, patient registries can provide a real-world view of clinical practice, patient outcomes, safety, and comparative effectiveness. This user's guide primarily focuses on practical design and operational issues, evaluation principles, and best practices. Where topics are well covered in other materials, references and/or links are provided. The goal of this document is to provide stakeholders in both the public and private sectors with information they can use to guide the design and implementation of patient registries, the analysis and interpretation of data from patient registries, and the evaluation of the quality of a registry or one of its components. Where useful, case examples have been incorporated to illustrate particular points or challenges.

The term *registry*[1] is defined both as the act of recording or registering and as the record or entry itself. Therefore, "registries" can refer to both programs that collect and store data and the records that are so created.

The term *patient registry* is generally used to distinguish registries focused on health information from other record sets, but there is no consistent definition in current use. E.M. Brooke, in a 1974 publication of the World Health Organization, further delineated registries in health information systems as "a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a predetermined purpose." [2]

The National Committee on Vital and Health Statistics[3] describes registries used for a broad range of purposes in public health and medicine as "an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have either a particular disease, a condition (e.g., a risk factor) that predisposes [them] to the occurrence of a health-related event, or prior exposure to substances (or circumstances) known or suspected to cause adverse health effects."

Other terms also used to refer to patient registries include clinical registries, clinical data registries, disease registries, and outcomes registries.[4, 5]

This user's guide focuses on patient registries that are used for evaluating patient outcomes. It is not intended to address several other types of or uses for registries (although many of the principles may be applicable), such as geographically based population registries (not based on a disease, condition, or exposure); registries created for public health reporting without tracking outcomes (e.g., vaccine registries); or listing registries that are used solely to identify patients with particular diseases in clinical practices but are not used for evaluating outcomes. This user's guide is also not intended to address the wide range of studies that use secondary analyses of data collected for other purposes.

Many of these other types of registries are included in the Registry of Patient Registries (RoPR) effort. RoPR is a central listing of patient registries established in 2012 by the Agency for Healthcare Research and Quality (AHRQ) in collaboration with the National Library of Medicine.[6] It is designed to improve transparency and reduce redundancy in registry-based research. Inclusion of all types of patient registries is important to achieve RoPR's goals, and the system therefore defines patient registries broadly.

In contrast to RoPR, this user's guide focuses on the subset of patient registries used for evaluating patient outcomes, defined as follows:

- A patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes.

- The patient registry database describes a file (or files) derived from the registry.

Based on these definitions, the user's guide focuses on patient registries in which the following are true (although exceptions may apply):

- The data are collected in a naturalistic manner, such that the management of patients is determined by the caregiver and patient together and not by the registry protocol.
- The registry is designed to fulfill specific purposes, and these purposes are defined before collecting and analyzing the data. In other words, the data collection is purpose driven rather than the purpose being data driven (meaning limited to or derived from what is already available in an existing data set).
- The registry captures data elements with specific and consistent data definitions.
- The data are collected in a uniform manner for every patient. This consideration refers to both the types of data and the frequency of their collection.
- The data collected include data derived from and reflective of the clinical status of the patient (e.g., history, examination, laboratory test, or patient-reported data). Registries include the types of data that clinicians would use for the diagnosis and management of patients.
- At least one element of registry data collection is active, meaning that some data are collected specifically for the purpose of the registry (usually collected from the patient or clinician) rather than inferred from sources that are collected for another purpose (administrative, billing, pharmacy databases, etc.). This definition does not exclude situations where registry data collection is a specific, but not the exclusive, reason data are being collected, such as might be envisioned with future uses of electronic health records, as described in Chapter 15. This definition also does not exclude the incorporation of other data sources. Registries can be enriched by linkage with extant databases (e.g., to determine deaths and other outcomes or to assess pharmacy use or resource utilization), as discussed in Chapter 6.

Data from patient registries are generally used for studies that address the purpose for which the registry was created. In some respects, such as the collection of detailed clinical and longitudinal followup data, studies derived from the patient registries described in this user's guide resemble traditional observational cohort studies. Beyond traditional cohort studies, however, some registry-based studies may be more flexible in that the scope and focus of the data collection activity of the registry may be adapted over time to address additional needs. For example, new studies, such as cluster-randomized studies or case-control studies, may be nested within an ongoing registry, and the database derived from the registry may be used to support secondary studies, such as studies that link the registry database with other data sources to explore new questions.

2. Current Uses for Patient Registries

A patient registry can be a powerful tool to observe the course of disease; to understand variations in treatment and outcomes; to examine factors that influence prognosis and quality of life; to describe care patterns, including appropriateness of care and disparities in the delivery of care; to assess effectiveness; to monitor safety and harm; and to measure quality of care. Through functionalities such as feedback of data, registries are also being used to study quality improvement.[7]

Different stakeholders perceive and may benefit from the value of registries in different ways. For example, for a clinician, registries can collect data about disease presentation and outcomes on large numbers of patients rapidly, thereby producing a real-world picture of disease, current treatment practices, and outcomes. For a physician organization, a registry might provide data that can be used to assess the degree to which clinicians are managing a disease in accordance with evidence-based guidelines, to focus attention on specific aspects of a particular disease that might otherwise be overlooked, or to provide data for clinicians to

compare themselves with their peers.[8] For patients and patient advocacy organizations, a registry may increase understanding of the natural history of a disease, contribute to the development of treatment guidelines, or facilitate research on treatment.[9, 10] From a payer's perspective, registries can provide detailed information from large numbers of patients on how procedures, devices, or pharmaceuticals are actually used and on their effectiveness in different populations. This information may be useful for determining coverage policies.[11] For a drug or device manufacturer, a registry-based study might demonstrate the performance of a product in the real world, meet a postmarketing commitment or requirement,[12] develop hypotheses, or identify patient populations that will be useful for product development, clinical trials design, and patient recruitment. The U.S. Food and Drug Administration (FDA) has noted that "through the creation of registries, a sponsor can evaluate safety signals identified from spontaneous case reports, literature reports, or other sources, and evaluate the factors that affect the risk of adverse outcomes such as dose, timing of exposure, or patient characteristics."[13]

The use of patient registries varies by priority condition, with cancer and cardiovascular disease having a large number of registries and areas such as developmental delays or dementia, far fewer. Overall, the use of patient registries appears to be active and growing. For example, a review of ClinicalTrials.gov in the area of cancer reveals over 270 large (more than 2,000 patients) observational studies that would meet the criteria for a patient registry. Of these studies, 4 have more than 100,000 patients, and 27 have more than 10,000. In some cases, the drivers for these registries have been Federal stakeholders. For example, since 2005, the FDA Center for Devices and Radiological Health has called for some 160 postapproval studies, many of which use new or existing registries to study the real-world effectiveness of specific devices in community practice.[14] The establishment of RoPR provides a new resource for tracking registry development and use by condition, purpose, type, and multiple other factors.[6]

2.1 Evaluating Patient Outcomes

Studies from patient registries and randomized controlled trials (RCTs) have important and complementary roles in evaluating patient outcomes.[15] Ideally, patient registries collect data in a comprehensive manner (with few excluded patients) and therefore produce outcome results that may be generalizable to a wide range of patients. They also evaluate care as it is actually provided, because care is not assigned, determined, or even recommended by a protocol. As a result, the outcomes reported may be more representative of what is achieved in real-world practice. Patient registries also offer the ability to evaluate patient outcomes when clinical trials are not practical (e.g., very rare diseases), and they may be the only option when clinical trials are not ethically acceptable. They are a powerful tool when RCTs are difficult to conduct, such as in surgery or when very long-term outcomes are desired.

RCTs are controlled experiments designed to test hypotheses that can ultimately be applied to real-world care. Because RCTs are often conducted under strict constraints, with detailed inclusion and exclusion criteria (and the need for subjects who are willing to be randomized), they are sometimes limited in their generalizability. If RCTs are not generalizable to the populations to which the information will be applied, they may not be sufficiently informative for decisionmaking. Conversely, patient registries that observe real-world clinical practice may collect all of the information needed to assess patient outcomes in a generalizable way, but interpreting this information correctly requires analytic methodology geared to address the potential sources of bias that challenge observational studies. Interpreting patient registry data also requires checks of internal validity and sometimes the use of external data sources to validate key assumptions (such as comparing the key characteristics of registry participants with external sources in order to demonstrate the comparability of registry participants with the ultimate reference population). Patient registries, RCTs, other study designs, and other data sources should all be considered tools in the toolbox for evidence development, each with its own advantages and limitations.[16]

2.2 Hierarchies of Evidence

One question that arises in a discussion of this type is where to place studies derived from patient registries within the hierarchies of evidence that are frequently used in developing guidelines or decisionmaking. While the definition of patient registry used in this user's guide is intentionally broad, the parameters of quality described in Chapter 25 are intended to help the user evaluate and identify registries that are sufficiently rigorous observational studies for use as evidence in decisionmaking. Many registries are, or include, high-quality studies of cohorts designed to address a specific problem and hypothesis. Still, even the most rigorously conducted registries, like prospective observational studies, are traditionally placed in a subordinate position to RCTs in some commonly used hierarchies, although equal to RCTs in others.[17-19] Debate continues in the evidence community regarding these traditional methods of grading levels of evidence, their underlying assumptions, their shortcomings in assessing certain types of evidence (e.g., benefit vs. harm), and their interscale consistency in evaluating the same evidence.[16, 20, 21]

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group has proposed a more robust approach that addresses some of the decisionmaking issues described in this user's guide. As noted by the GRADE collaborators:

> [R]andomised trials are not always feasible and, in some instances, observational studies may provide better evidence, as is generally the case for rare adverse effects. Moreover, the results of randomised trials may not always be applicable—for example, if the participants are highly selected and motivated relative to the population of interest. It is therefore essential to consider study quality, the consistency of results across studies, and the directness of the evidence, as well as the appropriateness of the study design.[22]

AHRQ has also developed a guidance system for grading the strength of evidence that recommends a careful assessment of the potential value of observational studies. The guidance, which is designed to support the systematic reviews conducted by the Evidence-based Practice Center (EPC) program, is conceptually similar to the GRADE system.[23] When using the AHRQ approach, reviewers typically give evidence from observational studies a low starting grade and evidence from RCTs a high starting grade. These initial grades can then be raised or lowered depending on the strength of the five required evidence domains (study limitations, directness, consistency, precision, and reporting bias).[24] For example, the reviewers may find that observational studies are particularly relevant for some systematic review questions. The report notes:

> EPCs may act on the judgment that, for certain outcomes such as harms, observational studies have less risk of bias than do RCTs or that the available RCTs have a substantial risk of bias. In such instances, the EPC may move up the initial grade for strength of evidence based on observational studies to moderate or move down the initial rating based on RCTs to moderate.[23]

Reviewers may also raise or lower evidence grades based on a secondary set of domains (dose-response association, existence of confounding that would diminish an observed effect, and strength of association). These secondary domains supplement the required domains and are used when relevant to the systematic review question. The report explains that the secondary domains "may increase strength of evidence and are especially relevant for observational studies where one may begin with a lower overall strength of evidence grade based on study limitations."[23]

As the methods for grading evidence for different purposes continue to evolve, this user's guide can serve as a guide to help such evaluators understand study quality and identify well-designed registries. Beyond the evidence hierarchy debate, users of evidence understand the value of registries for providing complementary information that can extend the results of clinical trials to populations not studied in those trials, for demonstrating the real-world effects of treatments outside of the research setting and potentially in large subsets of affected patients, and for providing long-term

followup when such data are not available from clinical trials.

2.3 Defining Patient Outcomes

The focus of this user's guide is the use of registries to evaluate patient outcomes. An outcome may be thought of as an end result of a particular health care practice or intervention. According to AHRQ, end results include effects that people experience and care about.[25] The National Cancer Institute further clarifies that "final" endpoints are those that matter to decisionmakers: patients, providers, private payers, government agencies, accrediting organizations, or society.[26, 27] Examples of these outcomes include biomedical outcomes, such as survival and disease-free survival, health-related quality of life, satisfaction with care, and economic burden.[28] Although final endpoints are ultimately what matter, it is sometimes more practical when creating registries to collect intermediate outcomes (such as whether processes or guidelines were followed) and clinical outcomes (such as whether a tumor regressed or recurred) that predict success in improving final endpoints.

In *Crossing the Quality Chasm*,[29] the Institute of Medicine (IOM) describes the six guiding aims of health care as providing care that is safe, effective, efficient, patient-centered, timely, and equitable. (The last three aims focus on the delivery and quality of care.) While these aims are not outcomes per se, they generally describe the dimensions of results that matter to decisionmakers in the use of a health care product or service: Is it safe? Does it produce greater benefit than harm? Is it clinically effective? Does it produce the desired effect in real-world practice? Does the right patient receive the right therapy or service at the right time? Is it cost effective or efficient? Does it produce the desired effect at a reasonable cost relative to other potential expenditures? Is it patient oriented, timely, and equitable? Most of the patient outcomes that registries evaluate reflect one or more of the IOM guiding aims. For example, a patient presenting with an ischemic stroke to an emergency room has a finite window of opportunity to receive a thrombolytic drug, and the patient outcome, whether or not the patient achieves full recovery, is dependent not only on the product's dissolving the clot but also on the timeliness of its delivery.[30, 31]

2.4 Purposes of Registries

As discussed throughout this user's guide, registries should be designed and evaluated with respect to their intended purpose(s). Registry purposes can be broadly described in terms of patient outcomes. While there are a number of potential purposes for registries, this handbook primarily discusses four major purposes: (1) describing the natural history of disease, (2) determining clinical and/or cost-effectiveness, (3) assessing safety or harm, and (4) measuring or improving quality of care. Other purposes of patient registries mentioned but not discussed in detail in this user's guide are for public health surveillance and disease control. An extensive body of literature from the last half century of experience with cancer and other disease surveillance registries is available.

2.4.1 Describing Natural History of Disease

Registries may be established to evaluate the natural history of a disease, meaning its characteristics, management, and outcomes with and/or without treatment. The natural history may be variable across different groups or geographic regions, and it often changes over time. In many cases, the natural histories of diseases are not well described. Furthermore, the natural histories of diseases may change after the introduction of certain therapies. As an example, patients with rare diseases, such as the lysosomal storage diseases, who did not previously survive to their 20s, may now be entering their fourth and fifth decades of life, and this uncharted natural history is being first described through a registry.[32] The role of registries in rare diseases is explored in Chapter 20.

2.4.2 Determining Effectiveness

Registries may be developed to determine clinical effectiveness or cost-effectiveness in real-world clinical practice. Multiple studies have demonstrated disparities between the results of clinical trials and results in actual clinical practice.[33, 34] Furthermore, efficacy in a clinical

trial for a well-defined population may not be generalizable to other populations or subgroups of interest. As an example, many important heart failure trials have focused on a predominantly white male population with a mean age of approximately 60 years, whereas actual heart failure patients are older, more diverse, and have a higher mortality rate than the patients in these trials.[35] Similarly, underrepresentation of older patients has been reported in clinical trials of 15 different types of cancer (e.g., studies with only 25 percent of patients age 65 years and over, while the expected rate is greater than 60 percent).[36] Data from registries have been used to fill these gaps for decisionmakers. For example, the FDA used the American Academy of Ophthalmology's intraocular lens registry to expand the label for intraocular lenses to younger patients.[37] Registries may also be particularly useful for tracking effectiveness outcomes for a longer period than is typically feasible with clinical trials. For example, some growth hormone registries have tracked children well into adulthood.

In addition to clinical effectiveness, registries can be used to assess cost-effectiveness. Registries can be designed to collect cost data and effectiveness data for use in modeling cost-effectiveness.[38] Cost-effectiveness is a means to describe the comparative value of a health care product or service in terms of its ability to achieve a desired outcome for a given unit of resources.[39] A cost-effectiveness analysis examines the incremental benefit of a particular intervention and the costs associated with achieving that benefit. Cost-effectiveness studies compare costs with clinical outcomes measured in units such as life expectancy or disease-free periods. Cost-utility studies compare costs with outcomes adjusted for quality of life (utility), such as quality-adjusted life years (QALYs). Utilities allow comparisons to be made across conditions because the measurement is not disease specific.[40] It should be noted that for both clinical effectiveness and cost-effectiveness, differences between treatments are indirect and must be inferred from data analysis, simulation modeling, or some mixture.

With improvement in methodologies for using observational research for comparative effectiveness research (CER), including better methods for managing bias and better understanding of the limitations,[41] there is both increasing interest and investment in registries for CER across a number of stakeholders. Reports from the IOM and the Congressional Budget Office in 2007 cited the importance of patient registries in developing comparative effectiveness evidence.[42, 43] The Federal Coordinating Council for Comparative Effectiveness Research, in its Report to the President and the Congress (June 30, 2009), defined CER as "the conduct and synthesis of research comparing benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in 'real world' settings."[44] The report specifically identifies patient registries as a core component of CER data infrastructure.

More recently, the newly formed Patient-Centered Outcomes Research Institute (PCORI) has identified registries as an important potential source of data to support patient-centered outcomes research (PCOR). PCOR "assesses the benefits and harms of preventive, diagnostic, therapeutic, palliative, or health delivery system interventions to inform decisionmaking, highlighting comparisons and outcomes that matter to people; is inclusive of an individual's preferences, autonomy and needs, focusing on outcomes that people notice and care about such as survival, function, symptoms, and health related quality of life; incorporates a wide variety of settings and diversity of participants to address individual differences and barriers to implementation and dissemination; and investigates (or may investigate) optimizing outcomes while addressing burden to individuals, availability of services, technology, and personnel, and other stakeholder perspectives."[45]

Similar to their function in CER, registries are expected play an important role in this new area of research in part because of their ability to provide information on 'real-world' settings and broad patient populations. PCORI included minimum standards for the use of registries for PCOR in the Methodology Report.[46] While some registries are

designed explicitly to examine questions of comparative effectiveness or patient-centered outcomes research, many others are designed for different objectives yet still collect data that are useful for these analyses. Registries that were not explicitly designed for CER or PCOR may need to be augmented or linked to other data sources—for example, to obtain long-term outcomes data in the case of an in-hospital registry using linkage to claims data to evaluate blood pressure medications.[47]

2.4.3 Measuring or Monitoring Safety and Harm

Registries may be created to assess safety versus harm. Safety here refers to the concept of being free from danger or hazard. One goal of registries in this context may be to quantify risk or to attribute it properly. Broadly speaking, patient registries can serve as an active surveillance system for the occurrence of unexpected or harmful events for products and services. Such events may range from patient complaints about minor side effects to severe adverse events such as fatal drug reactions or patient falls in the hospital.

Patient registries offer multiple advantages for active surveillance. First, the current practice of spontaneous reporting of adverse events relies on a nonsystematic recognition of an adverse event by a clinician and the clinician's active effort to make a report to manufacturers and health authorities. Second, these events are generally reported without a denominator (i.e., the exposed or treated population), and therefore an incidence rate is difficult to determine. Because patient registries can provide systematic data on adverse events and the incidence of these events, they are being used with increasing frequency in the areas of health care products and services. The role of registries in monitoring product safety is discussed in more detail in Chapter 19.

2.4.4 Measuring Quality

Registries may be created to measure quality of care. The IOM defines quality as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge." Quality-focused registries are being used increasingly to assess differences between providers or patient populations based on performance measures that compare treatments provided or outcomes achieved with "gold standards" (e.g., evidence-based guidelines) or comparative benchmarks for specific health outcomes (e.g., risk-adjusted survival or infection rates). Such programs may be used to identify disparities in access to care, demonstrate opportunities for improvement, establish differentials for payment by third parties, or provide transparency through public reporting. There are multiple examples of such differences in treatment and outcomes of patients in a range of disease areas.[48-53] Quality improvement registries are described further in Chapter 22.

2.4.5 Multiple Purposes

Many registries will be developed to serve more than one of these purposes. Registries developed for one purpose may also be modified to serve additional purposes as the research, practice, or policy environment changes. While registries often serve more than one purpose, their original or primary purpose generally guides their design and, as a result, more care is needed in evaluating results for secondary or additional purposes.

3. Taxonomy for Patient Registries

Even limited to the definitions described above, the breadth of studies that might be included as patient registries is large. Patients in a registry are typically selected based on a particular disease, condition (e.g., a risk factor), or exposure. This user's guide uses these common selection criteria to develop a taxonomy or classification based on how the populations for registries are defined. Three general categories with multiple subcategories and combinations account for the majority of registries that are developed for evaluating patient outcomes. These categories include observational studies in which the patient has had an exposure to a product or service, or has a particular disease or condition, or various combinations thereof.

3.1 Product Registries

In the case of a product registry, the patient is exposed to a health care product, such as a drug or a device. The exposure may be brief, as in a single dose of a pharmaceutical product, or extended, as in an implanted device or chronic usage of a medication.

Device registries may include all, or a subset, of patients who receive the device. A registry for all patients who receive an implantable cardioverter defibrillator, a registry of patients with hip prostheses, or a registry of patients who wear contact lenses are all examples of device registries. Biopharmaceutical product registries similarly have several archetypes, which may include all, or subsets, of patients who receive the biopharmaceutical product. For example, the British Society for Rheumatology established a national registry of patients on biologic therapy.[54] Again, the duration of exposure may range from a single event to a lifetime of use. Eligibility for the registry includes the requirement that the patient received the product or class of products (e.g., COX-2 inhibitors). In some cases, public health authorities mandate such registries to ensure safe use of medications. Examples include registries for thalidomide, clozapine, and isotretinoin.

Pregnancy registries represent a separate class of biopharmaceutical product registries that focus on possible exposures during pregnancy and the neonatal consequences. The FDA has a specific guidance focused on pregnancy exposure registries.[55] This guidance uses the term "pregnancy exposure registry" to refer to "a prospective observational study that actively collects information on medical product exposure during pregnancy and associated pregnancy outcomes." Pregnancy registries are discussed in more detail in Chapter 21.

3.2 Health Services Registries

In the context of evaluating patient outcomes, another type of exposure that can be used to define registries is exposure to a health care service. Health care services that may be used to define inclusion in a registry include individual clinical encounters, such as office visits or hospitalizations, procedures, or full episodes of care. Examples include registries enrolling patients undergoing a procedure (e.g., carotid endarterectomy, appendectomy, or primary coronary intervention) or admitted to a hospital for a particular diagnosis (e.g., community-acquired pneumonia). In these registries, one purpose of the registry is to evaluate the health care service with respect to the outcomes. Health care service registries are sometimes used to evaluate the processes and outcomes of care for quality measurement purposes (e.g., Get With The Guidelines® of the American Heart Association, National Surgical Quality Improvement Program of the Department of Veterans Affairs and the American College of Surgeons).

3.3 Disease or Condition Registries

Disease or condition registries use the state of a particular disease or condition as the inclusion criterion. In disease or condition registries, the patient may always have the disease (e.g., a rare disease such as cystic fibrosis or Pompe disease, or a chronic illness such as heart failure, diabetes, or end-stage renal disease) or may have the disease or condition for a more limited period of time (e.g., infectious diseases, some cancers, obesity). These registries typically enroll the patient at the time of a routine health care service, although patients also can be enrolled through voluntary self-identification processes that do not depend on utilization of health care services (such as Internet recruiting of volunteers). In other disease registries, the patient has an underlying disease or condition, such as atherosclerotic disease, but is enrolled only at the time of an acute event or exacerbation, such as hospitalization for a myocardial infarction or ischemic stroke.

3.4 Combinations

Complicating this classification approach is the reality that these categories can be overlapping in many registries. For example, a patient with ischemic heart disease may have an acute myocardial infarction and undergo a primary coronary intervention with placement of a drug-eluting stent and postintervention management with clopidogrel. This patient could be enrolled in an ischemic heart disease registry tracking all patients with this disease over time, a myocardial

infarction registry that is collecting data on patients who present to hospitals with acute myocardial infarction (cross-sectional data collection), a primary coronary intervention registry that includes management with and without devices, a coronary artery stent registry limited to ischemic heart disease patients, or a clopidogrel product registry that includes patients undergoing primary coronary interventions.

3.5 Duration of Observation

The duration of the observational period for a registry is also a useful descriptor. Observation periods may be limited to a single episode of care (e.g., a hospital discharge registry for diverticulitis), or they may extend for as long as the lifetime of patients with a chronic disease (e.g., cystic fibrosis or Pompe disease) or patients receiving a novel therapy (e.g., gene therapy). The period of observation or followup depends on the outcomes of interest.

3.6 From Registry Purpose to Design

As will be discussed extensively in this document, the purpose of the registry defines the registry focus (e.g., product vs. disease) and therefore the registry type. A registry created for the purpose of evaluating outcomes of patients receiving a particular coronary artery stent might be designed as a single product registry if, for example, the purpose is to systematically collect adverse event information on the first 10,000 patients receiving the product. However, the registry might alternatively be designed as a health care service registry for primary coronary intervention if a purpose is to collect comparative effectiveness or safety data on other treatments or products within the same registry.

4. Patient Registries and Policy Purposes

In addition to the growth of patient registries for scientific and clinical purposes, registries are receiving increased attention for their potential role in policymaking or decisionmaking.[56] As stated earlier, registries may offer a view of real-world health care that is typically inaccessible from clinical trials or other data sources and may provide information on the generalizability of the data from clinical trials to populations not studied in those trials.

The utility of registry data for decisionmaking is related to three factors: the stakeholders, the primary scientific question, and the context. The stakeholders are those associated with the disease or procedure that may be affected from a patient, provider, payer, regulator, or other perspective. The primary scientific question for a registry may relate to effectiveness, safety, or practice patterns. The context includes the scientific context (e.g., previous randomized trials and modeling efforts that help to more precisely define the primary scientific question), as well as the political, regulatory, funding, and other issues that provide the practical parameters around which the registry is developed. In identifying the value of information from registries, it is essential to look at the data with specific reference to the purpose and focus of the registry.

From a policy perspective, there are several scenarios in which the decision to develop a registry may arise. One possible scenario is as follows. An item or service is considered for use. Stakeholders in the decision collaboratively define "adequate data in support of the decision at hand." Here, "adequate data" refers to information of sufficient relevance and quality to permit an informed decision. An evidence development strategy is selected from one of many potential strategies (RCT, practical clinical trial, registry, etc.) based on the quality of the evidence provided by each design, as well as the burden of data collection and the cost that is imposed. This tradeoff of the quality of evidence versus cost of data collection for each possible design is termed the "value of information" exercise (Figure 1–1). Registries should be preferred in those circumstances where they provide sufficiently high-quality information for decisionmaking at a sufficiently low cost (relative to other "acceptable" designs).

One set of policy determinations that may be informed by a patient registry centers on the area of payment for items or services. For example, the Centers for Medicare & Medicaid Services (CMS)

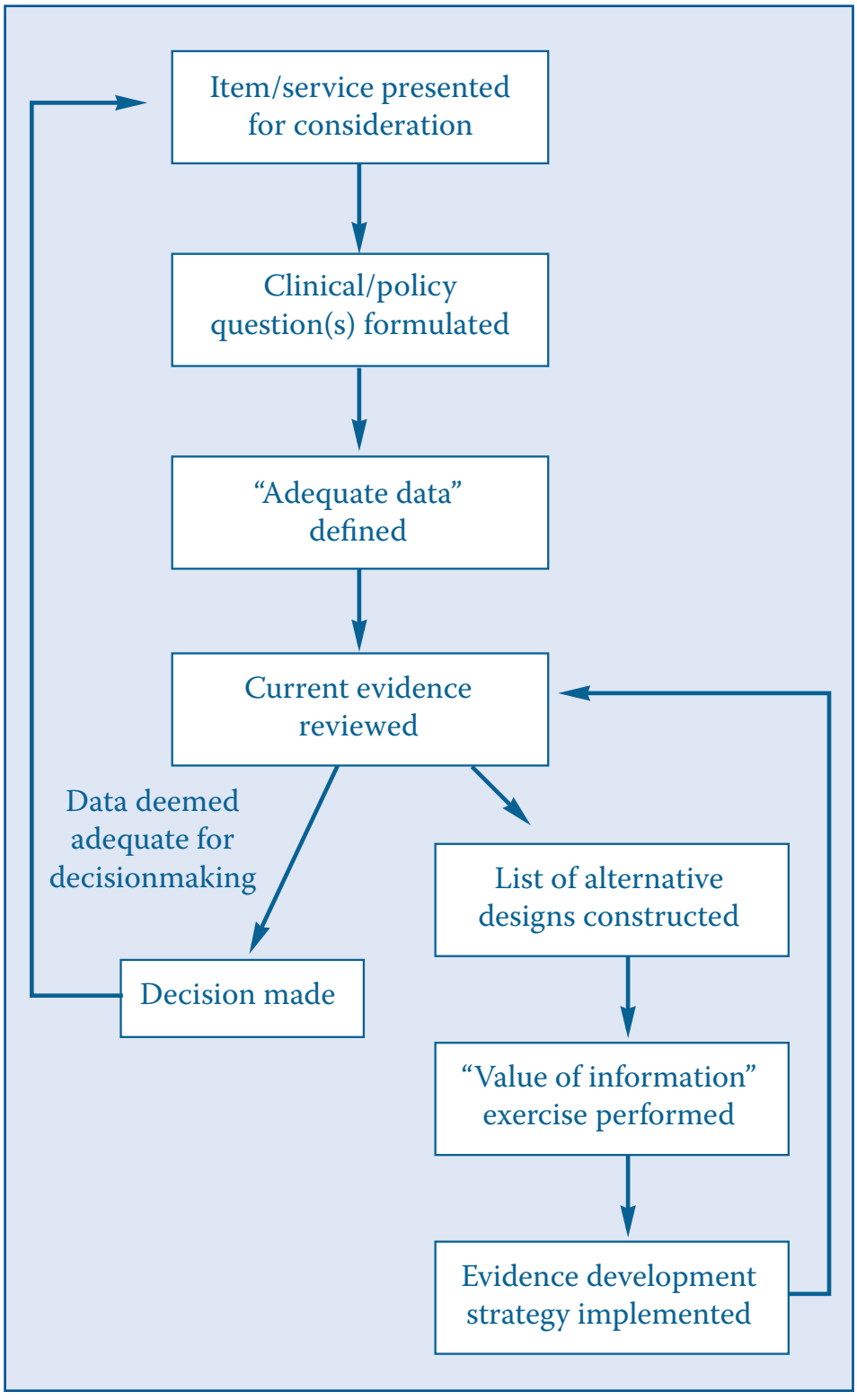
issued Guidance on National Coverage Determinations With Data Collection as a Condition of Coverage in 2006. That original guidance document (which has undergone subsequent revisions, including an additional draft guidance published in 2012[57]) provided several examples of how data collected in a registry might be used in the context of coverage determinations. As described in the Guidance:

> [T]he purpose of CED [Coverage with Evidence Development] is to generate data on the utilization and impact of the item or service evaluated in the NCD [National Coverage Determination], so that Medicare can (a) document the appropriateness of use of that item or service in Medicare beneficiaries under current coverage; (b) consider future changes in coverage for the item or service; (c) generate clinical information that will improve the evidence base on which providers base their recommendations to Medicare beneficiaries regarding the item or service.[56]

The Guidance provided insight into when registry data may be useful to policymakers. These purposes range from demonstrating that a particular item or service was provided appropriately to patients meeting specific characteristics, to collecting new information that is not available from existing clinical trials. CED based on registries may be especially relevant when current data do not address relevant outcomes for beneficiaries, off-label or unanticipated uses, important patient subgroups, or operator experience or other qualifications. Registry-based studies may also be important when an existing treatment is being reconsidered. (An RCT may not be possible under such circumstances.) Registry-based studies are also being used increasingly in fulfillment of postmarketing commitments and requirements.

In many countries, policy determinations on payment rely on cost-effectiveness and cost-utility data and therefore can be informed by registries as well as clinical trials.[58] These data are used and reviewed in a variety of ways. In some countries, there may be a threshold above which a payer is willing to pay for an improvement in patient outcomes.[59] In these scenarios—particularly for rare diseases, when it can be difficult to gather clinical effectiveness data together with quality-of-life data in a utility format—the establishment of disease-specific data registries has been recommended to facilitate the process of technology assessment and improving patient care.[60] In fact, the use of new or existing registries to assess health technology or risk-sharing arrangements is growing in such countries as the United Kingdom, France, Germany, and Australia, and in conditions ranging from bariatric surgery to stroke care.[61-66]

Figure 1–1. Deciding when to develop a registry: The "value of information" exercise



Consider the clinical question of carotid endarterectomy surgery for patients with a high degree of stenosis of the carotid artery. Randomized trials, using highly selected patients and surgeons, indicate a benefit of surgery over medical management in the prevention of stroke. However, that benefit may be exquisitely sensitive to the surgical complication rates; a relatively small increase in the rate of surgical complications is enough to make medical management the preferred strategy instead. In addition, the studies of surgical performance in a variety of hospitals may suggest substantial variation in surgical mortality and morbidity for this procedure. In such a case, a registry to evaluate treatment outcomes, adjusted by hospital and surgeon, might be considered to support a policy decision as to when the procedure should be reimbursed (e.g., only when performed in medical centers resembling those in the various randomized trials, or only by surgeons or facilities with an acceptably low rate of complications).[67]

5. Global Registries

As many stakeholders have international interests in diseases, conditions, and health care products and services, it is not surprising that interest in global patient registries is growing. While some of the specific legal and regulatory discussions in this user's guide are intended for and limited to the United States, most of the concepts and specifics are more broadly applicable to similar activities worldwide. Chapter 7 (ethics, data ownership, and privacy), Chapter 9 (protection of registry data), and Chapter 12 (adverse event detection, processing, and reporting) are perhaps the most limited in their applicability outside the United States. There may be additional considerations in data element selection and patient-reported outcome measure selection (Chapters 4 and 5) stemming from differences ranging from medical training to use of local remedies; the types of data sources that are available outside the United States (Chapter 6); the requirements for informed consent (Chapter 8); the issues surrounding clinician and patient recruitment and retention in different health systems and cultures (Chapter 10); specific data collection and management options and complexities (Chapter 11), ranging from available technologies to languages; and specific requirements for mandated pregnancy registries (Chapter 21).

6. Summary

A patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure and that serves predetermined scientific, clinical, or policy purpose(s). Studies derived from well-designed and well-performed patient registries can provide a real-world view of clinical practice, patient outcomes, safety, and clinical, comparative, and cost-effectiveness, and can serve a number of evidence development and decisionmaking purposes. In the chapters that follow, this user's guide presents practical design and operational issues, evaluation principles, and good registry practices.

References for Chapter 1

1. Webster's English Dictionary. http://www.m-w.com. Accessed August 12, 2012.
2. Brooke EM. The current and future use of registers in health information systems. Geneva: World Health Organization; 1974. Publication No. 8.
3. National Committee on Vital and Health Statistics. Frequently Asked Questions About Medical and Public Health Registries. http://ncvhs.hhs.gov/9701138b.htm. Accessed August 14, 2012.
4. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. Am Heart J. 2009 Jun;157(6):971-82. PMID: 19464406.
5. Hammill BG, Hernandez AF, Peterson ED, et al. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. Am Heart J. 2009 Jun;157(6):995-1000. PMID: 19464409. Pubmed Central PMCID: 2732025.

6. Developing a Registry of Patient Registries (RoPR). Project Abstract. Agency for Healthcare Research and Quality. http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=690&pageaction=displayproduct. Accessed August 14, 2012.
7. LaBresh KA, Gliklich R, Liljestrand J, et al. Using "get with the guidelines" to improve cardiovascular secondary prevention. Jt Comm J Qual Saf. 2003 Oct;29(10):539-50. PMID: 14567263.
8. Kennedy L, Craig AM. Global registries for measuring pharmacoeconomic and quality-of-life outcomes: focus on design and data collection, analysis and interpretation. Pharmacoeconomics. 2004;22(9):551-68. PMID: 15209525.
9. Charrow J, Esplin JA, Gribble TJ, et al. Gaucher disease: recommendations on diagnosis, evaluation, and monitoring. Arch Intern Med. 1998 Sep 14;158(16):1754-60. PMID: 9738604.
10. Bushby K, Lynn S, Straub T, et al. Collaborating to bring new therapies to the patient--the TREAT-NMD model. Acta Myol. 2009 Jul;28(1):12-5. PMID: 19772190. PMCID: 2859629.
11. Dhruva SS, Phurrough SE, Salive ME, et al. CMS's landmark decision on CT colonography--examining the relevant data. N Engl J Med. 2009 Jun 25;360(26):2699-701. PMID: 19474419.
12. Postmarketing studies and clinical trials – implementation of Section 505(o) of the Federal Food, Drug and Cosmetic Act. FDA Guidance for Industry. Draft guidance. July 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf. Accessed August 14, 2012.
13. U.S. Food and Drug Administration. Guidance for Industry. Good pharmacovigilance practices and pharmacoepidemiologic assessment. March 2005. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126834.pdf. Accessed August 14, 2012.
14. U.S. Food and Drug Administration. FDA Post-Approval Studies. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma_pas.cfm?utm_campaign=Google2&utm_source=fdaSearch&utm_medium=website&utm_term=post-approval%20studies&utm_content=1. Accessed August 14, 2012.
15. Dreyer NA, Garner S. Registries for robust evidence. JAMA. 2009 Aug 19;302(7):790-1. PMID: 19690313.
16. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000 Jun 22;342(25):1887-92. PMID: 10861325. PMCID: 1557642.
17. Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. JAMA. 1995 Dec 13;274(22):1800-4. PMID: 7500513.
18. Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Version 1.0 [Draft posted Oct. 2007]. Rockville, MD. http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf. Accessed August 14, 2012.
19. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med. 2008 Mar 11;5(3):e67. PMID: 18336067. PMCID: 2265762.
20. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res. 2004 Dec 22;4(1):38. PMID: 15615589. PMCID: 545647.
21. Rawlins M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions. Clin Med. 2008 Dec;8(6):579-88. PMID: 19149278.
22. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004 Jun 19;328(7454):1490. PMID: 15205295. PMCID: 428525.
23. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions – AHRQ and the Effective Health Care Program: An Update. Draft released for public comment. http://www.effectivehealthcare.ahrq.gov/ehc/products/457/1163/GradingTheStrengthofEvidence_DraftMethodsChapter_20120625.pdf. Accessed October 12, 2012.

24. Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters LM, Santaguida PL, Shamliyan T, Singh K, Tsertsvadze A, Treadwell JR. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. AHRQ Publication No. 12-EHC047-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2012. http://www.effectivehealthcare.ahrq.gov/ehc/products/322/998/MethodsGuideforCERs_Viswanathan_IndividualStudies.pdf. Accessed October 12, 2012.

25. Clancy CM, Eisenberg JM. Outcomes research: measuring the end results of health care. Science. 1998 Oct 9;282(5387):245-6. PMID: 9841388.

26. Lipscomb J, Snyder CF. The Outcomes of Cancer Outcomes Research: focusing on the National Cancer Institute's quality-of-care initiative. Med Care. 2002 Jun;40(6 Suppl):III3-10. PMID: 12064755.

27. National Cancer Institute. The Emerging Field of Outcomes Research. http://outcomes.cancer.gov/aboutresearch/index.html. Accessed August 14, 2012.

28. Lipscomb J, Donaldson MS, Hiatt RA. Cancer outcomes research and the arenas of application. J Natl Cancer Inst Monogr. 2004 (33):1-7. PMID: 15504917.

29. Institute of Medicine, Committee on Quality of Heallth Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC: National Academy Press, Institute of Medicine; 2001.

30. Schwamm LH, LaBresh KA, Pan WQ, et al. Get With The Guidelines – Stroke produces sustainable improvements in hospital-based acute stroke care. Stroke. 2006;37(2):624.

31. Schwamm LH, LaBresh KA, Pan WQ, et al. Get With The Guidelines – Stroke improves the rate of "defect-free" acute stroke care. Stroke. 2006;37(2):659.

32. Barranger JA, O'Rourke E. Lessons learned from the development of enzyme therapy for Gaucher disease. J Inherit Metab Dis. 2001;24 Suppl 2:89-96; discussion 87-8. PMID: 11758684.

33. Wennberg DE, Lucas FL, Birkmeyer JD, et al. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. JAMA. 1998 Apr 22-29;279(16):1278-81. PMID: 9565008.

34. MacIntyre K, Capewell S, Stewart S, et al. Evidence of improving prognosis in heart failure: trends in case fatality in 66 547 patients hospitalized between 1986 and 1995. Circulation. 2000 Sep 5;102(10):1126-31. PMID: 10973841.

35. Konstam MA. Progress in heart failure Management? Lessons from the real world. Circulation. 2000 Sep 5;102(10):1076-8. PMID: 10973832.

36. Hutchins LF, Unger JM, Crowley JJ, et al. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. N Engl J Med. 1999 Dec 30;341(27):2061-7. PMID: 10615079.

37. Opthalmic devices and clinical epidemiology. In: Brown SL, Bright RA, Tavris DR, eds. Medical device epidemiology and surveillance: John Wiley & Sons, Ltd.; 2007:chapter 28.

38. Lipscomb J, Yabroff KR, Brown ML, et al. Health care costing: data, methods, current applications. Med Care. 2009 Jul;47(7 Suppl 1):S1-6. PMID: 19536004.

39. Eichler HG, Kong SX, Gerth WC, et al. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? Value Health. 2004 Sep-Oct;7(5):518-28. PMID: 15367247.

40. Palmer AJ. Health economics--what the nephrologist should know. Nephrol Dial Transplant. 2005 Jun;20(6):1038-41. PMID: 15840678.

41. Good ReseArch for Comparative Effectiveness. http://www.graceprinciples.org. Accessed August 14, 2012.

42. Institute of Medicine. Learning what works best: the Nation's need for evidence on comparative effectiveness in health care. Washington DC: National Academy Press; 2007. p. 1-80.

43. Congressional Budget Office. Research on the Comparative Effectiveness of Medical Treatments: Issues and Options for an Expanded Federal Role. 2007. http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/88xx/doc8891/12-18-comparativeeffectiveness.pdf. Accessed August 14, 2012.

44. Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress. June 30, 2009. U.S. Department of Health and Human Services. http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf. Accessed August 14, 2012.

45. Patient-Centered Outcomes Research Institute. "Patient-Centered Outcomes Research.". http://www.pcori.org/what-we-do/pcor/. Accessed August 28, 2012.

46. Public comment draft report of the Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee presented on July 23, 2012, and revised thereafter. http://pcori.org/assets/MethodologyReport-Comment.pdf. Accessed August 28, 2012.

47. Agency for Healthcare Research and Quality. Bridging Knowledge Gaps in the Comparative Effectiveness of ACE Inhibitors and ARBs. Draft abstract. 2008. . http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=296&pageaction=displayproduct. Accessed August 14, 2012.

48. Hodgson DC, Fuchs CS, Ayanian JZ. Impact of patient and provider characteristics on the treatment and outcomes of colorectal cancer. J Natl Cancer Inst. 2001 Apr 4;93(7):501-15. PMID: 11287444.

49. Reeves MJ, Fonarow GC, Zhao X, et al. Quality of care in women with ischemic stroke in the GWTG program. Stroke. 2009 Apr;40(4):1127-33. PMID: 19211482.

50. Fonarow GC, Abraham WT, Albert NM, et al. Influence of a performance-improvement initiative on quality of care for patients hospitalized with heart failure: results of the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients With Heart Failure (OPTIMIZE-HF). Arch Intern Med. 2007 Jul 23;167(14):1493-502. PMID: 17646603.

51. Greene FL, Gilkerson S, Tedder P, et al. The role of the hospital registry in achieving outcome benchmarks in cancer care. J Surg Oncol. 2009 Jun 15;99(8):497-9. PMID: 19466739.

52. Schweikert B, Hunger M, Meisinger C, et al. Quality of life several years after myocardial infarction: comparing the MONICA/KORA registry to the general population. Eur Heart J. 2009 Feb;30(4):436-43. PMID: 19019995.

53. Lane K, Kempf A, Magno C, et al. Regional differences in the use of sentinel lymph node biopsy for melanoma: a potential quality measure. Am Surg. 2008 Oct;74(10):981-4. PMID: 18942627.

54. Griffiths I, Silman A, Symmons D, et al. BSR Biologics Registry. Rheumatology (Oxford). 2004 Dec;43(12):1463-4. PMID: 15328420.

55. U.S. Food and Drug Administration. Guidance for Industry: Establishing Pregnancy Exposure Registries. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071639.pdf. Accessed August 15, 2012.

56. Centers for Medicare & Medicaid Services. Guidance for the Public, Industry, and CMS Staff: National Coverage Determinations with Data Collection as a Condition of Coverage: Coverage with Evidence Development. July 12, 2006.

57. Centers for Medicare & Medicaid Services. Draft Guidance for the Public, Industry, and CMS Staff: Coverage with Evidence Development in the context of coverage decisions. http://www.cms.gov/medicare-coverage-database/details/medicare-coverage-document-details.aspx?MCDId=23. Accessed February 20, 2013.

58. Connock M, Burls A, Frew E, et al. The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review. Health Technol Assess. 2006 Jul;10(24):iii-iv, ix-136. PMID: 16796930.

59. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ. 2004 May;13(5):437-52. PMID: 15127424.

60. Connock M, Juarez-Garcia A, Frew E, et al. A systematic review of the clinical effectiveness and cost-effectiveness of enzyme replacement therapies for Fabry's disease and mucopolysaccharidosis type 1. Health Technol Assess. 2006 Jun;10(20):iii-iv, ix-113. PMID: 16729919.

61. Chalkidou K, Tunis S, Lopert R, et al. Comparative effectiveness research and evidence-based health policy: experience from four countries. Milbank Q. 2009 Jun;87(2):339-67. PMID: 19523121. Pubmed Central PMCID: 2881450.

62. National Institute for Health and Clinical Excellence (UK). Final appraisal determination: alteplase for the treatment of acute ischaemic stroke. 2007. http://www.nice.org.uk/nicemedia/live/11617/33971/33971.pdf. Accessed August 14, 2012.

63. Graves SE, Davidson D, Ingerson L, et al. The Australian Orthopaedic Association National Joint Replacement Registry. Med J Aust. 2004 Mar 1; 180(5 Suppl):S31-4. PMID: 14984361.

64. Owen A, Spinks J, Meehan A, et al. A new model to evaluate the long-term cost effectiveness of orphan and highly specialised drugs following listing on the Australian Pharmaceutical Benefits Scheme: the Bosentan Patient Registry. J Med Econ. 2008;11(2):235-43. PMID: 19450082.

65. Haute Autorité de Santé. Heart surgery with or without extracorporeal circulation: role of the second surgeon. 2008. http://www.has-sante.fr/portail/upload/docs/application/pdf/abstract_heart_surgery.pdf. Accessed August 14, 2012.

66. Haute Autorité de Santé. Interview du docteur Jean-François Thébaut, cardiologue libéral et Président du Conseil National Professionnel de Cardiologie. [French.]. http://www.has-sante.fr/portail/jcms/c_777504/interview-du-docteur-jean-francois-thebautcardiologue-liberal-et-president-du-conseil-nationalprofessionnel-de-cardiologie?portal=c_63456. Accessed August 14, 2012.

67. Matchar DB, Oddone EZ, McCrory DC, et al. Influence of projected complication rates on estimated appropriate use rates for carotid endarterectomy. Appropriateness Project Investigators. Academic Medical Center Consortium. Health Serv Res. 1997 Aug;32(3):325-42. PMID: 9240284. PMCID: 1070194.

Chapter 2. Planning a Registry

1. Introduction

There is tremendous variability in size, scope, and resource requirements for registries. Registries may be large or small in terms of numbers of patients or participating sites. They may target rare or common conditions and exposures. They may require the collection of limited or extensive amounts of data, operate for short or long periods of time, and be funded generously or operate with limited financial support. In addition, the scope and focus of a registry may be adapted over time to reach broader or different populations, assimilate additional data, focus on or expand to different geographical regions, or address new research questions. While this degree of flexibility confers enormous potential, registries require good planning in order to be successful.

When planning a registry, it is desirable to follow these initial steps: (1) articulate the purpose of the registry; (2) determine if a registry is an appropriate means to achieve the purpose; (3) identify key stakeholders; and (4) assess the feasibility of a registry.

Once a decision is made to proceed, the next considerations in planning are to (5) build a registry team; (6) establish a governance and oversight plan; (7) define the scope and rigor needed; (8) define the data set, patient outcomes, and target population; (9) develop a study plan or protocol; and (10) develop a project plan. Of course, the planning for a registry is often not a linear process. Many of the steps described in this chapter occur in parallel.

Registry planners should also recognize the importance of periodic critical evaluations of the registry by key stakeholders to ensure that the objectives are being met. This is particularly important for patient registries that collect data over many years. When registry objectives are no longer being met or when clinical or other changes affect the registry (e.g., changes in treatment practices, the introduction of a new therapy), the registry may need to be adapted, or the registry may stop collecting new data. Registries may undergo a transition or cease collecting new data for many reasons. These considerations are fully discussed in Chapter 14.

The *Guidelines for Good Pharmacoepidemiology Practice* from the International Society of Pharmacoepidemiology is a useful resource for registry planners.[1] The *Updated Guidelines for Evaluating Public Health Surveillance Systems* may also be useful, especially the appendixes, which provide various checklists.[2] A *Guide to the Project Management Body of Knowledge (PMBOK® Guide)* and the GRACE principles for comparative effectiveness (www.graceprinciples.org) may also be useful resources to registry planners.[3, 4]

2. Steps in Planning a Registry

2.1 Articulate the Registry's Purpose

One of the first steps in planning a registry is articulating its purpose. Having a clearly defined goal and/or purpose and supporting rationale makes it easier to evaluate whether a registry is the right approach for capturing the information of interest.[5, 6] In addition, a clearly defined purpose helps clarify the need for certain data. Conversely, having a clear sense of how the data may be used will help refine the stated purpose. Attempts to be all inclusive may add cost but not value, resulting in overly burdensome data collection that can reduce quality and erode compliance.

A registry may have a singular purpose or several purposes.[7] In either case, the overall purpose should be translated into specific objectives or questions to be addressed through the registry. This process needs to take into account the interests of those collaborating in the registry and the key audiences to be reached.[8] Clear objectives are essential to define the structure and process of data collection and to ensure that the registry effectively addresses the important questions through the appropriate outcomes analyses. Specific objectives also help the registry to avoid

collecting large amounts of data of limited value. The time and resources needed to collect and process data from a registry can be substantial.[9] The identification of a core data set is essential. The benefits of any data element included in the registry must outweigh the costs of including it.

Registry planners should establish specific objectives by considering what key questions the registry needs to answer. Critical consideration should be given to defining the key questions in order to evaluate how best to proceed, as these questions will help to establish the type of registry (e.g., whether single focus or comparative), the data elements to be captured, and the types of analysis to be undertaken. Examples of key or driving questions are listed below:

- What is the natural course of a disease, and how does geographic location affect the course?
- Does a treatment lead to long-term benefits or harm, including delayed complications?
- How is disease progression affected by available therapies?
- What are significant predictors of poor outcomes?
- What is the safety profile of a specific therapy?
- Is a specific product or therapy teratogenic?
- How do clinical practices vary, and what are the best predictors of treatment practices?
- Are there disparities in the delivery and/or outcomes of care?
- What characteristics or practices enhance compliance and adherence?
- Do quality improvement programs affect patient outcomes, and, if so, how?
- What process and outcomes metrics should be incorporated to track quality of patient care?
- Should a particular procedure or product be a covered benefit in a particular population?
- Was an intervention program or risk-management activity successful?
- What are the resources used/economic parameters of actual use in typical patients?

2.2 Determine if a Registry Is an Appropriate Means To Achieve the Purpose

Two key questions to consider are whether a registry (or other study) is needed to address the purpose and, if the answer is yes, whether prospective data collection through a registry is an appropriate means of accomplishing the scientific objectives. Every registry developer should consider the following questions early in the planning process:

- Do these data already exist?
- If so, are they of sufficient quality to answer the research question?
- Are they accessible, or does an entirely new data collection effort need to be initiated?

For example, could the necessary data be extracted from electronic medical records or administrative health insurance claims data? In such cases, registries might avoid re-collecting data that have already been collected elsewhere and are accessible. Thought should be given to adapting the registry (based on extant data) and/or linking to other relevant data sources (including "piggybacking" onto other registries). The Registry of Patient Registries (RoPR), developed by the Agency for Healthcare Research and Quality, is a resource for finding patient registries.[10] When the required data have not been sufficiently collected or are not accessible for the desired purpose, it is appropriate to consider creating a new registry.

The next step is to consider whether the purpose would be well served by a registry. When making this decision, it is important to fully define the specific research question(s) of interest and to consider the state of current knowledge and gaps in evidence. Other factors that may influence this decision include the breadth of the target population of interest, the complexity of the current treatment patterns, the length of an observational period needed to achieve the objective, the scope and variety of treatments used, the approximate amount of funding available to address these objectives, and the urgency of decisions that will be made based on the resulting

evidence. Registries may be the most appropriate choice for some research questions. For example, registries are particularly useful in situations where a comprehensive, flexible research design is needed,[11, 12] or when the purpose is to discover how a product works in a wide variety of subgroups. (See Chapter 3, Section 2 for a discussion of research questions appropriate for registries.)

Other research questions, such as ones that might be used to petition a regulatory agency for a new indication, will require different approaches, such as traditional randomized controlled trials. In some cases, a hybrid approach, such as a registry that incorporates data collected retrospectively as well as prospectively, will be required. A research strategy, as opposed to a single study, may be necessary to address some research questions. For example, some research questions may require an interventional approach to address concerns about efficacy combined with an observational approach to examine long-term outcomes and quality of life in a broad patient population. When making a decision about study design, it is important to select the approach or combination of approaches best able to answer the specific research questions, from both scientific and practical standpoints. A careful evaluation of the possibilities for data collection and registry design, the degree of certainty required, and the timeframe in which this certainty is expected can help in selecting an appropriate study design.

Historically, there has been a lack of consensus standards for conducting and reporting methods and results for registries. Therefore, registries have been more variable in implementation and more difficult to assess for quality than randomized controlled trials. In recent years, advances in epidemiological and biostatistical methods have broadened the scope of questions that can be addressed through observational studies such as registries. Stratification, propensity score matching, and risk adjustment are increasingly useful approaches for addressing confounding issues and for creating comparably homogeneous subgroups for analysis within registry data sets, and advances in bias analysis are being used to help interpret results from observational studies such as registries.[13-15] (See Chapters 3, 13, and 18.) These techniques may allow registries to be used to support investigations of comparative safety and effectiveness. Following good registry practices, as described in this user's guide, can strengthen scientific rigor. (See Chapter 25.)

2.3 Identify Key Stakeholders

As a means of identifying potential stakeholders, it is important to consider to whom the research questions matter. It is useful to identify these stakeholders at an early stage of the registry planning process, as they may have important input into the type and scope of data to be collected, they may ultimately be users of the data, and/or they may have a key role in disseminating the results of the registry.

One or more parties could be considered stakeholders of the registry. These parties could be as specific as a regulatory agency that will be monitoring postmarketing studies or as broad as the general population, or simply those patients with the conditions of interest. Often, a stakeholder's input directly influences whether development of a registry can proceed, and it can have a strong influence on how a registry is conducted. A regulatory agency looking for management of a therapeutic product with a known toxicity profile may require a different registry design than a manufacturer with general questions about how a product is being used.

Typically, there are primary and secondary stakeholders for any registry. A primary stakeholder is usually responsible for creating and funding the registry. The party that requires the data, such as a regulatory authority, may also be considered a primary stakeholder. A secondary stakeholder is a party that would benefit from knowledge of the data or that would be impacted by the results but is not critical to establishing the registry. Treating clinicians and their patients could be considered secondary stakeholders. A partial list of possible stakeholders, both primary and secondary, follows:

- Public health or regulatory authorities
- Product manufacturers
- Health care service providers

- Payer or commissioning authorities
- Patients and/or advocacy groups
- Treating clinician groups
- Academic institutions or consortia
- Professional societies

Although interactions with potential stakeholders will vary, the registry will be best supported by defined interactions and communications with these parties. Defining these interactions during the planning stage will ensure that adequate dialog occurs and appropriate input is received to support the overall value of the registry. Interactions throughout the entire duration of the registry can also assure stakeholders that the registry is aligned with the purposes and goals that were set out during the planning stages and that the registry complies with all required guidances, rules, and/or regulations.

2.4 Assess Feasibility

A key element in determining the feasibility of developing a new registry relates to funding. Registries that meet the attributes described in this user's guide will most likely require significant funding. The degree of expense incurred will be determined by the scope of the registry, the rigor of data collection, and any audits that may be required. The larger the number of sites, the number of patients, and the scope of data collected, and the greater the need for representation of a wide variety of patient characteristics, the greater the expense will be. In addition, the method of data collection will contribute to expense. Historically, electronic data collection has been more expensive to implement, but generally less expensive to maintain, than forms that are faxed and scanned or mailed;[16] however, the cost difference for startup has been lessening. Funding will be affected by whether other relevant data sources and/or infrastructures exist that capture some of the information of interest; whether the registry adapts to new issues over time; and whether multiple funding sources participate. Funding needs should also be examined in terms of the projected life of the registry and/or its long-term sustainability.

There are many potential funding sources for registries. Funding sources are likely to want to share in planning and to provide input for the many choices that need to be made in the implementation plans. Funding sources may negotiate to receive access to deidentified data as a condition for their participation. Funding models for registries may vary significantly, and there is no preferred approach. Rather, the funding model for a registry should be dictated by the needs of the registry. Potential sources of funding include:

- *Foundations:* Nonprofit disease foundations may be interested in a registry to track the natural history of the disease of interest as well as the impact of therapeutic interventions. Registries may be used to track practice patterns and outcomes for quality improvement initiatives. Ongoing registries can sometimes serve the additional purpose of assisting in recruitment for clinical trials.[17]
- *Government:* Federal agencies, such as the National Institutes of Health (NIH), Centers for Disease Control and Prevention (CDC), Centers for Medicare & Medicaid Services (CMS), Agency for Healthcare Research and Quality (AHRQ), U.S. Food and Drug Administration (FDA), and State agencies, may be interested in a registry to determine long-term outcomes of agents, devices, groups of drugs, or procedures. While the pharmaceutical industry or device manufacturers collect most long-term data on drug and device safety, many research questions arise that could potentially be suitable for government funding, ranging from clinical or comparative effectiveness to natural history of disease to the performance of health care providers based on accepted measures of quality of care. To determine if an agency might be interested in funding a registry, look for Requests for Proposals (RFPs) on its Web site. An RFP posting or direct communication with the appropriate agency staff may provide a great deal of specific information as to how a submission will be judged and what criteria would be needed in order for a proposal to be favorably ranked. Even if an RFP is not posted, contacting the

appropriate agency staff may uncover potential interest in a registry to fill an unmet need.

- *Health plan providers:* Under certain circumstances, health plan providers may be interested in funding a registry, since practical clinical research is increasingly viewed as a useful tool for providing evidence for health coverage and health care decisions.[18]
- *Patient groups:* Patients may be able to contribute funding to focus on rare diseases or patient subgroups of interest for more common conditions. They may also contribute value in-kind.
- *Private funding:* Private philanthropic individuals or charitable foundations and trusts may have an interest in furthering research to better understand the effects of a particular intervention or sets of interventions on a disease process.
- *Product manufacturers:* Product manufacturers may be interested in studying the natural history of the disease for which they have (or are developing) a product; demonstrating the effectiveness and/or safety of existing products in real-world use through Risk Evaluation and Mitigation Strategy (REMS) programs as part of postmarketing commitments or requirements, or through studies; or assisting providers in evaluating or improving quality of care.
- *Professional societies:* Health care professional associations are increasingly participating in developing or partnering with registries for scientific and quality measurement or improvement purposes.
- *Professional society/pharmaceutical industry "hybrids":* Situations may exist in which a product manufacturer funds a registry designed and implemented by a professional society to gain insight into a set of research questions.
- *Multiple sponsors:* Registries may meet the goals of multiple stakeholders, and such stakeholders may have an interest in sharing the funding. Registries for isotretinoin and antiretrovirals in pregnancy are examples, as is INTERMACS™, a registry for patients who are receiving mechanical circulatory support device therapy to treat advanced heart failure.[19] While multiple sponsorship can decrease the costs for each funding source, their varied interests and needs almost always increase the complexity and overall cost of the registry.

A public-private partnership is a service or business venture that is funded and operated through a partnership (contractual agreement) between a public agency (Federal, State, or local) and a private-sector entity or entities.[20] While some true public-private partnerships for registries currently exist (e.g., State-level immunization registries, bioterrorism surveillance efforts),[21-23] there is great potential for growth in this approach. Both government and private sources have shown increasing interest in registries for improved safety monitoring, for comparative effectiveness goals, and for streamlining the costs of the drug development process.[24-29] Several legislative actions have stated or suggested the role of public-private partnerships for activities such as registry development.[30] There are many good reasons for multiple stakeholders, including government agencies, providers, and industry, to work together for certain purposes. Thus, it is anticipated that shared funding mechanisms are likely to become more common. Chapter 24 provides more detail on the use of public-private partnerships to support registries.

2.5 Build a Registry Team

Several different kinds of knowledge, expertise, and skills are needed to plan and implement a registry. In a small registry run by a single individual, consultants may be able to provide the critical levels of expertise needed to plan all components of the registry. In a large registry, a variety of individuals may work together as a team to contribute the necessary expertise. Depending on the size, scope, and purpose of the registry, few, some, or all of the individuals representing the components of expertise described below may be included at the time of the planning process. Whatever number of individuals is eventually assembled, it is important to build a group that can work together as a collegial team to accomplish the goals of the registry. Additionally, the team

participants must understand the data sources. By understanding the goals and data sources, the registry team will enable the data to be used in the most appropriate context for the most appropriate interpretation. The different kinds of expertise and experience that are useful include the following:

- *Project management:* Project management will be needed to coordinate the components of the registry; to manage timelines, milestones, deliverables, and budgets; and to ensure communication with sites, stakeholders, oversight committees, and funding sources. Ongoing oversight of the entire process will require a team approach. (See Section 2.6, "Establish a Governance and Oversight Plan.")
- *Subject matter:* A registry must be designed so that it contains the appropriate data to meet its goals as well as the needs of its stakeholders. For example, experts in the treatment of the clinical disease to be studied who are also familiar with the potential toxicities of the treatment(s) to be studied are critical to the success of the registry. Clinical experts must be able to apply all of the latest published clinical, toxicity, and outcome data to components of the registry and determine which elements are necessary, desirable, or superfluous. Depending on the outcomes and registry purpose, it is often useful to have patient representatives or advocates.
- *Registry science:* Epidemiology and biostatistics expertise specific to the subtleties of patient registries and observational research is very important in the design, implementation, and analysis of registry data. Epidemiologists can provide the study design and can work in collaboration with biostatisticians to develop a mutual understanding of the research objectives and data needed. Health outcomes researchers and economics researchers can also lend valuable expertise to the registry team. These scientists should work with the subject matter experts to ensure that appropriate analytic methods are being used to address the clinical issues relevant to achieving the goals of the registry.
- *Data collection and database management:* The decision to include various data elements can be made in consultation with experts in this field to place "critical fields" in a prominent and logical position on the data form for both paper-based and electronic data collection tools. (A final determination of what is usable and workable for data collection tools should be approved by all members of the team.) These experts may also need to write specific programs so that the data received from the registry are grouped, stored, and identified. They may generate reports for individuals who track registry participation, and they may provide data downloads periodically to registry analysts. This team will also be responsible for implementing and maintaining firewalls to protect the data according to accepted levels of security for similar collections of sensitive data.
- *Legal issues/patient privacy:* It is critical that either information that identifies individual patients be excluded or applicable legal requirements for the inclusion of patient identifiable information be met (e.g., obtaining informed consent or Health Insurance Portability and Accountability Act [HIPAA] authorization, where required). The complexities of this topic are dealt with in detail in Chapters 7, 8, and 9. Legal and privacy expertise is needed to protect the patients and the owners of the database by ensuring that the registry complies with all Federal and State laws applicable to patient information.
- *Quality assurance:* As discussed in Chapter 11, Section 3, quality assurance of procedures and data is another important component of registry success. Expertise in quality assurance will help in planning a good registry. The goals for quality assurance should be established for each registry, and the efforts made and the results achieved should be described.

2.6 Establish a Governance and Oversight Plan

Governance refers to guidance and high-level decisionmaking, including purpose, funding, execution, and dissemination of information. A goal of proper governance and oversight should be

transparency to stakeholders in operations, decisionmaking, and reporting of results.

The composition and relative mix of stakeholders and experts relate largely to the purpose of the registry. For example, if the purpose of the registry is to determine a comparative effectiveness or reimbursement policy, those impacted by the policy should not solely govern the registry. Broad stakeholder involvement in governance boards is most desirable when there are many stakeholders. Depending on the size of the registry, governance may be assumed by various oversight committees made up of interested individuals who are part of the design team (internal governance) or who remain external to the day-to-day operations of the registry (external governance). Differences in the nature of the study questions, the overall resources being consumed by the registry, the soundness of the underlying data sources, and many other factors will influence the degree of involvement and role of oversight groups. In other words, the purpose of the committee functions described below is to lay out the roles that need to be assumed by the governance structure of many registries, but these should be individualized for a particular registry. It is also possible, if methods are clear and transparent, that oversight requirements may be minimal.

Registries fulfill governance roles in a variety of ways. Many of the roles, for example, could be assumed by a single committee (e.g., a steering committee) in some registries. Whatever model is adopted, it must accommodate all of the working constituencies and provide a mechanism for these individuals to work together to achieve the goals of the registry.

All aspects of governance should be codified in a written format that can be reviewed, shared, and refined over time. In addition, governance is a dynamic process, subject to change in policy as evidence emerges that is likely to lead to improvements in the process.

Governance and oversight functions that may be considered include:

- *Executive or steering:* This function assumes responsibility for the major financial, administrative, legal/ethical, and scientific decisions that determine the direction of the registry. These decisions are made with appropriate input from legal, scientific, and administrative experts. Depending on their capabilities and the size and resources of the registry, the group serving the steering function may also assume some of the functions described below.
- *Scientific:* This function may include experts in areas ranging from database content, to general clinical research, to epidemiology and biostatistics. This function may determine the overall direction of database inquiries and recommend specific analyses to the executive or steering group. It is strongly desirable that the reports that emerge from a registry be scientifically based analyses that are independent and transparent.[31] To enhance credibility and in the interest of full disclosure, the role of all stakeholders in the publication process should be specified and any potential conflicts of interest identified.
- *Liaison:* In large registries, a function may be specified to focus on maintaining relationships with the funding source, health care providers, and patients who need access to registry information. The group serving this function may develop monitoring and satisfaction tools to ensure that the day-to-day operations of the registry remain healthy.
- *Adjudication:* Adjudication is used to review and confirm cases (outcomes) that may be difficult to classify. Individuals performing this function are generally blinded to the exposure (product or process) under study so that the confirmation of outcomes is made without knowledge of exposure.
- *External review:* External review committees and/or advisory boards can be useful for providing independent oversight throughout the course of the registry. The majority of registries will not require a data safety monitoring board (DSMB), since a DSMB is commonly used in situations where data are randomized and treatment status is blinded. However, there may be situations in which the registry is responsible for the primary accumulation of safety data on a particular intervention; in such

situations, an external committee or DSMB would be useful for conducting periodic reviews (e.g., annually).

- *Data access, use, and publications:* This function should address the process by which registry investigators access and perform analyses of registry data for the purpose of submitting abstracts to scientific meetings and developing manuscripts for peer-reviewed journal submission. Authorship (including that of registry sponsors) in scientific publications should satisfy the conditions of the Uniform Requirements for Manuscripts Submitted to Biomedical Journals.[32] The rules governing authorship may be affected by the funding source, as in the case of NIH or foundation funding, or by the biomedical journal. (See Case Examples 1 and 22.) Other investigators may request permission to access the data. For example, a Ph.D. candidate at an institution might seek registry-wide aggregate data for the purpose of evaluating a new scientific question. A process for reviewing and responding to such requests from other investigators or entities should be considered in some registries that may generate broad external interest if the registry stakeholders and participants are agreeable to such use.

2.7 Consider the Scope and Rigor Needed

2.7.1 Scope of Data

The scope of a registry may be viewed in terms of size, setting, duration, geography, and financing. The purpose and objectives of the registry should frame the scope, but other factors (aside from feasibility) may ultimately shape it. For example, the scope may be affected by:

- Regulatory requirements, such as those imposed by the FDA as a condition of product marketing.
- Reimbursement decisions, such as national coverage decisions by CMS or "Prior Authorization" requirements used by health insurers in some situations.
- National research interests, such as those driven by NIH.
- Public health policy, such as CDC policy and immunization policy.

The scope is also affected by the degree of uncertainty that is acceptable to the primary stakeholders, with that uncertainty being principally driven by the quantity, quality, and detail of the data collection balanced against its considered importance and value. Therefore, it is critical to understand the potential questions that may or may not be answerable because of the quantity and quality of the data. It should also be noted that the broader the audience of stakeholders, the broader the list of questions that may need to be included. This increased breadth can result in an increase in the number of patients who need to be enrolled and/or data points that need to be collected in order to meet the objective of the registry with an acceptable level of precision.

Some of the specific variables that can characterize the scope of a registry include:

- *Size:* This refers to the number and complexity of data points, the frequency of data collection, and the enrollment of investigators and patients. A registry with a large number of complex data points may allow for detailed and thoughtful analyses but may be so burdensome as to discourage investigator and patient enrollments. In turn, a small registry with few patients and data points may be easier to execute, but the data could lack depth and be less meaningful.[33] Size also determines the precision with which measures of risk or risk difference can be calculated.
- *Duration:* The planning of a registry must reflect the length of time that the registry is expected to collect the data in order to achieve its purpose and provide analysis of the data collected. Some registries are limited by commercial interests, such as when the product under study is approaching the end of its patent life.

- *Setting:* This refers to the specific setting through which the registry will recruit investigators and patients as well as collect data (e.g., hospital, doctor's office, pharmacy, home).
- *Geography:* A locally run registry is very different in scope from a global registry, in terms of setup, management, and analysis. A global registry poses challenges (e.g., language, cultural, time zone, regulatory) that must be taken into consideration in the planning process.
- *Cost:* The scope of a registry will determine the cost of creating, managing, and analyzing the registry. Budgetary constraints must be carefully considered before moving from conception to reality. Additionally, the value of the information is a factor in the financial decisions. Certain choices in planning, such as building on existing infrastructure and/or linking to data sources relevant to the purposes of the registry, may increase the net return.
- *Richness of clinical data needed:* In some situations, the outcome may be relatively simple to characterize (e.g., death). In other cases, the focus of interest may be a complex set of symptoms and measurements (e.g., for Churg-Strauss Syndrome) or may require specialized diagnostic testing or tissue sampling (e.g., sentinel node in melanoma). Some outcomes may require assessment by an independent third party. Depending on the objectives of the registry, collection and storage of biological samples may be considered. (See Section 2.7.3 below.) The collection of biosamples itself is a rapidly evolving field, and registry developers should consult both technical and legal sources regarding how to include biosamples in a registry.

2.7.2 When Data Need To Be Available for Analysis

Meaningful data on disease progression or other long-term patient outcomes may not be available through a registry for many years, whereas safety data could be examined periodically over time. Therefore, the type of data on patient outcomes and when they will be available for analysis should be addressed from the perspective of the intended uses of the data in both the short term and long term. For industry-sponsored registries, if planning begins at an early stage, it may be possible to consider whether to align registry questions with those from the clinical trial (where appropriate) so that some data can carry over for more comprehensive longitudinal analyses.

2.7.3 Scientific Rigor

The content of the data to be collected should be driven by the scientific analyses that are planned for the registry, which, in turn, are determined by the specific objectives of the registry. A registry designed primarily for monitoring safety will contain different data elements from one designed primarily for monitoring effectiveness. Similarly, the extent to which data need to be validated will depend on the purpose of the registry and the complexity of the clinical information being sought. For some outcomes, clinical diagnosis may be sufficient; for others, supporting documents from hospitalizations, referrals, or biopsies may be needed; and for others, formal adjudication by a committee may be required. Generally, registries that are undertaken for regulatory decisionmaking will require increased attention toward diagnostic confirmation (i.e., enhanced scientific rigor).

2.8 Define the Core Data Set, Patient Outcomes, and Target Population

2.8.1 Core Data Set

Elements of data to be included must have potential value in the context of the current scientific and clinical climate and must be chosen by a team of experts, preferably with input from experts in biostatistics and epidemiology. Each data element should relate to the purpose and specific objectives of the registry. Ideally, each data element should address the central questions for which the registry was designed. It is useful to consider the generalizability of the information collected, as appropriate. For example, when seeking information on cost-effectiveness, it may be preferable to collect data on resource utilization rather than actual costs of this utilization, since the broader descriptor can be more easily generalized to other settings and cost structures. While a certain number of "speculative" fields may be

desired to generate and explore hypotheses, these must be balanced against the risk of overburdening sites with capturing superfluous data. A plan for quality assurance should be considered in tandem with developing the core data set.

The core data set variables ("need to know") define the information set needed to address the critical questions for which the registry was created. At a minimum, when calculating the resource needs and overall design of the registry, registry planners must account for these fields. If additional noncore variables ("nice to know") are included, such as more descriptive or exploratory variables, it is important that such data elements align with the goals of the registry and take into account the burden of data collection and entry at the site level. A parsimonious use of "nice to know" variables is important for several reasons.

First, when data elements change, there is a cascade effect on all dependent components of the registry process and outputs. For example, the addition of new data elements may require changes to the data collection system, retraining of site personnel on data definitions and collection practices, adjustments to the registry protocol, and amendment submissions to institutional review boards. Such changes often require additional financial resources. Ideally, the registry would both limit the total number of data elements and include, at the outset, data elements that might change from "nice to know" to "need to know" during the course of the registry. In practice, this is a difficult balance to achieve, so most registries should plan adequate resources to be used for change management.

Second, a registry should avoid attempting to accomplish too many goals, or its burden will outweigh its usefulness to the clinical sites and researchers. Examples exist, however, of registries that serve multiple purposes successfully without overburdening clinicians. (See Case Example 1.)

Third, even "need-to-know" variables can sometimes be difficult to collect reliably (e.g., use of illegal substances) or without substantial burden (e.g., unusual laboratory tests). Even with a limited core data set, feasibility must still be considered. (See Chapter 4)

Fourth, it is useful to consider what data are already available and/or collected and what additional data need to be collected. When determining additional data elements, it is imperative to consider whether the information desired is consistent with general practice or whether it might be more intensive or exceeding usual practice. For some purposes, collecting specific laboratory results or additional visits may be necessary, but could change how the registry is perceived by institutional review boards or ethics committees. The distinction between "interventional" and "observational" is straightforward in terms of random assignment to treatment, but some registries with requirements that exceed a threshold of usual practice—in Europe, for example—may be subject to additional requirements more typical of "interventional" research. In Chapter 1.7.1 of Volume 9A of the Rules Governing Medicinal Products in the European Union,[34] it has been clarified that registries may "collect a battery of information using standardized questionnaires in a prospective fashion" and "questionnaires, by themselves, are not considered interventional." These rules also state that

- "[T]he assignment of a patient to a particular strategy is not decided in advance by a [trial] protocol but falls within the current practice…"
- "[N]o additional diagnostic or monitoring procedures shall be applied to patients."

This last requirement can be challenging to interpret since registries sometimes perform diagnostic tests that are consistent with general practice but that may be performed more frequently than would be the case in general practice. The determination that a registry should be considered "interventional" from a regulatory perspective can add significant burden and cost to the registry program, and, therefore, the tradeoffs must be carefully considered in planning schedules for registry visits and the collection of data and/or specimens.

Finally, it is important to consider patient privacy, national and international rules concerning ethics, and regulatory requirements to assure that the registry data requirements do not jeopardize

patient privacy or put institutional/ethics reviews and approvals at risk.

2.8.2 Patient Outcomes

The outcomes of greatest importance should be identified early in the concept phase of the registry. Delineating these outcomes (e.g., primary or secondary endpoints) will force registry designers to establish priorities. Prioritization of interests in the planning phase will help focus the work of the registry and will guide study size requirements. (See Chapter 3.) Identifying the patient outcomes of the greatest importance will also help to guide the selection of the data set. Avoiding the temptation to collect "nice to know" data that are likely of marginal value is of paramount importance, yet some registries do, in fact, need to collect large amounts of data to accomplish their purposes. Possessing adequate data in order to properly address potential confounders during analyses is one reason that extensive data collection is sometimes required.[35]

Methods to ascertain the principal outcomes should be clearly established. The diagnostic requirements, level of data detail, and level of data validation and/or adjudication should also be addressed. As noted below in the context of identifying a target population, relying on established guidelines and standards to aid in defining outcomes of interest has many benefits and should be considered.

The issues of ascertainment noted here are important to consider because they will have a bearing on some attributes by which registries may be evaluated.[36] These attributes include sensitivity (the extent to which the methods identify all outcomes of interest) and external validity (generalizability to similar populations), among others.

2.8.3 Target Population

The target population is the population to which the findings of the registry are meant to apply. It must be defined for two basic reasons. First, the target population serves as the foundation for planning the registry. Second, it also represents a major constituency that will be impacted by the results of the registry.

One of the goals for registry data may be to enable generalization of conclusions from clinical research on narrowly defined populations to broader ones, and therefore the inclusion criteria for most (although not all) registries are relatively broad. As an example, screening criteria for a registry may allow inclusion of elderly patients, patients with multiple comorbidities, patients on multiple therapies, patients who switch treatments during the period of observation, or patients who are using products "off label." The definition of the target population will depend on many factors (e.g., scope and cost), but ultimately will be driven by the purpose of the registry.

As with defining patient outcomes, target population criteria and/or definitions should be consistent with established guidelines and standards within the therapeutic area. Achieving this goal increases the potential utility of the registry by leveraging other data sources (historical or concurrent) with different information on the same target population and enhancing statistical power if similar information is collected on the target population.

In establishing target population criteria, consideration should be given to the feasibility of access to that population. One should try to distinguish the ideal from the real. Some questions to consider in this regard are:

- How common is the exposure or disease of interest?
- Can eligible people be readily identified?
- Are other sources competing for data on the same patients?
- Is care centralized or dispersed (e.g., in a referral or tertiary care facility)?
- How mobile is the target population?

Ultimately, methods to ascertain members of the target population should be carefully considered (e.g., use of screening logs that identify all potential patients and indicate whether they participate and, if not, why not), as should the use of sources outside the registry (e.g., patient groups). Greater accessibility to the target population will reap benefits in terms of enhanced representativeness and statistical power.

Lastly, thought should be given to comparison (control) groups either internal or external to the registry. Again, much of this consideration will be driven by the purpose and specific objectives of the registry. For example, natural history registries do not need controls, but controls are especially desirable for registries created to evaluate comparative effectiveness or safety.

2.9 Develop a Study Plan or Protocol

The study plan documents the objectives of the registry and describes how those objectives will be achieved. At a minimum, the study plan should include the registry objectives, the eligibility criteria for participants, and the data collection procedures. Ideally, a full study protocol will be developed to document the objectives, design, participant inclusion/exclusion criteria, outcomes of interest, data to be collected, data collection procedures, governance procedures, and plans for complying with ethical obligations and protecting patient privacy.

In addition to a study plan or protocol, registries may have statistical analysis plans. Chapters 13 and 25 discuss the importance of analysis plans.

2.10 Develop a Project Plan

Developing an overall project plan is critically important so that the registry team has a roadmap to guide their collective efforts. Depending on the complexity of the registry project, the project plan may include some or all of the following elements:

- Scope management plan to control the scope of the project. It should provide the approach to making changes to the scope through a clearly defined change-control system.
- Detailed timeline and schedule management plan to ensure that the project and its deliverables are completed on time.
- Cost management plan for keeping project costs within the budget. The cost management plan may provide estimates on cost of labor, purchases and acquisitions, compliance with regulatory requirements, et cetera. This plan should be aligned with the change-control system so that all changes to the scope will be reflected in the cost component of the registry project.
- Quality management plan to describe the procedures to be used to test project concepts, ideas, and decisions in the process of building a registry. Having a quality management plan in place can help in detecting design errors early, formulating necessary changes to the scope, and ensuring that the final product meets stakeholders' expectations.
- Staffing management plan to determine what skills will be needed and when to meet the project goals. (See Chapter 2, Section 2.5).
- Communication plan that includes who is responsible for communicating information and to whom it should be communicated. Considerations include different categories of information, frequency of communications, and methods of communication. The plan should also provide steps to escalate issues that cannot be resolved on a lower staff level.
- Procurement plan for external components or equipment and/or outsourced software development for the planned registry, if pertinent. Such a plan should describe how the procurement process would be managed within the organization. Decisions to procure products or services may have a direct impact on other components of the project plan, including the staffing plan and timeline.
- Risk management plan to identify and mitigate risks. Many project risks are predictable events, and therefore they can and should be assessed in the very early stages of registry planning. It is important to prioritize project risks by their potential impact on the specific objectives and to develop an adequate risk response plan for the most significant risks. Some predictable risks include—
  - Disagreement between stakeholders over the scope of specific tasks.
  - Inaccurate cost estimates.
  - Delays in the timeline.

3. Summary

In summary, planning a patient registry involves several key steps, including articulating its purpose, determining whether it is an appropriate means of addressing the research question, identifying stakeholders, defining the scope and target population, assessing feasibility, and securing funding. A registry team and advisors must be assembled to develop, coordinate, and guide the registry; these individuals should be selected based on their expertise and experience. Governance and oversight for the registry should also be addressed during the planning phase. While registries differ tremendously in size, scope, and resource requirements, the basic elements of planning described here are relevant for most, if not all registries, and can help to support the launch and operation of a successful registry.

Case Example for Chapter 2

Case Example 1. Creating a registry to fulfill multiple purposes and using a publications committee to review data requests

Description	The National Registry of Myocardial Infarction (NRMI) collected, analyzed, and disseminated data on patients experiencing acute myocardial infarction. Its goal was improvement of patient care at individual hospitals through the hospital team's evaluation of data and assessment of care delivery systems.
Sponsor	Genentech, Inc.
Year Started	1990
Year Ended	2006
No. of Sites	451 hospitals in the final phase of NRMI (NRMI 5). Over 2,150 hospitals participated in NRMI over 16 years.
No. of Patients	2,515,106

Challenge

Over the past 20 years, there have been significant changes in the treatment of acute myocardial infarction (AMI) patients. Evidence from large clinical trials has led to the introduction of new guidelines and therapies for treating AMI patients, including fibrinolytic therapy and percutaneous coronary intervention. While these treatments can improve both morbidity and mortality for AMI patients, they are time sensitive and must be administered very soon after hospital arrival in order to be most effective.

After the release of its first fibrinolytic therapy product in 1987, the sponsor's field representatives learned from their discussions with emergency department physicians, cardiologists, and hospital staff that most clinicians believed they were treating patients quickly, although there was no documentation or benchmarking to confirm this assumption or to identify and correct delays. At that time, many emergency departments did not have readily available diagnostic tools (such as angiography labs), and hospitals with AMI-specific decision pathways and treatment protocols were the exception rather than the rule.

In addition, since fibrinolytic therapy was being widely used for the first time, the sponsor wanted to gather safety information related to its use in real-world situations and in a broader range of patients than those treated in the controlled environment of a clinical trial.

Proposed Solution

The sponsor decided to create the registry to fulfill the multiple purposes of identifying treatment patterns, promoting time-to-treatment and other quality improvements, and gathering real-world safety data. The scope of the data collection necessary to meet these needs could have made such a registry impracticable, so the project team faced the sizable challenge of balancing the data needs with the feasibility of the registry.

Case Example 1. Creating a registry to fulfill multiple purposes and using a publications committee to review data requests (continued)

Proposed Solution (continued)

The sponsor formed a scientific advisory board with members representing the various clinical stakeholders (emergency department, cardiology, nursing, research, etc.). The scientific advisory board developed the data set for the registry, keeping a few guiding principles in mind. These principles emphasized maintaining balance between the clinical research and the feasibility of the registry. The first principle was to determine whether the proposed data element was necessary by asking several key questions: How will the data element be used in generating hospital feedback reports or research analyses? Is the data element already collected? If not, should it be collected? If it should be collected, is it feasible to collect those data? The second principle focused on using existing data standards whenever possible. If a data standard did not exist, the team tried to collect the data in the simplest possible way. The third principle emphasized data consistency and making the registry user-friendly by continually refining data element definitions until they were as clear as possible.

In 1990, the sponsor launched the registry. During the 16 years that the registry was conducted, it demonstrated that the advisory board's efforts to create a feasible multipurpose registry were successful. The registry collected data on the clinical presentation, treatment, and outcomes of over 2.5 million patients with AMI from more than 2,150 participating sites.

The success of the registry presented a new challenge for the registry team. The sponsor received a large volume of requests to analyze the registry data, often for research topics that fell outside of the standardized reports developed for the registry. As a guiding principle, the registry team was committed to making the data available for research projects, but it had limited resources. To support these requests, the team developed a process that would allow outside researchers to access the registry data without overburdening the registry team.

The registry team created a publication process to determine when another group could use the data for research. The team set high-level criteria for all data requests: the analysis had to be feasible given the data in the registry, and the request could not represent a duplication of another research effort.

The registry team involved its scientific advisory board, made up of cardiologists, emergency department physicians, nurses, research scientists, pharmacists, and reviewers with specialties in biostatistics and statistical programming, in creating a publication review committee. The review committee evaluated all research proposals to determine originality, interest to peers, feasibility, appropriateness, and priority. The review committee limited its review of research proposals to a set number of reviews per year, and scheduled the reviews and deadlines around the abstract deadlines for the major cardiology conferences. Research analyses had to be intended to result in peer-reviewed presentations and publications. Researchers were asked to submit proposals that included well defined questions and an analysis plan. If the proposal was accepted, the researchers discussed any further details with the biostatisticians and statistical programmers who performed the analyses (and who were employed at an independent clinical research organization). The results were sent directly to the researchers.

The scientific advisory board and review committee remained involved in the process after a data request had been granted. All authors submitted their abstracts to the review committee before sending them to conferences. The review committee offered constructive criticism to help the authors improve their abstracts. The review committee also reviewed manuscripts before journal submission to help identify any issues or concerns that the authors should address.

Case Example 1. Creating a registry to fulfill multiple purposes and using a publications committee to review data requests (continued)

Results

This publication process enabled the wealth of data collected in this registry to be used in over 150 scientific abstracts and 100 peer-reviewed articles, addressing each of the purposes of the registry as well as other research topics. By involving the scientific advisory board and providing independent biostatistical support, the registry team developed an infrastructure that enhanced the credibility of the research uses of this observational database.

Key Point

Registries can be developed to fulfill more than one purpose, but this added complexity requires careful planning to ensure that the final registry data collection burden and procedures are feasible. Making sure that the advisory board includes representatives with clinical and operational perspectives can help the board to maintain its focus on feasibility. As a registry database gains large amounts of data, the registry team will likely receive research proposals from groups interested in using the data. The registry team may want to set up a publication process during the registry design phase.

For More Information

Califf RM. The benefits of moving quality to a national level. Am Heart J. 2008;156(6):1019–22.

Rogers WJ, Frederick PD, Stoehr E, et al. NRMI Investigators; Trends in presenting characteristics and hospital mortality among patients with ST elevation and non-ST elevation myocardial infarction in the NRMI from 1990 to 2006. Am Heart J. 2008;156(6):1026–34.

Gibson CM, Pride YB, Frederick PD. et al. NRMI Investigators; Trends in reperfusion strategies, door-to-needle and door-to-balloon times, and in-hospital mortality among patients with ST-segment elevation myocardial infarction enrolled in the NRMI from 1990 to 2006. Am Heart J. 2008;156(6):1035–44.

Peterson ED, Shah BR, Parsons L. et al. NRMI Investigators; Trends in quality of care for patients with acute myocardial infarction in the NRMI from 1990 to 2006. Am Heart J. 2008;156(6):1045–55.

References for Chapter 2

1. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf. 2008 Feb;17(2):200-8. PMID: 17868186.
2. German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep. 2001 Jul 27;50(RR-13):1-35; quiz CE1-7. PMID: 18634202.
3. Project Management Institute. A Guide to the Project Management Body of Knowledge (PMBOK Guide). 4th ed. Newtown Square, PA: Project Management Institute; 2008.
4. Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. Am J Manag Care. 2010 Jun;16(6):467-71. PMID: 20560690.
5. Dreyer NA, Garner S. Registries for robust evidence. JAMA. 2009 Aug 19;302(7):790-1. PMID: 19690313.
6. Solomon DJ, Henry RC, Hogan JG, et al. Evaluation and implementation of public health registries. Public Health Rep. 1991 Mar-Apr;106(2):142-50. PMID: 1902306. PMCID: 1580226.
7. Glaser SL, Clarke CA, Gomez SL, et al. Cancer surveillance research: a vital subdiscipline of cancer epidemiology. Cancer Causes Control. 2005 Nov;16(9):1009-19. PMID: 16184466.
8. Kennedy L, Craig AM. Global registries for measuring pharmacoeconomic and quality-of-life outcomes: focus on design and data collection, analysis and interpretation. Pharmacoeconomics. 2004;22(9):551-68. PMID: 15209525.

9. Bookman MA. Using tumor registry resources in analyzing concordance with guidelines and outcomes. Oncology (Williston Park). 2000 Nov;14(11A):104-7. PMID: 11195403.

10. Agency for Healthcare Research and Quality. Developing a Registry of Patient Registries (RoPR). http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=690. Accessed August 17, 2012.

11. Avorn J. In defense of pharmacoepidemiology--embracing the yin and yang of drug research. N Engl J Med. 2007 Nov 29;357(22):2219-21. PMID: 18046025.

12. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med. 2008 Mar 11;5(3):e67. PMID: 18336067. PMCID: 2265762.

13. Alter DA, Venkatesh V, Chong A, et al. Evaluating the performance of the Global Registry of Acute Coronary Events risk-adjustment index across socioeconomic strata among patients discharged from the hospital after acute myocardial infarction. Am Heart J. 2006 Feb;151(2):323-31. PMID: 16442894.

14. Hernan MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002 Jan 15;155(2):176-84. PMID: 11790682.

15. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. New York, NY: Springer; 2009.

16. Retchin SM, Wenzel RP. Electronic medical record systems at academic health centers: advantages and implementation issues. Acad Med. 1999 May;74(5):493-8. PMID: 10353279.

17. Andersen MR, Schroeder T, Gaul M, et al. Using a population-based cancer registry for recruitment of newly diagnosed patients with ovarian cancer. Am J Clin Oncol. 2005 Feb;28(1):17-20. PMID: 15685029.

18. Tunis SR. A clinical research strategy to support shared decision making. Health Aff (Millwood). 2005 Jan-Feb;24(1):180-4. PMID: 15647229.

19. INTERMACS. http://www.uab.edu/intermacs/. Accessed August 14, 2012.

20. Wikipedia. Public-Private Partnerships. http://en.wikipedia.org/wiki/Public-private_partnership. Accessed August 14, 2012.

21. Riverside County Department of Public Health. Vaxtrack Immunization Registry. http://www.naccho.org/topics/modelpractices/database/practice.cfm?practiceID=139. Accessed August 14, 2012.

22. California Immunization Registry (CAIR). http://www.ca-siis.org. Accessed August 14, 2012.

23. North Carolina Department of Health and Human Services. N.C Public-Private Partnership Creates Statewide Bioterrorism Surveillance System. http://www.infectioncontroltoday.com/news/2004/07/public-private-partnership-creates-statewide-biot.aspx. Accessed August 14, 2012.

24. Ray WA, Stein CM. Reform of drug regulation--beyond an independent drug-safety board. N Engl J Med. 2006 Jan 12;354(2):194-201. PMID: 16407517.

25. Strom BL. How the US drug safety system should be changed. JAMA. 2006 May 3;295(17):2072-5. PMID: 16670415.

26. Okie S. Safety in numbers--monitoring risk in approved drugs. N Engl J Med. 2005 Mar 24;352(12):1173-6. PMID: 15788493.

27. Rawlins M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions. Clin Med. 2008 Dec;8(6):579-88. PMID: 19149278.

28. Lyratzopoulos G, Patrick H, Campbell B. Registers needed for new interventional procedures. Lancet. 2008 May 24;371(9626):1734-6. PMID: 18502284.

29. American Recovery and Reinvestment Act of 2009, H.R.1. http://www.gpo.gov/fdsys/pkg/BILLS-111hr1enr/pdf/BILLS-111hr1enr.pdf. Accessed August 14, 2012.

30. Food and Drug Administration Amendments Act of 2007., Pub. L. No. 110-85 (2007), Title VI. Sect. 603. Critical Path Public-Private Partnerships.http://www.gpo.gov/fdsys/pkg/PLAW-110publ85/html/PLAW-110publ85.htm. Accessed August 14, 2012.

31. Davidoff F, DeAngelis CD, Drazen JM, et al. Sponsorship, authorship, and accountability. Ann Intern Med. 2001 Sep 18;135(6):463-6. PMID: 11560460.

32. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication. http://www.icmje.org. Accessed August 14, 2012.

33. Woodward M. Sample size determination. Epidemiology: study design and data analysis. Boca Raton (FL): Chapman & Hall/CRC Press; 2005:chapter 8.

34. Guidelines on Pharmacovigilance of Medical Products for Human Use, Volume 9A of the Rules Governing Medicinal Products in the European Union. Sep2008. http://ec.europa.eu/health/documents/eudralex/vol-9/index_en.htm. Accessed August 14, 2012.

35. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006 Jan 26;354(4):353-65. PMID: 16436767.

36. Newton J, Garner S. Disease registers in England. Report commissioned by the Department of Health Policy Research Programme. Institute of Health Sciences. University of Oxford; Feb, 2002. http://www.sepho.org.uk/Download/Public/5445/1/disease_registers_in_england.pdf. Accessed August 27, 2012.

Chapter 3. Registry Design

1. Introduction

This chapter is intended as a high-level practical guide to the application of epidemiologic methods that are particularly useful in the design of registries that evaluate patient outcomes. Since it is not intended to replace a basic textbook on epidemiologic design, readers are encouraged to seek more information from textbooks and scientific articles. Table 3–1 summarizes the key considerations for study design that are discussed in this chapter. Throughout the design process, registry planners may want to discuss options and decisions with the registry stakeholders and relevant experts to ensure that sound decisions are made. The choice of groups to be consulted during the design phase generally depends on the nature of the registry, the registry funding source and funding mechanism, and the intended audience for registry reporting. More detailed discussions of registry design for specific types of registries are provided in Chapters 19, 20, 21, 22, and 23.

Table 3–1. Considerations for study design

Construct	Relevant Questions
Research question	What are the clinical and/or public health questions of interest?
Resources	What resources, in terms of funding, sites, clinicians, and patients, are available for the study?
Exposures and outcomes	How do the clinical questions of interest translate into measurable exposures and outcomes?
Data sources	Where can the necessary data elements be found?
Study design	What types of design can be used to answer the questions or fulfill the purpose?
Study population	What types of patients are needed for study? Is a comparison group needed? How should patients be selected for study?
Sampling	How should the study population be sampled, taking into account the target populations and study design?
Study size and duration	For how long should data be collected, and for how many patients?
Internal and external validity	What are the potential biases? What are the concerns about generalizability of the results (external validity)?

2. Research Questions Appropriate for Registries

The questions typically addressed in registries range from purely descriptive questions aimed at understanding the characteristics of people who develop the disease and how the disease generally progresses, to highly focused questions intended to support decisionmaking. Registries focused on determining clinical effectiveness or cost-effectiveness or assessing safety or harm are generally hypothesis driven and concentrate on evaluating the effects of specific treatments on patient outcomes. Research questions should address the registry's purposes, as broadly described in Table 3–2.

Observational studies derived from registries (or "registry-based studies") are an important part of the research armamentarium alongside interventional studies, such as randomized

controlled trials (RCTs) or pragmatic trials, and retrospective studies, such as studies derived exclusively from administrative claims data. Each of these study designs has strengths and limitations, and the selection of a study design should be guided by the research questions of interest. (See Chapter 2, Section 2.2, for a discussion of the factors that influence the study design decision.) In some cases, multiple studies with different designs or a hybrid study that combines study designs will be necessary to address a research question. In fact, this more comprehensive approach to evidence development is likely to become more common as researchers strive to address multiple questions for multiple stakeholders most efficiently. Observational studies and interventional studies are more complementary than competitive, precisely because some research questions are better answered by one method than the other. Interventional studies are considered by many to provide the highest grade evidence for evaluating whether a drug has the ability to bring about an intended effect in optimal or "ideal world" situations, a concept also known as "efficacy."[1] Observational designs, on the other hand, are particularly well suited for studying broader populations, understanding actual results (e.g., some safety outcomes) in real-world practice (see Case Example 2), and for obtaining more representative quality-of-life information. This is particularly true when the factors surrounding the decision to treat are an important aspect of understanding treatment effectiveness.[2]

Table 3–2. Overview of registry purposes

• Assessing natural history, including estimating the magnitude of a problem; determining the underlying incidence or prevalence rate of a condition; examining trends of disease over time; conducting surveillance; assessing service delivery and identifying groups at high risk; documenting the types of patients served by a health provider; and describing and estimating survival.
• Determining clinical effectiveness, cost-effectiveness, or comparative effectiveness of a test or treatment, including for the purpose of determining reimbursement.
• Measuring or monitoring safety and harm associated with the use of specific products and treatments, including conducting comparative evaluation of safety and effectiveness.
• Measuring or improving quality of care, including conducting programs to measure and/or improve the practice of medicine and/or public health.

In many situations, nonrandomized comparisons either are sufficient to address the research question or, in some cases, may be necessary because of the following issues with randomizing patients to a specific treatment:

- *Equipoise:* Can providers ethically introduce randomization between treatments when the treatments may not be clinically equivalent?
- *Ethics:* If reasonable suspicion about the safety of a product has become known, would it be ethical to conduct a trial that deliberately exposes patients to potential harm? For example, can pregnant women be ethically exposed to drugs that may be teratogenic? (See Chapter 21 and Case Examples 49, 50, 51, and 52.)
- *Practicality:* Will patients enroll in a study where they might not receive the treatment, or might not receive what is likely to be the best treatment? How can compliance and adherence to a treatment be studied, if not by observing what people do in real-world situations?

Registries are particularly suitable for some types of research questions, such as:

- Natural history studies where the goal is to observe clinical practice and patient experience but not to introduce any intervention.

- Measures of clinical effectiveness, especially as related to compliance, where the purpose is to learn about what patients and practitioners actually do and how their actions affect real-world outcomes. This is especially important for treatments that have poor compliance.
- Studies of effectiveness and safety for which clinician training and technique are part of the study of the treatment (e.g., a procedure such as placement of carotid stent).
- Studies of heterogeneous patient populations, since unlike randomized trials, registries generally have much broader inclusion criteria and fewer exclusion criteria. These characteristics lead to studies with greater generalizability (external validity) and may allow for assessment of subgroup differences in treatment effects.
- Followup for delayed or long-term benefits or harm, since registries can extend over much longer periods than most clinical trials (because of their generally lower costs to run and lesser burden on participants).
- Surveillance for rare events or of rare diseases.
- Studies for treatments in which randomization is unethical, such as intentional exposure to potential harm (as in safety studies of marketed products that are suspected of being harmful).
- Studies for treatments in which randomization is not necessary, such as when certain therapies are only available in certain places owing to high cost or other restrictions (e.g., proton beam therapy).
- Studies for which blinding is challenging or unethical (e.g., studies of surgical interventions, acupuncture).
- Studies of rapidly changing technology.
- Studies of conditions with complex treatment patterns and treatment combinations.
- Studies of health care access and barriers to care.
- Evaluations of actual standard medical practice. (See Case Example 58.)

Registry studies may also include embedded substudies as part of their overall design. These substudies can themselves have various designs (e.g., highly detailed prospective data collection on a subset of registry participants, or a case-control study focused on either incident or prevalent cases identified within the registry). (See Case Examples 3 and 47.) Registries can also be used as sampling frames for RCTs.

3. Translating Clinical Questions Into Measurable Exposures and Outcomes

The specific clinical questions of interest in a registry will guide the definitions of study subjects, exposure, and outcome measures, as well as the study design, data collection, and analysis. In the context of registries, the term "exposure" is used broadly to include treatments and procedures, health care services, diseases, and conditions.

The clinical questions of interest can be defined by reviewing published clinical information, soliciting experts' opinions, and evaluating the expressed needs of the patients, health care providers, and payers. Examples of research questions, key outcome and exposure variables, and sources of data are shown in Table 3–3. As these examples show, the outcomes (generally beneficial or deleterious outcomes) are the main endpoints of interest posed in the research question. These typically represent measures of health or onset of illness or adverse events, but also commonly include quality of life measures, and measures of health care utilization and costs.

Relevant exposures also derive from the main research question and relate to why a patient might experience benefit or harm. Evaluation of an exposure includes collection of information that affects or augments the main exposure, such as dose, duration of exposure, route of exposure, or adherence. Other variables of interest include independent risk factors for the outcomes of interest (e.g., comorbidities, age), as well as variables known as potential confounding variables, that are related to both the exposure and the outcome and are necessary for conducting

valid statistical analyses. Confounding can result in inaccurate estimates of association between the study exposure and outcome through mixing of effects. For example, in a study of asthma medications, prior history of treatment resistance should be collected or else results may be biased. The bias could occur because treatment resistance may relate both to the likelihood of receiving the new drug (meaning that doctors will be more likely to try a new drug in patients who have failed other therapies) and the likelihood of having a poorer outcome (e.g., hospitalization). Refer to Chapter 4 for a discussion of selecting data elements and Chapter 5 for a discussion of selecting patient-reported outcomes.

Table 3–3. Examples of research questions and key exposures and outcomes

Research Question	Key Exposure (source of data)	Key Outcome (source of data)
What is the expected time to rejection for first kidney transplants among adults, and how does that differ according to immunosuppressive regimen?	All immunosuppressants, including dosage and duration (clinician or medical record)	Organ rejection (clinician or medical record)
Are patients using a particular treatment better able to perform activities of daily living than others?	Treatments for disease of interest (clinician or medical record)	Ability to independently perform key activities related to daily living (patient)
Do patients undergoing gastric bypass surgery for weight loss use fewer health care resources in the year following surgery?	Surgery (clinician or medical record)	Number of inpatient and outpatient visits, medications dispensed, associated costs (administrative databases, clinician, or medical record)
Are patients using a particular drug more likely to have serious adverse pregnancy outcomes?	Drug use by mother during pregnancy (clinician, medical record, or patient)	Pregnancy outcome (clinician, medical record, or patient)

4. Finding the Necessary Data

The identification of key outcome and exposure variables and patients will drive the strategy for data collection, including the choice of data sources. A key challenge to registries is that it is generally not possible to collect all desired data. As discussed in Chapter 4, data collection should be both parsimonious and broadly applicable. For example, while experimental imaging studies may provide interesting data, if the imaging technology is not widely available, the data will not be available for enough patients to be useful for analysis. Moreover, the registry findings will not be generalizable if only sophisticated centers that have such technology participate. Instead, registries should focus on collecting relevant data with relatively modest burden on patients and clinicians. Registry data can be obtained from patients, clinicians, medical records, and linkage with other sources (in particular, extant databases), depending on the available budget. (See Chapters 6, 15, and 16.)

Examples of patient-reported data include health-related quality of life; utilities (i.e., patient preferences); symptoms; use of over-the-counter (OTC), complementary, and alternative medication; behavioral data (e.g., smoking and alcohol use); family history; and biological specimens. These data may rely on the subjective interpretation and reporting of the patient (e.g., health-related quality of life, utilities, symptoms such as pain or fatigue); may be difficult to otherwise track (e.g., use of complementary and alternative medication, smoking, and alcohol use);

or may be unique to the patient (e.g., biological specimens). Health care resource utilization is another important construct that reflects both cost of care (burden of illness) and health-related quality of life. For example, more frequent office visits, procedures, or hospitalizations may result in reduced health-related quality of life for the patient. The primary advantage of this form of data collection is that it provides direct information from the entity that is ultimately of the most interest—the patient. The primary disadvantages are that the patient is not necessarily a trained observer and that various forms of bias, such as recall bias, may influence subjective information. For example, people may selectively recall certain exposures because they believe they have a disease that was caused by that exposure, or their recall may be influenced by recent news stories claiming cause-and-effect relationships. (See Case Example 4.)

Examples of clinician data include clinical impressions, clinical diagnoses, clinical signs, differential diagnoses, laboratory results, and staging. The primary advantage of clinician data is that clinicians are trained observers. Even so, the primary disadvantages are that clinicians are not necessarily accurate reporters of patient perceptions, and their responses may also be subject to recall bias. Moreover, the time that busy clinicians can devote to registry data collection is often limited.

Medical records also are a repository of clinician-derived data. Certain data about treatments, risk factors, and effect modifiers are often not consistently captured in medical records of any type, but where available, can be useful. Examples of such data that are difficult to find elsewhere include OTC medications, smoking and alcohol use, complementary and alternative medicines, and counseling activities by the clinician on lifestyle modifications. Medical records are often relied upon as a source of detailed clinical information for adjudication by external reviewers of medical diagnoses corresponding to study endpoints.

Electronic medical records, increasingly available, improve access to the data within medical records. The increasing use of electronic health records has facilitated the development of a number of registries within large health plans. Kaiser Permanente has created several registries of patients receiving total joint replacement, bariatric surgery, and nonsurgical conditions (e.g., diabetes), all of which rely heavily on existing electronic health record data. As discussed further in Chapter 15, the availability of medical records data in electronic format does not, by itself, guarantee consistency of terminology and coding.

Examples of other data sources include health insurance claims, pharmacy data, laboratory data, other registries, and national data sets, such as Medicare claims data and the National Death Index. These sources can be used to supplement registries with data that may otherwise be difficult to obtain, subject to recall bias, not collected because of loss to followup, or likely inaccurate by self-report (e.g., in those patients with diseases affecting recall, cognition, or mental status). See Table 6–1 (in Chapter 6) for more information on data sources.

5. Resources and Efficiency

Ideally, a study is designed to optimally answer a research question of interest and funded adequately to achieve the objectives based on the requirements of the design. Frequently, however, finite resources are available at the outset of a project that constrain the approaches that may be pursued. Often, through efficiencies in the selection of a study design and patient population (observational vs. RCT, case-control vs. prospective cohort), selection of data sources (e.g., medical-records–based studies vs. information collected directly from clinicians or patients), restriction of the number of study sites, or other approaches, studies may be planned that provide adequate evidence for addressing a research question, in spite of limited resources.

Section 6 below discusses how certain designs may be more efficient for addressing some research questions.

6. Study Designs for Registries

Although studies derived from registries are, by definition, observational studies, the framework for how the data will be analyzed drives the data collection and choices of patients for inclusion in the study.

The study models of case series, cohort, case-control, and case-cohort are commonly applied to registry data and are described briefly here. When case-control or case-cohort designs are applied to registry data, additional data may be collected to facilitate examination of questions that arise. Before adding new data elements, whether in a nested substudy or for a new objective, several of the steps outlined in Chapter 2, including assessing feasibility, considering the necessary scope and rigor, and evaluating the regulatory/ethical impact, should be undertaken. Other models that are also useful in some situations, but are not covered here, include: case-crossover studies, which are efficient designs for studying the effects of intermittent exposures (e.g., use of erectile dysfunction drugs) on conditions with sudden onset, and quasi-experimental studies or "pragmatic trials." For example, in a pragmatic trial, providers may be randomized as to which intervention or quality improvement tools they use, but patients are observed without further intervention. Also, there has been recent interest in applying the concept of adaptive clinical trial design to registries. An adaptive design has been defined as a design that allows adaptations or modifications to some aspects of a clinical trial after its initiation without undermining the validity and integrity of the trial.[3] While many long-term registries are modified after initiation, the more formal aspects of adaptive trial design have yet to be applied to registries and observational studies.

Determining what framework will be used to analyze the data is important in designing the registry and the registry data collection procedures. Readers are encouraged to consult textbooks of epidemiology and pharmacoepidemiology for more information. Many of the references in Chapters 13 and 18 relate to study design and analysis.

6.1 Case Series Design

Using a registry population to develop case series is a straightforward application that does not require sophisticated analytics. Depending on the generalizability of the registry itself, case series drawn from the registry can be used to describe the characteristics to be used in comparison to other case series (e.g., from spontaneous adverse event reports). Self-controlled methods, including self-controlled case series, are a relatively new set of methods that lends itself well to registry analyses as it focuses on only those subjects who have experienced the event of interest and uses an internal comparison to derive the relative (not absolute) incidence of the event during the time the subject is "exposed" compared with the incidence during the time when they are "unexposed."[4] This design implicitly controls for all confounders that do not vary over the followup time (e.g., gender, genetics, geographic area), as the subject serves as his or her own control. The self-controlled case series design may also be very useful in those circumstances where a comparison group is not available. Self-controlled case series require that the probability of exposure is not affected by the occurrence of an outcome; in addition, for non-recurrent events, the method works only when the event risk is small and varies over the followup time. Derivative methods, grouped as self-controlled cohort methods, include observational screening[5] and temporal pattern discovery.[6] These methods compare the rate of events post-exposure with the rate of events pre-exposure among patients with at least one exposure.

6.2 Cohort Design

Cohort studies follow, over time, a group of people who possess a characteristic, to see if individuals in the group develop a particular endpoint or outcome. The cohort design is used for descriptive studies as well as for studies seeking to evaluate comparative effectiveness and/or safety or quality of care. Cohort studies may include only people with exposures (such as to a particular drug or class of drugs) or disease of interest. Cohort studies may also include one or more comparison groups for which data are collected using the same

methods during the same period. A single cohort study may in fact include multiple cohorts, each defined by a common disease or exposure. Cohorts may be small, such as those focused on rare diseases, but often they target large groups of people (e.g., in safety studies), such as all users of a particular drug or device. Some limitations of registry-based cohort studies may include limited availability of treatment data and underreporting of outcomes if a patient leaves the registry or is not adequately followed up.[7] These pitfalls should be considered and addressed when planning a study.

6.3 Case-Control Design

A case-control study gathers patients who have a particular outcome or who have suffered an adverse event ("cases") and "controls" who have not but are representative of the source population from which the cases arise.[8] If properly designed and conducted, it should yield results similar to those expected from a cohort study of the population from which the cases were derived. The case-control design is often employed for understanding the etiology of rare diseases[9] because of its efficiency. In studies where expensive data collection is required, such as some genetic analyses or other sophisticated testing, the case-control design is more efficient and cost effective than a cohort study because a case-control design collects information only from cases and a sample of noncases. However, if no de novo data collection is required, the use of the cohort design may be preferable since it avoids the challenge of selecting a suitable control group and the concomitant danger of introducing more bias.

Depending on the outcome or event of interest, cases and controls may be identifiable within a single registry. For example, in the evaluation of restenosis after coronary angioplasty in patients with end-stage renal disease, investigators identified both cases and controls from an institutional percutaneous transluminal coronary angioplasty registry; in this example, controls were randomly selected from the registry and matched by age and gender.[10] Alternatively, cases can be identified in the registry and controls chosen from outside the registry. Care must be taken, however, that the controls from outside the registry meet the requirement of arising from the same source population as the cases to which they will be compared. Matching in case-control designs—for example, ensuring that patient characteristics such as age and gender are similar in the cases and their controls—may yield additional efficiency, in that a smaller number of subjects may be required to answer the study question with a given power. However, matching does not eliminate confounding and must be undertaken with care. Matching variables must be accounted for in the analysis, because a form of selection bias similar to confounding will have been introduced by the matching.[11]

Properly executed, a case-control study can add efficiency to a registry if more extensive data are collected by the registry only for the smaller number of subjects selected for the case-control study. This design is sometimes referred to as a "nested" case-control study, since subjects are taken from a larger cohort. It is generally applied because of budgetary or logistical concerns relating to the additional data desired. Nested case-control studies have been conducted in a wide range of patient registries, from studying the association between oral contraceptives and various types of cancer using the Surveillance Epidemiology and End Results (SEER) program[12-14] to evaluating the possible association of depression with Alzheimer's disease. As an example, in the latter case-control study design, probable cases were enrolled from an Alzheimer's disease registry and compared with randomly selected nondemented controls from the same base population.[15]

Case-control studies present special challenges with regard to control selection. More information on considerations and strategies can be found in a set of papers by Wacholder.[16-18]

6.4 Case-Cohort Design

The case-cohort design is a variant of the case-control study. As in a case-control study, a case-cohort study enrolls patients who have a particular outcome or who have suffered an adverse event ("cases"), and "controls" who have not, but who are representative of the source population from

which the cases arise. In nested case-control studies where controls are selected via risk-set sampling, each person in the source population has a probability of being selected as a control that is, ideally, in proportion to his or her person-time contribution to the cohort. In a case-cohort study, however, each control has an equal probability of being sampled from the source population.[19] This allows for collection of pertinent data for cases and for a sample of the full cohort, instead of the whole cohort. For example, in a case-cohort study of histopathologic and microbiological indicators of chorioamnionitis, which included identification of specific microorganisms in the placenta, cases consisted of extreme preterm infants with cerebral palsy. Controls, which can be thought of as a randomly selected subcohort of subjects at risk of the event of interest, were selected from among all infants enrolled in a long-term study of preterm infants.[20]

With the assumptions that competing risks and loss to followup are not associated with the exposure or the risk of disease, the case-cohort design allows for the selection of one control group that can be compared with various case series since the controls are selected at the beginning of followup. Analogous to a cohort study where every subject in the source population is at risk for the disease at the start of followup, the control series in a case-cohort design represents a sample of the exposed and unexposed in the source population who are disease-free at the start of followup.

7. Choosing Patients for Study

The purpose of a registry is to provide information or describe events and patterns, and often to generate hypotheses about a specific patient population to whom study results are meant to apply. Studies can be conducted of people who share common characteristics, with or without the inclusion of comparison groups. For example, studies can be conducted of:

- People with a particular disease/outcome or condition. (These are focused on characteristics of the person.)
  - Examples include studies of the occurrence of cancer or rare diseases, pregnancy outcomes, and recruitment pools for clinical trials.
- Those with a particular exposure. (These exposures may be to a product, procedure, or other health service.)
  - Examples include general surveillance registries, pregnancy registries for particular drug exposures, and studies of exposure to medications and to devices such as stents.[21] They also include studies of people who were treated under a quality improvement program, as well as studies of a particular exposure that requires controlled distribution, such as drugs with serious safety concerns (e.g., isotretinoin, clozapine, natalizumab [Tysabri®]), where the participants in the registry are identified because of their participation in a controlled distribution/risk management program.
- Those who were part of a program evaluation, disease management effort, or quality improvement project.
  - An example is the evaluation of the effectiveness of evidence-based program guidelines on improving treatment.

7.1 Target Population

Selecting patients for registries can be thought of as a multistage process that begins with understanding the target population (the population to which the findings are meant to apply, such as all patients with a disease or a common exposure) and then selecting a sample of this population for study. Some registries will enroll all, or nearly all, of the target population, but most registries will enroll only a sample of the target population. The accessible study population is that portion of the target population to which the participating sites have access. The actual study population is the subset of those who can actually be identified and invited and who agree to participate.[22] While it is desirable for the patients who participate in a study to be representative of the target population, it is rarely possible to study groups that are fully representative from a

statistical sampling perspective, either for budgetary reasons or for reasons of practicality. An exception is registries composed of all users of a product (as in postmarketing surveillance studies where registry participation is required as a condition of receiving an intervention), an approach which is becoming more common to manage expensive interventions and/or to track potential safety issues.

Certain populations pose greater difficulties in assembling an actual study population that is truly representative of the target population. Children and other vulnerable populations present special challenges in recruitment, as they typically will have more restrictions imposed by institutional review boards and other oversight groups.

As with any research study, very clear definitions of the inclusion and exclusion criteria are necessary and should be well documented, including the rationale for these criteria. A common feature of registries is that they typically have few inclusion and exclusion criteria, which enhances their applicability to broader populations. Restriction, the strategy of limiting eligibility for entry to individuals within a certain range of values for a confounding factor, such as age, may be considered in order to reduce the effect of a confounding factor when it cannot otherwise be controlled, but this strategy may reduce the generalizability of results to other patients.

These criteria will largely be driven by the study objectives and any sampling strategy. For a more detailed description of target populations and their subpopulations, and how these choices affect generalizability and interpretation, see Chapter 13.

Once the patient population has been identified, attention shifts to selecting the institutions and providers from which patients will be selected. For more information on recruiting patients and providers, see Chapter 10.

7.2 Comparison Groups

Once the target population has been selected and the mechanism for their identification (e.g., by providers) is decided, the next decision involves determining whether to collect data on comparators (sometimes called parallel cohorts). Depending on the purpose of the registry, internal, external, or historical groups can be used to strengthen the understanding of whether the observed effects are real and in fact different from what would have occurred under other circumstances. Comparison groups are most useful in registries where it is important to distinguish between alternative decisions or to assess differences, the magnitude of differences, or the strength of associations between groups. Registries without comparison groups can be used for descriptive purposes, such as characterizing the natural history of a disease or condition, or for hypothesis generation. The addition of a comparison group may add significant complexity, time, and cost to a registry.

Although it may be appealing to use more than one comparison group in an effort to overcome the limitations that may result from using a single group, multiple comparison groups pose their own challenges to the interpretation of registry results. For example, the results of comparative safety and effectiveness evaluations may differ depending on the comparison group used. Generally, it is preferable to make judgments about the "best" comparison group for study during the design phase and then concentrate resources on these selected subjects. Alternatively, sensitivity analyses can be used to test inferences against alternative reference groups to determine the robustness of the findings. (See Chapter 13, Section 5.)

The choice of comparison groups is more complex in registries than in clinical trials. Whereas clinical trials use randomization to try to achieve an equal distribution of known and unknown risk factors that can confound the drug-outcome association, registry studies need to use various design and analytic strategies to control for the confounders that they have measured. The concern for observational studies is that people who receive a new drug or device have different risk factors for adverse events than those who choose other treatments or receive no treatment at all. In other words, the treatment choices are often related to demographic and lifestyle characteristics and the presence of coexisting conditions that affect clinician decisionmaking about whom to treat.[23]

One design strategy that is used frequently to ensure comparability of groups is individual matching of exposed patients and comparators with regard to key demographic factors such as age and gender. Compatibility is also achieved by inclusion criteria that could, for example, restrict the registry focus to patients who have had the disease for a similar duration or are receiving their first drug treatment for a new condition. These inclusion criteria make the patient groups more similar but may add constraints to the external validity by defining the target population more narrowly. Other design techniques include matching study subjects on the basis of a large number of risk factors, by using statistical techniques (e.g., propensity scoring) to create strata of patients with similar risks. As an example, consider a recent study of a rare side effect in coronary artery surgery for patients with acute coronary syndrome. In this instance, the main exposure of interest was the use of antifibrinolytic agents during revascularization surgery, a practice that had become standard for such surgeries. The sickest patients, who were most likely to have adverse events, were much less likely to be treated with antifibrinolytic agents. To address this, the investigators measured more than 200 covariates (by drug and outcome) per patient and used this information in a propensity score analysis. The results of this large-scale observational study revealed that the traditionally accepted practice (aprotinin) was associated with serious end-organ damage and that the less expensive generic medications were safe alternatives.[24] Incorporation of propensity scores in analysis is discussed further in Chapter 13, Section 5.

An internal comparison group refers to simultaneous data collection for patients who are similar to the focus of interest (i.e., those with a particular disease or exposure in common), but who do not have the condition or exposure of interest. For example, a registry might collect information on patients with arthritis who are using acetaminophen for pain control. An internal comparison group could be arthritis patients who are using other medications for pain control. Data regarding similar patients, collected during the same calendar period and using the same data collection methods, are useful for subgroup comparisons, such as for studying the effects in certain age categories or among people with similar comorbidities. However, the information value and utility of these comparisons depend largely on having adequate sample sizes within subgroups, and such analyses may need to be specified a priori to ensure that recruitment supports them. Internal comparisons are particularly useful because data are collected during the same observation period as for all study subjects, which will account for time-related influences that may be external to the study. For example, if an important scientific article is published that affects general clinical practice, and the publication occurs during the period in which the study is being conducted, clinical practice may change. The effects may be comparable for groups observed during the same period through the same system, whereas information from historical comparisons, for example, would be expected to reflect different practices.

An external comparison group is a group of patients similar to those who are the focus of interest, but who do not have the condition or exposure of interest, and for whom relevant data that have been collected outside of the registry are available. For example, the SEER program maintains national data about cancer and has provided useful comparison information for many registries where cancer is an outcome of interest.[25] External comparison groups can provide informative benchmarks for understanding effects observed, as well as for assessing generalizability. Additionally, large clinical and administrative claims databases can contribute useful information on comparable subjects for a relatively low cost. A drawback of external comparison groups is that the data are generally not collected the same way and the same information may not be available. The underlying populations may also be different from the registry population. In addition, plans to merge data from other databases require the proper privacy safeguards to comply with legal requirements for patient data; Chapter 7 covers patient privacy rules in detail.

A historical comparison group refers to patients who are similar to the focus of interest, but who do not have the condition or exposure of interest, and

for whom information was collected in the past (such as before the introduction of an exposure or treatment or development of a condition). Historical controls may actually be the same patients who later become exposed, or they may consist of a completely different group of patients. For example, historical comparators are often used for pregnancy studies since there is a large body of population-based surveillance data available, such as the Metropolitan Atlanta Congenital Defects Program (MACDP).[26] This design provides weak evidence because symmetry is not assured (i.e., the patients in different time periods may not be as similar as desired). Historical controls are susceptible to bias by changes over time in uncontrollable, confounding risk factors, such as differences in climate, management practices, and nutrition. Bias stemming from differences in measuring procedures over time may also account for observed differences.

An approach related to the use of historical comparisons is the use of Objective Performance Criteria (OPC) as a comparator. This research method has been described as an alternative to randomized trials, particularly for the study of devices.[27] OPC are "performance criteria based on broad sets of data from historical databases (e.g., literature or registries) that are generally recognized as acceptable values. These criteria may be used for surrogate or clinical endpoints in demonstrating the safety or effectiveness of a device."[28] A U.S. Food and Drug Administration guidance document on medical devices includes a description of study designs that should be considered as alternatives to randomized clinical trials and that may meet the statutory criteria for preapproval as well as postapproval evidence.[28] Registries serve as a source of reliable historical data in this context. New registries with safety or effectiveness endpoints may also be planned that will incorporate previously existing OPC as comparators (e.g., for a safety endpoint for a new cardiac device). Such registries might use prior clinical study data to set the "complication-free rate" for comparison.

There are several situations in which conventional prospective design for comparison selection is impossible and a historical comparison may be considered:

- When one cannot ethically continue the use of older treatments or practices, or when clinicians and/or patients refuse to continue their use, so that the researcher cannot identify relevant sites using the older treatments.
- When uptake of a new medical practice has been rapid, concurrent comparisons may differ so markedly from treated patients, in regard to factors related to outcomes of interest, that they cannot serve as valid comparison subjects due to intractable confounding.
- When conventional treatment has been consistently unsuccessful and the effect of new intervention is obvious and dramatic (e.g., first use of a new product for a previously untreatable condition).
- When collecting the comparison data is too expensive.
- When the Hawthorne effect (a phenomenon that refers to changes in the behavior of subjects because they know they are being studied or observed) makes it impossible to replicate actual practice in a comparison group during the same period.
- When the desired comparison is to usual care or "expected" outcomes at a population level, and data collection is too expensive due to the distribution or size of that population.

8. Sampling

Various sampling strategies for patients and sites can be considered. Each of these has tradeoffs in terms of validity and information yield. The representativeness of the sample, with regard to the range of characteristics that are reflective of the broader target population, is often a consideration, but representativeness mainly affects generalizability rather than the internal validity of the results. Representativeness should be considered in terms of patients (e.g., men and women, children, the elderly, different racial or ethnic groups) and sites (academic medical centers, community practices). For sites (health

care providers, hospitals, etc.), representativeness is often considered in terms of geography, practice size, and academic or private practice type. Reviewing and refining the research question can help researchers define an appropriate target population and a realistic strategy for subject selection.

To ensure that enough meaningful information will be available for analysis, registry studies often restrict eligibility for entry to individuals within a certain range of characteristics. Alternatively, they may use some form of sampling: random selection, systematic sampling, or a nonrandom approach. Often-used sampling strategies include the following:

- Probability sampling: Some form of random selection is used, wherein each person in the population must have a known (often equal) probability of being selected.[29-32]
  - Census: A census sample includes every individual in a population or group (e.g., all known cases). A census is not feasible when the group is large relative to the costs of obtaining information from individuals.
  - Simple random sampling: The sample is selected in such a way that each person has the same probability of being sampled.
  - Stratified random sampling: The group from which the sample is to be taken is first stratified into subgroups on the basis of an important, related characteristic (e.g., age, parity, weight) so that each individual in a subgroup has the same probability of being included in the sample, but the probabilities for different subgroups or strata are different. Stratified random sampling ensures that the different categories of characteristics that are the basis of the strata are sufficiently represented in the sample. However, the resulting data must be analyzed using more complicated statistical procedures (such as Mantel-Haenszel) in which the stratification is taken into account.
  - Systematic sampling: Every nth person in a population is sampled.
  - Cluster (area) sampling: The population is divided into clusters, these clusters are randomly sampled, and then some or all patients within selected clusters are sampled. This technique is particularly useful in large geographic areas or when cluster-level interventions are being studied.
  - Multistage sampling: Multistage sampling can include any combination of the sampling techniques described above.
- Nonprobability sampling: Selection is systematic or haphazard but not random. The following sampling strategies affect the type of inferences that can be drawn; for example, it would be preferable to have a random sample if the goal were to estimate the prevalence of a condition in a population. However, systematic sampling of "typical" patients can generate useful data for many purposes, and is often used in situations where probability sampling is not feasible.[33]
  - Case series or consecutive (quota) sampling: All consecutive eligible patients treated at a given practice or by a given clinician are enrolled until the enrollment target is reached. This approach is intended to reduce conscious or unconscious selection bias on the part of clinicians as to whom to enroll in the study, especially with regard to factors that may be related to prognosis.
  - Haphazard, convenience, volunteer, or judgmental sampling: This includes any sampling not involving a truly random mechanism. A hallmark of this form of sampling is that the probability that a given individual will be in the sample is unknown before sampling. The theoretical basis for statistical inference is lost, and the result is inevitably biased in unknown ways.
  - Modal instance: The most typical subject is sampled.
  - Purposive: Several predefined groups are deliberately sampled.
  - Expert: A panel of experts judges the representativeness of the sample or is the source that contributes subjects to a registry.

Individual matching of cases and controls is sometimes used as a sampling strategy for controls. Controls are matched with individual cases who have similar confounding factors, such as age, to reduce the effect of the confounding factors on the association being investigated.

Patients may be recruited in a fashion that allows for individual matching. For example, if a 69-year-old "case" participates in the registry, a control near in age will be sought. Individual matching for prospective recruitment is challenging and not customarily used. More often, matching is used to create subgroups for supplemental data collection for case-control studies and cohort studies when subjects are limited and/or stratification is unlikely to provide enough subjects in each stratum for meaningful evaluation.

A number of other sampling strategies have arisen from survey research (e.g., snowball, heterogeneity), but they are of less relevance to registries.

9. Registry Size and Duration

Precision in measurement and estimation corresponds to the reduction of random error; it can be improved by increasing the size of the study and modifying the design of the study to increase the efficiency with which information is obtained from a given number of subjects.[29]

During the registry design stage, it is critical to explicitly state how large the registry will be, how long patients should be followed, and what the justifications are for these decisions. These decisions are based on the overall purpose of the registry. For example, in addressing specific questions of product safety or effectiveness, the desired level of precision to confirm or rule out the existence of an important effect should be specified, and ideally should be linked to policy or practice decisions that will be made based on the evidence. For registries with aims that are descriptive or hypothesis generating, study size may be arrived at through other considerations.

The duration of registry enrollment and followup should be determined both by required sample size (number of patients or person-years to achieve the desired power) and by time-related considerations. The induction period for some outcomes of interest must be considered, and sufficient followup time allowed for the exposure under study to have induced or promoted the outcome. Biological models of disease etiology and causation usually indicate the required time period of observation for an effect to become apparent. Calendar time may be a consideration in studies of changes in clinical practice or interventions that have a clear beginning and end. The need for evidence to inform policy may also determine a timeframe within which the evidence must be made available to decisionmakers.

A detailed discussion of the topic of sample size calculations for registries is provided in Appendix A. For present purposes it is sufficient to briefly describe some of the critical inputs to these calculations that must be provided by the registry developers:

- The expected timeframe of the registry and the time intervals at which analyses of registry data will be performed.
- Either the size of clinically important effects (e.g., minimum clinically important differences) or the desired precision associated with registry-based estimates.
- Whether or not the registry is intended to support regulatory decisionmaking. If the results from the registry will affect regulatory action—for example, the likelihood that a product may be pulled from the market—then the precision of the overall risk estimate is important, as is the necessity to predict and account for attrition.

In a classical calculation of sample size, the crucial inputs that must be provided by the investigators include either the size of clinically important effects or their required precision. For example, suppose that the primary goal of the registry is to compare surgical complication rates in general practice with those in randomized trials. The inputs to the power calculations would include the complication rates from the randomized trials (e.g., 4 percent) and the complication rate in general practice, which would reflect a meaningful

departure from this rate (e.g., 6 percent). If, on the other hand, the goal of the registry is simply to track complication rates (and not to compare the registry with an external standard), then the investigators should specify the required width of the confidence interval associated with those rates. For example, in a large registry, the 95-percent confidence interval for a 5-percent complication rate might extend from 4.5 percent to 5.5 percent. If all of the points in this confidence interval lead to the same decision, then an interval of ±0.5 percent is considered sufficiently precise, and this is the input required for the estimation of sample size.

Specifying the above inputs to sample size calculations is a substantial matter and usually involves a combination of quantitative and qualitative reasoning. The issues involved in making this specification are essentially similar for registries and other study designs, though for registries designed to address multiple questions of interest, one or more primary objectives or endpoints must be selected that will drive the selection of a minimum sample size to meet those objectives.

Other considerations that should sometimes be taken into account when estimating sample sizes include—

- whether individual patients can be considered "independent," or whether they share factors that would lead to correlation in measures between them;
- whether multiple comparisons are being made and subjected to statistical testing; and
- whether levels of expected attrition or lack of adherence to therapy may require a larger number of patients to achieve the desired number of person-years of followup or exposure.

In some cases, patients under study who share some group characteristics, such as patients treated by the same clinician or practice, or at the same institution, may not be entirely independent from one another with regard to some outcomes of interest or when studying a practice-level intervention. To the extent they are not independent, a measure of interdependence, the intraclass correlation (ICC), and so-called "design effect" must be considered in generating the overall sample size calculation. A reference addressing sample size considerations for a study incorporating a cluster-randomized intervention is provided.[34] A hierarchical or multilevel analysis may be required to account for one or more levels of "grouping" of individual patients, discussed further in Chapter 13, Section 5. One approach to addressing multiple comparisons in the surgical complication rate example above is to use control chart methodology, a statistical approach used in process measurement to examine the observed variability and determine whether out-of-control conditions are occurring. Control chart methodology is also used in sample size estimation, largely for studies with repeated measurements, to adjust the sample size as needed and therefore maintain reasonably precise estimates of confidence limits around the point estimate. Accordingly, for registries that involve ongoing evaluation, sample size per time interval could be determined by the precision associated with the related confidence interval, and decision rules for identifying problems could then be based on control chart methodology.

Although most of the emphasis in estimating study size requirements is focused on patients, it is equally important to consider the number of sites needed to recruit and retain enough patients to achieve a reasonably informative number of person-years for analysis. Many factors are involved in estimating the number of sites needed for a given study, including the number of eligible patients seen in a given practice during the relevant time period, desired representativeness of sites with regard to geography, practice size, or other features, and the timeframe within which study results are required, which may also limit the timeframe for patient recruitment.

In summary, the aims of a registry, the desired precision of information sought, and the hypotheses to be tested, if any, determine the process and inputs for arriving at a target sample size and specifying the duration of followup.

Registries with mainly descriptive aims, or those that provide quality metrics for clinicians or medical centers, may not require the choice of a target sample size to be arrived at through power calculations. In either case, the costs of obtaining study data, in monetary terms and in terms of researcher, clinician, and patient time and effort, may set upper as well as lower limits on study size. Limits to study budgets and the number of sites and patients that could be recruited may be apparent at the outset of the study. However, an underpowered study involving substantial data collection that is ultimately unable to satisfactorily answer the research question(s) may prove to be a waste of finite monetary as well as human resources that could better be applied elsewhere.

10. Internal and External Validity

The potential for bias refers to opportunities for systematic errors to influence the results. Internal validity is the extent to which study results are free from bias, and the reported association between exposure and outcome is not due to unmeasured or uncontrolled-for variables. Generalizability, also known as external validity, is a concept that refers to the utility of the inferences for the broader population that the study subjects are intended to represent. In considering potential biases and generalizability, we discuss the differences between RCTs and registries, since these are the two principal approaches to conducting clinically relevant prospective research.

The strong internal validity that earns RCTs high grades for evidence comes largely from the randomization of exposures that helps ensure that the groups receiving the different treatments are similar in all measured or unmeasured characteristics, and that, therefore, any differences in outcome (beyond those attributable to chance) can be reasonably attributed to differences in the efficacy or safety of the treatments. However, it is worth noting that RCTs are not without their own biases, as illustrated by the "intent-to-treat" analytic approach, in which people are considered to have used the assigned treatment, regardless of actual compliance. The intent-to-treat analyses can minimize a real difference—generating a distortion known as "bias toward the null"—by including the experience of people who did not adhere to the recommended study product along with those who did.

Another principal difference between registries and RCTs is that RCTs are often focused on a relatively homogeneous pool of patients from which significant numbers of patients are purposefully excluded at the cost of external validity—that is, generalizability to the target population of disease sufferers. Registries, in contrast, usually focus on generalizability so that their population will be representative and relevant to decisionmakers.

10.1 Generalizability

The strong external validity of registries is achieved by the fact that they include typical patients, which often include more heterogeneous populations than those participating in RCTs (e.g., wide variety of age, ethnicity, and comorbidities). Therefore, registry data can provide a good description of the course of disease and impact of interventions in actual practice and, for some purposes, may be more relevant for decisionmaking than the data derived from the artificial constructs of the clinical trial. In fact, even though registries have more opportunities to introduce bias (systematic error) because of their nonexperimental methodology, well designed observational studies can approximate the effects of interventions observed in RCTs on the same topic[35, 36] and, in particular, in the evaluation of health care effectiveness in many instances.[37]

The choice of groups from which patients will be selected directly affects generalizability. No particular method will ensure that an approach to patient recruitment is adequate, but it is worthwhile to note that the way in which patients are recruited, classified, and followed can either enhance or diminish the external validity of a registry. Some examples of how these methods of patient recruitment and followup can lead to systematic error follow.

10.2 Information Bias

If the registry's principal goal is the estimation of risk, it is possible that adverse events or the number of patients experiencing them will be underreported if the reporter will be viewed negatively for reporting them. It is also possible for those collecting data to introduce bias by misreporting the outcome of an intervention if they have a vested interest in doing so. This type of bias is referred to as information bias (also called detection, observer, ascertainment, or assessment bias), and it addresses the extent to which the data that are collected are valid (represent what they are intended to represent) and accurate. This bias arises if the outcome assessment can be interfered with, intentionally or unintentionally. On the other hand, if the outcome is objective, such as whether or not a patient died or the results of a lab test, then the data are unlikely to be biased.

10.3 Selection Bias

A registry may create the incentive to enroll only patients who either are at low risk of complications or who are known not to have suffered such complications, biasing the results of the registry toward lower event rates. Those registries whose participants derive some sort of benefit from reporting low complication rates, for example, those with surgeons participating are at particularly high risk for this type of bias. Another example of how patient selection methods can lead to bias is the use of patient volunteers, a practice that may lead to selective participation from subjects most likely to perceive a benefit, distorting results for studies of patient-reported outcomes.

Enrolling patients who share a common exposure history, such as having used a drug that has been publicly linked to a serious adverse effect, could distort effect estimates for cohort and case-control analyses. Registries can also selectively enroll people who are at higher risk of developing serious side effects, since having a high-risk profile can motivate a patient to participate in a registry.

The term selection bias refers to situations where the procedures used to select study subjects lead to an effect estimate among those participating in the study that is different from the estimate that is obtainable from the target population.[38] Selection bias may be introduced if certain subgroups of patients are routinely included or excluded from the registry.

10.4 Channeling Bias (Confounding by Indication)

Channeling bias, also called confounding by indication, is a form of selection bias in which drugs with similar therapeutic indications are prescribed to groups of patients with prognostic differences.[39] For example, physicians may prescribe new treatments more often to those patients who have failed on traditional first-line treatments.

One approach to designing studies to address channeling bias is to conduct a prospective review of cases, in which external reviewers are blinded as to the treatments that were employed and are asked to determine whether a particular type of therapy is indicated and to rate the overall prognosis for the patient.[40] This method of blinded prospective review was developed to support research on ruptured cerebral aneurysms, a rare and serious situation. The results of the blinded review were used to create risk strata for analysis so that comparisons could be conducted only for candidates for whom both therapies under study were indicated, a procedure much like the application of additional inclusion and exclusion criteria in a clinical trial.

A computed "propensity score" (i.e., the predicted probability of use of one therapy over another based on medical history, health care utilization, and other characteristics measured prior to the initiation of therapy) is increasingly incorporated into study designs to address this type of confounding.[41, 42] Propensity scores may be used to create cohorts of initiators of two different treatments matched with respect to probability of use of one of the two therapies, for stratification or for inclusion as a covariate in a multivariate analysis. Studies incorporating propensity scores as part of their design may be planned prior to and implemented shortly following launch of a new drug as part of a risk management program, with matched comparators being selected over time, so

that differences in prescribing patterns following drug launch may be taken into account.[43]

Instrumental variables, or factors strongly associated with treatment but related to outcome only through their association with treatment, may provide additional means of adjustment for confounding by indication, as well as unmeasured confounding.[44] Types of instrumental variables include providers' preferences for one therapy over another—a variable which exploits variation in practice as a type of natural experiment; variation or changes in insurance coverage or economic factors (e.g., cigarette taxes) associated with an exposure; or geographic distance from a specific type of service.[45, 46] Variables that serve as effective instruments of this nature are not always available and may be difficult to identify. While use of clinician or study site may, in some specific cases, offer potential as an instrumental variable for analysis, the requirement that use of one therapy over another be very strongly associated with the instrument is often difficult to meet in real-world settings. In most cases, instrumental variable analysis provides an alternative for secondary analysis of study data. Instrumental variable analysis may either support the conclusions drawn on the basis of the initial analysis, or it may raise additional questions regarding the potential impact of confounding by indication.[42]

In some cases, however, differences in disease severity or prognosis between patients receiving one therapy rather than another may be so extreme and/or unmeasurable that confounding by indication is not remediable in an observational design.[47] This represents special challenges for observational studies of comparative effectiveness, as the severity of underlying illness may be a strong determinant of both choice of treatment and treatment outcome.

10.5 Bias from Study of Existing Rather Than New Product Users

If there is any potential for tolerance to affect the use of a product, such that only those who perceive benefit from it or are free from harm continue using it, the recruitment of existing users rather than new users may lead to the inclusion of only those who have tolerated or benefited from the intervention, and would not necessarily capture the full spectrum of experience and outcomes. Selecting only existing users may introduce any number of biases, including incidence/prevalence bias, survivorship bias, and followup bias. By enrolling new users (an inception or incidence cohort), a study ensures that the longitudinal experience of all users will be captured, and that the ascertainment of their experience and outcomes will be comparable.[48]

10.6 Loss to Followup

Loss to followup or attrition of patients and sites threatens generalizability as well as internal validity if there is differential loss; for example, loss of participants with a particular exposure or disease, or with particular outcomes. Loss to followup and attrition are generally a serious concern only when they are nonrandom (that is, when there are systematic differences between those who leave or are lost and those who remain). The magnitude of loss to followup or attrition determines the potential impact of any bias. Given that the differences between patients who remain enrolled and those who are lost to followup are often unknown (unmeasurable), preventing loss to followup in long-term studies to the fullest extent possible will increase the credibility and validity of the results.[49] Attrition should be considered with regard to both patients and study sites, as results may be biased or less generalizable if only some sites (e.g., teaching hospitals) remain in the study while others discontinue participation.

10.7 Assessing the Magnitude of Bias

Remaining alert for any source of bias is important, and the value of a registry is enhanced by its ability to provide a formal assessment of the likely magnitude of all potential sources of bias. Any information that can be generated regarding nonrespondents, missing respondents, and the like, is helpful, even if it is just an estimation of their raw numbers. As with many types of survey research, an assessment of differential response rates and patient selection can sometimes be undertaken when key data elements are available for both registry enrollees and nonparticipants. Such analyses can easily be undertaken when the

initial data source or population pool is that of a health care organization, employer, or practice that has access to data in addition to key selection criteria (e.g., demographic data or data on comorbidities). Another tool is the use of sequential screening logs, in which all subjects fitting the inclusion criteria are enumerated and a few key data elements are recorded for all those who are screened. This technique allows some quantitative analysis of nonparticipants and assessments of the effects, if any, on representativeness. Whenever possible, quantitative assessment of the likely impact of bias is desirable to determine the sensitivity of the findings to varying assumptions. A text on quantitative analysis of bias through validation studies, and on probabilistic approaches to data analysis, provides a guide for planning and implementing these methods.[50]

Qualitative assessments, although not as rigorous as quantitative approaches, may give users of the research a framework for drawing their own conclusions regarding the effects of bias on study results if the basis for the assessment is made explicit in reporting the results.

Accordingly, two items that can be reported to help the user assess the generalizability of research results based on registry data are a description of the criteria used to select the registry sites, and the characteristics of these sites, particularly those characteristics that might have an impact on the purpose of the registry. For example, if a registry designed for the purpose of assessing adherence to lipid screening guidelines requires that its sites have a sophisticated electronic medical record in order to collect data, it will probably report better adherence than usual practice because this same electronic medical record facilitates the generation of real-time reminders to engage in screening. In this case, a report of rates of adherence to other screening guidelines (for which there were no reminders), even if these are outside the direct scope of inquiry, would provide some insight into the degree of overestimation.

Finally, and most importantly, whether or not study subjects need to be evaluated on their representativeness depends on the purpose and kind of inference needed. For example, sampling in proportion to the underlying distribution in the population is not necessary to understand biological effects. However, if the study purpose were to estimate a rate of occurrence of a particular event, then sampling would be necessary to reflect the appropriate underlying distributions.

11. Summary

In summary, the key points to consider in designing a registry include study design, data sources, patient selection, comparison groups, sampling strategies, and considerations of possible sources of bias and ways to address them, to the extent that is practical and achievable.

Case Examples for Chapter 3

Case Example 2. Designing a registry for a health technology assessment

Description	The Nuss procedure registry was a short-term registry designed specifically for the health technology assessment of the Nuss procedure, a novel, minimally invasive procedure for the repair of pectus excavatum, a congenital malformation of the chest. The registry collected procedure outcomes, patient-reported outcomes, and safety outcomes.
Sponsor	National Institute for Health and Clinical Excellence (NICE), United Kingdom
Year Started	2004
Year Ended	2007
No. of Sites	13 hospitals
No. of Patients	260

Challenge

The Nuss procedure is a minimally invasive intervention for the repair of pectus excavatum. During a review of the evidence supporting this procedure conducted in 2003, the National Institute for Health and Clinical Excellence (NICE) determined that the existing data included relatively few patients and few quality of life outcomes, and did not sufficiently address safety concerns. NICE concluded in the 2003 review that the evidence was not adequate for routine use and that more evidence was needed to make a complete assessment of the procedure.

Proposed Solution

Gathering additional evidence through a randomized controlled trial was not feasible for several reasons. First, a blinded trial would be difficult because the other procedures for the repair of pectus excavatum produce much larger scars than the Nuss procedure. Surgeons also tend to perform either only the Nuss procedure or only another procedure, a factor that would complicate randomization efforts. In addition, only a small number of procedures are done in the United Kingdom. The sample for a randomized trial would likely be very small, making it difficult to detect rare adverse events.

Due to these limitations, NICE decided to develop a short-term registry to gather evidence on the Nuss procedure. The advantages of a registry were its ability to gather data on all patients undergoing the procedure in the United Kingdom to provide a more complete safety assessment, and its ability to collect patient-reported outcomes.

The registry was developed by an academic partner, with input from clinicians. Hospitals performing the procedure were identified and asked to enter into the registry data on all patients undergoing the intervention. Once the registry was underway, the cases in the registry were compared against cases included in the Hospital Episodes Statistics (HES) database, a nationwide source of routine data on hospital activity, and nonparticipating hospitals were identified and prompted to enter their data.

Results

NICE conducted a reassessment of the Nuss procedure in 2009, comparing data from the registry with other published evidence on safety and efficacy. The quantity of published literature had increased substantially between 2003 and 2009. The new publications primarily focused on technical and safety outcomes, while the registry included patient-reported outcomes. The literature and the registry reported similar rates of major adverse events such as bar displacement (from 2 to 10 percent). Based on the registry data and the new literature, the review committee found that the evidence was now sufficient to support routine use of the Nuss procedure, and no

Case Example 2. Designing a registry for a health technology assessment (continued)

Results (continued)

further review of the guidance is planned. Committee members considered that the registry made a useful contribution to guidance development.

Key Point

The Nuss registry demonstrated that a small, short-term, focused registry with recommended (but not automatic or mandatory) submission can produce useful data, both about safety and about patient-reported outcomes.

Case Example 3. Developing prospective nested studies in existing registries

Description	The Consortium of Rheumatology Researchers of North America (CORRONA) is a national disease registry of patients with rheumatoid arthritis (RA) and psoriatic arthritis (PsA).
Sponsor	CORRONA Investigators and Genentech
Year Started	2001
Year Ended	Ongoing
No. of Sites	Over 100 sites in the United States
No. of Patients	As of March 31 2012: 36,922 (31,701 RA patients and 5,221 PsA patients)

Challenge

In 2001, the CORRONA data collection program was established to collect longitudinal, physician- and patient-reported safety and effectiveness data for the treatment and management of RA and PsA. Any patient with RA or PsA upon diagnosis can participate in the registry, and participation in the registry is lifelong unless the patient withdraws consent. With an existing infrastructure and its representative, real-world nature, the disease registry can be used as a robust opportunity for nested trials at sites that have been trained in data collection and verification.

Proposed Solution

In collaboration with Genentech, the CORRONA investigators are utilizing the registry in two separate prospective, nested substudies: the Comparative Effectiveness Registry to study Therapies for Arthritis and Inflammatory CoNditions (CERTAIN) and the Treat to Target (T2T) study. Based on the study eligibility criteria and the capabilities of CORRONA sites, different patients and sites are being selected to participate in CERTAIN and T2T.

The CERTAIN study is a nested comparative effectiveness and safety study evaluating real-world differences in classes of biologic agents among RA patients initiating either tumor necrosis factor (TNF) antagonists or non-TNF-inhibitor biologic agents. The study is enrolling approximately 2,750 patients over three years to address comparative effectiveness questions. Long-term safety followup data will be collected through lifelong patient participation in the CORRONA registry after CERTAIN study completion. Data are collected at mandated 3-month intervals and include standard validated physician- and patient-derived outcomes and centrally-processed laboratory measures such as complete blood counts, metabolic panel, high sensitivity CRP, lipids with direct (nonfasting) LDL, immunoglobulin levels and serology (CCP and RF). Serum, plasma, DNA and RNA will be stored for future research. In addition, adverse event data are being obtained with inclusion of primary "source" documents, followed by a robust process of verification and adjudication.

Case Example 3. Developing prospective nested studies in existing registries (continued)

Proposed Solution (continued)

The T2T study is a cluster-randomized, open-label study comparing treatment acceleration (i.e., monthly visits with a change in therapeutic agent, dosage or route of administration in order to achieve a target metric of disease activity) against usual care (i.e., no mandated changes to therapy or visit frequencies beyond what the treating physician considers appropriate for the patient). This study will attempt to determine both the feasibility and outcomes of treating to target in a large U.S. population. This one-year study is enrolling 888 patients. Data collection includes standard measures of disease activity such as Clinical-Disease Activity Index (CDAI) score, Disease Activity Score-28 (DAS28), and Routine Assessment of Patient Index Data-3 (RAPID 3), as well as rates of acceleration, frequency of visits, and suspected RA-drug-related toxicities. The purpose of the trial is to test the hypothesis that accelerated aggressive therapy of RA correlates with better long-term patient outcomes.

Results

The CERTAIN and T2T studies, now in the enrollment phase, exemplify the key advantages and the unique operational synergies of successfully nesting studies within an existing disease registry. This design approach has the advantage of minimizing the usual study start-up and implementation challenges. The registry allows real-time identification of eligible patients typically seen in a U.S. clinical practice, a capability that can facilitate patient recruitment. Both CERTAIN and T2T have broad inclusion criteria to increase representativeness of the population enrolled. Established registry sites include investigators, staff, and patients already experienced with the registry questionnaires and research activities.

The two nested substudies require additional patient consent and site reimbursement, as they collect blood samples that increase the time required to complete a study visit. CORRONA collaborates with an academic institution to collect personal identifiers and patient consent to release medical records, thereby facilitating verification of serious adverse event for patients participating in CERTAIN. While this feature adds value to CERTAIN's ability to address long-term safety questions, it entailed establishing a new mechanism to ensure that the two databases (CORRONA and a database for personal identifiers) remain separate from each other in a highly secure way. New enrollment and screening instructions were developed for each substudy, with mandated completion of required training for participating physicians and research coordinators.

Key Point

Designing a prospective, nested study within an established disease registry has many benefits: the study leverages existing infrastructure, patient and site staff are familiar with the registry, and site relationships are already in place. Substudies need to be well planned and address a compelling clinical issue. Registry personnel must provide sufficient guidance, instructions, and rationale to sites to ensure that the transition is smooth and that the distinction from core registry operations is maintained in order to achieve the goal of high-quality research.

For More Information

Kremer J. The CORRONA database. Ann Rheum Dis. 2005 Nov;64 Suppl 4:iv37-41.

The Consortium of Rheumatology Researchers of North America, Inc. (CORRONA). http://www.corrona.org.

US National Institutes of Health, ClinicalTrials.gov. http://www.clinicaltrials.gov/ct2/show/NCT01407419?term=CORRONA&rank=2.

Case Example 4. Designing a registry to address unique patient enrollment challenges

Description	The Anesthesia Awareness Registry is a survey-based registry that collects detailed data about patient experiences of anesthesia awareness. Patient medical records are used to assess anesthetic factors associated with the patient's experience. An optional set of psychological assessment instruments measure potential trauma-related sequelae including depression and post-traumatic stress disorder (PTSD).
Sponsor	American Society of Anesthesiologists
Year Started	2007
Year Ended	Ongoing
No. of Sites	Not applicable
No. of Patients	265

Challenge

Anesthesia awareness is a recognized complication of general anesthesia, defined as the unintended experience and explicit recall of events during surgery. The incidence of anesthesia awareness has been estimated at 1–2 patients per 1,000 anesthetics and may result in development of serious and long-term psychological sequelae including PTSD. The causes of the phenomenon and preventive strategies have been studied, but there is disagreement in the scientific community about the effectiveness of monitoring devices for prevention of anesthesia awareness.

The population of patients experiencing anesthesia awareness is difficult to identify. Although standard short questionnaires designed to identify anesthesia awareness are sometimes administered to patients postoperatively, many patients experience delayed recollection and do not realize that they were awake during their procedure until several weeks later. These patients may or may not report their experience to their provider. In addition, because of the often unsettling and traumatic nature of their experience, even patients who recognize their anesthesia awareness before being discharged from the hospital may not feel comfortable reporting it to their surgeon or other health care providers.

With ongoing coverage in the media, anesthesiologists were facing increasing concern and fear about anesthesia awareness among their patients. The American Society of Anesthesiologists sought a patient-oriented approach to this problem.

Proposed Solution

Because this population of patients is not always immediately recognized in the health care setting, the registry was created to collect case reports of anesthesia awareness directly from patients. A patient advocate was invited to consult in the registry's development and provides ongoing advice from the patient perspective. The registry hosts a Web site that provides information about anesthesia awareness and directions for enrolling in the registry. Any patient who believes they have experienced anesthesia awareness may voluntarily submit a survey and medical records to the registry. Psychological assessments are optional. An optional open-ended discussion about the patient's anesthesia awareness experience provides patients with an opportunity to share information that may not be elicited through the survey.

Results

The registry has enrolled 265 patients since 2007. Patients who enroll are self-selected, and the sample is likely biased towards patients with emotional sequelae. While the information provided to potential enrollees clearly states that eligibility is restricted to awareness during general anesthesia, a surprising number of enrollments are patients who were supposed to be awake during regional anesthesia or sedation. This revealed a different side to the problem of anesthesia awareness: clearly, some patients did not understand the nature of the anesthetic that would be provided for their procedure, or patients had expectations that were not met by their

Case Example 4. Designing a registry to address unique patient enrollment challenges (continued)

Results (continued)

anesthesia providers. Most enrollees experienced long-term psychological sequelae regardless of anesthetic technique.

Key Point

Allowing the registry's purpose to drive its design produces a registry that is responsive to the expected patient population. Employing direct-to-patient recruitment can be an effective way of reaching a patient population that otherwise would not be enrolled in the registry, and can yield surprising and important insights into patient experience.

For More Information

http://www.awaredb.org

Domino, KB. Committee on Professional Liability opens anesthesia awareness registry. ASA Newsletter. 2007;71(3): 29, 34.

Domino, KB. Update on the Anesthesia Awareness Registry. ASA Newsletter 72(11): 32, 36, 2008.

Kent CD, Bruchas RR, Posner KL, et al. Anesthesia Awareness Registry Update. Anesthesiology. 2009;111:A1518.

Kent CD. Awareness during general anesthesia: ASA Closed Claims Database and Anesthesia Awareness Registry. ASA Newsletter. 2010; 74(2): 14-16.

Kent CD, Metzger NA, Posner KL, et al. Anesthesia Awareness Registry: psychological impacts for patients. Anesthesiology. 2011; A003.

Domino KB, Metzger NA, Mashour GA. Anesthesia Awareness Registry: patient responses to awareness. Br J Anaesth. 2012;108(2):338P.

References for Chapter 3

1. Strom BL. Pharmacoepidemiology. 3rd ed. Chichester, England: John Wiley; 2000.
2. Gliklich R. A New Framework For Comprehensive Evidence Development. IN VIVO. Oct 22, 2012. http://www.elsevierbi.com/publications/in-vivo/30/9/a-new-framework-for-comprehensive-evidence-development. Accessed June 17, 2013.
3. Chow SC, Chang M, Pong A. Statistical consideration of adaptive methods in clinical development. J Biopharm Stat. 2005;15(4):575-91. PMID: 16022164.
4. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. Biometrics. 1995 Mar;51(1):228-35. PMID: 7766778.
5. Ryan PB, Powell GE, Pattishall EN, et al. Performance of screening multiple observational databases for active drug safety surveillance. Providence, RI: International Society of Pharmacoepidemiology; 2009.
6. Norén GN, Hopstadius J, Bate A, et al. Temporal pattern discovery in longitudinal electronic patient records. Data Min Knowl Discov. 2010;20(3):361-87.
7. Travis LB, Rabkin CS, Brown LM, et al. Cancer survivorship—genetic susceptibility and second primary cancers: research strategies and recommendations. J Natl Cancer Inst. 2006 Jan 4;98(1):15-25. PMID: 16391368.
8. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology. Boston: Little, Brown and Company; 1985. p. 228.
9. Hennekens CH, Buring JE. Epidemiology in medicine. 1st ed. Boston: Little, Brown and Company; 1987.
10. Schoebel FC, Gradaus F, Ivens K, et al. Restenosis after elective coronary balloon angioplasty in patients with end stage renal disease: a case-control study using quantitative coronary angiography. Heart. 1997 Oct;78(4):337-42. PMID: 9404246. PMCID: 1892250.

11. Rothman K, Greenland S. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1998. p. 175–9.
12. Oral contraceptive use and the risk of endometrial cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. JAMA. 1983 Mar 25;249(12):1600-4. PMID: 6338265.
13. Oral contraceptive use and the risk of ovarian cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. JAMA. 1983 Mar 25;249(12):1596-9. PMID: 6338264.
14. Long-term oral contraceptive use and the risk of breast cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. JAMA. 1983 Mar 25;249(12):1591-5. PMID: 6338262.
15. Speck CE, Kukull WA, Brenner DE, et al. History of depression as a risk factor for Alzheimer's disease. Epidemiology. 1995 Jul;6(4):366-9. PMID: 7548342.
16. Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. Am J Epidemiol. 1992 May 1;135(9):1019-28. PMID: 1595688.
17. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. II. Types of controls. Am J Epidemiol. 1992 May 1;135(9):1029-41. PMID: 1595689.
18. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. III. Design options. Am J Epidemiol. 1992 May 1;135(9):1042-50. PMID: 1595690.
19. Rothman K, Greenland S. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1998. p. 108.
20. Vigneswaran R, Aitchison SJ, McDonald HM, et al. Cerebral palsy and placental infection: a case-cohort study. BMC Pregnancy Childbirth. 2004 Jan 27;4(1):1. PMID: 15005809. PMCID: 343280.
21. Ong AT, Daemen J, van Hout BA, et al. Cost-effectiveness of the unrestricted use of sirolimus-eluting stents vs. bare metal stents at 1 and 2-year follow-up: results from the RESEARCH Registry. Eur Heart J. 2006 Dec;27(24):2996-3003. PMID: 17114234.
22. Hulley SB, Cumming SR. Designing clinical research. Baltimore: Williams & Wilkins; 1988.
23. Hunter D. First, gather the data. N Engl J Med. 2006 Jan 26;354(4):329-31. PMID: 16436764.
24. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006 Jan 26;354(4):353-65. PMID: 16436767.
25. National Cancer Institute. Surveillance Epidemiology and End Results. http://seer.cancer.gov. Accessed August 27, 2012.
26. Metropolitan Atlanta Congenital Defects Program (MACDP). National Center on Birth Defects and Developmental Disabilities. Centers for Disease Control and Prevention. http://www.cdc.gov/ncbddd/birthdefects/MACDP.html. Accessed August 27, 2012.
27. Chen E, Sapirstein W, Ahn C, et al. FDA perspective on clinical trial design for cardiovascular devices. Ann Thorac Surg. 2006 Sep;82(3):773-5. PMID: 16928481.
28. U.S. Food and Drug Administration, Center for Devices and Radiological Health. The Least Burdensome Provisions of the FDA Modernization Act of 1997; Concept and Principles: Final Guidance for FDA and Industry. Document issued October 4, 2002. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm085994.htm#h3. Accessed August 14, 2012.
29. Cochran WG. Sampling Techniques. 3rd ed. New York: John Wiley & Sons; 1977.
30. Lohr SL. Sampling: Design and Analysis. Boston: Duxbury; 1999.
31. Sudman S. Applied Sampling. New York: Academic Press; 1976.
32. Henry GT. Practical Sampling. Newbury Park, CA: Sage; 1990.
33. Rothman K, Greenland S. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1998. p. 116.
34. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. Psychological Methods. 1997;2(2):173-85.
35. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000 Jun 22;342(25):1887-92. PMID: 10861325. PMCID: 1557642.

36. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med. 2000 Jun 22;342(25):1878-86. PMID: 10861324.

37. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996 May 11;312(7040):1215-8. PMID: 8634569. PMCID: 2350940.

38. Rothman K. Modern Epidemiology. Boston: Little Brown and Company; 1986. p. 83.

39. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med. 1991 Apr;10(4):577-81. PMID: 2057656.

40. Johnston SC. Identifying confounding by indication through blinded prospective review. Am J Epidemiol. 2001 Aug 1;154(3):276-84. PMID: 11479193.

41. Sturmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006 May;59(5): 437-47. PMID: 16632131. PMCID: 1448214.

42. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):253-9. PMID: 16611199. PMCID: 1790968.

43. Loughlin J, Seeger JD, Eng PM, et al. Risk of hyperkalemia in women taking ethinylestradiol/drospirenone and other oral contraceptives. Contraception. 2008 Nov;78(5):377-83. PMID: 18929734.

44. Brookhart, M. Alan. Instrumental Variables for Comparative Effectiveness Research: A Review of Applications. Slide Presentation from the AHRQ 2008 Annual Conference (Text Version); Rockville, MD: Agency for Healthcare Research and Quality; Jan, 2009. http://www.ahrq.gov/about/annualmtg08/090908slides/Brookhart.htm. Accessed August 14, 2012.

45. Evans WN, Ringel JS. Can higher cigarette taxes improve birth outcomes? J Public Economics. 1999;72(1):135-54.

46. Schneeweiss S, Seeger JD, Landon J, et al. Aprotinin during coronary-artery bypass grafting and risk of death. N Engl J Med. 2008 Feb 21;358(8):771-83. PMID: 18287600.

47. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2010 Jan;63(1):64-74. PMID: 19457638. PMCID: 2789188.

48. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003 Nov 1;158(9):915-20. PMID: 14585769.

49. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751-60. PMID: 15469032.

50. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data: Springer; 2009.

Chapter 4. Data Elements for Registries

1. Introduction

Selection of data elements for a registry requires a balancing of potentially competing considerations. These considerations include the importance of the data elements to the integrity of the registry, their reliability, their necessity for the analysis of the primary outcomes, their contribution to the overall response burden, and the incremental costs associated with their collection. Registries are generally designed for a specific purpose, and data elements not critical to the successful execution of the registry or to the core planned analyses should not be collected unless there are explicit plans for their analysis.

The selection of data elements for a registry begins with the identification of the domains that must be quantified to accomplish the registry purpose. The specific data elements can then be selected, with consideration given to clinical data standards, common data definitions, and the use of patient identifiers. Next, the data element list can be refined to include only those elements that are necessary for the registry purpose. Once the selected elements have been incorporated into a data collection tool, the tool can be pilot tested to identify potential issues, such as the time required to complete the form, data that may be more difficult to access than realized during the design phase, and practical issues in data quality (such as appropriate range checks). This information can then be used to modify the data elements and reach a final set of elements.

2. Identifying Domains

Registry design requires explicit articulation of the goals of the registry and close collaboration among disciplines, such as epidemiology, health outcomes, statistics, and clinical specialties. Once the goals of the study are determined, the domains most likely to influence the desired outcomes must be defined. Registries generally include personal, exposure, and outcomes information. The personal domain consists of data that describe the patient, such as information on patient demographics, medical history, health status, and any necessary patient identifiers. The exposure domain describes the patient's experience with the product, disease, device, procedure, or service of interest to the registry. Exposure can also include other treatments that are known to influence outcome but are not necessarily the focus of the study, so that their confounding influence can be adjusted for in the planned analyses. The outcomes domain consists of information on the patient outcomes that are of interest to the registry; this domain should include both the primary endpoints and any secondary endpoints that are part of the overall registry goals.

In addition to the goals and desired outcomes, it is necessary to consider the need to create important subsets when defining the domains. Measuring potential confounding factors (variables that are linked with both the exposure and outcome) should be taken into account in this stage of registry development. Collecting data on potential confounders will allow for analytic or design control. (See Chapters 3 and 13.)

Understanding the time reference for all variables that can change over time is critical in order to distinguish cause-and-effect relationships. For example, a drug taken after an outcome is observed cannot possibly have contributed to the development of that outcome. Time reference periods can be addressed by including start and stop dates for variables that can change; they can also be addressed categorically, as is done in some quality improvement registries. For example, the Paul Coverdell National Acute Stroke Registry organized its patient-level information into categories to reflect the timeframe of the stroke event from onset through treatment to followup. In this case, the domains were categorized as prehospital, emergency evaluation and treatment, in-hospital evaluation and treatment, discharge information, and postdischarge followup.[1]

3. Selecting Data Elements

Once the domains have been identified, the process of selecting data elements begins with identification of the data elements that best quantify that domain and the source(s) from which those data elements can be collected. When selecting data elements, gaining consensus among the registry stakeholders is important, but this must be achieved without undermining the purpose of the registry by including elements solely to please a stakeholder. Each data element should support the purpose of the registry and answer an explicit scientific question or address a specific issue or need. The most effective way to select data elements is to start with the study purpose and objective, and then decide what types of groupings, measurements, or calculations will be needed to analyze that objective. Once the plan of analysis is clear, it is possible to work backward to define the data elements necessary to implement that analysis plan. This process keeps the group focused on the registry purpose and limits the number of extraneous ("nice to know") data elements that may be included.[2] (See Case Example 5.)

The data element selection process can be simplified if clinical data standards for a disease area exist. (See Case Example 7.) While there is a great need for common core data sets for conditions, there are few consensus or broadly accepted sets of standard data elements and data definitions for most disease areas. Thus, different studies of the same disease state may use different definitions of fundamental concepts, such as the diagnosis of myocardial infarction or the definition of worsening renal function.

To address this problem and to support more consistent data elements so that comparisons across studies can be more easily accomplished, some specialty societies and organizations are beginning to compile clinical data standards. For example, the American College of Cardiology (ACC) has created clinical data standards for acute coronary syndromes, heart failure, and atrial fibrillation.[3-5] These are used by registries such as the National Cardiovascular Data Registry (NDCR)® ICD Registry ™ for implantable cardioverter defibrillators and leads, which derived their publically posted data elements and definitions from the American College of Cardiology/American Heart Association (ACC/AHA) Key Data Elements and Definitions for Electrophysiological Studies and Procedures.[6] The National Cancer Institute (NCI) provides the Cancer Data Standards Registry and Repository (caDSR), which includes the caBIG® (Cancer Biomedical Informatics Grid®)–NCI data standards and the Cancer Therapy Evaluation Program (CTEP) common data element initiative.[7, 8] The North American Association of Central Cancer Registries (NAACCR) has developed a set of standard data elements and a data dictionary, and it promotes and certifies the use of these standards.[9] The American College of Surgeons National Cancer Database (NCDB) considers its data elements to be nationally standardized and open source.[10]

To a lesser extent, other disease areas also have begun to catalog data element lists and definitions. In the area of trauma, the International Spinal Cord Society has developed an International Spinal Cord Injury Core data set to facilitate comparison of studies from different countries,[11] and the National Center for Injury Prevention and Control has developed Data Elements for Emergency Department Systems (DEEDS), which are uniform specifications for data entered into emergency department patient records.[12] In the area of neurological disorders, the National Institute of Neurological Disorders and Stroke (NINDS) maintains a list of several hundred data elements and definitions (Common Data Elements).[13] In the area of infection control, the National Vaccine Advisory Committee (NVAC) in 2007 approved a new set of core data elements for immunization information systems, which are used as functional standards by groups such as the American Immunization Registry Association (AIRA).[14, 15] Currently, there are more than one set of lists for some conditions (e.g., cancer) and no central method to search broadly across disease areas.

Some standards organizations are also working on core data sets. The Clinical Data Interchange Standards Consortium (CDISC) Clinical Data

Acquisition Standards Harmonization (CDASH) is a global, consensus-based effort to recommend minimal data sets in 16 domains. While developed primarily for clinical trials, these domains have significant utility for patient registries. They comprise adverse events, comments, prior and concomitant medications, demographics, disposition, drug accountability, electrocardiogram test results, exposure, inclusion and exclusion criteria, laboratory test results, medical history, physical examination, protocol deviations, subject characteristics, substance abuse, and vital signs. The CDASH Standards information also includes a table on best practices for developing case report forms.[16]

The use of established data standards, when available, is essential so that registries can maximally contribute to evolving medical knowledge. Standard terminologies—and to a greater degree, higher level groupings into core data sets for specific conditions—not only improve efficiency in establishing registries but also promote more effective sharing, combining, or linking of data sets from different sources. Furthermore, the use of well-defined standards for data elements and data structure ensures that the meaning of information captured in different systems is the same. This is critical for "semantic" interoperability between information systems, which will be increasingly important as health information system use grows. This is discussed more in Chapter 15, Section 6.2.

Clinical data standards are important to allow comparisons between studies, but when different sets of standards overlap (i.e., are not harmonized), the lack of alignment may cause confusion during analyses. To consolidate and align standards that have been developed for clinical research, CDISC, the HL7 (Health Level 7) Regulated Clinical Research Information Management Technical Committee (RCRIM TC), NCI, and the U.S. Food and Drug Administration (FDA) have collaborated to create the Biomedical Research Integrated Domain Group (BRIDG) model. The purpose of this project is to provide an overarching model that can be used to harmonize standards between the clinical research domain and the health care domain. BRIDG is a domain analysis model (DAM), meaning that it provides a common representation of the semantics of protocol-driven clinical and preclinical research, along with the associated data, resources, rules, and processes used to formally assess a drug, treatment, or procedure.[17] The BRIDG model is freely available to the public as part of an open-source project at www.bridgemodel.org. It is hoped that the BRIDG model will guide clinical researchers in selecting approaches that will enable their data to be compared with other clinical data, regardless of the study phase or data collection method.[18]

In cases where clinical data standards for the disease area do not exist, established data sets may be widely used in the field. For example, United Network of Organ Sharing (UNOS) collects a large amount of data on organ transplant patients. Creators of a registry in the transplant field should consider aligning their data definitions and data element formats with those of UNOS to simplify the training and data abstraction process for sites.

Other examples of widely used data sets are the Joint Commission and the Centers for Medicare & Medicaid Services (CMS) data elements for hospital data submission programs. These data sets cover a range of procedures and diseases, from heart failure and acute myocardial infarction to pregnancy and surgical infection prevention. Hospital-based registries that collect data on these conditions may want to align their data sets with the Joint Commission and CMS. However, one limitation of tying elements and definitions to another data collection program rather than a fixed standard is that these programs may change their elements or definitions. With Joint Commission core measure elements, for example, this has occurred with some frequency.

If clinical data standards for the disease area and established data sets do not exist, it is still possible to incorporate standard terminology into a registry. This will make it easier to compare the registry data with the data of other registries and reduce the training needs and data abstraction burden on sites. Examples of several standard terminologies used to classify important data elements are listed in Table 4–1. Standard terminologies and suggestions for minimal data sets specific to pregnancy registries are provided in Chapter 21.

Table 4–1. Standard terminologies

Standard	Acronym	Description and Web Site	Developer
Billing-related			
Current Procedural Terminology	CPT®	Medical service and procedure codes commonly used in public and private health insurance plans and claims processing. Web site: http://www.ama-assn.org/ama/pub/category/3113.html	American Medical Association
International Classification of Diseases	ICD, ICD-O, ICECI, ICF, ICPC	International standard for classifying diseases and other health problems recorded on health and vital records. ICD-9-CM, a modified version of the ICD-9 standard, is used for billing and claims data in the United States, which will transition to ICD-10-CM in 2014. The ICD is also used to code and classify mortality data from death certificates in the United States. ICD adaptations include ICD-O (oncology), ICECI (External Causes of Injury), ICF (Functioning, Disability and Health), and ICPC-2 (Primary Care, Second Edition). Web site: http://www.who.int/classifications/icd/en	World Health Organization
Clinical			
Systemized Nomenclature of Medicine	SNOMED CT	Clinical health care terminology that maps clinical concepts with standard descriptive terms. Formerly SNOMED RT and SNOP. Web site: http://www.ihtsdo.org/snomed-ct	International Health Terminology Standards Development Organization
Unified Medical Language System	UMLS	Database of 100 medical terminologies with concept mapping tools.[19] Web site: http://www.nlm.nih.gov/research/umls/	National Library of Medicine
Classification of Interventions and Procedures	OPCS-4	Code for operations, surgical procedures, and interventions. Mandatory for use in National Health Service (England). Web site: http://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp	Office of Population, Censuses, and Surveys
Diagnostic and Statistical Manual	DSM	The standard classification of mental disorders used in the United States by a wide range of health and mental health professionals. The version currently in use is the DSM-IV. Web site: http://www.psych.org/MainMenu/Research/DSMIV.aspx	American Psychiatric Association
Drugs			
Medical Dictionary for Regulatory Activities	MedDRA	Terminology covering all phases of drug development, excluding animal toxicology. Also covers health effects and malfunctions of devices. Replaced COSTART (Coding Symbols for a Thesaurus of Adverse Reaction Terms). Web site: http://www.meddramsso.com	International Conference on Harmonisation (ICH)

Table 4–1. Standard terminologies (continued)

Standard	Acronym	Description and Web Site	Developer
VA National Drug File Reference Terminology	NDF-RT	Extension of the VA National Drug File; used for modeling drug characteristics, including ingredients, chemical structure, dose form, physiologic effect, mechanism of action, pharmacokinetics, and related diseases. Web site not available.	U.S. Department of Veterans Affairs
National Drug Code	NDC	Unique 3-segment number used as the universal identifier for human drugs. Web site: http://www.fda.gov/cder/ndc/	U.S. Food and Drug Administration
RxNorm	RxNorm	Standardized nomenclature for clinical drugs. The name of a drug combines its ingredients, strengths, and/or form. Links to many of the drug vocabularies commonly used in pharmacy management and drug interaction software. Web site: http://www.nlm.nih.gov/research/umls/rxnorm/	National Library of Medicine
World Health Organization Drug Dictionary	WHODRUG	International drug dictionary. Web site: http://www.who-umc.org/DynPage.aspx?id=98105&mn1=7347&mn2=7252&mn3=7254&mn4=7338	World Health Organization
Lab-Specific			
Logical Observation Identifiers Names and Codes	LOINC®	Concept-based terminology for lab orders and results.[19] Web site: http://www.regenstrief.org/loinc/	Regenstrief Institute for Health Care
Other			
HUGO Gene Nomenclature Committee	HGNC	Recognized standard for human gene nomenclature. Web site: http://www.genenames.org/	Human Genome Organization
Dietary Reference Intakes	DRIs	Nutrient reference values developed by the Institute of Medicine to provide the scientific basis for the development of food guidelines in Canada and the United States. Web site: http://fnic.nal.usda.gov/dietary-guidance/dietary-reference-intakes/dri-tables	Institute of Medicine Food and Nutrition Board
Substance Registry Services	SRS	The central system for standards identification of, and information about, all substances tracked or regulated by the Environmental Protection Agency. Web site: http://iaspub.epa.gov/sor_internet/registry/substreg/searchandretrieve/searchbylist/search.do	Environmental Protection Agency

In addition to these standard terminologies, numerous useful commercial code listings target specific needs, such as proficiency in checking for drug interactions or compatibility with widely used electronic medical record systems. Mappings between many of these element lists are also increasingly available. For example, SNOMED CT® (Systemized Nomenclature of Medicine Clinical Terminology) can currently be mapped to ICD-9-CM (International Classification of Diseases, 9th Revision, Clinical Modification), and mapping between other standards is planned or underway.[20]

After investigating clinical data standards, registry planners may find that there are no useful standards or established data sets for the registry, or that these standards comprise only a small portion of the data set. In these cases, the registry will need to define and select data elements with the guidance of its project team, which may include an advisory board.

When selecting data elements, it is often helpful to gather input from statisticians, epidemiologists, psychometricians, and experts in health outcomes assessment who will be analyzing the data, as they may notice potential analysis issues that need to be considered at the time of data element selection. Data elements may also be selected based on performance or quality measures in a clinical area. (See Case Examples 6 and 53.)

When beginning the process of defining and selecting data elements, it can be useful to start by considering the registry design. Since many registries are longitudinal, sites often collect data at multiple visits. In these cases, it is necessary to determine which data elements can be collected once and which data elements should be collected at every visit. Data elements that can be collected once are often collected at the baseline visit.

In other cases, the registry may be collecting data at an event level, so all of the data elements will be collected during the course of the event rather than in separate visits. In considering when to collect a data element, it is also important to determine the most appropriate order of data collection. Data elements that are related to each other in time (e.g., dietary information and a fasting blood sample for glucose or lipids) should be collected in the same visit rather than in different visit case report forms.

International clinician and patient participation may be required to meet certain registry data objectives. In such situations, it is desirable to consider the international participation when selecting data elements, especially if it will be necessary to collect and compare data from individual countries. Examination and laboratory test results or units may differ among countries, and standardization of data elements may become necessary at the data-entry level. Data elements relating to cost-effectiveness studies may be particularly challenging, since there is substantial variation among countries in health care delivery systems and practice patterns, as well as in the cost of medical resources that are used as "inputs."

Alternatively, if capture of internationally standardized data elements is not desirable or cannot be achieved, registry stakeholders should consider provisions to capture data elements according to local standards. Later, separate data conversions and merging outside the database for uniform reporting or comparison of data elements captured in multiple countries can be evaluated and performed as needed if the study design ensures that all data necessary for such conversions have been collected.

Table 4–2 provides examples of possible baseline data elements. The actual baseline data elements selected for a specific registry will vary depending on the design, nature, and goals of the registry. Examples listed include patient identifiers (e.g., for linkage to other databases), contact information (e.g., for followup), and residence location of enrollee (e.g., for geographic comparisons). Other administrative data elements that may be collected include the source of enrollment, enrollee sociodemographic characteristics, and information on provider locations.

Table 4–2. Examples of possible baseline data elements	
Enrollee contact information	• Enrollee contact information for registries with direct-to-enrollee contact • Another individual who can be reached for followup (address, telephone, email)
Enrollment data elements	• Patient identifiers (e.g., name [last, first, middle initial], date of birth, place of birth, Social Security number) • Permission/consent • Source of enrollment (e.g., provider, institution, phone number, address, contact information) • Enrollment criteria • Sociodemographic characteristics, including race, gender, and age or date of birth • Education and/or economic status, insurance, etc. • Preferred language • Place of birth • Location of residence at enrollment • Source of information • Country, State, city, county, ZIP Code of residence

Depending on the purpose of a registry, other sets of data elements may be required. Table 4–3 provides examples of possible additional data elements; again, the data elements selected for a specific registry will vary and should be driven by the design and purpose of the registry.

Table 4–3. Examples of possible additional enrollee, provider, and environmental data elements	
Pre-Enrollment History	
Medical history	• Morbidities/conditions • Onset/duration • Severity • Treatment history • Medications • Adherence • Health care resource utilization • Diagnostic tests and results • Procedures and outcomes • Emergency room visits, hospitalizations (including length of stay), long-term care, or stays in skilled nursing facilities • Genetic information • Comorbidities • Development (pediatric/adolescent)
Environmental exposures	• Places of residence

Table 4–3. Examples of possible additional enrollee, provider, and environmental data elements (continued)	
Patient characteristics	• Functional status (including ability to perform tasks related to daily living), quality of life, symptoms • Health behaviors (alcohol, tobacco use, physical activity, diet) • Social history • Marital status • Family history • Work history • Employment, industry, job category • Social support networks • Economic status, income, living situation • Sexual history • Foreign travel, citizenship • Legal characteristics (e.g., incarceration, legal status) • Reproductive history • Health literacy • Individual understanding of medical conditions and the risks and benefits of interventions • Social environment (e.g., community services) • Enrollment in clinical trials (if patients enrolled in clinical trials are eligible for the registry)
Provider/system characteristics	• Geographical coverage • Access barriers • Quality improvement programs • Disease management, case management • Compliance programs • Information technology use (e.g., computerized physician order entry, e-prescribing, electronic medical records)
Financial/economic information	• Disability, work attendance (days lost from work), or absenteeism/presenteeism Out-of-pocket costs • Health care utilization behavior, including outpatient visits, hospitalizations (and length of stay), and visits to the emergency room or urgent care • Patients' assessments of the degree to which they avoid health care because of its costs • Patients' reports of insurance coverage to assist/cover the costs of outpatient medications • Destination when discharged from a hospitalization (home, skilled nursing facility, long-term care, etc.) • Medical costs, often derived from data clinician office visits, hospitalizations (especially length of stay), and/or procedures
Followup	
Key primary outcomes	• Safety: adverse events (see Chapter 12) • Quality measurement/improvement: key selected measures at appropriate intervals • Effectiveness and value: intermediate and endpoint outcomes; health case resource use and hospitalizations, diagnostic tests and results. Particularly important are outcomes meaningful to patients, including survival, symptoms, function, and patient-reported outcomes, such as health-related quality-of-life measures. • Natural history: progression of disease severity; use of health care services; diagnostic tests, procedures, and results; quality of life; mortality; cause/date of death

Table 4–3. Examples of possible additional enrollee, provider, and environmental data elements (continued)	
Key secondary outcomes	• Economic status • Social functioning
Other potentially important information	• Changes in medical status • Changes in patient characteristics • Changes in provider characteristics • Changes in financial status • Residence • Changes to, additions to, or discontinuation of exposures (medications, environment, behaviors, procedures) • Changes in health insurance coverage • Sources of care (e.g., where hospitalized) • Changes in individual attitudes, behaviors

In addition, data elements that may be needed for specific types of registries are outlined here:

- For registries examining questions of safety for drugs, vaccines, procedures, or devices, key information includes history of the exposure and data elements that will permit analysis of potential confounding factors that may affect observed outcomes, such as enrollee characteristics (e.g., comorbidities, concomitant therapies, socioeconomic status, ethnicity, environmental and social factors) and provider characteristics. For drug exposures, data on use (start and stop dates), as well as data providing continuing evidence that the drug was actually used (data on medication persistence and/or adherence), may be important. In some instances, it is also useful to record reasons for discontinuation and whether pills were split or shared with others. Refer to Chapter 19 for more information on using registries for product safety assessments. For registries designed to study devices, unique device identifier information may be collected. See Chapter 23 for more information on issues specific to medical devices.
- For registries examining questions of effectiveness and cost-effectiveness, key information includes the history of exposure and data elements that will permit analysis of potential confounding factors that may affect observed outcomes. It may be particularly useful to collect information to assess confounding by indication, such as the reason for prescribing a medication. In addition to the data elements mentioned above for safety, data elements may include individual behaviors and provider and/or system characteristics. For assessment of cost-effectiveness, information may be recorded on the financial and economic burden of illness, such as office visits, visits to urgent care or the emergency room, and hospitalizations, including length of stay. Information on indirect or productivity costs (such as absenteeism and disability) may also be collected. For some studies, a quality-of-life instrument that can be analyzed to provide quality-adjusted life years or similar comparative data across conditions may be useful.
- For registries assessing quality of care and quality improvement, data that categorize and possibly differentiate among the services provided (e.g., equipment, training, or experience level of providers, type of health care system) may be sought, as well as information that identifies individual patients as potential candidates for the treatment (Chapter 22). In addition, patient-reported outcomes are valuable to assess the patients' perception of quality of care (Chapter 5).
- For registries examining the natural history of a condition, the selection of data elements would be similar to those of effectiveness registries.

If one goal of a registry is to identify patient subsets that are at higher risk for particular outcomes, more detailed information on patient and provider characteristics should be collected, and a higher sample size also may be required. This information may be important in registries that look at the usage of a procedure or treatment. Quality improvement registries also use this information to understand how improvement differs across many types of institutions.

Another question that may arise during data element selection relates to endpoint adjudication. Some significant endpoints may either be difficult to confirm without review of the medical record (e.g., stroke) or may not be specific to a single disease and therefore difficult to attribute without such review (e.g., mortality). While clinical trials commonly use an adjudication process for such endpoints to better assess the endpoint or the most likely cause, this is much less common in registries. The use of adjudication for endpoints will depend on the purpose of the registry.

3.1 Patient Identifiers

When selecting patient identifiers, there are a variety of options to use (e.g., the patient's name, date of birth, or some combination thereof) that are subject to legal and security considerations. When the planned analyses require linkage to other data (such as medical records), more specific patient information may be needed, depending on the planned method of linkage (e.g., probabilistic or deterministic). (For more information on linkage considerations, see Chapter 16.) In selecting patient identifiers, some thought should be given to the possibility that patient identifiers may change during the course of the registry. For example, patients may change their names during the course of the registry following marriage/divorce, or patients may move or change their telephone numbers. Patient identifiers can also be inaccurate because of intentional falsification by the patient (e.g., for privacy reasons in a sexually transmitted disease registry), unintentional misreporting by the patient or a parent (e.g., wrong date of birth), or typographical errors by clerical staff. In these cases, having more than one patient identifier for linking patient records can be invaluable. In addition, identifier needs will differ based on the registry goals. For example, a registry that tracks children will need identifiers related to the parents, and registries that are likely to include twins (e.g., immunization registries) should plan for the duplication of birth dates and other identifiers. In selecting patient identifiers for use in a registry, registry planners will need to determine what data are necessary for their purpose and plan for potential inaccurate and changing data.

Generally, patient identifiers can simplify the process of identifying and tracking patients for followup. Patient identifiers also allow for the possibility of identifying patients who are lost to followup due to death (i.e., through the National Death Index) and linking to birth certificates for studies in children. In addition, unique patient identifiers allow for analysis to remove duplicate patients.

When considering the advantages of patient identifiers, it is important to take into account the potential challenges that collecting patient identifiers can present and the privacy and security concerns associated with the collection and use of patient identifiers. Obtaining consent for the use of patient-identifiable information can be an obstacle to enrollment, as it can lead to the refusal of patients to participate. Chapter 7 contains more information on the ethical and legal considerations of using patient identifiers.

In addition to the data points related to primary and secondary outcomes, it is important to plan for patients who will leave the registry. While the intention of a registry is generally for all patients to remain in the study until planned followup is completed, planning for patients to leave the study before completion of full followup may reduce analysis problems. By designing a final study visit form, registry planners can more clearly document when losses to followup occurred and possibly collect important information about why patients left the study. Not all registries will need a study discontinuation form, as some studies collect data on the patient only once and do not include followup information (e.g., in-hospital procedure registries).

3.2 Data Definitions

Creating explicit data definitions for each variable to be collected is essential to the process of selecting data elements. This is important to ensure internal validity of the proposed study so that all participants in data collection are acquiring the requisite information in the same reproducible way. (See Chapter 11.) The data definitions should include the ranges and acceptable values for each individual data element, as well as the potential interplay of different data elements. For example, logic checks for the validity of data capture may be created for data elements that should be mutually exclusive.

When deciding on data definitions, it is important to determine which data elements are required and which elements may be optional. This is particularly true in cases where the registry may collect a few additional "nice to know" data elements. The determination will differ depending on whether the registry is using existing medical record documentation to obtain a particular data element or whether the clinician is being asked directly. For example, the New York Heart Association Functional Class for heart failure is an important staging element but is often not documented.[21] However, if clinicians are asked to provide the data point prospectively, they can readily do so. Consideration should also be given to accounting for missing or unknown data. In some cases, a data element may be unknown or not documented for a particular patient, and followup with the patient to answer the question may not be possible. Including an option on the form for "not documented" or "unknown" will allow the person completing the case report form to provide a response to each question rather than leaving it blank. Depending on the analysis plans for the registry, the distinction between undocumented data and missing data may be important.

3.3 Patient-Reported Outcomes

When collecting data for patient outcomes analysis, it is important to use patient-reported outcomes (PROs) that are valid, reliable, responsive, interpretable, and translatable. PROs reflect the patients' perceptions of their status and their perspective on health and disease. PROs have become an increasingly important avenue of investigation, particularly in light of the 2001 Institute of Medicine report calling for a more patient-centered health care system.[22] The FDA also noted the importance of PRO data in understanding certain treatment effects in its 2009 guidance document.[23] The use of PROs in registries is discussed in more detail in Chapter 5.

When using an instrument to gather data on PROs, it is important both to collect the individual question responses and to calculate the summary or composite score. The summary score, which may be for the entire instrument or for individual domains, is ultimately used to report results. However, if the registry collects only the summary score, it will not be possible to examine how the patients scored on different components of the instrument during the registry analysis phase.

4. Registry Data Map

Once data elements have been selected, a data map should be created. The data map identifies all sources of data (Chapter 6) and explains how the sources of data will be integrated. Data maps are useful to defend the validity and/or reliability of the data, and they are typically an integral part of the data management plan (Chapter 11, Section 2.5).

5. Pilot Testing

After the data elements have been selected and the data map created, it is important to pilot test the data collection tools to determine the time needed to complete the form and the resulting subject/abstractor burden. For example, through pilot testing, registry planners might determine that it is wise to collect certain data elements that are either highly burdensome or only "nice to know" in only a subset of participating sites (nested registry) that agree to the more intensive data collection, so as not to endanger participation in the registry as a whole. Pilot testing should also help to identify the rate of missing data and any validity issues with the data collection system.

The burden of form collection is a major factor determining a registry's success or failure, with major implications for the cost of participation and for the overall acceptance of the registry by hospitals and health care personnel. Moreover, knowing the anticipated time needed for patient recruitment/enrollment will allow better communication to potential sites regarding the scope and magnitude of commitment required to participate in the study. Registries that obtain information directly from patients include the additional issue of participant burden, with the potential for participant fatigue, leading to failure to answer all items in the registry. Highly burdensome questions can be collected in a prespecified subset of subjects. The purpose of these added questions should be carefully considered when determining the subset so that useful and accurate conclusions can be achieved.

Pilot testing the registry also allows the opportunity to identify issues and make refinements in the registry-specific data collection tools, including alterations in the format or order of data elements and clarification of item definitions. Alterations to validated PRO measures are generally not advised unless they are revalidated. Validated PRO measures that are not used in the validated format may be perceived as invalid or unreliable.

Piloting may also uncover problems in registry logistics, such as the ability to accurately or comprehensively identify subjects for inclusion. A fundamental aspect of pilot testing is evaluation of the accuracy and completeness of registry questions and the comprehensiveness of both instructional materials and training in addressing these potential issues. Gaps in clarity concerning questions can result in missing or misclassified data, which in turn may cause bias and result in inaccurate or misleading conclusions. For example, time points, such as time to radiologic interpretation of imaging test, may be difficult to obtain retrospectively and, if they do exist in the chart, may not be consistently documented. Without additional instruction, some hospitals may indicate the time the image was read by the radiologist and others may use the time when the interpretation was recorded in the chart. The two time points can have significant variation, depending on the documentation practices of the institution.

Pilot testing ranges in practice from ad hoc assessments of the face validity of instruments and materials in clinical sites, to trial runs of the registry in small numbers of sites, to highly structured evaluations of inter-rater agreement. The level of pilot testing is determined by multiple factors. Accuracy of data entry is a key criterion to evaluate during the pilot phase of the registry. When a "gold standard" exists, the level of agreement with a reference standard (construct validity) may be measured.[24] Data collected by seasoned abstractors or auditors following strict operational criteria can serve as the gold standard by which to judge accuracy of abstraction for chart-based registries.[25]

In instances where no reference standard is available, reproducibility of responses to registry elements by abstractors (inter-rater reliability) or test-retest agreement of subject responses may be assessed.[26] Reliability and/or validity of a data element should be tested in the pilot phase whenever the element is collected in new populations or for new applications. Similar mechanisms to those used during the pilot phase can be used during data quality assurance (Chapter 11, Section 3). A kappa statistic measure of how much the level of agreement between two or more observers exceeds the amount of agreement expected by chance alone is the most common method for measuring reliability of categorical and ordinal data. The intraclass correlation coefficient, or inter-rater reliability coefficient, provides information on the degree of agreement for continuous data. It is a proportion that ranges from zero to one. Item-specific agreement represents the highest standard for registries; it has been employed in cancer registries and to assess the quality of data in statewide stroke registries. Other methods, such as the Bland and Altman method,[26] may also be chosen, depending upon the type of data and registry purpose.

6. Summary

The selection of data elements requires balancing such factors as their importance for the integrity of the registry and for the analysis of primary outcomes, their reliability, their contribution to the overall burden for respondents, and the incremental costs associated with their collection. Data elements should be selected with consideration for established clinical data standards, common data definitions, and whether patient identifiers will be used. It is also important to determine which elements are absolutely necessary and which are desirable but not essential. Once data elements have been selected, a data map should be created, and the data collection tools should be pilot tested. Overall, the choice of data elements should be guided by parsimony, validity, and a focus on achieving the registry's purpose.

Case Examples for Chapter 4

Case Example 5. Selecting data elements for a registry

Description	The Dosing and Outcomes Study of Erythropoiesis-stimulating Therapies (DOSE) Registry was designed to understand anemia management patterns and clinical, economic, and patient-reported outcomes in oncology patients treated in outpatient oncology practice settings across the United States. The prospective design of the DOSE Registry enabled data capture from oncology patients treated with erythropoiesis-stimulating therapies.
Sponsor	Centocor Ortho Biotech Services, LLC
Year Started	2003
Year Ended	2009
No. of Sites	71
No. of Patients	2,354

Challenge

Epoetin alfa was approved for patients with chemotherapy-induced anemia in 1994. In 2002, the U.S. Food and Drug Administration approved a second erythropoiesis-stimulating therapy (EST), darbepoetin alfa, for a similar indication. While multiple clinical trials described outcomes following intervention with ESTs, little information was available on real-world practice patterns and outcomes in oncology patients. The registry team determined that a prospective observational effectiveness study in this therapeutic area was needed to gain this information. The three key challenges were to make the study representative of real-world practices and settings (e.g., hospital-based clinics, community oncology clinics); to collect data elements that were straightforward so as to minimize potential data collection errors; and to collect sufficient data to study effectiveness, while ensuring that the data collection remained feasible and time efficient for outpatient oncology clinics.

Proposed Solution

The registry team began selecting data elements by completing a thorough literature review. Because this would be one of the first prospective observational studies in this therapeutic area, the team wanted to ensure that study results could be presented to health care professionals and decisionmakers in a manner consistent with clinical trials, of which there were many. The team also intended to make the data reports from this study comparable with clinical trial reports. To meet these objectives, data elements (e.g., baseline demographics, dosing patterns, hemoglobin levels) similar to those in clinical trials were selected whenever possible, based on a thorough literature review.

Case Example 5. Selecting data elements for a registry (continued)

Proposed Solution (continued)

For the patient-reported outcomes component of the registry, the team incorporated standard validated instruments. This decision allowed the team to avoid developing and validating new instruments and supported consistency with clinical trial literature, as many trials had incorporated these instruments. To capture patient-reported data, the team selected two instruments, the Functional Assessment of Cancer Therapy–Anemia (FACT-An) and the Linear Analog Scale Assessment (LASA) tool. The FACT-An tool, developed from the FACT-General scale, had been designed and validated to measure the impact of anemia in cancer patients. The LASA tool enables patients to report their energy level, activity level, and overall quality of life on a scale of 0 to 100. Both tools are commonly used to gather patient-reported outcomes data for cancer patients.

Following the literature review, an advisory board was convened to discuss the registry objectives, data elements, and study execution. The advisory board included representatives from the medical and nursing professions. The multidisciplinary board provided insights into both the practical and clinical aspects of the registry procedures and data elements. Throughout the process, the registry team remained focused on both the overall registry objectives and user-friendly data collection. In particular, the team worked to make each question clear and unambiguous in order to minimize confusion and enable a variety of site personnel, as well as the patients, to complete the registry data collection.

Results

The registry was launched in 2003 as one of the first prospective observational effectiveness studies in this therapeutic area. Seventy-one sites and 2,354 patients enrolled in the study. The sites participating in the registry represented a wide geographic distribution and a mixture of outpatient practice settings.

Key Point

Use of common data elements, guided by a literature review, and validated patient-reported outcomes instruments enhanced data generalizability and comparability with clinical trial data. A multidisciplinary advisory board also helped to ensure collection of key data elements in an appropriate manner from both a clinical and practical standpoint.

For More Information

Larholt K, Burton TM, Hoaglin DC. et al. Clinical and patient-reported outcomes based on achieved hemoglobin levels in chemotherapy-treated cancer patients receiving erythropoiesis-stimulating agents. Commun Oncol. 2009;6: 403–8.

Larholt K, Pashos CL, Wang Q. et al. Dosing and Outcomes Study of Erythropoiesis-Stimulating Therapies (DOSE): a registry for characterizing anaemia management and outcomes in oncology patients. Clin Drug Invest. 2008;28(3):159–67.

Case Example 6. Understanding the needs and goals of registry participants

Description	The Prospective Registry Evaluating Myocardial Infarction: Events and Recovery (PREMIER) studied the health status of patients for one year after discharge for a myocardial infarction. The registry focused on developing a rich understanding of the patients' symptoms, functional status, and squality of life by collecting extensive baseline data in the hospital and completing followup interviews at 1, 6, and 12 months.
Sponsor	CV Therapeutics and CV Outcomes
Year Started	2003
Year Ended	2004
No. of Sites	19
No. of Patients	2,498

Challenge

With the significant advances in myocardial infarction (MI) care over the past 20 years, many studies have documented the improved mortality and morbidity associated with these new treatments. These studies typically have focused on in-hospital care, with little to no followup component. As a result, information on the transition from inpatient to outpatient care has been lacking, as have data on health status outcomes.

PREMIER was designed to address these gaps by collecting detailed information on MI patients during the hospital stay and through followup telephone interviews conducted at 1, 6, and 12 months. The goal of the registry was to provide a rich understanding of patients' health status (their symptoms, function, and quality of life) 1 year after an acute MI. The registry also proposed to quantify the prevalence, determinants, and consequences of patient and clinical factors in order to understand how the structures and processes of MI care affect patients' health status.

To develop the registry data set, the team began by clearly defining the phases of care and recovery and identifying the clinical characteristics that were important in each of these phases. These included patient characteristics upon hospital arrival, details of inpatient care, and details of outpatient care. The team felt that information on each of these phases was necessary, since the variability of any outcome over 1 year may be explained by patient, inpatient treatment, or outpatient factors. Health status also includes many determinants beyond the clinical status of disease, such as access to care, socioeconomic status, and social support; the registry needed to collect these additional data in order to fully understand the health status outcomes.

Proposed Solution

While registries often try to include as many eligible patients and sites as possible by reducing the burden of data entry, this registry took an alternative approach. The team designed a data set that included more than 650 baseline data elements and more than 200 followup interview-assessed data elements. Instead of allowing retrospective chart abstraction, the registry required hospitals to complete a five-page patient interview while the patient was in the hospital. The registry demanded significant resources from the participating sites. For each patient, the registry required about 4 hours of time, with 15 minutes for screening, 2 hours for chart abstraction, 45 minutes for interviews, 45 minutes for data entry, and 15 minutes of a cardiologist's time to interpret the electrocardiograms and angiograms. A detailed, prespecified sampling plan was developed by each site and approved by the data coordinating center to ensure that the patients enrolled at each center were representative of all of the patients seen at that site.

The registry team developed this extremely detailed data set and data collection process through extensive consultations with the registry participants. The coordinators and steering

Case Example 6. Understanding the needs and goals of registry participants (continued)

Proposed Solution (continued)

committees reviewed the data set multiple times, with some sites giving extensive feedback. Throughout the development process, there was an ongoing dialog among the registry designers, the steering committee, and the registry sites.

The registry team also used standard definitions and established instruments whenever possible to enable the registry data to be cross-referenced to other studies and to minimize the training burden. The team used the American College of Cardiology Data Standards for Acute Coronary Syndromes for data definitions of any overlapping fields. To measure other areas of the patient experience, the team used the Patient Health Questionnaire to examine depression, the ENRICHD Social Support Inventory to measure social support, the Short Form-12 to quantify overall mental and physical health, and the Seattle Angina Questionnaire (SAQ) to understand the patients' perspective on how coronary disease affects their life.

Results

The data collection burden posed some challenges. Two of the 19 sites dropped out of the registry soon after it began. Two other sites fell behind on their chart abstractions. Turnover of personnel and multiple commitments at participating sites also delayed the study.

Despite these challenges, the registry experienced very little loss of enthusiasm or loss of sites once it was up and running. The remaining 17 sites completed the registry and collected data on nearly 2,500 patients. In return for this data collection, sites enjoyed the academic productivity and collaborative nature of the study. The data coordinating center created a Web site that offered private groups for the principal investigators, so that each investigator had access to all of the abstract ideas and all of the research that was being done. This structure provided nurturing and support for the investigators, and they viewed the registry as a way to engage themselves and their institutions in research with a prominent, highly respected team.

On the patient side, the registry met followup goals. More than 85 percent of participants provided 12-month followup information. The registry team attributed this followup rate to the strong rapport that the interviewers developed with the patients during the course of the followup period.

Key Point

This example illustrates that there is no maximum or minimum number of data elements for a successful registry. Instead, a registry can best achieve its goals by ensuring that sufficient information is collected to achieve the purpose of the registry while remaining feasible for the participants. An open, ongoing dialog with the participants or a subgroup of participants can help determine what is feasible for a particular registry and to ensure that the registry will retain the participants for the life of the study.

For More Information

Spertus JA, Peterson E, Rumsfeld JS. et al. The Prospective Registry Evaluating Myocardial Infarction: Events and Recovery (PREMIER)–evaluating the impact of myocardial infarction on patient outcomes. Am Heart J. 2006;151(3): 589–97.

Case Example 7. Using standardized data elements in a registry

Description	The Caris Registry is a national, multicenter, Web-based registry that tracks long-term outcomes for patients who have undergone Caris Molecular Intelligence™ Services.
Sponsor	Caris Life Sciences
Year Started	2009
Year Ended	Ongoing
No. of Sites	96
No. of Patients	>1400

Challenge

Molecular biomarker data may be valuable in guiding treatment decisions for cancer patients. Caris Life Sciences offers a commercial molecular profiling service (Caris Molecular Intelligence™) that combines biomarker analysis of a patient's tumor with an analysis of the published scientific literature in order to report personalized, evidence-based treatment options. These data may impact the physicians' and patients' treatment decisions at one point in time, but collection of longitudinal data would allow for correlation of treatment recommendations to clinical outcomes. In addition, longitudinal data could support collaborative investigator-initiated research that may be focused on using molecular profiling as a tool to improve treatment selection and associated outcomes for patients with cancer.

Proposed Solution

The Caris Registry employs a scientifically valid and regulatory-compliant protocol that is intended to capture clinical disease, treatment, and outcome data over the course of five years from patients who have had Caris Molecular Intelligence™ Services performed. Medical history, disease status, treatments, and outcomes are captured at enrollment (defined as the date of the report) and every 9 months for 5 years. The registry is maintained as a limited data set and all biological and laboratory data is de-identified.

During the planning phase of the registry, the sponsor elected to use standardized data elements wherever possible, in order to maintain flexibility and to anticipate multiple future uses of registry data. The National Cancer Institute's Cancer Data Standards Registry and Repository (caDSR) standardized data dictionary contains common data elements (CDEs) that can be reused for multiple purposes. The registry used some of these CDEs exactly as they appear in the caDSR (e.g., demographics). Other data elements that the sponsor wished to collect were not present in the caDSR (e.g., "Did the patient receive molecular-guided therapy?"). For these elements, the sponsor collaborated with the Center for Biomedical Informatics and Information Technology group to create new CDEs that were incorporated into the caDSR data dictionary. Of the 100 clinical data elements in the registry, 87 were incorporated directly from the caDSR data dictionary and 13 were added to the data dictionary through collaboration with the National Cancer Institute.

Results

To date, 1,400 patients from 96 centers across the United States have been enrolled in the Caris Registry. At least 1,124 of these patients have followup data capturing disease status, treatments and clinical outcomes and 500 of those have completed end of study reports capturing vital status and cancer related deaths.

In the first half of 2013, Caris agreed to participate in a retrospective study of registry data titled, "A Retrospective Investigation To Evaluate The Use Of Target Now™ Assay in Selecting Treatment in Patients with Advanced Stage Metastatic Cancer."

Key Point

Common data elements endorsed by recognized standards organizations are available for registry planners and may be useful for registries in some disease areas. Use of CDEs can increase opportunities for standardized collaboration, linkage, and additional exploratory analysis.

Case Example 7. Using standardized data elements in a registry (continued)

For More Information

Sanders S, Schroeder W, Wright A, et al. The Caris registry: Building a biomarker-focused database to advance patient care. Abstract presented at the 2012 ASCO Annual Meeting. June 1–5, 2012. https://asco.confex.com/asco/2012/sci/papers/viewonly.cgi?username=97155&password=340328. Accessed August 27, 2012.

http://www.carislifesciences.com/caris-registry

References for Chapter 4

1. Wattigney WA, Croft JB, Mensah GA, et al. Establishing data elements for the Paul Coverdell National Acute Stroke Registry: Part 1: proceedings of an expert panel. Stroke. 2003 Jan;34(1):151-6. PMID: 12511767.
2. Good PI. A manager's guide to the design and conduct of clinical trials. New York: John Wiley & Sons, Inc.; 2002.
3. Cannon CP, Battler A, Brindis RG, et al. American College of Cardiology key data elements and definitions for measuring the clinical management and outcomes of patients with acute coronary syndromes. A report of the American College of Cardiology Task Force on Clinical Data Standards (Acute Coronary Syndromes Writing Committee). J Am Coll Cardiol. 2001 Dec;38(7):2114-30. PMID: 11738323.
4. McNamara RL, Brass LM, Drozda JP, Jr., et al. ACC/AHA key data elements and definitions for measuring the clinical management and outcomes of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Data Standards on Atrial Fibrillation). Circulation. 2004 Jun 29;109(25):3223-43. PMID: 15226233.
5. Radford MJ, Arnold JM, Bennett SJ, et al. ACC/AHA key data elements and definitions for measuring the clinical management and outcomes of patients with chronic heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Heart Failure Clinical Data Standards): developed in collaboration with the American College of Chest Physicians and the International Society for Heart and Lung Transplantation: endorsed by the Heart Failure Society of America. Circulation. 2005 Sep 20;112(12):1888-916. PMID: 16162914.
6. National Cardiovascular Disease Registry. "ICD Registry Data Collection.". https://www.ncdr.com/webncdr/icd/home/datacollection. Accessed July 16, 2013.
7. National Cancer Institute. Cancer Data Standards Registry and Repository (caDSR). http://ncicb.nci.nih.gov/infrastructure/cacore_overview/cadsr. Accessed August 15, 2012.
8. National Cancer Institute. CTEP Common Data Elements. https://wiki.nci.nih.gov/display/caDSR/CTEP+Common+Data+Elements. Accessed July 16, 2013.
9. The North American Association of Central Cancer Registries. Data Standards & Data Dictionary (Volume II). http://www.naaccr.org/StandardsandRegistryOperations/VolumeII.aspx. Accessed August 15, 2012.
10. The American College of Surgeons Commission on Cancer. National Quality Forum Endorsed Commission on Cancer Measures for Quality of Cancer Care for Breast and Colorectal Cancers. http://www.facs.org/cancer/qualitymeasures.html. Accessed August 15, 2012.
11. DeVivo M, Biering-Sorensen F, Charlifue S, et al. International Spinal Cord Injury Core Data Set. Spinal Cord. 2006 Sep;44(9):535-40. PMID: 16955073.
12. National Center for Injury Prevention and Control. DEEDS – Data Elements for Emergency Department Systems. http://www.cdc.gov/ncipc/pub-res/deedspage.htm. Accessed August 15, 2012.
13. National Institute of Neurological Disorders and Stroke. Common Data Elements. http://www.commondataelements.ninds.nih.gov/#page=Default. Accessed August 15, 2012.

14. Centers for Disease Control and Prevention. Vaccines & Immunizations. IIS Recommended Core Data Elements. http://www.cdc.gov/vaccines/programs/iis/core-data-elements.html. Accessed August 15, 2012.

15. American Immunization Registry Association. Immunization Registry Functional Standards. http://www.immregistries.org/resources/standards/functional-standards. Accessed July 16, 2013.

16. Clinical Data Interchange Standards Consortium. Clinical Data Acquisition Standards and Harmonization (CDASH). http://www.cdisc.org/cdash. Accessed August 15, 2012.

17. Biomedical Research Integrated Domain Group (BRIDG). http://www.bridgmodel.org. Accessed August 15, 2012.

18. HL7. HL7 and CDISC mark first anniversary of renewed associate charter agreement, joint projects result from important healthcare-clinical research industry collaboration [press release]. http://www.hl7.org/documentcenter/public_temp_A7EA2C6E-1C23-BA17-0CA05816D86FD311/pressreleases/20051012b.pdf. Accessed August 15, 2012.

19. Kim K. iHealthReports. California HealthCare Foundation. Clinical data standards in health care: five case studies. http://www.chcf.org/topics/view.cfm?itemID=112795. Accessed August 15, 2012.

20. Imel M. A closer look: the SNOMED clinical terms to ICD-9-CM mapping. J AHIMA. 2002 Jun;73(6):66-9; quiz 71-2. PMID: 12066386.

21. Yancy CW, Fonarow GC, Albert NM, et al. Influence of patient age and sex on delivery of guideline-recommended heart failure care in the outpatient cardiology practice setting: findings from IMPROVE HF. Am Heart J. 2009 Apr;157(4):754-62 e2. PMID: 19332206.

22. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC: National Academy Press, Institute of Medicine; 2001.

23. U.S. Food and Drug Administration. Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development and Labeling Claims. December 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed August 15, 2012.

24. Goldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Rev. 1980;2:210-20. PMID: 7000537.

25. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol. 1996 Apr;25(2):435-42. PMID: 9119571.

26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;1(8476):307-10. PMID: 2868172.

Chapter 5. Use of Patient-Reported Outcomes in Registries

1. Introduction

As the medical system refocuses on delivering patient-centered care, the importance of measuring and reporting those aspects of health and well-being that are best described by patients themselves, whether related to disease, treatment, or both is increasingly recognized.[1-4] Discrepancies exist between patient and clinician estimates of both the prevalence and severity of patients' symptoms as well as functional impairments, highlighting the importance of direct patient reporting.[3, 5-9] According to the U.S. Food and Drug Administration (FDA), a patient-reported outcome (PRO) is defined as a measurement based on a report that comes directly from the patient (i.e., the study subject) about the status of a patient's health condition, without amendment or interpretation of the patient's response by a clinician or anyone else.[10] (See Table 5–1.) PROs are a subgroup of patient outcomes. The latter category is more general and refers to any outcome related to a patient, whether reported by the patient or described by a third party (e.g., by an imaging report, laboratory evaluation, or clinician assessment).

Over the past 20 years, an expanding body of literature has demonstrated that PROs are associated with traditional outcomes, such as overall survival[11-16] and tumor response.[17] PROs themselves are also increasingly recognized as valid outcomes (e.g., quality of life [QOL], pain, breathlessness, physical functioning).[18-27] Systematic collection of PROs in clinical trials, patient registries, and usual clinical care is feasible and efficient.[28-32] PROs are more reflective of underlying health status than physician reporting[33] and facilitate discussion of important symptoms and QOL with clinicians.[34] Additionally, they have been shown to serve as supporting documentation,[29] improve symptom management,[35] and potentially impact clinical decisionmaking,[30, 36] all of which are viewed favorably.[30] As a matter of terminology, the term "health-related quality of life" (HRQOL) has emerged as the preferential choice in recent literature, and there are cogent arguments surrounding its use. However, the more general "QOL" reflects the fact that health status affects numerous aspects of daily life and influences overall QOL. Thus, further discussions in this chapter will consistently use the term QOL.

While widespread adoption of PROs as a key component in clinical research has not occurred, there is increasing recognition of their role in complementing traditional clinical and administrative data. To this end, the importance of incorporating PROs into clinical research has been highlighted by a number of national policymaking organizations.[2, 37] Recently, the FDA identified PROs as the regulatory standard for supporting subjective endpoints, like symptoms, in drug approval and labeling, and their updated guidance distributed in December 2009 (hereafter referred to as "the FDA guidance document") provides clear instructions on PRO measurement in drug development trials.[10] While the purposes of PROs in registry studies are not for supporting labeling claims, the guidance provided by the FDA has helped refine the definition of PROs and expand the sphere of interest surrounding their use. Most importantly, the FDA guidance document[10] has established a benchmark, albeit a high one, for PRO data and has been the focus of much recent PRO-related literature (references too numerous to list). For this reason, the standards set by the FDA are heavily referenced in the following discussion.

Presently, there are no evidence-based guidelines for inclusion of PROs in registries; consequently, there is substantial heterogeneity in capture and reporting of PROs in this setting (see, for example, the review about some large registries in rheumatoid arthritis).[38] Recent initiatives to define how PROs should be used in oncology comparative effectiveness research (CER) are instructive,[39] as they reflect current, collaborative opinions of many different stakeholders, and may serve as a template for inclusion of PROs in registries (Table 5–2).

Table 5–1. Definitions of commonly encountered terms within PRO-related literature

Term	Definition
Ability to detect change	Evidence that a patient-reported outcome (PRO) instrument can identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept.[10]
Clinician reported outcome (ClinRO)	Outcomes that are either observed by the physician (e.g., cure of infection and absence of lesions) or require physician interpretation (e.g., radiologic results and tumor response). In addition, ClinROs may include formal or informal scales completed by the physician using information about the patient.[40]
Concept	The specific measurement goal, or the thing that is measured by a PRO.[10]
Conceptual framework	Explicitly defines the concepts measured by the instrument in a diagram that presents a description of the relationships between items, domain (subconcepts), and concepts measured and the scores produced by a PRO instrument.[10]
Construct validity	The degree to which what was measured reflects the a priori conceptualization of what should be measured.[41]
Content validity	The extent to which the instrument actually measures the concepts of interest.[42]
Criterion validity	The extent to which the scores of PRO measure reflect the gold standard measure of the same concept.[10]
Domain	A subconcept represented by a score of an instrument that measures a larger concept that includes multiple domains.[10]
HRQOL	The subjective assessment of the impact of disease and treatment across the physical, psychological, social, and somatic domains of functioning and well-being.[43]
Instrument	A means to capture data (e.g., a questionnaire) plus all the information and documentation that supports its use. Generally, this includes clearly defined methods and instruction for administration or responding, a standard format for data collection, and well-documented methods for scoring, analysis, and interpretation of results in the target population.[10]
Item	An individual question, statement, or task (and its standardized response options) that is evaluated by the patient to address a particular concept.[10]
Item bank	A comprehensive collection of questions (and their response options) designed to measure an underlying construct across its entire continuum.[44]
Metadata	Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information source.[45]
PRO	A measurement based on a report that comes directly from the patient (i.e., the study subject) about the status of a patient's health condition, without amendment or interpretation of the patient's response by a clinician or anyone else.[10]
Proxy-reported outcome	A measurement based on a report by someone other than the patient reporting as if he or she is the patient.[10]
QOL	An individual's perception of their position in life in the context of the culture and the value systems in which they live and in relation to their goals, expectations, standards, and concerns. It is a broad-ranging concept affected in a complex way by the person's physical health, psychological state, level of independence, social relationships, personal beliefs, and relationship to salient features of the environment.[46]
Recall period	The period of time patients are asked to consider in responding to a PRO item or question.[10]

Table 5–1. Definitions of commonly encountered terms within PRO-related literature (continued)

Term	Definition
Reliability	The ability of an instrument to yield the same result on serial administrations when no change in the concept being measured is expected.[42]
Scale	The system of numbers of verbal anchors by which a value or score is derived for an item. Examples include VAS, Likert scales, and rating scales.[10]
Score	A number derived from a patient's response to items in a questionnaire. A score is computed based on a prespecified, validated scoring algorithm and is subsequently used in statistical analyses of clinical results.[10]

Table 5–2. Example of guidelines for PRO incorporation into product-labeling claims in oncology[a]

Selection of Measures	
1.	Include PROs in all prospectively designed CER and post-marketing studies in adult oncology (including registries, observational cohorts, and controlled trials).
2.	Include systematic assessment of the following 14 patient-reported symptoms (the "Core" symptom set) in all CER and postmarketing clinical studies in adult oncology: anorexia, anxiety, constipation, depression, diarrhea, dyspnea, fatigue, insomnia, mucositis, nausea, pain, sensory neuropathy, rash, vomiting.
3.	Include additional patient-reported symptoms as appropriate to a specific study's population, intervention, context, objectives, and setting (in addition to the Core symptom set), and incorporate a process that allows individual patients to report unsolicited symptoms.
4.	Measure QOL, either via a single-item or multi-item questionnaire, in all prospective CER and post-marketing clinical studies. Inclusion of a measure that enables cost-utility analysis is encouraged.
5.	Selected measures to assess symptoms or QOL should have demonstrated content validity (based on direct patient input), criterion validity, reliability, and sensitivity in the intended patient population (including assessment of the meaningfulness of specific score changes and the ability to detect change over time), as well as an appropriate recall period. Translations from other languages should be conducted in accordance with existing methodological standards.
Implementation Methods	
6.	Limit PRO data collection so that the average patient can complete the process within 20 minutes at the initial (baseline) visit and within 10 minutes at any subsequent time points.
7.	Collect PROs as frequently as necessary to meet research objectives without overburdening patients. When using PROs to assess potential treatment benefits, collection of PROs at baseline and following treatment completion or study withdrawal as well as at selected long-term time points should be considered a minimum standard. When using PROs to assess treatment toxicities/harms or comparative tolerability, more frequent assessment is merited, such as at baseline and every 1-4 weeks during active therapy as well as at selected long-term time points.
8.	Collect PROs via electronic means whenever possible.
9.	Establish measurement equivalence when mixing modes of PRO measure administration in a study (e.g., Web, telephone/interactive voice response [IVRS], handheld device, and/or paper).
10.	Employ methods to minimize missing PRO data, including education of local site personnel, training of patients, and real-time monitoring of adherence with backup data collection.

Table 5–2. Example of guidelines for PRO incorporation into product-labeling claims in oncology (continued)	
Data Analysis and Reporting	
11.	Include in the protocol a plan for analyzing and reporting missing PRO data.
12.	For each PRO measure, report the proportion of patients experiencing a change from baseline demonstrated as being meaningful to patients.
13.	Evaluate the cumulative distribution of responses for each PRO measure and include cumulative distribution curves in reports and publications.
14.	Include a mechanism for alerting clinical staff in real-time about symptoms of concern reported by patients during study participation.
15.	Analyze and publish results of PRO analyses simultaneously with other clinical outcomes.

[a] Adapted from the Center for Medical Technology Policy (CMTP), Effectiveness Guidance Document: Recommendations for Incorporating Patient-Reported Outcomes into the Design of Post-Marketing Clinical Trials in Adult Oncology.[39] Used with permission. Copyright restrictions apply.

2. The Role of PROs in Registries

2.1 Relationship Between PROs and CER

CER was recently defined by the Institute of Medicine as:

> "… the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels."[47]

Central to this definition is that the information generated by CER should assist consumers of health care (i.e., patients) in making decisions. Of great interest to patients are factors like QOL, symptom burden, and functional status, which are best described directly by patients, thereby implicitly emphasizing the importance of PROs to CER.[48, 49] The strength of this relationship is furthered by the term patient-centered outcomes research (PCOR), which has emerged after passage of the Patient Protection and Affordable Care Act that established the Patient-Centered Outcomes Research Institute (PCORI). According to PCORI:

> Patient-centered outcomes research helps people make informed health care decisions and allows their voice to be heard in assessing the value of health care options. This research answers patient-focused questions: (1) "Given my personal characteristics, conditions and preferences, what should I expect to happen to me?" (2) "What are my options and what are the benefits and harms of those options?" (3) "What can I do to improve the outcomes that are most important to me?" (4) "How can the health care system improve my chances of achieving the outcomes I prefer?"[50]

By definition, PCOR is impossible to pursue without including the patient voice, and PROs are an important tool for capturing the patient voice. As PCOR is effectively a subset of CER (and will not be referred to independently from this point), PROs are therefore critical components of CER as well. The importance of PROs in CER is highlighted by the interest in the patient experience of the multiple stakeholders who ultimately use results of CER.[49]

2.2 Relationship Between CER and Registries

While clinical trials are generally felt to represent the gold standard of evidence to support clinical decisions, many clinical trials are conducted under conditions that limit generalizability or do not

emphasize factors that are important to patients and clinicians in the course of actual practice. Clinicians and patients face challenging decisions regarding treatment choices and toxicity profiles that are unaddressed by traditional clinical trials, and these are exactly the types of questions that CER is intended to address. Registries are important tools for answering such questions. They can evaluate effects in a more "real-world" population, improving generalizability. In uncommon diseases, where traditional clinical trials are unrealistic because of small numbers, registries can help fill the information void on any number of issues, including treatment options and responses, natural history, and QOL. Registries can be designed to answer specific questions that affect clinical practice but were unaddressed by pivotal clinical trials. Importantly, when partnered with electronic health records (EHRs), registries can capitalize on the massive amounts of data collected as part of routine clinical care to create data sets that more realistically represent the array of inputs that clinicians and patients assimilate in almost every clinical encounter. Electronic PRO instruments that are directly incorporated into routine clinical care, and thus directly into an EHR, are potentially important sources of PRO data for registry studies. Collection and analysis of such data sets, in the form of registries, offers the opportunity to inform clinical care in ways that are meaningful to all stakeholders in the health care system.

2.3 Importance of PROs in Registries

Given the centrality of PROs to CER and the role of registries in CER, the importance of PROs to registries is apparent. Inclusion of PROs in prospectively collected registries is almost always appropriate. PROs contribute information across the spectrum of registry purposes described in Chapter 1: describing the natural history of disease, determining effectiveness, measuring or monitoring safety or harm, and measuring quality. As one reads down the list of nominated purposes of registries, the substantive role of PROs in registry design becomes increasingly clear.

2.3.1 Describing the Natural History of Disease

A requirement of registries intended to describe natural history of disease is adequate information about symptom burden and related QOL trajectories, especially in the setting of rare diseases, inherited diseases with increasing life span (e.g., cystic fibrosis, sickle cell disease), and heterogeneous diseases (e.g., chronic obstructive pulmonary disease, breast cancer). Registries can provide useful information on the expected course of health even in the absence of treatment—information which could provide useful information regarding the need for and timing of treatment. Understanding how new therapies impact patient experience can also be captured under this rubric. For example, metastatic renal cell carcinoma is a relatively uncommon malignancy for which the FDA has approved six targeted therapies within the past decade. All have different toxicity profiles and different symptom alleviation profiles; insufficient information can be derived from the pivotal clinical trials to develop optimal strategies for the sequencing and timing of these therapies.[51] Registries of patients receiving routine care with these different agents (i.e., "real-world" registries), especially when containing PRO data, can help inform sequencing, timing, and impact of treatments, providing critical information where there is an explosion of treatment options but a dearth of comparative information.

2.3.2 Determining Effectiveness

In registries designed to determine effectiveness, PROs also figure prominently, especially considering the importance placed upon the patient experience as a meaningful outcome in the IOM's definition of CER. Beyond traditional outcome measures such as overall survival and risk reduction, QOL is a valid marker of efficacy by itself and is best captured by PRO measures. Patient-reported symptoms can be indicators of adverse consequences of therapy (e.g., toxicity monitoring), targets for meaningful intervention (e.g., symptom control intervention), and means of understanding how patient perceptions of toxicities or effectiveness impact effectiveness (e.g., through adherence behavior). Consider a prospective

registry intended to support CER for the management of early-stage prostate cancer. For these patients, differentiating between and comparing surgery and radiation is best achieved from patient-reported information on symptoms of radiation proctitis, sexual health, pain, and urinary function, as well as the relationship of these factors to overall QOL and patient preference.

Within the area of toxicity monitoring, PROs are likely to take a place on center stage. The National Cancer Institute has recently developed a patient-reported version of its Common Terminology Criteria for Adverse Events, PRO-CTCAE,[52] for use in cancer clinical trials. Pharmacovigilance studies provide another fertile area for PRO implementation. Perhaps even more powerful are efforts to link PROs to genomic and proteomic data in order to understand the biologic basis for toxicity phenotype. Registries intended for safety monitoring offer potential for a much more robust understanding of long-term safety than typical clinical efficacy trials, and when coupled with data on effectiveness may help answer difficult questions such as "Was the intervention worth it?" especially as viewed through the patient's lens.

2.3.3 Quality Measurement

Registries intended to measure quality can incorporate PROs in numerous ways, and PROs can contribute to quality assessment. In some instances, established quality standards do not exist, and registries can be used to establish realistic and acceptable standards. For example, there is an impetus to initiate quality monitoring in palliative medicine programs, but the evidence base is insufficient to establish benchmarks to define quality.[53] In such a setting, registries incorporating PROs would serve an important role in establishing definitions for quality, and could then be used in real-time to monitor quality. Some quality metrics focused on the patient experience already exist. For example, in the American Society of Clinical Oncology's Quality Oncology Practice Initiative assessment and management of pain, nausea/vomiting and dyspnea are core metrics; this requires both PRO assessment and response to findings.[54, 55]

2.4 PROs in Prospective Registries Versus Retrospective Studies

Having established the role for PROs across a spectrum of registries, it is important to consider the roles of PROs in prospective registries and retrospective studies. Patients' experiences are transient and are best captured "in the moment." They cannot be recreated or recalled precisely—a limitation that highlights the need to routinely and systemically capture PROs for prospective registries. Further, abundant evidence demonstrates that third party assessments (most notably those of clinicians) do not adequately reflect patients' subjective experience with their care.[7, 8, 28, 56] For example, in patients with lung cancer receiving chemotherapy, Basch and colleagues showed that, when compared with physician assessments, patient reports of symptoms were more reflective of daily health status as measured by EuroQoL EQ-5D.[33] As rapid-learning health care systems[57-59] become standard, routine capture of longitudinal and systematic PROs will happen as part of routine care, thereby making it easier to prospectively capture PROs for registry support.

In contrast to prospective registries, which can be designed to collect PROs as data accrue, studies constructed by manual chart extraction or from EHR queries should not attempt to retrospectively add PRO data that was not originally collected. Additionally, researchers should not ask patients to provide recalled/recreated PROs for missing data in such studies, as this may introduce recall bias. The exact length of time over which recall bias develops is unclear, and seems to vary for different experiences.[60] For pain, single-item assessments reflecting the prior week do not seem to represent actual pain levels or a mean of daily pain levels collected for the same 1-week period.[61] Thus, asking patients to precisely recall their symptom experience associated with a clinic visit at some arbitrary point in the past is fraught with pitfalls.

2.5 Other General Considerations on Inclusion of PROs in Registries

Including PROs in registries offers numerous advantages. First, incorporation of the patient voice helps keep care and research patient-

centered, acknowledging the balance and tension between traditional outcomes and PROs. Further, symptom burden, QOL, and satisfaction with care are dynamic variables that cannot be recreated accurately through retrospection; they are essentially lost if not captured "in the moment." For this reason, routine, systematic, and longitudinal collection is recommended and should be a standard of practice. The importance of longitudinal collection cannot be overstated; it allows patients to serve as their own controls; that is, each patient serves as his or her own comparator over time. Changes from baseline are tracked over time and linked to other interventions, such as initiation or discontinuation of a drug; or to other outcomes, such as change in disease status (e.g., cancer progression, cardiac event). Serial PROs address a number of critical issues. They: (1) improve our understanding of the trajectory of individual patient's symptom burden and QOL over the course of disease (or treatment); (2) remind clinicians of the variability between patients; (3) provide information on the value that the individual patient places on their health state; and, (4) are central to the efforts of CER, pharmacovigilance studies, and quality monitoring. When routine and systematic collection of PROs is incorporated into registries, the health care community can improve the efficiency of routine care through support of billing and clinical documentation functions.

Certainly, including PROs in registries poses challenges. Collection of PROs can generate significant amounts of data and adds another layer of complexity to already complex data sets. Clinician acceptance may lag slightly for several reasons.[62] Although the history (patient reports filtered through a clinician's lens) and physical exam are central to clinical diagnosis and decisionmaking,[63] long-standing and deeply ingrained beliefs persist that clinician assessment alone is objective and unbiased, casting doubt upon the value and validity of unfiltered, direct patient reports. Regardless, collection of PROs generates more data for clinicians to consider and incorporate into care, which could be viewed as onerous and burdensome, especially since PROs are not yet ubiquitous or the standard of care. More importantly, it is largely unclear how PROs collected within the context of clinical research should be used to inform care and change daily practice patterns. Without appropriate infrastructure for responding to critical reports, collection of PROs may pose a liability if critical data do not receive appropriate clinician attention and response. For example, significant liability could result if a patient reports a constellation of symptoms known to be strongly associated with suicidal behavior and there is inadequate clinical intervention. Further, it is possible that PROs could lead to decreased satisfaction with care if patients expect that their PROs will be reviewed and addressed, but they are unmet or unacknowledged in the clinical encounter.[29]

3. What Methods Are Available To Collect PROs and Which Is Best?

Often, choice of PRO instrument and mode of administration are considered jointly; however, they need not be, as administration methods simply provide a platform for collecting and presenting information. There are two main ways of collecting PRO data—on paper and electronically.

3.1 Paper-Based Methods

Historically, PROs were collected via paper forms and were developed based on this collection method. From a practical standpoint, collection of PRO data via paper-based methods is relatively straightforward. After selecting the instrument(s) to be used (discussed further in Section 4), consistency is the guiding principle. Items should be presented in the same order for every collection. If the PRO measurement selected is a single-item tool, this is automatic, but if multiple instruments are employed, presenting them in the same order is important. Patients should complete forms in a confidential space, without fear that "wandering eyes" will see responses. Once forms are completed, they should be reviewed multiple times for completeness. For those instruments completed in clinic, this review should be done by staff collecting the instruments, nurses involved in patient intake and rooming, and clinicians

reviewing responses. Once forms are submitted to the research team for data entry, completeness should be reassessed. Patients who fail to complete a pre-defined percentage of questions (there is no consensus on an acceptable percentage), should receive a followup telephone inquiry to attempt to minimize missing data. Finally, data should be entered into electronic forms using double data entry or similar techniques to enhance transcription accuracy. These techniques should ideally be augmented with near real-time exploratory analyses to examine the believability of the data within the clinical context.[64]

Paper forms are the historical gold standard for PRO collection. For this reason, patients are inherently familiar with them. Their use is not limited by unfamiliarity or unease with new technologies, although unfamiliarity with new technology dissipates quickly and patients are increasingly familiar with technology as advances continue to disseminate. Moreover, paper forms do not require significant upfront capital investment, in terms of devices or software. There are many measurement instruments across a variety of disease states that have been extensively evaluated and are available for immediate use.

However, paper forms have many limitations. They require research personnel to sort, distribute, and collect, introducing risk for inconsistencies and a source of ongoing cost. Paper forms collected as part of routine/scheduled clinic visits are generally straightforward, but this approach systematically misses participants unwilling or unable to attend a clinic appointment. Collection between visits is logistically difficult with paper forms; delivery of the paper forms either requires that participants take paper booklets home with them or that research personnel coordinate timely delivery of booklets through the postal service. With either approach, obtaining a time/date stamp for at-home, paper-based administration remains a challenge. Relying on at-home paper booklets risks participants completing multiple days of reporting all at once (i.e., the so-called "parking lot" effect[65] in which all responses for the past month are completed immediately before a visit while sitting in the parking lot). Paper forms often include illegible or uninterpretable responses and require manual data entry, which is administratively burdensome and subject to transcription errors. Manual entry also generates a lag time in monitoring response rates, complicating the process of reducing missing data.[66] Overall, there is a threshold beyond which the continuing data collection and quality assurance costs of paper-based PROs surpass the upfront technology costs for electronic data capture, making electronic PROs the more efficient and reliable approach.

3.2 Electronic Capture Methods

With the advent of portable and more cost-effective electronic capture methods, the presence of such methods within the literature has grown. Similar to traditional paper-based collection, electronic collection begins with instrument(s) selection. Integral to the choice of instruments is the choice of platform, as not all instruments are tested across multiple platforms, nor is every instrument amenable to every platform. Electronic PRO (ePRO) capture has been demonstrated on a variety of platforms, including Web-based, electronic tablets, interactive voice response system (IVRS), handheld device, and digital pen. For ePRO collection using tablet computers or handheld devices in the clinic setting, patients are provided the device at the time of check-in to clinic with pre-loaded PRO measures such that patients simply select their response to each item as it is presented. With the digital pen, patients select responses on a specially designed paper survey, with responses electronically recorded by the pen. With IVRS, patients call a telephone number and are prompted, via an automated transcript, to select a preferred language, provide an identifier and then are guided through the PRO measure, providing verbal responses to each item. Access to Web-based platforms can be provided at "confidential" computer stations in clinic waiting rooms, or in the exam room itself, as well as from any web-enabled device including home computers, handheld devices, and mobile telephones. Regardless of platform, data are transmitted to a central, secure repository immediately upon submission and can be accessed for "real-time" incorporation into routine care, if desired. Both Web-based and IVRS collection platforms can extend beyond the clinic and capture

PROs between visits. Factors influencing platform selection include budget and technical support, technology literacy of the registry's target population, collection logistics (in-clinic, between-visit, or combination), and the instrument(s) chosen.[2, 66]

Electronic methods of PRO capture have been widely shown to be feasible in a variety of practice settings, disease states, and age ranges.[29, 30, 67] Recently developed PRO measures have either been created specifically for electronic data capture or include features to capitalize on electronic capture technologies, such as the Patient-Reported Outcomes Measurement Information System (PROMIS),[68-70] the PRO-CTCAE,[8, 52] and the Patient Care Monitor, version 2 (PCM).[71] The PROMIS and PRO-CTCAE tools take advantage of electronic functionalities such as skip logic or computerized adaptive testing, which can reduce the number of items patients have to complete, while the PCM also fulfills clinical documentation needs for clinical review of systems and provides triggers for accompanying patient education.

As to obtaining hardware or software for these purposes, hardware often requires an upfront investment. Again, the size of the investment depends largely upon the scope and scale of the registry. Some software packages are publicly available (e.g., PROMIS Initiative items) while others are proprietary. Third-party commercial vendors specializing in design and implementation of PROs offer a variety of products. The decision to involve a commercial vendor depends upon factors like the rationale for including PROs in the registry, the size of the registry and number of involved sites, local technological expertise and support, whether the data will be collected as part of routine care or just for research purposes, and the degree of psychometric analysis needed. Although registry studies are not viewed as sufficiently rigorous for product labeling, exploratory analyses of PROs from a registry may serve as the basis for a subsequent trial for labeling purposes, in which case having a sound PRO measure in the registry could simplify the trial process. In such a scenario, it would be prudent to use a commercial vendor to ensure adequate audit trails and compliance with all FDA guidance for PROs. Alternatively, consider, for example, a health care system with an extensive EHR that plans a registry to monitor the impact of a series of clinical pathways to lessen the debilitation following major abdominal surgery. Such an organization may elect to develop or modify a PRO system to be directly integrated with their EHR without involving an ePRO vendor.

Compared with paper methods, delivery of ePROs can be automated, minimizing the risk of inconsistent presentation of materials or mishandling paper forms. Electronic collection of responses provides immediate and accurate time/date stamps, and facilitates real-time monitoring of response rates and review for missing data.[66] Additionally, electronic platforms may provide a safer environment for patients to disclose sensitive concerns, such as sexual function.[72]

Not all PRO measures were developed for, or have been tested on, electronic administration platforms. The transition of paper-based measures to electronic platforms is referred to as "migration," and guidelines were recently developed to assess the equivalence of measures that have migrated from one collection mode to another.[73] In general, paper-to-electronic migration yields between-mode equivalence comparable to the test-retest reliability of the original mode, but this is not always the case and should be tested.[74] When incorporating a migrated PRO measure into a registry, registry developers should verify that the ePRO measure has demonstrated validity in the intended mode of administration or reasonable equivalence with the mode for which validity, reliability, and sensitivity were initially demonstrated.[39]

Although electronic capture provides substantive advantages over paper-based methods, enthusiasm must be tempered on several fronts. First, completion of electronically delivered PRO measures requires some level of comfort with and access to newer technologies, which may prove challenging in certain situations. For example, in rural areas, using Web-based methods to collect PROs between visits may be impractical due to unpredictable Internet access, while some geriatric populations may be uncomfortable with tablet or

handheld technologies. Second, if paper-electronic equivalence has not already been verified for a migrated PRO instrument, the process of documenting equivalence can be time-consuming and expensive. Finally, electronic methods require greater up-front investment in terms of the devices and software, electronic storage (meeting appropriate security standards), training, and technical support. Depending upon the scale of the registry, these issues may render electronic methods too burdensome.

Software selection is a common question. While outside the scope of this chapter, some broad advice can be provided. First, there are many companies that offer software to collect ePROs. Publically available software is also in production (e.g., PROMIS) or being developed (e.g., ePRO CTCAE). The software solution itself is relatively simple and expensive systems are not needed, unless specific features are required (e.g., the requirement to be compliant with the FDA's CFR Part 11). Software should be from a credible vendor, with available security features that will support compliance with applicable privacy and security laws.

In general, patients should report few items or ideally one item per screen, the screen should be clear and move to the next item when the answer is provided, and there should not be any software delays between questions. Visually, the software should present questions and response "buttons" in large enough font for easy reading by mildly visually impaired individuals. Validation code and verifications should be built into the software, as well as any required clinical triggers. The software should be easily adaptable, and easily integrated into the registry workflow. Reports (e.g., for clinicians) should be visually appealing, efficient, and informative. Whenever possible, software should connect into the EHR workflow, including embedding data into the EHR for clinical documentation and/or contributing to an enterprise data warehouse.

Finally, it is important to ensure that the software has been tested before full-scale implementation with the registry. Testing documentation should be requested from the vendor, who should have completed it. Both usability and feasibility should be considered, and testing should be conducted with the population planned for the registry. As elaborated on http://www.usability.gov/, usability is not a single, one-dimensional property of the interface, but rather a synthesis of the following elements:

- *Ease of learning*: How fast can a user who has never seen the user interface before learn it sufficiently well to accomplish basic tasks?
- *Efficiency of use*: Once experienced users have learned to use the system, how fast can they accomplish tasks?
- *Memorability*: If a user has used the system before, can he or she remember enough to use it effectively the next time or does the user have to start over again learning everything?
- *Error frequency and severity*: How often do users make errors while using the system? How serious are these errors, and how do users recover from these errors?
- *Subjective satisfaction*: How much does the user like using the system?

The degree of usability testing should match the complexity of the task. For an ePRO system, this process minimally includes documentation of respondents' ability to navigate the electronic platform, follow instructions, and answer questions, with an overall goal of demonstrating that respondents can complete the computerized assessment as intended. Generally, fewer than ten representative patients are required to verify usability. If the system is not usable, then it should be iteratively updated until it is usable.

Feasibility extends usability and establishes the practical implementation of the software system in the local setting (e.g., clinic, home, hospital). Assessment approaches are similar and the software goes through iterative updates until feasible. During this process, patients can contribute critical advice for the "help" manual and instruction sets.

Although most often associated with questionnaire development, cognitive debriefing is also appropriate for usability and feasibility assessment through verbal probing by the interviewer (e.g., "What does the instruction 'skip item' mean to you

here?") and "thinking aloud" in which the interviewer asks the respondent to verbalize whatever comes to mind as they conduct a task. Incorporated in usability and feasibility testing, cognitive debriefing helps to assess whether the ePRO system influences the way respondents interpret the questions, decide on an answer, and respond. In addition, it can help to determine whether the instructions are clear or if anything is confusing.

3.3 Which Method Is Best?

As with most other aspects of involving PROs in registries, the choice of PRO capture method is highly dependent upon the design and purpose of the registry. Both paper-based and electronic platforms offer advantages and disadvantages, as outlined above. Ideally, when either method is shown to be valid for an instrument, both methods of PRO data collection should be available in a study. Providing an interface familiar to or preferred by particular patients or populations may reduce missing data not at random. Modes may be mixed across patients in a study (e.g., each patient selects a specific mode at baseline and continues to report via that mode throughout a study), or within patients (e.g., a patient reports by Web until he becomes symptomatically ill, at which point IVRS becomes preferable). One mode may be preferred at a particular site, for example in multinational studies where IVRS or Web access are heterogeneous across countries. "Real-world" registries are likely to enroll patients from a variety of settings (e.g., home, hospital, assisted living facility) and circumstances (e.g., independent, caregiver-assisted), such that flexibility in mode of administration facilitates capturing a broad mix of patients. Mixing modes is generally viewed as acceptable if a reasonable level of between-mode equivalence has been demonstrated.[39]

In general, electronic capture is preferred to paper because of its flexibility and its ability to reduce the chance that the PRO data in a registry will be missing. In contemporary research, paper methods are usually most cost-effective until registries start to grow in size or number of sites. When the registry is going to be intentionally small (e.g., fewer than 100 patients), paper methods will likely suffice. When the registry is going to be large, upfront investments in electronic approaches will realize substantial downstream gains in efficiency, cost, and data quality. Regardless of the ultimate choice of administration method, clear documentation of the rationale for the choice and clear evidence of appropriate psychometric assessment is strongly recommended. Assistance with this process may be available from internal expertise (as in many academic institutions) or may rely upon input from a commercial vendor, whose involvement can range from consulting only to nearly full control of the development and implementation process.

4. Which PRO Measure(s) Should Be Selected?

The process of choosing which PRO measure(s) to include in a registry can be challenging, largely because the plethora of available measures is overwhelming. In 2007, a PubMed search for PRO instrument development articles since 1995 resulted in more than 2000 citations.[75]

Existing PRO measures assume a variety of forms:

- General assessment scales (e.g., health-related QOL)
- Disease-specific scales (e.g., chronic obstructive pulmonary disease, cancer [including scales for individual tumor types], arthritis, or psoriasis)
- Symptom-specific scales (e.g., pain, breathlessness, distress)
- Evaluations of functioning across a variety of domains (e.g., physical, social, emotional)
- Scales assessing satisfaction with care received
- Other (e.g., adherence with therapy)

Some PRO measures are extensive, with dozens of items related to a single concept (e.g., breathlessness), while others have 80 or more items reflecting many different patient-reported concerns constituting an entire clinical review of systems, and yet others are single-item instruments measuring a single construct in a single question.

Further, there is extensive literature describing the important characteristics (i.e., conceptual framework, content validity, reliability, ability to detect change) of PRO measures, but consolidating this information into practical guidance for selecting among existing PRO measures is difficult. The FDA guidance document has outlined a standard for evaluating PRO measures for labeling claims that encompasses the salient points regarding development history, conceptual framework, and psychometric evaluation.[10] The standards outlined by the FDA may be more stringent than is necessary for certain registry purposes, but nevertheless serve as an important and well-conceived framework for discussion and conform to accepted best practices.[28] While a comprehensive review of PRO development and psychometric evaluation is beyond the scope of this chapter, below is a concise overview of the process and concepts. For more information, several texts provide detailed descriptions.[26, 27, 76]

4.1 Getting Started and the Importance of Clarity

The key to successfully navigating this process is to clearly define the following aspects of the registry:

- Population of interest (e.g., cancer patients receiving radiotherapy for painful bony metastases, individuals with oxygen-dependent chronic obstructive pulmonary disease, children with rhinoconjunctivitis, U.S. veterans with rheumatoid arthritis)
- Outcomes of interest, also known as the concept (e.g., specific symptom severity, overall symptom burden, treatment-related toxicities, physical functioning, social functioning, QOL)
- Intended users of the registry (e.g., clinicians, patient advocacy groups, pharmaceutical companies, insurance companies, governmental agencies)
- The purpose(s) of the registry (e.g., pharmacovigilance, establishment of symptom trajectories, correlation of survival benefit with QOL or symptom benefit).

As with any research activity, a priori specific aims and hypotheses to be tested must be outlined up front, and PRO selection appropriately aligned. Registry studies, in particular, are susceptible to poorly defined outcomes; PRO instruments may be chosen because they are general in nature and capture a broad range of patient-reported concerns, meeting a target goal of demonstrating that PROs are captured rather than capturing specific PRO concepts of interest. If the objectives of the registry, intended hypotheses, and outcomes of interest are clearly defined, the desired characteristics of the PRO instrument become more clearly delineated, facilitating a search of existing measurement instruments.

4.2 Potential Sources for Identifying PRO Instruments

Once these issues are clearly defined, identification of candidate PRO measures can begin in earnest. In general, the process of PRO development is time- and resource-intensive and using existing measures whenever possible is best. It is highly unlikely that any existing instrument will perfectly suit the needs of a registry study, or that a "perfect" instrument can be developed, further underscoring the importance of clearly defining the population, outcomes of interest, and purpose of the registry. Such clarity will allow more appropriate assessment of the relative strengths and weaknesses of existing PRO measures. In many cases, modifications to existing measures will improve the measure for use in a registry. These modifications can include changes in wording or order of questions, adding specific questions, or altering the method of administration. In general, such modifications require some degree of psychometric reassessment, though the degree to which instrument modification requires psychometric reassessment varies and is discussed by Snyder et al.[77]

Traditional literature searches can yield results, but may be quite time-consuming. The Mapi Institute maintains the Patient-Reported Outcome and Quality of Life Instruments Database (http://www.proqolid.org/), allowing users to search a large and relatively comprehensive database for PRO instruments that best address the specific needs

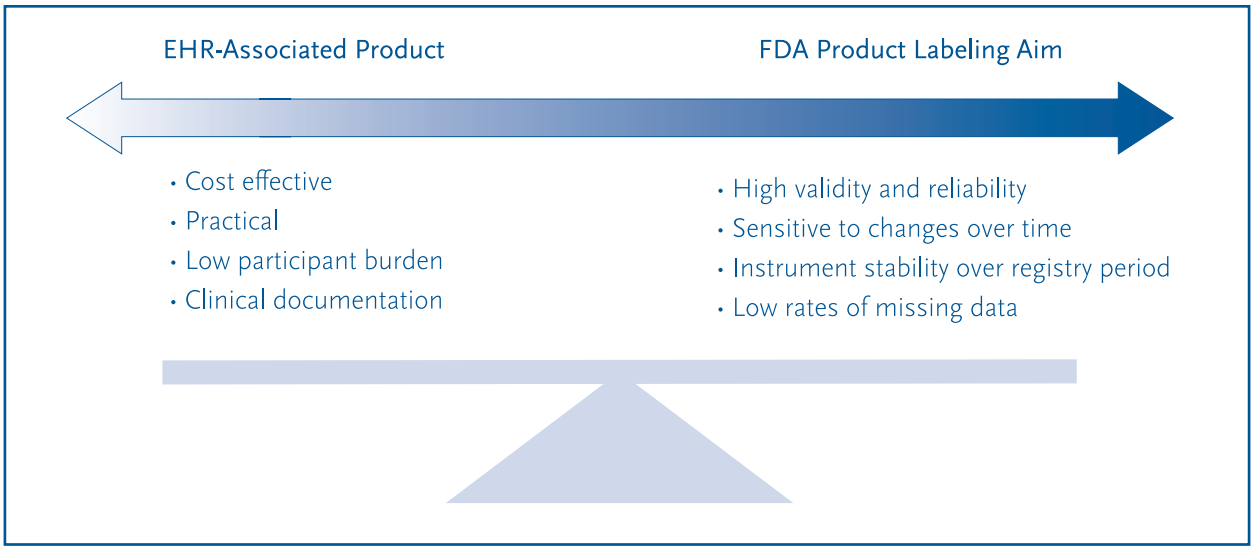
identified. The Online Guide to Quality-of-life Assessment (http://www.olga-qol.com/), is another database of existing QOL instruments. Additionally, the U.S. National Institutes of Health PROMIS Initiative (hhtp://www.nihpromis.org/) has been tasked with developing rigorously tested item banks across a broad range of domains and subdomains (functioning, disability, symptoms, distress, and role participation).[68] The PROMIS Initiative is also actively evaluating methods to achieve brevity in instruments through techniques such as computer adaptive testing. Importantly, these measures are publicly available through the PROMIS Assessment Center (http://www.assessmentcenter.net/). Commercial vendors can also aid in identifying appropriate measures; as with selecting a mode for administering the PRO measure, the decision to involve a commercial vendor is multifactorial, depending on the factors described in Section 3.2.

Item banks represent another option for developing PRO surveys. In general, item banks contain comprehensive collections of items that pertain to a particular construct (e.g., dyspnea).[44] Item banks generally rely on item response theory (IRT), in which the unit of focus is the item rather than the entire instrument. As such, instruments can be constructed using IRT that employ only those items which provide the most useful and relevant information, eliminating questions with little added value, without compromising psychometric qualities.[78] The PROMIS Initiative is an example of an item bank.[44] Item banks may represent the future of PRO collection, but they are currently limited by logistical issues, questions about whether IRT-based item banks represent an improvement over existing PRO instruments, concerns over regulatory acceptance, and limited data about psychometric properties of item banks in specific populations.[44] However, IRT-based item banks represent a promising approach, especially in light of the emphasis on limiting respondent burden.

4.3 Choice of the Best PRO for the Registry

Section 4.4 describes many of the properties of PRO instruments that should be considered when choosing the appropriate instrument for each unique registry scenario. Whether to adhere closely to the conservative FDA recommendations is a frequent question, if not a source of frank tension. While there is no formal avenue through which registries can support product-labeling claims, if the registry is in any way tied to trials with aspirations of product-labeling, then the answer is straightforward and the FDA PRO guidance should be followed. Anchoring the FDA threshold as a "maximally conservative" (and therefore usually least practical) state, there is a continuum of scenarios and a continuum of practical allowances to the ideal state where the need for precision and reduction of bias is balanced with the need for practical solutions and the reduction of missing data. Figure 5–1 shows that the tension between psychometric desirability and logistical considerations of PRO collection in registries requires a careful balance, driven primarily by the goals of the registry. Explicitly outlining the registry objectives, population, outcomes, and intended uses as described in Section 4.1 will help to define where the registry is on the continuum and guide decisionmaking.

Figure 5–1. Psychometric properties and logistical considerations exist along a spectrum



EHR = Electronic Health Record; FDA = U.S. Food and Drug Administration.

4.4 Development History and Conceptual Framework

The PRO development history and conceptual framework are inextricably linked and are discussed in close proximity for this reason.

4.4.1 Development History

The FDA guidance document strongly recommends transparency with respect to development history. "Development history" explicitly refers to the entire process of developing and psychometrically evaluating a PRO measure, including the conceptual framework, item development and revision history, and evidence of patient input. For newly developed PRO instruments, clearly documenting the development history is straightforward and can be integrated into the development process. This is in contrast to using an existing measure, where the development history may be very difficult, if not impossible, to obtain. Ideally, the development history is well vetted in the literature, but if the history is somewhat opaque, the FDA has indicated that demonstration of content validity with specific examples, including direct patient input from the appropriate population, is an acceptable alternative. For newly developed PROs, it is imperative, from an FDA and product-labeling standpoint, that the entire development history be well documented. The cornerstone of the development history is the conceptual framework.

4.4.2 Conceptual Framework

Clear identification of the target population, purpose of the registry, and outcomes of interest greatly facilitates developing a conceptual framework. According to the FDA Guidance document, a conceptual framework "explicitly defines the concepts measured by the instrument in a diagram that presents a description of the relationships between items, domain (subconcepts), and concepts measured and the scores produced by a PRO instrument."[10] Initially, the conceptual framework arises out of expert opinion and literature review. The framework is then refined by qualitative methods of patient input, such as patient interviews and focus groups, which ensures that a priori hypotheses are consistent with patient experiences and descriptions. The conceptual framework will be modified iteratively.[79] For complex concepts, such as breathlessness, multiple domains affect the overall concept, so identifying appropriate domains and then assessing these is paramount to assessing the overarching concept.

4.5 Psychometric Properties

Entire texts are written on psychometrics and there is an extensive literature on psychometric properties of PRO measures. An excellent series arising from the Mayo/FDA Patient-Reported Outcomes Consensus Group focused on PRO development in advance of the anticipated FDA guidance; it was published in a special supplement of the November/December 2007 issue of the journal Value in Health and provides more detailed descriptions of processes and procedures needed to implement PRO systems to meet FDA expectations.[80]

Almost every guideline regarding the use of PRO measures recommends selecting measures that have demonstrated content validity, criterion validity, reliability, and sensitivity (including the ability to detect change over time) in the target population.[39] It is important to note that psychometric properties are not dichotomous and that instruments are not completely "valid" or "reliable." These properties are continuous variables relaying incremental information. Additionally, it is inappropriate to refer to an instrument as "validated," as this simply means it has been subjected to psychometric analysis, but conveys no information regarding the measure's performance.[42] For this reason, instruments are reflected at varying points on our continuum in Figure 5–1 to demonstrate that differing states of reliability and validity may be appropriate depending upon the context of the registry and the PROs to be captured within it. The goal, ultimately, is to identify or develop instruments with acceptable psychometric properties in the population of interest.

4.5.1 Validity

From a psychometric standpoint, validity has three main forms: content, construct, and criterion validity. Content validity is the extent to which the instrument actually measures the concepts of interest. The FDA guidance document[10] understandably places significant emphasis on content validity, consistent with other groups,[81] even stating that without adequate content validity, labeling claims cannot be supported. At face value, the importance of content validity is intuitive; it is important that an instrument assess those concepts it was designed to measure. In general, qualitative evidence, in the form of documented patient input through focus groups, is an important standard in the view of the FDA.[42] Construct validity describes the degree to which what was measured reflects the a priori conceptualization of what should be measured.[41] Subcomponents of construct validity are convergent and discriminant validity, which assess the degree of similarity between measures that are theoretically similar (convergent validity) or the extent to which measures that are theoretically different actually differ (discriminant validity). For example, a new measure of anxiety would be expected to have high convergent validity with the anxiety subscale of the Hospital Anxiety and Depression Scale.[82] To that end, the FDA would expect comparisons of new PRO measures with similar existing measures to support construct validity. Criterion validity describes the extent to which the scores of PRO measure reflect the gold standard measure of the same concept.[10] Criterion validity is often difficult to assess in the PRO arena because it is difficult to identify gold standard measures for many PRO concepts. The FDA therefore deemphasizes criterion validity.

4.5.2 Reliability

Reliability reflects the ability of an instrument to yield the same result on serial administrations when no change in the concept being measured is expected. The reliability of an instrument is typically assessed via test-retest methods and by measuring the internal consistency.[42] Accordingly, two aspects of reliability can be distinguished:

- Test-retest reliability describes the ability of an instrument to generate the same results in the same respondent over a period of time during which no change is reasonably expected.[2, 10, 42] Thus, test-retest reliability assesses the intra-individual variability. Identifying the optimal timeframe for retesting can be challenging, and may vary by disease state and target population.[42]
- Internal consistency reliability reflects the degree to which items within a scale measure the same concept. It can be quantitatively

assessed with Cronbach's alpha, which measures the internal consistency of an instrument. Well-established thresholds for interpreting Cronbach's alpha are available; in general, coefficient alpha greater than 0.7 is the minimum acceptable threshold for comparisons between groups.[42]

4.5.3 Ability To Detect Change

The ability of a PRO measure to detect change is intuitively important. Demonstration of this ability, according to the FDA, requires that changes in the PRO instrument parallel changes in other factors that indicate a change in the status of the concept of interest. For example, in patients receiving a new treatment for opioid-induced constipation, changes in a PRO instrument designed to assess overall bowel health may be linked with use of certain other bowel products, such as enemas, to establish the ability to detect change. The measure must demonstrate ability to detect both improvements and losses in health status. Further, it is important to detect changes throughout the range of possible values. In registry studies, where longitudinal collection and analysis are critical, understanding the concept of minimally important change detected,[83] rather than establishing that number explicitly, may be sufficient.

4.5.4 Areas of Controversy

The emphasis placed upon content validity has generated some controversy as PRO developers attempt to improve content validity, in part by meticulously wording items and instructions to minimize variations in interpretation between patients. However, the ability to improve content validity is likely asymptotic, in that individual variability undoubtedly influences interpretation of questions in ways that cannot be accounted for, meaning that responses to an instrument capture the patient's true (and unique) perceptions. There are concerns that in the pursuit of greater content validity, other important characteristics of PRO instruments may be underdeveloped or underappreciated.[41] For example, in pursuing greater content validity, the constraints placed upon questions may actually limit patient perspective by forcing some degree of conformity, or may result in misinterpretation of results.

Consider a registry of patients with advanced cancer designed to assess the impact of certain interventions upon the development of disability. Upon entering the registry, a patient rates his disability as severe because his reference point is a previously healthy state. Four months later, he rates his disability as mild, though on more open-ended questioning, he notes that he can simply sit on the front porch and watch his grandchildren, as he knows that any other activities are unrealistic and that his goal is to simply make it to the front porch. Even though the instrument measures disability from the view of the patient and would thus have adequate content validity, the interpretation regarding the merits of the intervention would be erroneous, as the patient has clearly become more disabled, but has shifted his frame of reference, a fact which is not captured by content validity. This phenomenon is commonly referred to as "response shift" and has long been recognized as a challenge in QOL research.[84] Alternatively, all measures with marginal content validity may be cast aside without consideration of other properties. Consider two new measures for the same concept tested in different studies with different methodologies, resulting in different content validities. The measure with higher content validity is likely to propagate, even if it is more flawed, simply because of methodological issues.

These arguments on content validity are not intended to undermine the standards established by the FDA, nor should they be viewed as rationale for not adhering to these standards, but are meant to prompt careful consideration of all the psychometric properties of PRO measures, especially in the context of the specific registry. Remember first principles—before anything else, it needs to make good sense, have face validity, be doable, and limit patient burden.

4.6 Non-Psychometric Considerations

Beyond identifying a PRO instrument with desirable psychometric properties, consideration must be given to the people that are closely tied to completing and acting upon PRO data and the tension that can exist between impacts on people and psychometric desirability.

4.6.1 Patient Factors

In designing registries and considering PROs for inclusion, it is important to consider the burden to the patient the PRO measures represent. For instance, lengthy questionnaires may result in increasing missing data over time, as patients grow weary of serially completing such questionnaires. The capacity to answer lengthy instruments cannot be predicted a priori and differs between groups. At Duke Cancer Institute, patients in a variety of solid tumor clinics routinely complete [80-86] item instruments without significant fatigue or burnout;[7, 71] median time to complete the survey is 11 minutes, reducing to under 8 minutes after several visits in the clinic using the same instrument. While the FDA did not offer specific recommendations on questionnaire length, a guidance document from the Center for Medical Technology Policy recommends that, for patients with cancer, completion of PRO instruments take no more than 20 minutes at the initial visit and fewer than 10 minutes at subsequent visits.[39] Patients should be offered a private space for completing instruments, to minimize concerns regarding confidentiality, especially for sensitive questions. Instructions should be provided for every item, even if only to frame the recall period. The instrument should be delivered with adequate font size and at appropriate literacy levels. Additionally, physical assistance should be provided if needed, such as reading items aloud to patients with visual impairments. While most pilot studies of PRO instruments provide a small amount of remuneration,[29, 71] these studies have demonstrated that the collection of PROs made patients feel encouraged that their clinicians were seeking additional information and felt that the ePRO instrument facilitated communication between patient and clinician.[29] Outside the pilot testing phase, it is not advisable to provide remuneration to patients for completing PRO instruments, even in the setting of a registry study. PRO responses should be shared with clinicians, as this has been shown to be an important aspect of PROs to patients.[29]

4.6.2 Clinician Factors

Even within the research setting, assessing the impact of PRO collection on routine care is important. Will the PRO results be made available immediately as part of routine care or only available to research personnel? Whether or not PRO data are shared with clinicians in real time should be explicitly addressed in the informed consent process. If data are to be made available to clinicians, are appropriate support services available to assist in managing newly identified concerns or issues? Are there mechanisms to support incorporation of PRO data into clinical care, if it will be made available, or will it be "one more thing" for which clinicians are responsible? What will be the impact of the PRO collection on workflow?

Many recent guidelines recommend providing clinician feedback regarding patient-reported information of concern, such as reports of new chest pain. The thresholds for triggering a clinical alert, components of the alert message, and method of delivering the notice to the clinician must be carefully considered. What are the risk management concerns? How will the clinician's response be verified? Though often mundane, these factors are important to consider in the implementation phase. Teams experienced in embedding PROs into registries and clinical workflow can provide sage advice as to how to navigate these pathways (e.g., Duke Cancer Care Research Program, http://www.cancer.duke.edu/dccrp/); clear guidelines do not exist. See further discussion in Section 4.7.

4.6.3 Ensuring Data Quality

Collecting quality data is an implicit necessity of any registry. Although assessing data quality can assume many forms, for the purposes of registries, two concepts are critical. The first is to minimize missing data. Missing data inevitably undermine the quality of the information collected, thereby decreasing its analytic potential. It is essential to anticipate missing data and to plan interventions to minimize the problem. This is especially important in registry studies where time horizons may be long and the potential for missing data great. There

are a number of steps that can be taken to minimize missing data during the implementation phase of the registry. The most important step is to make sure that the PRO instrument chosen is meaningful, and that its role in the registry and related work is well described, especially to patients and families. Ideally, the PRO measures should be implemented as standard of care, such that they become ubiquitous and desired, not only by patients but also by clinicians.[58] If this occurs, the amount of missing data should decrease. Electronic data collection practically supports real-time, or near-real-time, quality monitoring of information being collected in order to identify patterns of missing data, leading to development of targeted interventions to reduce missing data. Additionally, with near-real-time quality analysis, backup data collection methods can—and should—be deployed. For example, a central telephone interviewer can contact individuals who did not respond to items (either individually or entire instruments) to both obtain the data and ascertain why the item was omitted. Analytic approaches must include a plan for managing the unavoidable occurrence of missing data; importantly, a "last observation carried forward" approach to handling missing data should be avoided.

The second issue related to data quality is consistency. In registries with long time horizons, it is not uncommon for measurement items, or instruments, to evolve or change entirely. Unfortunately, it is equally uncommon for notations of such changes to be embedded within the data structure as metadata, such that future analyses can quickly and readily identify which iteration of an instrument was completed at which point in time. Metadata is essentially data about data. More precisely, it is "…structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information source."[45] Consider a long-term registry where the primary measurement instrument undergoes an iterative update to "version 2" to reflect new knowledge in the field and is quickly implemented into the registry. Though the two versions are likely very similar, they also likely have slightly different questions (in terms of structure or order), psychometric properties, and scoring algorithms. In such a scenario, it is imperative that the version of the instrument completed at any given point in time be identified within the data set. Further, there may be cases where the person completing the questionnaire may not always be the patient (see discussion in Section 4.6.4.). For example, in a palliative care registry, patients are not always able to complete a PRO instrument, even with assistance. The ability of the person to complete the instrument may change over time as cognition wanes. In these settings, proxy-reports involving close family or caregivers may become the only available measures and the only available data to be incorporated into registries; therefore, it is essential to identify, via metadata, who is completing the instrument.

4.6.4 Special Populations: Are Proxy-Reports Ever Appropriate?

There are numerous situations in which patients are not physically or cognitively able to provide direct assessment of their experience. Obvious examples include infants and small children, individuals with significant cognitive impairment (congenital or acquired), and those at the end of life. In such settings, proxy-reports of QOL are often collected,[20] though the literature suggests that proxy-reports demonstrate moderate agreement, at best, with patient-reports.[85-87] Nevertheless, proxy-reports are viewed as valuable in many of these settings because caregiver or family perception is also an important consideration. The FDA strongly discourages proxy-reports in product-labeling claims.[10] Unfortunately, such an extreme stance leaves these vulnerable populations marginalized. By not considering proxy-reports, symptom-based research and other lines of inquiry in these populations face considerable obstacles with a potential end-result that drugs or products that could improve symptom burden or QOL never have the opportunity to gain FDA approval for such indications. The FDA's position on proxy-reports is emphasized because of the rigorous standard the FDA guidance document establishes, but that position should not devalue the potential role for proxy-reports. Ideally, the extent of agreement between patient- and proxy-reports can

be established in advance of use of proxy-reports. The PROMIS Initiative is investigating application of existing methods for PROs to proxy-reports to improve performance.[88, 89]

4.7 Implementation Issues

Upon successful navigation of the challenging process of selecting PRO instruments and their mode of administration comes the daunting task of implementing the selected instruments. Below is a practical framework for successful implementation, centered on achieving data quality and consistency.

Just as with mode of administration, implementing PRO data collection is best achieved if consistency is a central tenet, especially if the registry study is multicenter. In this setting, consistency refers to processes. Standard operating procedures should be established for each site of data collection that delineate, to the extent possible, how patients, researchers, and clinicians interact with the collection system (paper or electronic). As part of standard operating procedures, specific training should be provided, with accessible and easy-to-use manuals available (preferably in both text and video format). Every aspect of the process that can be standardized should be standardized, including the data set itself. That is, the data sets should include metadata that describe key components important for subsequent analyses and end-users, including who completed the instrument (patient or proxy), where it was completed (e.g., outpatient clinic, home, inpatient ward), which version was administered, and a flag for irregularities identified as part of internal quality control.

Ideally, for multisite studies these standard operating procedures are the same at each site, with another set of standard operating procedures for the central repository (or coordinating) site that delineates how often data from cooperating sites should be transmitted, how it should be compiled and stored, how often it should undergo quality assessment, and how it should be accessed and distributed for analysis. Within multisite registries, and even within some single-site registries, it may be necessary to select an instrument that has been translated into and validated in other languages besides English. It is not adequate to simply translate an instrument into another language, as the psychometric properties obtained within an American population of patients with disease X are unlikely to be reproduced in a population of Japanese patients with the same disease. Thus, formal assessment of the psychometric properties of the instrument is necessary when translating to another language.

Another aspect of consistency in this setting reflects administering the same instrument over the lifespan of the registry. The strength of this recommendation depends partly upon the purpose of the registry; for registries comparing effectiveness, this consistency is essential, while for a registry focused on quality and embedded within an EHR, this recommendation is less stringent. Nevertheless, if the data are collected prospectively, the strong preference is for consistency in PRO instrument administered. Regardless of purpose, collected data should include metadata labels.

Further, involving the entire health care team (physicians, mid-level providers, nurses, administrators, and other support staff) in the development process is essential, especially with respect to integrating the PRO instruments into the clinical workflow and providing clinician feedback. As part of this integration, clinical triggers should be established (and standardized) that explicitly force acknowledgement of a patient report by a provider (e.g., a pain score of 8 out of 10) or initiate some standardized intervention (e.g., a patient reporting a high distress level might be automatically contacted by a psychosocial care support team).[88] Such standardized triggers will only be embraced if there is inclusion of the health care team in the implementation process. This inclusive implementation process will also help shape the perception of the PRO data, in that buy-in from the health care team will make the PRO collection process a necessary and desired component of care, rather than simply an extra task to complete.[88]

Finally, explicitly including the patient voice in the form of PROs has been shown to improve patient well-being and enhance patient-provider communication.[35] Building on this premise, inclusion of PROs in observational studies may

improve patient engagement, recruitment, and retention, though there are no data directly supporting this. The experience of the Duke Cancer Care Research Program with ePRO collection as part of routine cancer care has shown remarkable response and participation rates, with rates of missing data routinely less than 5 percent, even for sensitive questions such as level of sexual enjoyment. Certainly, more rigorous documentation of improved long-term patient participation with inclusion of PROs is needed before more ardent assertions can be made.

4.8 Summary Regarding Selecting PRO Instruments

Selecting PRO instruments for inclusion in registry studies is not a one-size-fits-all process. The Center for Medical Technology Policy prepared a guidance document for inclusion of PROs in adult oncology trials[39] and these recommendations are included as an example (Table 5–2). Clear and careful definition of the target population, concept to be measured, and purpose of the registry is an important first step. For a given population or context, even in a registry, it is important to have some a priori hypotheses and justification for outcomes being measured, or the study risks becoming a prospective fishing expedition. As such, there needs to be a systematic approach to selecting salient outcomes (to the extent possible in a registry, which admittedly is sometimes exploratory by nature). In CER, the process of identifying meaningful outcomes requires upfront patient input. But regardless of how the outcomes are selected, there must be a systematic approach to determining whether an outcome is best reported by a patient (i.e., if information about a particular symptom or overall health state or satisfaction is sought, it is best reported from the patient/surrogate perspective; thus, a PRO instrument is appropriate). Far too frequently, the tail wags the dog in registry studies; that is, PRO instruments are selected first, prior to identifying outcomes of interest. Thus, the rational identification of outcomes of interest early in the process of registry development is important. Such an approach will quickly enable the researchers to determine if PROs are appropriate, and will produce a sound base for evaluating PRO instruments and administration methods. If this process is navigated effectively, the stage will be set for successful incorporation of PROs into the registry.

After the arduous process of clearly defining the population and outcomes of interest is complete, researchers should search for existing PRO instruments that will assess the outcomes of interest. (See Case Examples 8, 9, 10, and 11.) If a suitable measure is not identified, options include modifying an existing measure or developing a new measure. (See Case Example 8.) In general, development of new PRO instruments is resource-intensive, so it is preferable to use an existing measure whenever possible. After identifying (or developing) a measure, administration mode should be selected. Electronic administration is preferred, but not all instruments have been evaluated using electronic administration, though this can be accomplished. Important to the scientific basis of the registry are the psychometric properties of the instrument. While the FDA highly values content validity, it is possible to effectively use an instrument with modest content validity, depending on the purpose of the registry, highlighting the importance of understanding and defining the purpose of the registry.

In most registry studies, the purposes of the study and outcomes of interest will necessitate inclusion of PRO data. Careful planning is essential, in identifying appropriate PRO instruments for inclusion, selecting modes of instrument administration, and implementing the PRO collection system. When carried out effectively, this planning process generally produces more complete data sets that truly include the voices of all stakeholders in the health care system and are meaningful to all stakeholders.

5. Example of PRO Use in a Registry

Consider the division of pulmonary medicine at an academic university. Within the division there is a growing multidisciplinary cystic fibrosis (CF) program with a large catchment area and approximately 250 patients ranging in age from 21 to 65 years, though most patients are younger than

age 35. As the program develops, the team plans to implement a series of initiatives for patients with CF, targeting not only improved survival, but also improved functioning. Proposed interventions include routine endocrinology consultations for all CF-related diabetes mellitus, improved psychological services, and standardized exercise regimens during hospitalizations. The outcomes of interest for these interventions are equally broad-ranging, but include traditional measures such as pulmonary function (as measured by pulmonary function tests), end-organ damage (diabetes, chronic kidney disease), resistant organism colonization rates, hospitalization utilization, symptom burden (including breathlessness, weight change, worry, and fatigue) and quality of life (QOL). The team plans to use a registry for this project because they do not feel that they can reasonably test the effectiveness of these interventions through parallel or sequential randomized, controlled trials, but do wish to systematically capture outcomes of interest in a longitudinal manner as the interventions are introduced.

In considering the outcomes of interest, symptom burden and health-related QOL merit closer inspection for inclusion of PROs. Certainly, patients are better positioned than clinicians to report breathlessness, worry, fatigue, and QOL. In fact, most argue that patients are the only valid source of information on these issues, thus inclusion of PROs in this registry is appropriate.

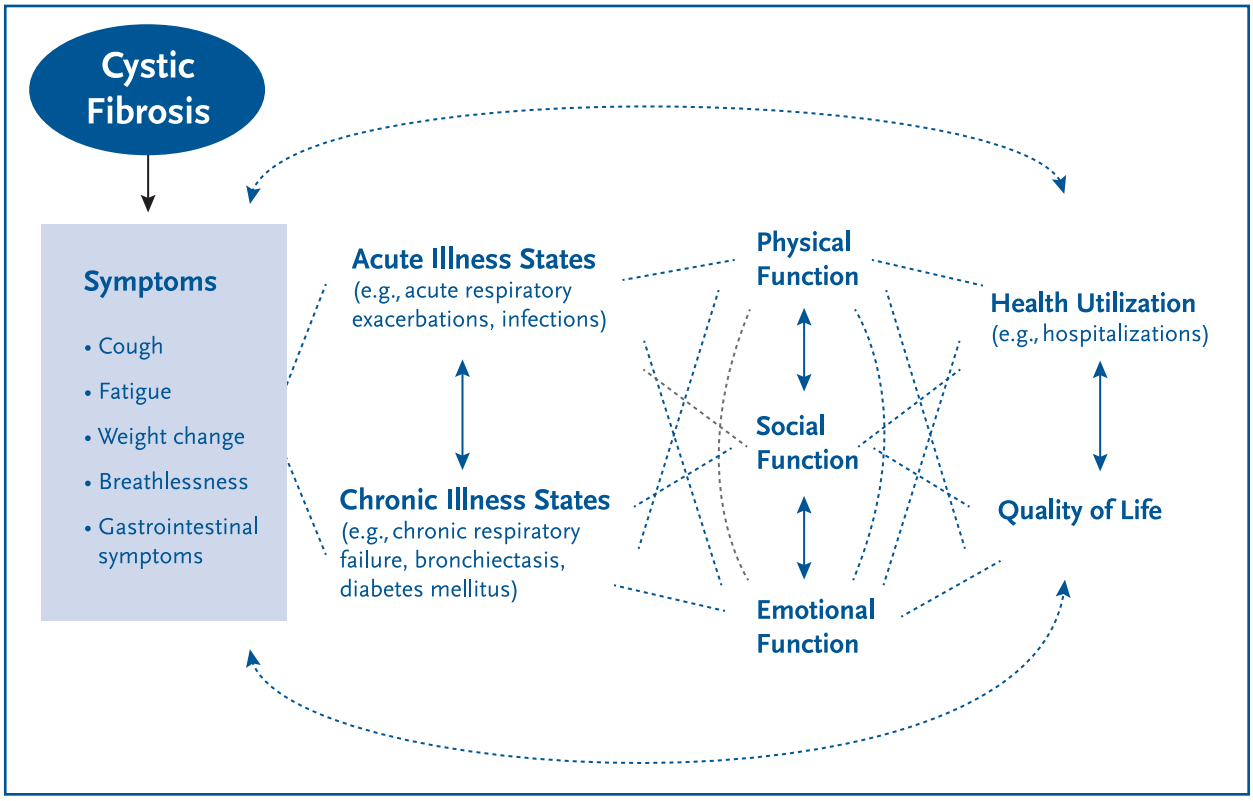
In considering which instruments to use, it is important for the team to consider the relationships between the symptoms under consideration and QOL. Figure 5–2 illustrates some of the relationships that exist around health-related QOL in CF. Specifically, the influence of symptom burden on QOL must be weighed carefully, to help determine if a series of single-item instruments is most appropriate or if a multi-item, disease-specific instrument (of which there are several in CF) or another approach is most appropriate.

Because the team plans to use this registry in a longitudinal fashion for numerous planned interventions and wants to understand how specific interventions affect certain domains impacted by CF, it selects an established, multi-item, multi-domain, CF-specific measure that incorporates an aggregate assessment of QOL, as well as several component domains of well-being.

Since the planned settings of intervention include both inpatient and outpatient settings, and since there are travel issues related to the catchment area, the team also plans to capture reports between visits, such that no more than two months elapse between sessions of PRO data collection. For this reason, the team prefers to use electronic methods, but the instrument it selected has only been psychometrically assessed via paper-based methods. The team collaborates with the institutional expert on PROs to document paper-electronic equivalence, and to perform usability and feasibility testing for Web-based administration. This pilot study demonstrates that it is reasonable to use a Web-based approach for PRO assessments.

From a health-care team standpoint, implementation goes smoothly, since the entire CF team was involved in developing the registry and PRO system. Missing data are minimal for inpatient and clinic appointment collection, as the team heavily advertised the PRO collection system to the patients prior to implementation, provided in-clinic teaching, and used reports during the clinical visits; the instrument quickly becomes viewed as a necessary component to the health care encounter. However, as data accumulates, the team identifies patterns of missing data for between-visit administrations. It determines that at-home Internet access remains a problem for a small but significant portion of patients. It receives a grant from the local CF foundation to support Internet access for vulnerable patients, with subsequent reduction in missing data.

Figure 5–2. Simplified concept map for cystic fibrosis



This example highlights several key points: (1) the importance of understanding the target population; (2) the need to identify outcomes of interest prior to selecting PRO instruments as the outcomes of interest should dictate the instrument, not vice versa; and (3) the benefit of incorporating PRO instruments into longitudinal, routine care.

Case Examples for Chapter 5

Case Example 8. Developing and validating a patient-administered questionnaire

Description	The Benign Prostatic Hypertrophy (BPH) Registry and Patient Survey was a multicenter, prospective, observational registry examining the patient management practices of primary care providers and urologists, and assessing patient outcomes, including symptom amelioration and disease progress. The registry collected patient-reported and clinician-reported data at multiple clinical visits.
Sponsor	sanofi-aventis
Year Started	2004
Year Ended	2007
No. of Sites	403
No. of Patients	6,928

Challenge

Lower urinary tract symptoms associated with benign prostatic hyperplasia (LUTS/BPH) have a strong relationship to sexual dysfunction in aging males. Sexual dysfunction includes both erectile dysfunction (ED) and ejaculatory dysfunction (EjD), and health care providers treating patients with symptoms of BPH should evaluate men for both types of dysfunction. Providers can use the Male Sexual Health Questionnaire (MSHQ), a validated, self-administered, sexual function scale, to assess dysfunction, but the 25-item scale can be perceived as too long. To assess EjD more efficiently, it was necessary to develop a brief, patient-administered, validated questionnaire.

Proposed Solution

The team used representative, population-based samples to develop a short-form scale for assessing EjD. The team administered the 25-item MSHQ to three populations: a sample of men from the Men's Sexual Health Population Survey, a subsample of men from the Urban Men's Health Study, and a sample of men enrolled in the observational registry.

Using the data from the sample populations, the team conducted a series of analyses to develop the scale. The team used factor analysis to help select the items from the scale that had the highest correlations with the principal factors. Using conventional validation, the team examined reliability (both internal consistency and test-retest repeatability). To assess validity, tests of repeatability and discriminant/convergent validity were used to determine that the short form successfully discriminated between men with no to mild LUTS/BPH and those with moderate to severe LUTS/BPH. Lastly, the team examined the correlation between the 7-item ejaculation domain of the 25-item MSHQ and the new short-form scale, using data from the observational registry.

Results

Based on the results of these analyses, the team selected three ejaculatory function items and one ejaculation bother item for inclusion in the new MSHQ-EjD Short Form. The new scale demonstrates a high degree of internal consistency and reliability, and it provides information to identify men with no to mild LUTS/BPH and those with moderate to severe LUTS/BPH.

Key Point

Developing new instruments for collecting patient-reported outcomes requires careful testing of the new tool in representative populations to ensure validity and reliability. Registries can provide a large sample population for validating new instruments.

For More Information

Rosen RC, Catania JA, Althof SE. et al. Development and validation of four-item version of Male Sexual Health Questionnaire to assess ejaculatory dysfunction. Urology. 2007;69(5):805–9.

Rosen R, Altwein J, Boyle P. et al. Lower urinary tract symptoms and male sexual dysfunction: the Multinational Survey of the Aging Male. Eur Urol. 2003;44:637–49.

Case Example 9. Using validated measures to collect patient-reported outcomes

Description	The Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD) is a household panel registry designed to assess the prevalence and incidence of diabetes mellitus and cardiovascular disease; disease burden and progression; risk predictors; and knowledge, attitudes, and behaviors regarding health in the U.S. population. The study involves three distinct phases: an initial screening survey, a baseline survey, and yearly followup surveys for 5 years.
Sponsor	AstraZeneca Pharmaceuticals LP
Year Started	2004
Year Ended	2009
No. of Sites	Not applicable
No. of Patients	More than 211,000 individuals were included in the screening survey; approximately 15,000 individuals were followed for 5 years.

Challenge

The SHIELD registry used survey methodologies to collect health information from a large sample of adults. The goal of the study was to capture participants' perspectives and views on diabetes and cardiovascular disease, risk factors for the diseases, and burden of the diseases. The study investigators, noting that treatment for diabetes and cardiovascular disease relies heavily on patient self-management, felt that it was particularly important to gather information on activities, weight control, health attitudes, quality of life, and other topics directly from the participant, without a physician as an intermediary. The investigators also wanted to follow participants over time to better understand disease progression and changes in health behaviors or activities.

To achieve the study goals, the registry needed to collect health-related data directly from participants in such a way that the data would be reliable, valid, and comparable across participant groups and over time.

Proposed Solution

The investigators decided to use validated patient-reported outcomes measures (PROs) to collect information on health status and behaviors. The PROs allowed the data from the registry to be compared with data collected in other registries to assess the generalizability of data on the study population. In addition, the PROs already took into account issues such as recall bias and interpretability of the questions, and self-administered instruments eliminated the possibility of introducing interviewer bias.

The registry included seven PROs: (1) the 12-item Short Form Health Survey (SF-12) and European Quality of Life (EuroQoL) EQ-5D instrument, to assess health-related quality of life; (2) the Sheehan Disability Scale, to assess the level of disruption in work, social life, and family/home life; (3) the 9-item Patient Health Questionnaire, to assess depression; (4) the Work Productivity and Activity Impairment Questionnaire: General Health, to assess work productivity and absenteeism; (5) the Diet and Health Knowledge Survey; (6) the Press-Ganey Satisfaction questionnaire; and (7) the International Physical Activity Questionnaire, to assess health-related physical activity and sedentary behaviors.

The investigators considered many factors, such as length, ease of use, format, and scoring system, when selecting the PROs to include in the survey. For example, a major reason for selecting the SF-12 rather than the SF-36 as a measure of quality of life was the length of the forms (12 vs. 36 items). The survey is entirely paper-based, with participants mailing back completed forms. The validated scoring algorithms are used to account for missing or illegible values on the completed forms. All participants must be able to read and write English.

Case Example 9. Using validated measures to collect patient-reported outcomes (continued)

Results

The registry had a generally high response rate for the surveys. The response rates were 63.7 percent for the screening survey, 71.8 percent for the baseline survey, and between 71 and 75 percent for the annual surveys. In terms of missing data, participants who return the survey forms tended to complete all of the questions in the appropriate manner. However, the registry is missing longitudinal data from some participants. For example, a participant may have returned the completed form in 2005, failed to return the form in 2006, and returned the form again in 2007. The investigators must account for the missing 2006 values when conducting longitudinal analyses. The data from the survey have been sufficient to support comparisons over time and across participant groups, leading to several publications.

Key Point

Utilization of standardized, validated instruments in a registry can offer many benefits, including enhanced scientific rigor, the ability to compare patient views over time, and the ability to compare registry data with data from other sources to assess the representativeness of the registry population. It should be noted that significant initial planning is necessary to identify appropriate PROs, obtain the necessary permissions, and include them in a registry. Issues with missing data must be considered in the planning phases for a registry. This registry considered missing data within returned survey questionnaires. In addition, an acceptable followup rate should be stated a priori so that response rates can be better interpreted with respect to their potential for introducing bias.

For More Information

Gavin JR III, Rodbard HW, Fox KM, et al. Association of overweight and obesity with health status, weight management, and exercise behaviors among individuals with type 2 diabetes mellitus or with cardiometabolic risk factors. Risk Management and Healthcare Policy 2009;2:1−7.

Grandy S, Chapman RH, Fox KM, et al. Quality of life and depression of people living with type 2 diabetes mellitus and those at low and high risk for type 2 diabetes: findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). Int J Clin Pract 2008;62:562–8.

Grandy S, Fox KM. EQ-5D visual analog scale and utility index values in individuals with diabetes and at risk for diabetes: findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). Health Qual Life Outcomes 2008;6:18.

Grandy S, Fox KM, Bazata DD, et al. Association of self-reported weight change and quality of life, and exercise and weight management behaviors among adults with type 2 diabetes mellitus. Cardiol Res Pract 2012;2012:892564.

Rodbard HW, Bays HE, Gavin JR III, et al. Rate and risk predictors for development of self-reported type 2 diabetes mellitus over a 5-year period: the SHIELD study. Int J Clin Pract 2012;66:684-691.

Case Example 10. Challenges in the collection of PROs in a longitudinal registry

Description	A longitudinal registry of men with metastatic castrate-resistant prostate cancer is being conducted among men receiving outpatient care at the Memorial Sloan-Kettering Cancer Center, Oregon Health and Science University, and John Hopkins School of Medicine.
Sponsor	U.S. Department of Defense
Year Started	2012
Year Ended	Ongoing
No. of Sites	3
No. of Patients	Planned enrollment is 400 men with castrate-resistant prostate cancer

Challenge

Regulatory and government agencies and cancer organizations, including the U.S. Food and Drug Administration (FDA), National Cancer Institute, and the American Cancer Society, recommend collecting patient-reported outcomes (PRO) data to capture cancer patients' perspectives on and experiences of their symptoms, disease status and functioning, and health-related quality of life. Collecting PRO data is important for diseases such as cancer, in order to fully evaluate the benefit and risk profile of potentially toxic oncology treatments. To this end, industry sponsors and the FDA wish to include symptom endpoints in clinical trials, but currently lack sufficient information about optimal methods to design robust endpoints. Many challenges to designing PRO endpoints exist, such as identifying appropriate and validated instruments, ensuring interpretability of the data's clinical significance, and having information on the variability of symptoms in order to accurately determine necessary sample sizes. The collection of PRO data presents additional challenges, including identifying the optimal mode(s) of administration, minimizing patient burden, and minimizing incomplete or missing PRO questionnaires.

Proposed Solution

Investigators are using a registry to evaluate longitudinal PRO data among men with metastatic castrate-resistant prostate cancer (mCRPC). mCRPC is characterized by disease progression (continued elevated prostate-specific antigen [PSA] or radiographic progression) regardless of first-line androgen depletion therapy. Many patients with bone metastases experience debilitative symptoms, such as bone pain, in addition to treatment toxicities. The purpose of the registry is to collect PRO and clinical data to inform the development of pain endpoints for future oncology clinical trials that conform to regulatory standards of the FDA and the European Medicines Agency. The registry does not seek to develop a new PRO, but instead to advance methods for administering and interpreting PRO results, as well as to validate a PRO measure of analgesic medication use. Specifically, the goals of the registry are to identify (1) the clinical significance of pain score changes, (2) the average time to pain progression, (3) the proportion of men with pain starting new lines of treatment, (4) appropriate recall periods for pain assessment, and (5) the comparison of methods for quantifying analgesic use.

Results

Registry participants will include 400 patients with metastatic castrate-resistant prostate cancer who are receiving outpatient care between 2012 and 2014 at one of the three institutions. Clinical data, including diagnosis, treatment, and resource utilization, will be abstracted from medical records every 3 months. Patient-reported data, including pain, analgesic use, and other symptoms, will be collected every 6 weeks by an automated telephone survey. A key feature of this registry is the use of a single centralized survey platform that includes a phone survey completed by patients and a Web interface through which study staff at participating sites can enter patient medical record data on a quarterly basis. The integrated system of data collection is intended to

Case Example 10. Challenges in the collection of PROs in a longitudinal registry (contiued)

Results (continued)

reduce the burden of data management. The registry is designed in a way which addresses the challenges of collecting longitudinal PRO data, taking into account (1) the importance of electronic PRO assessment and choosing the best mode (e.g., phone, Web-based) of assessment, (2) the importance of choosing the optimal frequency and length of each assessment, and (3) the use of automated reminders, clear instruction sheets, and survey questions relevant to patients to ensure high survey completion rates.

Key Point

To ensure appropriate clinical interpretation and quality of PRO data, registries can be used to evaluate PRO instrument characteristics and collect data necessary to develop PRO endpoints for use in clinical trials. Addressing operational barriers such as mode of administration, instrument length and frequency, and missing PRO data will improve patient-reported data for use in endpoint development, clinical trials, treatment decisionmaking, and routine patient care.

For More Information

Basch E, Abernethy AP. Supporting clinical practice decisions with real-time patient-reported outcomes. J Clin Oncol. 2011 Mar 10; 29(8): 954-6.

Bennett AV, Jensen RE, Basch EM. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin. 2012 Sep-Oct;62(5):337-47.

Case Example 11. Collecting PRO data in a sensitive patient population

Description	The Cedars-Sinai Psychiatric Treatment Outcome Registry (CS-PTR) is a single-site patient registry that tracks the outcomes of psychiatric interventions in a naturalistic clinical setting using measurement-based care and patient-reported outcomes.
Sponsor	Cedars Sinai Medical Center
Year Started	2005
Year Ended	2012
No. of Sites	1
No. of Patients	2,600

Challenge

Psychiatric disorders are strongly associated with grave impairments in functioning and quality of life (QOL), but most previous research has focused on symptom improvement and has not specifically investigated the extent to which treatment can improve functioning and QOL outcomes.

The Department of Psychiatry and Behavioral Neurosciences at Cedars Sinai Medical Center enrolled consecutive patients presenting for psychiatric evaluation in a patient registry. Demographic information, DSM-IV diagnosis, and current psychiatric comorbidities were obtained by the provider using structure interviewing (the Mini International Neuropsychiatric Interview). Patients completed "self-assessment questionnaires" during their baseline visit and during quarterly followup visits. Validated patient-reported outcome (PRO) tools included questionnaires that collected information on depressive symptom severity (Quick Inventory of Depressive Symptomatology), functioning (Work and Social Adjustment Scale), and QOL (Quality of Life Enjoyment and Satisfaction Questionnaire).

The registry often encountered significant barriers to obtaining self-reported data from psychiatric patients. For example, the baseline and followup questionnaires took approximately 20 to 30 minutes for patients to complete, and many patients were resistant to spending that amount of time completing the questionnaires, as they did not see the value in completing them.

Case Example 11. Collecting PRO data in a sensitive patient population (continued)

Proposed Solution

Staff at the clinic educated the patients about the value of the self-assessment questionnaires. They explained to patients that the results of the questionnaires would be used to inform their providers' decisions about diagnoses, appropriate treatment, and treatment progress. The staff implemented an appointment scheduling system that built in a 30-minute block of time before patients were seen by a provider, to allow time for them to complete the self-assessment questionnaires. For patients who had trouble completing the written questionnaire independently, clinicians worked with them to help them complete the questionnaire verbally and recorded the answers themselves. During quarterly followup visits, clinicians were expected to review the answers and PRO scores with patients, including any trends in symptom severity, functioning, or QOL changes following treatment initiation.

Results

A total of 2,600 patients were enrolled in the registry over the course of seven years. At baseline, patients reported a wide range of symptom severity, which is to be expected given the consecutive enrollment of patients in the registry with no exclusion criteria. Psychiatric patients tended to report severely low QOL levels, especially if they were older, of Hispanic ethnicity, or had been diagnosed with mood disorders. Baseline analysis also showed that although symptom severity and functional impairments are significantly correlated with lower reported levels of QOL, they only explained a moderate amount of the variance in QOL. The findings point to the critical need to go beyond symptom severity monitoring and include functioning and QOL measures during the course of assessment, treatment, and research of psychiatric disorders. Analysis of followup data is ongoing.

Key Point

PROs are important tools for informing treatment decisions. In patient populations where it is particularly difficult to obtain PRO data, operational steps, such as discussing the benefits of PROs, changing appointment time frames, and offering support in completing PRO tools from clinicians or other trained professionals, can be taken to minimize the burden on patients. Regularly reviewing PRO data with patients and informing patients of the value of PRO data to their treatment may increase patients' participation.

For More Information

IsHak WW, Balayan K, Bresee C, et al. A descriptive analysis of quality of life using patient-reported measures in major depressive disorder in a naturalistic outpatient setting. Qual Life Res. 2013 Apr;22(3):585-96.

References for Chapter 5

1. Fairclough DL. Design and Analysis of Quality of Life Studies in Clinical Trials. Chapman & Hall/CRC Interdisciplinary Statistics. 2nd ed: Chapman and Hall/CRC; 2010. p. 424.
2. Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. CA Cancer J Clin. 2007 Sep-Oct;57(5):278-300. PMID: 17855485.
3. Hewlett SA. Patients and clinicians have different perspectives on outcomes in arthritis. J Rheumatol. 2003 Apr;30(4):877-9. PMID: 12672220.
4. Flynn KE, Lin L, Ellis SJ, et al. Outcomes, health policy, and managed care: relationships between patient-reported outcome measures and clinical measures in outpatients with heart failure. Am Heart J. 2009 Oct;158(4 Suppl):S64-71. PMID: 19782791. PMCID: 2805910.
5. Cleeland CS, Sloan JA, Group AO. Assessing the Symptoms of Cancer Using Patient-Reported Outcomes (ASCPRO): searching for standards. J Pain Symptom Manage. 2010 Jun;39(6): 1077-85. PMID: 20538189.

6. Bushmakin AG, Cappelleri JC, Taylor-Stokes G, et al. Relationship between patient-reported disease severity and other clinical outcomes in osteoarthritis: a European perspective. J Med Econ. 2011;14(4):381-9. PMID: 21574904.
7. Pakhomov SV, Jacobsen SJ, Chute CG, et al. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. 2008 Aug;14(8):530-9. PMID: 18690769. PMCID: 2581509.
8. Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. Lancet Oncol. 2006 Nov;7(11):903-9. PMID: 17081915.
9. Stacy M, Bowron A, Guttman M, et al. Identification of motor and nonmotor wearing-off in Parkinson's disease: comparison of a patient questionnaire versus a clinician assessment. Mov Disord. 2005 Jun;20(6):726-33. PMID: 15719426.
10. U.S. Food and Drug Administration. Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development and Labeling Claims. December 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed August 15, 2012.
11. Norekval TM, Fridlund B, Rokne B, et al. Patient-reported outcomes as predictors of 10-year survival in women after acute myocardial infarction. Health Qual Life Outcomes. 2010;8:140. PMID: 21108810. PMCID: 3004873.
12. Wang XS, Shi Q, Lu C, et al. Prognostic value of symptom burden for overall survival in patients receiving chemotherapy for advanced nonsmall cell lung cancer. Cancer. 2010 Jan 1;116(1):137-45. PMID: 19852033. PMCID: 3148693.
13. Singh JA, Nelson DB, Fink HA, et al. Health-related quality of life predicts future health care utilization and mortality in veterans with self-reported physician-diagnosed arthritis: the veterans arthritis quality of life study. Semin Arthritis Rheum. 2005 Apr;34(5):755-65. PMID: 15846592.
14. Cunningham WE, Crystal S, Bozzette S, et al. The association of health-related quality of life with survival among persons with HIV infection in the United States. J Gen Intern Med. 2005 Jan;20(1):21-7. PMID: 15693923. PMCID: 1490035.
15. Mapes DL, Lopes AA, Satayathum S, et al. Health-related quality of life as a predictor of mortality and hospitalization: the Dialysis Outcomes and Practice Patterns Study (DOPPS). Kidney Int. 2003 Jul;64(1):339-49. PMID: 12787427.
16. Gotay CC, Kawamoto CT, Bottomley A, et al. The prognostic significance of patient-reported outcomes in cancer clinical trials. J Clin Oncol. 2008 Mar 10;26(8):1355-63. PMID: 18227528.
17. Victorson D, Soni M, Cella D. Metaanalysis of the correlation between radiographic tumor response and patient-reported outcomes. Cancer. 2006 Feb 1;106(3):494-504. PMID: 16353212.
18. Rolfson O, Karrholm J, Dahlberg LE, et al. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011 Jul;93(7):867-75. PMID: 21705555.
19. Castellano D, del Muro XG, Perez-Gracia JL, et al. Patient-reported outcomes in a phase III, randomized study of sunitinib versus interferon-{alpha} as first-line systemic therapy for patients with metastatic renal cell carcinoma in a European population. Ann Oncol. 2009 Nov;20(11):1803-12. PMID: 19549706. PMCID: 2768734.
20. Thornburg CD, Calatroni A, Panepinto JA. Differences in health-related quality of life in children with sickle cell disease receiving hydroxyurea. J Pediatr Hematol Oncol. 2011 May;33(4):251-4. PMID: 21516020.
21. Molnar-Varga M, Molnar MZ, Szeifert L, et al. Health-related quality of life and clinical outcomes in kidney transplant recipients. Am J Kidney Dis. 2011 Sep;58(3):444-52. PMID: 21658828.
22. Pronzato P, Cortesi E, van der Rijt CC, et al. Epoetin alfa improves anemia and anemia-related, patient-reported outcomes in patients with breast cancer receiving myelotoxic chemotherapy: results of a European, multicenter, randomized, controlled trial. Oncologist. 2010;15(9):935-43. PMID: 20798194. PMCID: 3228044.

23. Abernethy AP, McDonald CF, Frith PA, et al. Effect of palliative oxygen versus room air in relief of breathlessness in patients with refractory dyspnoea: a double-blind, randomised controlled trial. Lancet. 2010 Sep 4;376(9743):784-93. PMID: 20816546. PMCID: 2962424.

24. Deshpande AD, Sefko JA, Jeffe DB, et al. The association between chronic disease burden and quality of life among breast cancer survivors in Missouri. Breast Cancer Res Treat. 2011 Oct;129(3):877-86. PMID: 21519836. PMCID: 3250926.

25. Sprangers MA. Disregarding clinical trial-based patient-reported outcomes is unwarranted: Five advances to substantiate the scientific stringency of quality-of-life measurement. Acta Oncol. 2010;49(2):155-63. PMID: 20059312.

26. Lipscomb J, Gotay CC, Snyder C. Outcomes assessment in cancer. Cambridge Univ Pr; 2005. p. 662.

27. Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. 4th ed: Oxford University Press; 2008. p. 428.

28. Basch E, Iasonos A, Barz A, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. J Clin Oncol. 2007 Dec 1;25(34):5374-80. PMID: 18048818.

29. Abernethy AP, Herndon JE, 2nd, Wheeler JL, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. J Pain Symptom Manage. 2009 Jun;37(6):1027-38. PMID: 19394793.

30. Basch E, Artz D, Dulko D, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. J Clin Oncol. 2005 May 20;23(15):3552-61. PMID: 15908666.

31. Snyder CF, Jensen R, Courtin SO, et al. PatientViewpoint: a website for patient-reported outcomes assessment. Qual Life Res. 2009 Sep;18(7):793-800. PMID: 19544089. PMCID: 3073983.

32. Dudgeon DJ, Knott C, Chapman C, et al. Development, implementation, and process evaluation of a regional palliative care quality improvement project. J Pain Symptom Manage. 2009 Oct;38(4):483-95. PMID: 19699607.

33. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. 2009 Dec 2;101(23):1624-32. PMID: 19920223. PMCID: 2786917.

34. Berry DL, Blumenstein BA, Halpenny B, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. J Clin Oncol. 2011 Mar 10;29(8):1029-35. PMID: 21282548. PMCID: 3068053.

35. Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol. 2004 Feb 15;22(4):714-24. PMID: 14966096.

36. Detmar SB, Muller MJ, Schornagel JH, et al. Role of health-related quality of life in palliative chemotherapy treatment decisions. J Clin Oncol. 2002 Feb 15;20(4):1056-62. PMID: 11844830.

37. National Quality Forum. A National Framework and Preferred Practices for Palliative and Hospice Care Quality: A Consensus Report. Oxford University Press, USA; 2006:124. www.qualityforum.org/Publications/2006/12/A_National_Framework_and_Preferred_Practices_for_Palliative_and_Hospice_Care_Quality.aspx. Accessed August 17, 2012.

38. Curtis JR, Jain A, Askling J, et al. A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. Semin Arthritis Rheum. 2010 Aug;40(1):2-14 e1. PMID: 20674669. PMCID: 2958101.

39. Center for Medical Technology Policy (CMTP). Effectiveness Guidance Document: Recommendations for Incorporating Patient-Reported Outcomes into the Design of Post-Marketing Clinical Trials in Adult Oncology. May 2012. www.cmtpnet.org/wp-content/uploads/downloads/2012/05/PRO-EGD.pdf. Accessed August 20, 2012.

40. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. Control Clin Trials. 2004 Dec;25(6):535-52. PMID: 15588741.

41. McClimans LM, Browne J. Choosing a patient-reported outcome measure. Theor Med Bioeth. 2011 Feb;32(1):47-60. PMID: 21110121.

42. Frost MH, Reeve BB, Liepa AM, et al. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health. 2007 Nov-Dec;10 Suppl 2:S94-S105. PMID: 17995479.

43. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res. 2000;9(8):887-900. PMID: 11284208.

44. Flynn KE, Dombeck CB, DeWitt EM, et al. Using item banks to construct measures of patient reported outcomes in clinical trials: investigator perceptions. Clin Trials. 2008;5(6):575-86. PMID: 19029206. PMCID: 2662709.

45. NISO. Understanding Metadata. Bethesda, MD: NISO Press; 2004. p. 1-20. www.niso.org/publications/press/UnderstandingMetadata.pdf. Accessed July 16, 2013.

46. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med. 1995 Nov;41(10):1403-9. PMID: 8560308.

47. Institute of Medicine. Committee on Comparative Effective Research Prioritization. Initial National Priorities for Comparative Effectiveness Research. Washington DC: National Academy Press; 2009. p. 227.

48. Wu AW, Snyder C, Clancy CM, et al. Adding the patient perspective to comparative effectiveness research. Health Aff (Millwood). 2010 Oct;29(10):1863-71. PMID: 20921487.

49. Hirsch BR, Giffin RB, Esmail LC, et al. Informatics in action: lessons learned in comparative effectiveness research. Cancer J. 2011 Jul-Aug;17(4):235-8. PMID: 21799331.

50. Patient-Centered Outcomes Research Institute. www.pcori.org/patient-centered-outcomes-research. Accessed August 15, 2012.

51. Plimack ER, Hudes GR. Selecting targeted therapies for patients with renal cell carcinoma. J Natl Compr Canc Netw. 2011 Sep 1;9(9):997-1006; quiz 7. PMID: 21917624.

52. Basch EM, Reeve BB, Mitchell SA, et al. Electronic toxicity monitoring and patient-reported outcomes. Cancer J. 2011 Jul-Aug;17(4):231-4. PMID: 21799330.

53. Kamal AH, Currow DC, Ritchie C, et al. The value of data collection within a palliative care program. Curr Oncol Rep. 2011 Aug;13(4):308-15. PMID: 21556703.

54. Neuss MN, Desch CE, McNiff KK, et al. A process for measuring the quality of cancer care: the Quality Oncology Practice Initiative. J Clin Oncol. 2005 Sep 1;23(25):6233-9. PMID: 16087948.

55. Campion FX, Larson LR, Kadlubek PJ, et al. Advancing performance measurement in oncology: quality oncology practice initiative participation and quality outcomes. J Oncol Pract. 2011 May;7(3 Suppl):31s-5s. PMID: 21886517. PMCID: 3092462.

56. Fromme EK, Eilers KM, Mori M, et al. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol. 2004 Sep 1;22(17):3485-90. PMID: 15337796.

57. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. J Clin Oncol. 2010 Sep 20;28(27):4268-74. PMID: 20585094. PMCID: 2953977.

58. Abernethy AP, Herndon JE, 2nd, Wheeler JL, et al. Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data. Health Serv Res. 2008 Dec;43(6):1975-91. PMID: 18761678. PMCID: 2613994.

59. Etheredge LM. A rapid-learning health system. Health Aff (Millwood). 2007 Mar-Apr;26(2):w107-18. PMID: 17259191.

60. Stone AA, Broderick JE, Schwartz JE. Validity of average, minimum, and maximum end-of-day recall assessments of pain and fatigue. Contemp Clin Trials. 2010 Sep;31(5):483-90. PMID: 20620239.

61. Stone AA, Broderick JE, Kaell AT. Single momentary assessments are not reliable outcomes for clinical trials. Contemp Clin Trials. 2010 Sep;31(5):466-72. PMID: 20580945. PMCID: 3014612.

62. Bainbridge D, Seow H, Sussman J, et al. Multidisciplinary health care professionals' perceptions of the use and utility of a symptom assessment system for oncology patients. J Oncol Pract. 2011 Jan;7(1):19-23. PMID: 21532805. PMCID: 3014504.

63. Paley L, Zornitzki T, Cohen J, et al. Utility of clinical examination in the diagnosis of emergency department patients admitted to the department of medicine of an academic hospital. Arch Intern Med. 2011 Aug 8;171(15):1394-6. PMID: 21824956.

64. Day S, Fayers P, Harvey D. Double data entry: what value, what price? Control Clin Trials. 1998 Feb;19(1):15-24. PMID: 9492966.

65. Rand CS, Wise RA, Nides M, et al. Metered-dose inhaler adherence in a clinical trial. Am Rev Respir Dis. 1992 Dec;146(6):1559-64. PMID: 1456575.

66. Rose M, Bezjak A. Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples. Qual Life Res. 2009 Feb;18(1):125-36. PMID: 19152119.

67. Salaffi F, Gasparini S, Grassi W. The use of computer touch-screen technology for the collection of patient-reported outcome data in rheumatoid arthritis: comparison with standardized paper questionnaires. Clin Exp Rheumatol. 2009 May-Jun;27(3):459-68. PMID: 19604439.

68. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care. 2007 May;45(5 Suppl 1):S3-S11. PMID: 17443116. PMCID: 2829758.

69. Garcia SF, Cella D, Clauser SB, et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. J Clin Oncol. 2007 Nov 10;25(32):5106-12. PMID: 17991929.

70. Reeve BB, Burke LB, Chiang YP, et al. Enhancing measurement in health outcomes research supported by Agencies within the US Department of Health and Human Services. Qual Life Res. 2007;16 Suppl 1:175-86. PMID: 17530449.

71. Abernethy AP, Zafar SY, Uronis H, et al. Validation of the Patient Care Monitor (Version 2.0): a review of system assessment instrument for cancer patients. J Pain Symptom Manage. 2010 Oct;40(4):545-58. PMID: 20579839.

72. Dupont A, Wheeler J, Herndon JE, 2nd, et al. Use of tablet personal computers for sensitive patient-reported information. J Support Oncol. 2009 May-Jun;7(3):91-7. PMID: 19507456.

73. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health. 2009 Jun;12(4):419-29. PMID: 19900250.

74. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health. 2008 Mar-Apr;11(2):322-33. PMID: 18380645.

75. Turner RR, Quittner AL, Parasuraman BM, et al. Patient-reported outcomes: instrument development and selection issues. Value Health. 2007 Nov-Dec;10 Suppl 2:S86-93. PMID: 17995478.

76. Fayers P, Machin D. Quality of Life: The Assessment, Analysis and Interpretation of Patient-reported Outcomes. 2nd ed: Wiley; 2007. p. 566.

77. Snyder CF, Watson ME, Jackson JD, et al. Patient-reported outcome instrument selection: designing a measurement strategy. Value Health. 2007 Nov-Dec;10 Suppl 2:S76-85. PMID: 17995477.

78. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010 Nov;63(11):1179-94. PMID: 20685078. PMCID: 2965562.

79. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA. 1995 Jan 4;273(1):59-65. PMID: 7996652.

80. Sloan JA, Halyard MY, Frost MH, et al. The Mayo Clinic Manuscript Series Relative to the Discussion, Dissemination, and Operationalization of the Food and Drug Administration Guidance on Patient-Reported Outcomes. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2007;10:S59-S63.

81. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34-42. PMID: 17161752.

82. Snaith RP. The Hospital Anxiety And Depression Scale. Health Qual Life Outcomes. 2003;1:29. PMID: 12914662. PMCID: 183845.

83. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003 May;41(5):582-92. PMID: 12719681.

84. Sprangers MA, Schwartz CE. The challenge of response shift for quality-of-life-based clinical oncology research. Ann Oncol. 1999 Jul;10(7):747-9. PMID: 10470418.

85. Varni JW, Thissen D, Stucky BD, et al. PROMIS(R) Parent Proxy Report Scales: an item response theory analysis of the parent proxy report item banks. Qual Life Res. 2012 Sep;21(7): 1223-40. PMID: 21971875.

86. Kutner JS, Bryant LL, Beaty BL, et al. Symptom distress and quality-of-life assessment at the end of life: the role of proxy response. J Pain Symptom Manage. 2006 Oct;32(4):300-10. PMID: 17000347.

87. Jones JM, McPherson CJ, Zimmermann C, et al. Assessing agreement between terminally ill cancer patients' reports of their quality of life and family caregiver and palliative care physician proxy ratings. J Pain Symptom Manage. 2011 Sep;42(3):354-65. PMID: 21454041.

88. Basch E, Abernethy AP. Supporting clinical practice decisions with real-time patient-reported outcomes. J Clin Oncol. 2011 Mar 10;29(8): 954-6. PMID: 21282536.

89. Cleeland CS, Wang XS, Shi Q, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. J Clin Oncol. 2011 Mar 10;29(8):994-1000. PMID: 21282546. PMCID: 3068055.

Chapter 6. Data Sources for Registries

1. Introduction

Identification and evaluation of suitable data sources should be completed within the context of the registry purpose and availability of the data of interest. A single registry may have multiple purposes and integrate data from various sources. While some data in a registry are collected directly for registry purposes (primary data collection), important information also can be transferred into the registry from existing databases. Examples include demographic information from a hospital admission, discharge, and transfer system; medication use from a pharmacy database; and disease and treatment information, such as details of the coronary anatomy and percutaneous coronary intervention from a catheterization laboratory information system, electronic medical record, or medical claims databases. In addition, observational studies can generate as many hypotheses as they test, and secondary sources of data can be merged with the primary data collection to allow for analyses of questions that were unanticipated when the registry was conceived.

This chapter will review the various sources of both primary and secondary data, comment on their strengths and weaknesses, and provide some examples of how data collected from different sources can be integrated to help answer important questions.

2. Types of Data

The types of data to be collected are guided by the registry design and data collection methods. The form, organization, and timing of required data are important components in determining appropriate data sources. Data elements can be grouped into categories identifying the specific variable or construct they are intended to describe. One framework for grouping data elements into categories follows:

- *Patient identifiers*—Some registries may use patient identifiers to link data. In these registries, data elements are linked to the specific patient through a unique patient identifier or registry identification number. The use of patient identifiers may not be possible in all registries due to the additional legal requirements that usually apply to the use and disclosure of such data. (See Chapter 7.)
- *Patient selection criteria*—The eligibility criteria in a registry protocol or study plan determine the group that will be included in the registry. These criteria may be very broad or restrictive, depending on the purpose. Criteria often include demographics (e.g., target age group), a disease diagnosis, a treatment, or diagnostic procedures and laboratory tests. Health care provider, health care facility or system, and insurance criteria may also be included in certain types of registries (e.g., following care patterns of specific conditions at large medical centers compared with small private clinics).
- *Treatments and tests*—Treatments and tests are necessary to describe the natural history of patients. Treatments can include pharmaceutical, biotechnology, or device therapies, or procedures such as surgery or radiation. Evaluation of the treatment itself is often a primary focus of registries (e.g., treatment safety and effectiveness over 5 years). Results of laboratory testing or diagnostic procedures may be included as registry outcomes and may also be used in defining a diagnosis or condition of interest.
- *Confounders*—Confounders are elements or factors that have an independent association with the outcomes of interest. These are particularly important because patients are typically not randomized to therapies in registries. Confounders such as comorbidities (disease diagnoses and conditions) can confuse analysis results and interpretation of causality. Information on the health care provider,

treatment facility, concomitant therapies, or insurance may also be considered. Unknown confounders, or those not recorded in the registry, pose particular challenges for the analysis of patient outcomes. If external, or linked, data sources may provide values for these confounder variables otherwise not in the registry, they may ultimately help reduce bias in the analysis and interpretation of patient outcomes.

- *Outcomes*—The focus of this document is on patient outcomes. Outcomes are end results and are defined for each condition. Outcomes may include patient-reported outcomes (PROs). In some registries, surrogate markers, such as biomarkers or other interim outcomes (e.g., hemoglobin A1c levels in diabetes) that are highly reflective of the longer term end results are used.

Before considering the potential sources for registry data, it is important to understand the types of data that may be collected in a registry. Several types of data that may be gathered from other sources in some registries are described below.

Cost/resource utilization—Cost and/or resource utilization data may be necessary to examine the cost-effectiveness of a treatment. Resource utilization data reflect the resources consumed (both services and products), while cost data reflect a monetary value assigned to those resources. Examples include the actual cost of the treatment (e.g., medication, screening, procedure) and the associated costs of the intervention (e.g., treatment of side effects, expenses incurred traveling to and from clinicians' appointments). Costs that are avoided due to the treatment (e.g., the cost to treat the avoided disease) and costs related to lost workdays may also be important to collect, depending on the objectives of the study. Registries that collect cost data over long periods of time (i.e., many years) may need to adjust costs for inflation during the analysis phase of the study. The types of data elements included in this framework are further described in Chapter 4 and below with respect to their source or the utility of the data for linking to other sources. Many of these may be available through data sources outside of the registry system.

Patient identifiers—Depending on the data sources required, some registries may use certain personal identifiers for patients in order to locate them in other databases and link the data. For example, Social Security numbers (SSNs) in combination with other personal identifiers can be used to identify individuals in the National Death Index (NDI). Patient contact information, such as address and phone numbers, may be collected to support tracking of participants over time. Information for additional contacts (e.g., family members) may be collected to support followup in cases where the patient cannot be reached. In many cases, patient informed consent and appropriate privacy authorizations are required so that personal identifiers can be used for registry purposes, and the use of personal identifiers may not be possible in some registries; Chapter 7 discusses the legal requirements for including patient identifiers. Systems and processes must be in place to manage security and confidentiality of these data. Confidentiality can be enhanced by assigning a registry-specific identifier via a crosswalk algorithm, as discussed below. Demographics, such as date of birth (to calculate age at any time point), gender, and ethnicity, are typically collected and may be used to stratify the registry population.

Disease/condition—Disease or condition data include those related to the disease or condition of focus for the registry and may incorporate comorbidities. Elements of interest related to the confirmation of a diagnosis or condition could be date of diagnosis and the specific diagnostic results that were used to make the diagnosis, depending on the purpose of the registry. Disease or condition is often a primary eligibility or outcome variable in registries, whether the intent is to answer specified treatment questions (e.g., measure effectiveness or safety) or to describe the natural history. This information may also be collected in constructing a medical history for a patient. In addition to "yes" or "no" to indicate presence or absence of the diagnosis, it may be important to capture responses such as "missing" or "unknown."

Treatment/therapy—Treatment or therapy data include specific identifying information for the primary treatment (e.g., drug name or code, biologic, device product or component parts, or

surgical intervention, such as organ transplant or coronary artery bypass graft) and may include information on concomitant treatments. Dosage (or parameters for devices), route of administration, and prescribed exposure time, such as daily or 3 times weekly for 4 weeks, should be collected. Pharmacy data may include dispensing information, such as the primary date of dispensation and subsequent refill dates. Data in device registries can include the initial date of dispensation or implantation and subsequent dates and specifics of required evaluations or modifications. Compliance data may also be collected if pharmacy representatives or clinic personnel are engaged to conduct and report pill counts or volume measurements on refill visits or return visits for device evaluations and modifications.

Laboratory/procedures—Laboratory data include a broad range of testing, such as blood, tissue, catheterization, and radiology. Specific test results, units of measure, and laboratory reference ranges or parameters are typically collected. Laboratory databases are becoming increasingly accessible for electronic transfer of data, whether through a system-wide institutional database or a private laboratory database. Diagnostic testing or evaluation may include procedures such as psychological or behavioral assessments. Results of these procedures and clinician exam procedures may be difficult to obtain through data sources other than the patient medical record.

Biosamples—The increased collection, testing, and storage of biological specimens as part of a registry (or independently as a potential secondary data source such as those described further below) provides another source of information that includes both information from genetic testing (such as genetic markers) and actual specimens.

Health care provider characteristics—Information on the health care provider (e.g., physician, nurse, or pharmacist) may be collected, depending on the purpose of the registry. Training, education, or specialization may account for differences in care patterns. Geographic location has also been used as an indicator of differences in care or medical practice.

Hospital/clinic/health plan—System interactions include office visits, outpatient clinic visits, emergency room visits, inpatient hospitalizations, procedures, and pharmacy visits, as well as associated dates. Data on all procedures as defined by the registry protocol or plan (e.g., physical exam, psychological evaluation, chest x-ray, CAT scan), including measurements, results, and units of measure where applicable, should be collected. Cost accounting data may also be available to match these interactions and procedures. Descriptive information related to the points of care may be useful in capturing differences in care patterns and can also be used to track patterns of referral of care (e.g., outpatient clinic, inpatient hospital, academic center, emergency room, pharmacy).

Insurance—The insurance system or payer claims data can provide useful information on interactions with the health care systems, including visits, procedures, inpatient stays, and costs associated with these events. When using these data, it is important to understand what services were covered under the various insurance plans at the time the data were collected, as this may affect utilization patterns.

3. Data Sources

Data sources are classified as primary or secondary based on the relationship of the data to the registry purpose. Primary data sources incorporate data collected for direct purposes of the registry (i.e., primarily for the registry). Primary data sources are typically used when the data of interest are not available elsewhere or, if available, are unlikely to be of sufficient accuracy and reliability for the planned analyses and uses. Primary data collection increases the probability of completeness, validity, and reliability because the registry drives the methods of measurement and data collection. (See Chapter 4.) These data are prospectively planned and collected under the direction of a protocol or study plan, using common procedures and the same format across all registry sites and patients. The data are readily integrated for tracking and analyses. Since the data

entered can be traced to the individual who collected them, primary data sources are more readily reviewed through automated checks or followup queries from a data manager than is possible with many secondary data sources.

Secondary data sources are comprised of data originally collected for purposes other than the registry under consideration (e.g., standard medical care, insurance claims processing). Data that are collected as primary data for one registry are considered secondary data from the perspective of a second registry if linking was done. These data are often stored in electronic format and may be available for use with appropriate permissions. Data from secondary sources may be used in two ways: (1) the data may be transferred and imported into the registry, becoming part of the registry database, or (2) the secondary data and the registry data may be linked to create a new, larger data set for analysis. This chapter primarily focuses on the first use for secondary data, while Chapters 16, 17, and 18 discuss the complexities of linking registries with other databases.

When considering secondary data sources, it is important to note that health professionals are accustomed to entering the data for defined purposes, and additional training and support for data collection are not required. Often, these data are not constrained by a data collection protocol and they represent the diversity observed in real-world practice. However, there may be increased probability of errors and underreporting because of inconsistencies in measurement, reporting, and collection. Staff changes can further complicate data collection and may affect data quality. There may also be increased costs for linking the data from the secondary source to the primary source and dealing with any potential duplicate or unmatched patients.

Sufficient identifiers are also necessary to accurately match data between the secondary sources and registry patients. The potential for mismatch errors and duplications must be managed. (See Case Example 40.) The complexity and obligations inherent in the collection and handling of personal identifiers have previously been mentioned (e.g., obligations for informed consent, appropriate data privacy, and confidentiality procedures).

Some of the secondary data sources do not collect information at a specific patient level but are anonymous and intended to reflect group or population estimates. For example, census tract or ZIP-Code-level data are available from the Census Bureau and can be merged with registry data. These data can be used as "ecological variables" to support analyses of income or education when such socioeconomic data are missing from registry primary data collection. The intended use of the data elements will determine whether patient-level information is required.

The potential for data completeness, variation, and specificity must be evaluated in the context of the registry and intended use of the data. It is advisable to have a solid understanding of the original purpose of the secondary data collection, including processes for collection and submission, and verification and validation practices. Questions to ask include: Is data collection passive or active? Are standard definitions or codes used in reporting data? Are standard measurement criteria or instruments used (e.g., diagnoses, symptoms, quality of life)? The existence and completeness of claims data, for example, will depend on insurance company coverage policies. One company may cover many preventive services, whereas another may have more restricted coverage. One company may cover a treatment without restriction, while another may require prior authorization by the physician or require that the patient must have first failed on a previous, less expensive treatment. Also, coverage policies can change over time. These variations must be known and carefully documented to prevent misinterpretation of use rates. Additionally, secondary data may not all be collected in the format (e.g., units of measure) required for registry purposes and may require transformation for integration and analyses.

An overview of some secondary data sources that may be used for registries is given below. Table 6–1 identifies some key strengths and limitations of the identified data sources.

Table 6–1. Key data sources—strengths and limitations

Data Source	Strength and Uses	Limitations
Patient-reported data	• Patient and/or caregiver outcomes. • Unique perspective. • Obtaining information on treatments not necessarily prescribed by clinicians (e.g., over-the-counter drugs, herbal medications). • Obtaining intended compliance information. • Useful when timing of followup may not be concordant with timing of clinical encounter.	• Literacy, language, or other barriers that may lead to underenrollment of some subgroups. • Validated data collection instruments may need to be developed. • Loss to followup or refusal to continue participation. • Limited confidence in reporting clinical information and utilization information.
Clinician-reported data	• More specific information than available from coded data or medical record.	• Clinicians are highly sensitive to burden. • Consistency in capture of patient signs, symptoms, use of nonprescribed therapy varies.
Medical chart abstraction	• Information on routine medical care, with more clinical context than coded claims. • Potential for comprehensive view of patient medical and clinical history. • Use of abstraction and strict coding standards (including handling missing data) increases the quality and interpretation of data abstracted.	• The underlying information is not collected in a systematic way. For example, a diagnosis of bacterial pneumonia by one physician may be based on a physical exam and patient report of symptoms, while another physician may record the diagnosis only in the presence of a confirmed laboratory test. • It is difficult to interpret missing data. For example, does absence of a specific symptom in the visit record indicate that the symptom was not present or that the physician did not actively inquire about this specific symptom or set of symptoms? • Data abstraction is resource intensive. • Complete medical and clinical history may not be available (e.g., new patient to clinic).

Table 6–1. Key data sources—strengths and limitations (continued)

Data Source	Strength and Uses	Limitations
Electronic health records (EHRs)	• Information on routine medical care and practice, with more clinical context than coded claims. • Potential for comprehensive view of patient medical and clinical history. • Efficient access to medical and clinical data. • Use of data transfer and coding standards (including handling of missing data) will increase the quality of data abstracted.	• Underlying information from clinicians is not collected using uniform decision rules. (See example under "Medical chart abstraction.") • Consistency of data quality and breadth of data collected varies across sites. • Difficult to handle information uploaded as text files into the EHRs (e.g., scanned clinician reports) vs. direct entry into data fields. • Historical data capture may require manual chart abstraction prior to implementation date of medical records system. • Complete medical and clinical history may not be available (e.g., new patient to clinic). • EHR systems vary widely. If data come from multiple systems, the registry should plan to work with each system individually to understand the requirements of the transfer.
Institutional or organizational databases	• Diagnostic and treatment information (e.g., pharmacy, laboratory, blood bank, radiology). • Resource utilization (e.g., days in hospital). • May incorporate cost data (e.g., billed and/or paid amounts from insurance claims submissions).	• Important to be knowledgeable about coding systems used in entering data into the original systems. • Institutional or organizational databases vary widely. The registry should plan to work with each system individually to understand the requirements of the transfer.

Table 6–1. Key data sources—strengths and limitations (continued)

Data Source	Strength and Uses	Limitations
Administrative databases	• Useful for tracking health care resource utilization and cost-related information. • Range of data includes anything that is reimbursed by health insurance, generally including visits to physicians and allied health providers, most prescription drugs, many devices, hospitalization(s), if a lab test was performed, and in some cases, actual lab test results for selected tests (e.g., blood test results for cholesterol, diabetes). • In some cases, demographic information (e.g., gender, date of birth from billing files) can be uploaded. • Potential for efficient capture of large populations.	• Represents clinical cost drivers vs. complete clinical diagnostic and treatment information. • Important to be knowledgeable about the process and standards used in claims submission. For example, only primary diagnosis may be coded and secondary diagnoses not captured. In other situations, value-laden claims may not be used (e.g., an event may be coded as a "nonspecific gynecologic infection" rather than a "sexually transmitted disease"). • Important to be knowledgeable about data handling and coding systems used when incorporating the claims data into the administrative systems. • Can be difficult to gain the cooperation of partner groups, particularly in regard to receiving the submissions in a timely manner.
Death indexes	• Completeness—death reporting is mandated by law in the United States. • Strong backup source for mortality tracking (e.g., patient lost to followup). • National Death Index (NDI)—centralized database of death records from State vital statistics offices; database updated annually. • NDI causes of death relatively reliable (93–96%) compared with State death certificates. • Social Security Administration's (SSA) Death Master File—database of deaths reported to SSA; database updated weekly.	• Time delay—indexes depend on information from other data sources (e.g., State vital statistics offices), with delays of 12 to 18 months or longer (NDI). It is important to understand the frequency of updates of specific indexes that may be used. • Absence of information in death indexes does not necessarily indicate "alive" status at a given point in time. • Most data sources are country specific and thus do not include deaths that occurred outside of the country. • As of November 2011, Death Master File no longer includes protected State records.

Table 6–1. Key data sources—strengths and limitations (continued)

Data Source	Strength and Uses	Limitations
U.S. Census Bureau databases	• Population data. • Core census survey conducted every decade. • Wide range in specificity of information from U.S. population down to neighborhood and household level. • Useful in determining population estimates (e.g., numbers, age, family size, education, employment status).	• Targets participants via survey sampling methodology and estimates. • Does not provide subject-level data.
Existing registries	• Can be merged with another data source to answer additional questions not considered in the original registry protocol or plan. • May include specific data not generally collected in routine medical practice. • Can provide historical comparison data. • Reduces data collection burden for sites, thereby encouraging participation.	• Important to understand the existing registry protocol or plan to evaluate data collected for element definitions, timing, and format, as it may not be possible to merge data unless many of these aspects are similar. • Creates a reliance on the other registry. Other registry may end. • Other registry may change data elements (which highlights the need for regular communication). • Some sites may not participate in both. Must rely on the data quality of the other registry.

Medical chart abstraction—Medical charts primarily contain information collected as a part of routine medical care. These data reflect the practice of medicine or health care in general and at a specific level (e.g., geographical, by specialty care provider). Charts also reflect uncontrolled patient behavior (e.g., noncompliance). Collection of standard medical practice data is useful in looking at treatments and outcomes in the real world, including all of the confounders that affect the measurement of effectiveness (as distinguished from efficacy) and safety outside of the controlled conditions of a clinical trial. Chart documentation is often much poorer than one might expect, and there may be more than one patient-specific medical record (e.g., hospital and clinical records). A pilot collection is recommended for this labor-intensive method of data collection to explore the availability and reproducibility of the data of interest. It is important to recognize that physicians and other clinicians do not generally use standardized data definitions in entering information into medical charts, meaning that one clinician's documented diagnosis of "chronic sinusitis" or "osteoarthritis" or description of "pedal edema" may differ from that of another clinician.

Electronic health records—The use of electronic health records (EHRs), sometimes called electronic medical records (EMRs), is increasing. EHRs have an advantage over paper medical records because the data in some EHRs can be readily searched and integrated with other information (e.g., laboratory data). The ease with which this is accomplished depends on whether

the information is in a relational database[a] or exists as scanned documents. An additional challenge relates to terminology and relationships. For example, including the term "fit" in a search for patients with epilepsy can yield a record for someone who was noted as "fit," meaning "healthy." Relationships can also be difficult to identify through searches (e.g., "Patient had breast cancer" vs. "Patient's mother had breast cancer"). The quality of the information has the same limitations as described in the paragraph above. Both the availability and standardization of EHR data have grown significantly in recent years, and this trend is expected to continue. As of 2009, some data suppliers cited individual data sets exceeding 10 million lives.[1] More recently, data suppliers are reporting 20 million[2] to 35 million[3] patients in their data sets. Further, it is anticipated that more significant standardization of EHR data will result from the "EHR certification" requirements being developed in phases under the American Recovery and Reinvestment Act of 2009 (ARRA). Such standardization should increase not only the availability and utility of EHR records, but also the ability to aggregate them into larger data sources.

Institutional or organizational databases—Institutional or organizational databases may be evaluated as potential sources of a wide variety of data. System-wide institutional or hospital databases are central data repositories, or data warehouses, that are highly variable from institution to institution. They may include a portion of everything from admission, discharge, and transfer information to data reflecting diagnoses and treatment, pharmacy prescriptions, and specific laboratory tests. Laboratory test data might be chemistry or histology laboratory data, including patient identifiers with associated dates of specimen collection and measurement, results, and standard "normal" or reference ranges. Catheterization laboratory data for cardiac registries may be accessible and may include details on the coronary anatomy and percutaneous coronary intervention. Other organizational examples are computerized order entry systems, pharmacies, blood banks, and radiology departments.

Administrative databases—Private and public medical insurers collect a wealth of information in the process of tracking health care, evaluating coverage, and managing billing and payment. Information in the databases includes patient-specific information (e.g., insurance coverage and copays; identifiers such as name, demographics, SSN or plan number, and date of birth) and health care provider descriptive data (e.g., identifiers, specialty characteristics, locations). Typically, private insurance companies organize health care data by physician care (e.g., physician office visits) and hospital care (e.g., emergency room visits, hospital stays). Data include procedures and associated dates, as well as costs charged by the provider and paid by the insurers. Amounts paid by insurers are often considered proprietary and unavailable. Standard coding conventions are used in the reporting of diagnoses, procedures, and other information. Coding conventions include the Current Procedure Terminology (CPT) for physician services and International Classification of Diseases (ICD) for diagnoses and hospital inpatient procedures. The databases serve the primary function of managing and implementing insurance coverage, processing, and payment. (See Case Example 12.)

Medicare and Medicaid claims files are two examples of commonly used administrative databases. The Medicare program covers over 43 million people in the United States, including almost everyone over the age of 65, people under the age of 65 who qualify for Social Security Disability, and people with end-stage renal disease.[4] The Medicaid program covers low-income children and their mothers; pregnant women; and blind, aged, or disabled people. As of 2007, approximately 40 million people were covered by Medicaid.[5] Medicare and Medicaid claims files, maintained by the Centers for Medicare & Medicaid Services (CMS), can be

[a]In a relational database, information is presented in tables with rows and columns. Data within a table may be related by a common concept, and the related data may be retrieved from the database. See: A Relational Database Overview. http://docs.oracle.com/javase/tutorial/jdbc/overview/database.html. Accessed July 16, 2013.

obtained for inpatient, outpatient, physician, skilled nursing facility, durable medical equipment, and hospital services. As of 2006, Medicare claim files for prescription drugs can also be obtained. The claims files generally contain person-specific data on providers, beneficiaries, and recipients, including individual identifiers that would permit the identity of a beneficiary or physician to be deduced. Data with personal identifiers are clearly subject to privacy rules and regulations. As such, the information is confidential and to be used only for reasons compatible with the purpose(s) for which the data are collected. The Research Data Assistance Center (ResDAC), a CMS contractor at the University of Minnesota, provides assistance to academic, government, and nonprofit researchers interested in using Medicare and/or Medicaid data for their research.[6]

Death and birth records—Death indexes are national databases tracking population death data (e.g., the NDI[7] and the Death Master File [DMF] of the Social Security Administration [SSA][8]). Data include patient identifiers, date of death, and attributed causes of death. These indexes are populated through a variety of sources. For example, the DMF includes death information on individuals who had an SSN and whose death was reported to the SSA. Reports may come in to the SSA by different paths, including from survivors or family members requesting benefits or from funeral homes. Because of the importance of tracking Social Security benefits, all States, nursing homes, and mortuaries are required to report all deaths to the SSA. Prior to 2011, the DMF contained virtually 100-percent complete mortality ascertainment for those eligible for SSA benefits. As of November 2011, however, the DMF no longer includes protected State death records. In practical terms, this means that approximately 4.2 million records were removed from the historical public DMF (which contained 89 million records), and some 1 million fewer records will be added to the DMF each year.[9] The NDI can be used to provide both fact of death and cause of death, as recorded on the death certificate. Cause-of-death data in the NDI are relatively reliable (93–96 percent) compared with death certificates.[10, 11] Time delays in death reporting should be considered when using these sources, and vital status should not be assumed to be "alive" by the absence of information at a recent point in time. These indexes are valuable sources of data for death tracking. Of course, mortality data can be accessed directly through queries of State vital statistics offices and health departments when targeting information on a specific patient or within a State. Likewise, birth certificates are available through State departments and may be useful in registries of children or births.

Area-level databases—Two sources of area-level data are the U.S. Census and the Area Health Resources Files (AHRF). The U.S. Census Bureau databases[12] provide population-level data utilizing survey sampling methodology. The Census Bureau conducts many different surveys, the main one being the population census. The primary use of the data is to determine the number of seats assigned to each State in the House of Representatives, although the data are used for many other purposes. These surveys calculate estimates through statistical processing of the sampled data. Estimates can be provided with a broad range of granularity, from population numbers for large regions (e.g., specific States), to ZIP Codes, all the way down to a household level (e.g., neighborhoods identified by street addresses). Information collected includes demographic, gender, age, education, economic, housing, and work data. The data are not collected at an individual level but may serve other registry purposes, such as understanding population numbers in a specific region or by specific demographics. The AHRF is maintained by the Health Resources and Services Administration, which is part of the Department of Health and Human Services. The AHRF includes county-level data on health facilities, health professions, measures of resource scarcity, health status, economic activity, health training programs, and socioeconomic and environmental characteristics.[13]

Provider-level databases—Data on medical facilities and physicians may be important for categorizing registry data or conducting subanalyses. Two sources of such data are the American Hospital Association's Annual Survey

Data and the American Medical Association's Physician Masterfile Data Collection. The Annual Survey Data is a longitudinal database that collects 700 data elements, covering organizational structure, personnel, hospital facilities and services, and financial performance, from more than 6,000 hospitals in the United States.[14] Each hospital in the database has a unique ID, allowing the data to be linked to other sources; however, there is a data lag of about 2 years, and the data may not provide enough nuanced detail to support some analyses of cost or quality of care. The Physician Masterfile Data Collection contains current and historic data on nearly one million physicians and residents in the United States. Data on physician professional medical activities, hospital and group affiliations, and practice specialties are collected each year.

Encounter-level databases—Databases of individual patient encounters (e.g., physician office visits, emergency department visits, hospital inpatient stays), generally do not contain individual patient identifiers and thus may not be linkable to patient registries, but nevertheless provide valuable insight into the makeup of the registry's target population. This is particularly true for data from nationally representative surveys, such as AHRQ's Health Care Utilization Project (H-CUP) Nationwide Inpatient Sample (NIS) and the suite of surveys by the Centers for Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS), including the National Ambulatory Medical Care Survey (NAMCS), the National Hospital Ambulatory Medicare Care Survey (NHAMCS), and the National Hospital Discharge Survey (NHDS).

Existing registry and other databases—There are numerous national and regional registries and other databases that may be leveraged for incorporation into other registries (e.g., disease-specific registries managed by nonprofit organizations, professional societies, or other entities). An example is the National Marrow Donor Program (NMDP),[15] a global database of cord blood units and volunteers who have consented to donate marrow and blood cells. Databases maintained by the NMDP include identifiers and locators in addition to information on the transplants, such as samples from the donor and recipient, histocompatibility, and outcomes. NMDP actively encourages research and utilization of registry data through a data application process and submission of research proposals.

The Registry of Patient Registries (RoPR) may become a useful resource for finding existing registries (https://patientregistry.ahrq.gov). RoPR is a database of registry-specific information intended to promote collaboration, reduce redundancy, and improve transparency in registry-based research. The database contains information on existing registries, such as the registry description, classification, and purpose, as well as the registry sponsor's interest in collaboration opportunities. Registry planners may be able to use RoPR to identify relevant registries to contact about data sharing or research collaborations.

In accessing data from one registry for the purposes of another, it is important to recognize that data may have changed during the course of the source registry, and this may or may not have been well documented by the providers of the data. For example, in the United States Renal Data System (USRDS),[16] a vital part of personal identification is CMS 2728, an enrollment form that identifies the incident data for each patient as well as other pertinent information, such as the cause of renal failure, initial therapy, and comorbid conditions. Originally created in 1973, this form is in its third version, having been revised in 1995 and again in 2005. Consequently, there are data elements that exist in some versions and not others. In addition, the coding for some variables has changed over time. For example, race has been redefined to correspond with Office of Management and Budget directives and Census Bureau categories. Furthermore, form CMS 2728 was optional in the early years of the registry, so until 1983 it was filled out for only about one-half of the subjects. Since 1995, it has been mandatory for all people with end-stage renal disease. These changes in form content, data coding, and completeness would not be evident to most researchers trying to access the data.

4. Other Considerations for Secondary Data Sources

The discussion below focuses on logistical and data issues to consider when incorporating data from other sources. Chapter 11 fully explores data collection, management, and quality assurance for registries.

Before incorporating a secondary data source into a registry, it is critical to consider the potential impact of the data quality of the secondary data source on the overall data quality of the registry. The potential impact of quality issues in the secondary data sources depends on how the data are used in the primary registry. For example, quality would be significant for secondary data that are intended to be populated throughout the registry (i.e., used to populate specific data elements in the entire registry over time), particularly if these populated data elements are critical to determining a primary outcome. Quality of the secondary data will have less effect on overall registry quality if the secondary data are to be linked to registry data only for a specific analytic study (see Chapter 18). For more information on data quality, see Chapter 11.

The importance of patient identifiers for linking to secondary data sources cannot be overstated. Multiple patient identifiers should be used, and primary data for these identifiers should not be entered into the registry unless the identifying information is complete and clear. While an SSN is very useful, high-quality probabilistic linkages can be made to secondary data sources using various combinations of such information as name (last, middle initial, and first), date of birth, and gender. For example, the NDI will make possible matches when at least one of seven matching conditions is met (e.g., one matching condition is "exact month and day of birth, first name, and last name"). However, the degree of success in such probabilistic and deterministic matching generally is enhanced by having many identifiers to facilitate matching. As noted earlier, the various types of data (e.g., personal history, adverse events, hospitalization, and drug use) have to be linked through a common identifier. A discussion of both statistical and privacy issues in linkage is provided in Chapter 16, and a discussion of managing patient identity across systems is provided in Chapter 17.

The best identifier is one that is not only unique but has no embedded personal identification, unless that information is scrambled and the key for unscrambling it is stored remotely and securely. The group operating the registry should have a process by which each new entry to the registry is assigned a unique code and there is a crosswalk file to enable the system to append this identifier to all new data as they are accrued. The crosswalk file should not be accessible by people or entities outside the management group.

In addition, consideration should be given to the fact that a registry may need to accept and link data sets from more than one outside organization. Each institution contributing data to the registry will have unique requirements for patient data, access, privacy, and duration of use. While having identical agreements with all institutions would be ideal, this may not always be possible from a practical perspective. Yet all registries have resource constraints, and decisions about including certain institutions have to be determined based on the resources available in order to negotiate specialized agreements or to maintain specialized requirements. Agreements should be coordinated as much as possible so that the function of the registry is not greatly impaired by variability among agreements. All organizations participating in the registry should have a common understanding of the rules regarding access to the data. Although exceptions can be made, it should be agreed that access to data will be based on independent assessment of research protocols and that participating organizations will not have individual veto power over access.

When data from secondary sources are used, agreements should specify ownership of the source data and clearly permit data use by the recipient registry. The agreements should also specify the roles of each institution, its legal responsibilities, and any oversight issues. It is critical that these issues and agreements be put in place before data are transferred so that there are no ambiguities or unforeseen restrictions on the recipient registry later on.

Some registries may wish to incorporate data from more than one country. In these cases, it is important to ensure that the data are being collected in the same manner in each country or to plan for any necessary conversion. For example, height and weight data collected from sites in Europe will likely be in different units than height and weight data collected from sites in the United States. Laboratory test results may also be reported in different units, and there may be variations in the types of pharmaceutical products and medical devices that are approved for use in the participating countries. Understanding these issues prior to incorporating secondary data sources from other countries is extremely important to maintain the integrity and usefulness of the registry database.

When incorporating other data sources, consideration should also be given to the registry update schedule. A mature registry will usually have a mix of data update schedules. The registry may receive an annual update of large amounts of data, or there could be monthly, weekly, or even daily transfers of data. Regardless of the schedule of data transfer, routine data checks should be in place to ensure proper transfer of data. These should include simple counts of records as well as predefined distributions of key variables. Conference calls or even routine meetings to go over recent transfers will help avoid mistakes that might not otherwise be picked up until much later.

An example of the need for regular communication is a situation that arose with the United States Renal Data System a few years ago. The United Network for Organ Sharing (UNOS) changed the coding for donor type in their transplant records. This resulted in an apparent 100-percent loss of living donors in a calendar year. The change was not conveyed to USRDS and was not detected by USRDS staff. After USRDS learned about the change, standard analysis files that had been sent to researchers with the errors had to be replaced.

Distributed data networks are another model for sharing data. In a distributed data network, data sharing may be limited to the results of analyses or aggregated data only. There is much interest in the potential of distributed data networks, particularly for safety monitoring or public health surveillance (see Chapter 15, Section 11). However, the complexities of data sharing within a distributed data network are still being addressed, and it is premature to discuss good practice for this area.

5. Summary

In summary, a registry is not a static enterprise. The management of registry data sources requires attention to detail, constant feedback to all participants, and a willingness to make adjustments to the operation as dictated by changing times and needs.

Case Example for Chapter 6

Case Example 12. Using claims data along with patient-reported data to identify patients

Description	The National Amyotrophic Lateral Sclerosis (ALS) Registry is a rare disease registry created by the Agency for Toxic Substances and Disease Registry (ATSDR) within the U.S. Department of Health and Human Services (HHS). The purpose of the registry is to quantify the incidence and prevalence of ALS in the United States, describe the demographics of people with ALS, and examine potential risk factors for the disease.
Sponsor	U.S. Department of Health and Human Services and Agency for Toxic Substances and Disease Registry, through funding from the "ALS Registry Act" (U.S. Congress Public Law 110-373).
Year Started	2010
Year Ended	Ongoing
No. of Sites	All 50 States, including U.S. territories; data from national administrative databases are combined with patient self-enrollment data.
No. of Patients	The first registry report is anticipated for release in spring 2014.

Challenge

Amyotrophic lateral sclerosis (ALS) is a progressive, fatal neurodegenerative disorder of both the upper and lower motor neurons. Many knowledge gaps exist in the understanding of ALS, including uncertainty about the disease's incidence and prevalence, misdiagnosis of ALS in patients with other motor neuron disorders, and the role of environmental exposures in the etiology of ALS. Because ALS is a non-reportable disease in the United States (except for the Commonwealth of Massachusetts), previous attempts to estimate ALS incidence and prevalence using nonspecific mortality data have faced many challenges and at best overestimated disease frequency. Identifying patients through site recruitment for research purposes poses additional challenges, as access to patient medical records can be limited, costly, and time-consuming to obtain. Patient recruitment issues are compounded by the complexities of this rare disease, in which the average timeframe from diagnosis to death is 2–5 years. U.S. governmental agencies acknowledged that a national, structured data collection program for ALS was greatly needed, and that alternative data sources and recruitment strategies would need to be identified.

Proposed Solution

In 2008, President Bush signed the ALS Registry Act into law, allowing ATSDR to create the National ALS Registry. The registry is the only Congressionally mandated population-based ALS registry in the United States. As a first step in developing the registry, a workshop of international experts in neurological and autoimmune conditions was convened to discuss approaches to creating a national database. Based on feedback from these experts, the registry uses a two-pronged approach to identify all U.S. cases of ALS. The first approach uses national administrative databases, including those of Medicare, Medicaid, the Veterans Health Administration, and the Veterans Benefit Administration, to identify prevalent cases based on an algorithm developed through pilot projects. These administrative databases cover approximately 90 million Americans, and the algorithm identifies 80 to 85 percent of all true ALS cases when applied to these databases. The second approach uses a secure Web portal to allow patients to self-enroll voluntarily. Data

Case Example 12. Using claims data along with patient-reported data to identify patients (continued)

Proposed Solution (continued)

from the two approaches are combined into the registry database, and duplicate patients are identified and removed so that each person with ALS is counted only once in the registry.

Results

The registry data will support several research projects. The Web portal for self-enrolled participants contains brief surveys that collect information on potential risk factors, such as socio-demographic characteristics, occupational history, military history, cigarette smoking, alcohol consumption, physical activity, family history of neurodegenerative diseases, and disease progression. ATSDR is also currently implementing active surveillance projects that will allow population-based case estimates of ALS in certain smaller geographic areas (i.e., at the State and metropolitan levels) to help ATSDR evaluate the completeness of the registry. In addition, ATSDR has developed a system to inform people with ALS about new research (e.g., clinical trials, epidemiological studies) for which they may be eligible. Lastly, ATSDR is funding a feasibility study for the creation of a national biospecimen repository that would be open to all U.S. residents with ALS who are enrolled in the registry. This proposed biorepository will help researchers better understand the disease because it will pair biospecimens (e.g., blood, brain tissue) with existing risk-factor data from patients.

Key Point

Combining multiple data sources, such as administrative databases and patient-reported information, is a novel approach and can be an effective way to successfully identify patients with a rare disease and to better understand the prevalence, incidence, and etiology of the disease. However, using alternative approaches requires a strong understanding of the nuances of the individual data sources; pilot testing is also helpful to identify potential issues with data sources prior to registry launch.

For More Information

http://wwwn.cdc.gov/als

References for Chapter 6

1. Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress. June 30, 2009. U.S. Department of Health and Human Services. http://www.med.upenn.edu/sleepctr/documents/FederalCoordinatingCouncilforCER_2009.pdf. Accessed August 14, 2012.
2. GE Healthcare. "Medical Quality Improvement Consortium.". http://www3.gehealthcare.com/en/Products/Categories/Healthcare_IT/Clinical_Knowledge_Solutions/MQIC. Accessed August 15, 2012.
3. Practice Fusion. "Practice Fusion Releases EMR Data set, Launches Health Data Challenge with Kaggle." http://www.practicefusion.com/pages/pr/health-data-initiative-forum-challenge-2012.html. Accessed August 15, 2012.
4. Centers for Medicare and Medicaid Services. "Medicare Coverage – General Information." http://www.cms.gov/Medicare/Coverage/CoverageGenInfo/index.html. Accessed August 15, 2012.
5. DeNavas-Walt C, Proctor BD, Smith JC. Current Population Reports, P60-235, Income, Poverty, and Health Insurance Coverage in the United States: 2007. Washington, DC: U.S. Government Printing Office; 2008. http://www.census.gov/prod/2008pubs/p60-235.pdf. Accessed July 16, 2013.
6. Research Data Assistance Center. http://www.resdac.org. Accessed August 15, 2012.
7. National Center for Health Statistics. National Death Index. http://www.cdc.gov/nchs/ndi.htm. Accessed July 16, 2013.
8. Social Security Administration. Death Master File. National Technical Information Service. http://www.ntis.gov/products/ssa-dmf.aspx. Accessed August 15, 2012.

9. National Technical Information Service. Important Notice: Change in Public Death Master File Records. http://www.ntis.gov/pdf/import-change-dmf.pdf. Accessed August 14, 2012.

10. Doody MM, Hayes HM, Bilgrad R. Comparability of national death index plus and standard procedures for determining causes of death in epidemiologic studies. Ann Epidemiol. 2001 Jan;11(1):46-50. PMID: 11164119.

11. Sathiakumar N, Delzell E, Abdalla O. Using the National Death Index to obtain underlying cause of death codes. J Occup Environ Med. 1998 Sep;40(9):808-13. PMID: 9777565.

12. U.S. Bureau of the Census. http://www.census.gov. Accessed August 15, 2012.

13. Health Resources and Services Administration. Area Health Resources Files (AHRF). http://arf.hrsa.gov. Accessed August 15, 2012.

14. American Hospital Association. AHA Data and Directories. http://www.aha.org/research/rc/stat-studies/data-and-directories.shtml. Accessed August 15, 2012.

15. National Marrow Donor Program. http://www.bethematch.org. Accessed August 15, 2012.

16. United States Renal Database. http://www.usrds.org. Accessed August 15, 2012.

Section II

Legal and Ethical Considerations for Registries

Chapter 7. Principles of Registry Ethics, Data Ownership, and Privacy

1. Introduction

This chapter covers the ethical and legal considerations that should accompany the development and use of all health information registries, including patient registries as defined in this document, for the purposes of public health activities, governmental health program oversight, quality improvement/assurance (I/A), and research. These considerations apply generally accepted ethical principles for scientific research involving human subjects to health information registries. Related topics include issues of transparency in the operation of registries, oversight of registry activities, and property rights in health care information and registries.

Section 2.1 of this chapter discusses the ethical concerns and considerations involved with obtaining and using confidential health information in registries. Section 2.2 describes the transformation of ethical concerns into the legal regulation of human subjects research and the privacy of individually identifiable health information. In Section 3, an overview is presented of these regulatory requirements and their interactions as they specifically relate to registries. Section 4 makes recommendations about registry transparency and oversight, based on the need to ensure the independence, integrity, and credibility of biomedical research, while preserving and improving the utility of registry data. Finally, property rights in health information and registries are briefly discussed. Table 7-1, at the end of this chapter, provides an overview of the applicable regulatory requirements based on the type of registry developer and the extent to which registry data are identifiable.

The purpose of this chapter is solely to provide information that will help readers understand the issues, not to provide specific legal opinions or regulatory advice. Legal advisors should always be consulted to address specific issues and ensure that all applicable Federal, State, and local laws and regulations are followed. The discussion below about legal protections for the privacy of health information focuses solely on U.S. law. Health information is also legally protected in European and some other regions by distinctly different rules, none of which are discussed in this chapter.[1] If registry developers intend to obtain health information from outside of the United States or transfer to or share their information with registries outside the United States, they should consult legal counsel early in the registry planning process for the necessary assistance. It also should be noted that the rules and regulations described here are to protect patients and research participants, not to prevent legitimate research. While the requirements may seem daunting, they are not insurmountable barriers to research. With careful planning and legal guidance, registries can be designed and operated in compliance with applicable rules and regulations.

In the context of this chapter, health information is broadly construed to include any information created or used by health care providers and insurance plans that relates to an individual's health condition, the provision of health care services to an individual, or payment for health care services provided to an individual.[2] As a result, health information may include a broad range of demographic information and personal characteristics, such as socioeconomic and marital status, the extent of formal education, developmental disability, cognitive capacities, emotional stability, and gender, age, and race, all of which may affect health status or health risks. Typically, health information includes information about family members, so it also can have an impact on the privacy of third parties. Patients widely regard health information as private and thus expect confidentiality to be maintained.

Concerns about potential risks to individual privacy have led to Federal legal requirements for prospective review of research projects and conditions on the use or disclosure of health information for research and other purposes. The creation and use of patient registries for a research

purpose ordinarily constitute "research involving human subjects" as defined by regulations applicable to research activities funded by the U.S. Department of Health and Human Services[3] (HHS) and certain other Federal agencies. Moreover, Federal privacy regulations promulgated under the Health Insurance Portability and Accountability Act of 1996 (HIPAA)[4] and modified by the Health Information Technology for Economic and Clinical Health Act (HITECH – part of the American Recovery and Reinvestment Act of 2009) specifically apply to the use and disclosure of certain individually identifiable health information for research and other purposes.

The term *human subjects* is used throughout this chapter for consistency with applicable Federal law. Some may prefer the term *research participants*.

This chapter provides a general guide to Federal legal requirements in the United States. (Legal requirements in other countries may also be relevant and may be different from those in this country, but a discussion of any applicable international rules is beyond the scope of this document.) These legal requirements may influence registry decisions involving the selection of data elements and data verification procedures, and may also affect subsequent uses of registry data for secondary research purposes. State laws also may apply to the use of health information for research purposes. The purpose of a registry, the status of its developer, and the extent to which registry data are identifiable largely determine applicable regulatory requirements. This chapter reviews the most common of these arrangements. The complexity and sophistication of registry structures and operations vary widely, with considerable variability also observed in the processes registry stewards use to obtain data. Nonetheless, common ethical and legal principles are associated with the creation and use of registries. These commonalities are the focus of this chapter.

Ethical concerns about the conduct of biomedical research, especially research involving the interaction of the clinical research community with its patients and commercial funding agencies, have produced an impetus to make financial and other arrangements more public. The discussion of transparency in this chapter includes recommendations for the public disclosure of registry operations as a means of maintaining public trust and confidence in the use of health information for registry purposes. Reliance on a standing advisory committee is recommended to registry developers as a way to provide expert technical guidance for registry operations and to firmly establish the independence of the registry from committed or conflicted interests, as described in Chapter 2. This discussion of transparency in methods is not intended to discourage private investments in registries that produce proprietary information in some circumstances. Neither the funding source nor the generation of proprietary information from a registry determines whether a registry exercises and adheres to the good practices described in this guide.

Health care providers and health insurance plans have plausible claims of ownership to health and claims information, although the public perspective on these claims has not been tested. Registry developers should anticipate negotiating access to health and claims information, especially when it is maintained in electronic form. Registry developers also are likely to encounter licensing requirements, including processing and use fees, in obtaining health and claims information. The processes for use of registry data sets, especially in multiple analyses by different investigators, should be publicly disclosed to assure the public that registries are appropriately protecting the confidentiality of health information.

2. Ethical Concerns Relating to Health Information Registries

2.1 Application of Ethical Principles

The Belmont Report[5] is a summary of the basic principles and guidelines developed to assist in resolving ethical problems in conducting research using human subjects. It was the work product of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral

Research, which was created by the National Research Act of 1974.[6]

The Belmont Report identifies three fundamental principles for the ethical conduct of scientific research that involves human subjects. These principles are respect for persons as autonomous agents (self-determination), beneficence (do good, do no harm, protect from harm), and justice (fairness, equitable distribution of benefits and burdens, equal treatment). Together, they provide a foundation for the ethical analysis of human subjects research, including the use of health information in registries developed for scientific purposes with a prospect of producing social benefits. These principles are substantively the same as those identified by the Council for International Organizations of Medical Sciences in its international guidelines for the ethical review of epidemiologic studies.[7]

Nevertheless, the application of these principles to specific research activities can result in different conclusions about what comprises ethical design and conduct of the research in question. These different conclusions frequently occur because the principles are assigned different values and relative importance when more than one person performs the ethical analysis. In most of these situations, however, a generally supported consensus position on the ethical design and conduct of the research is a desired and achievable goal. This goal does not preclude re-analysis as social norms or concerns about research activities change over time in response to new information, new technologies or persistent ethical questioning.

The ethical principle of *respect for persons* supports the practice of obtaining individuals' consent to the use of their health information for research purposes related or unrelated to the clinical and insurance reasons for creating the information. In connection with research registries, consent may have multiple components: (1) consent to registry creation by the compilation of patient information; (2) consent to the initial research purpose and uses of registry data; and (3) consent to subsequent use of registry data by the registry developer or others for the same or different research purposes. The consent process should adequately describe registry purposes and operations to inform potential subjects' decisions about participation in a research registry. In some defined circumstances, the principle of *respect for persons* may be subordinate to other ethical principles and values, with the result that an explicit consent process for participation in the registry may not be necessary. A waiver of informed consent requirements may apply to the registry and be ethically acceptable. (See discussion of waivers of informed consent requirements in Section 3.3.5.) In these situations, alternatives to an explicit consent process for each individual contributing health information to the registry may be adequate. For example, the registry might provide readily accessible, publicly available information about its activities as an alternative to individual informed consent.

A general ethical requirement for consent clearly implies that human subjects voluntarily permit the use of their health information in a registry, unless a specific exception to voluntary participation applies to the registry. One such exception is a legally mandated, public health justification for the compilation of health information (e.g., certain infectious disease reporting). Voluntary agreement to the use of health information in a registry necessarily allows a subsequent decision to discontinue participation. Any limitation on an individual's ability to withdraw information from the registry (e.g., once incorporation into aggregated data has occurred) should be clearly communicated in the consent process as a condition of initial participation. The consent process should also include instructions about the procedures for withdrawal at any time from participation in the registry unless a waiver of consent applies to the registry. Incentives for registry use of health information (e.g., insurance coverage of payments for health care services) should be carefully evaluated for undue influence both on the individuals whose health information is sought for registry projects and on the health care providers of those services.[8, 9]

Conflicts of interest also may result in undue influence on patients and may compromise voluntary participation. One potential source of conflict widely identified within clinical research is the use of recruitment incentives paid by

funding agencies to health care providers.[10] Some professional societies and research organizations have established policy on the use of recruitment incentives. Many entities have characterized as unethical incentives that are significantly beyond fair market value for the work performed by the health care provider; others require disclosure to research subjects of any conflicting interest, financial or nonfinancial.[11] Federal law now requires manufacturers of certain drugs, devices, or medical supplies to report, for public display, the amounts of remuneration paid to physicians for research purposes.[12] Some States, including Massachusetts, have similar laws in effect.[13] Research organizations, particularly grantees of Federal research funding, may have systematic policies and procedures in place that registry developers can rely on for managing employee conflicts of interest. Nonetheless, in their planning, registry developers should specify and implement recruitment practices that protect patients against inappropriate influences.

Applying the principle of *respect for persons* to the research use of health information generates additional ethical concerns about preserving the privacy and dignity of patients, protecting the confidentiality of health information, and minimizing potential harms. These concerns have intensified as health care services, third-party payment systems, and health information systems have become more complex. Legal standards for the use and disclosure of health information create a baseline of required privacy protections for individually identifiable health information. However, depending on the particular health condition or population of interest, safeguards for the confidentiality of registry data beyond applicable legal requirements may be ethically necessary to protect the privacy and dignity of those individuals contributing health information to the registry.

The principle of *beneficence* ethically obligates developers of health information registries for research purposes to minimize potential harms to the individuals or groups[14] whose health information is included in the registry. There are usually no apparent benefits to offset potential harm to the individuals or groups whose health information is used in the registry. Exceptions to this arise when a registry is designed to provide benefits to the human subjects as individuals, such as longitudinal reports on treatment effects or health status or quality-of-care reports. Risks to privacy and dignity are minimized by conscientious protection of the confidentiality of the health information included in the registry[15] through the use of appropriate physical, technical, and administrative safeguards for data in the operations of the registry. These safeguards should include controls on access to registry data, including access to individual identifiers that may be included in registry data. Minimization of risks also requires a precise determination of what information is necessary for the research purposes of the registry and limiting the information collected accordingly.

Certain populations of patients may be vulnerable to social, economic, or psychological harms as a result of a stigmatizing health condition. Developers of registries compiling this health information must make special efforts to protect the identities of the human subjects contributing data to the registry. Additional legal protections may apply if HHS-supported research is being conducted through or in connection with the registry. Additional protections also apply to populations such as pregnant women, human fetuses, neonates, prisoners, and children, who are considered vulnerable to undue influence and coercion during the consent process. In particular, data obtained from pediatric and adolescent populations may lead to ethical concerns if there is the potential for lifelong discrimination that may effectively exclude them from educational opportunities and other social benefits[16] (e.g., health insurance, although under the Affordable Care Act health insurers may not discriminate against individuals on the basis of pre-existing conditions).

In an analysis applying the principle of *beneficence*, research involving human subjects that is unlikely to produce valid scientific information is unethical. This conclusion is based on the lack of social benefit to offset even minimal risks imposed by the research on participating individuals. Health information registries should

incorporate an appropriate design (including, where appropriate, calculation of the patient sample as described in Chapter 3) and data elements, written operating procedures, and documented methodologies, as necessary, to ensure the fulfillment of a valid scientific purpose.[17]

An ethical analysis employing the principle of *justice* also yields candid recognition of the potential risks to those who contribute health information to a registry, and the probable lack of benefit to those individuals (except in the cases where registries are specifically constructed to provide benefit to those individuals). The imbalance of burden and benefit to individuals reinforces the need to minimize the risks from registry use of health information. Precise and well-developed scientific reasons for inclusion (or exclusion) of defined health information in a registry help ensure that the burden placed on individuals as a result of their participation is fair and equitable.

The above analysis refers to research activities. However, the ethical concerns expressed may also apply to other activities involving the use or disclosure of individuals' health information for nonresearch purposes. Public health, oversight of the delivery of health care services through government programs, and quality I/A activities all can evoke the same set of ethical concerns as research activities about the protection of patient self-determination, privacy, and dignity; the maintenance of the confidentiality of individually identifiable health information to avoid potential harms; and the imposition of a risk of harm on some individuals to the benefit of others not at risk. In the past, different assignments of social value to these activities and different potential for the social benefits and harms they produce have created different levels of social acceptance and formal oversight for these activities compared with research activities. Nonetheless, these activities may include a research component in addition to their primary stated objectives, a circumstance that implicates the ethical concerns discussed above and produces additional concerns about compliance with the legal requirements for research activities. Registry developers should prospectively apply careful scrutiny to the proposed purposes for and activities of a registry, in consultation with appropriate institutional officials, to avoid both ethical and compliance issues that may undermine achievement of the registry's objectives.

Registry developers also must consider confidentiality and/or proprietary concerns with regard to the identity of the health care providers, at the level of both individual professionals and institutions, and the health care insurance plans from which they obtain registry data. Information about health care providers and insurance plans can also identify certain patient populations and, in rare circumstances, individual patients. Moreover, the objectives of any registry, broadly speaking, are to enhance the value of the health care services received, not to undermine the credibility and thus the effectiveness of health care providers and insurance plans in their communities. Developers of registries created for public health investigations, health system oversight activities, and quality I/A initiatives to monitor compliance with recognized clinical standards must consider whether safeguards for the identity of service professionals and institutions are appropriate. At the same time, however, any confidentiality safeguards should permit certain disclosures, as permitted by applicable law and designated by the service professionals and institutions, for the reporting of performance data, which are increasingly associated with payment from payers.

2.2 Transformation of Ethical Concerns Into Legal Requirements

Important ethical concerns about the creation, maintenance, and use of patient registries for research purposes include risks of harm to human subjects resulting from unauthorized access to registry data and inappropriate use of the compiled health information. These concerns about harms arise from public expectations of confidentiality for health information and the importance of that confidentiality in preserving the privacy and dignity of individual patients as well as the clinician/patient relationship.

Over the last decade, two rapid technological developments have intensified these ethical

concerns. One of these advances was DNA sequencing, replication, recombination, and the concomitant application of this technology to biomedical research activities in human genetics. Widespread anticipation of potential social benefits produced by biomedical research as a result of these technologies was accompanied by ethical concern about the potential for affronts to personal dignity and economic, social, or psychological harms to individuals or related third parties.

In addition to specific ethical concerns about the effect of technological advances in biomedical research, general social concerns about the privacy of patient information have accompanied the advance of health information systems technology and electronic information processing, as applied to the management and communication of health information. These social concerns produced legal privacy protections, first in Europe and later in the United States. The discussion below about legal protections for the privacy of health information focuses solely on U.S. law.

2.2.1 The Common Rule

International and domestic concerns about the protection, respect, and privacy of human subjects resulted in a uniform set of regulations from the Federal agencies that fund such research known as the "Common Rule."[18, 19] The legal requirements of the Common Rule apply to research involving human subjects conducted or supported by the 17 Federal departments and agencies that adopted the Rule. Some of these agencies may require additional legal protections for human subjects. The HHS regulations will be used for all following references to the Common Rule.

Among these requirements is a formal written agreement, from each institution engaged in such research, to comply with the Common Rule. For human subjects research conducted or supported by most of the Federal entities that apply the Common Rule, the required agreement is called a Federalwide Assurance (FWA).[20] Research institutions may opt in their FWA to apply Common Rule requirements to all human subjects research activities conducted within their facilities or by their employees and agents, regardless of the source of funding. The application of Common Rule requirements to a particular registry depends on the institutional context of the registry developer, relevant institutional policies, and whether the health information contributed to the registry maintains patient identifiers.

The Office for Human Research Protections (OHRP) administers the Common Rule as it applies to human subjects research conducted or supported by HHS. Guidance published by OHRP discusses research use of identifiable private health information. This guidance makes clear that OHRP considers the creation of health information registries—containing individually identifiable, private information—for research purposes to be human subjects research for the institutions subject to its jurisdiction.[21] The applicability of the Common Rule to research registries is discussed in more detail in Section 4.

OHRP regulations for human subjects protection require prospective review and approval of the research by an institutional review board (IRB) and the informed consent (usually written) of each of the human subjects involved in the research, unless an IRB expressly grants a waiver of informed consent requirements.[22] A research project must satisfy certain regulatory conditions to obtain IRB approval of a waiver of the informed consent requirements. (See Section 3.3.5 for discussion of waivers of informed consent requirements.) A registry plan is the research "protocol" reviewed by the IRB. At a minimum, the protocol should identify (1) the research purpose of a health information registry, (2) detailed arrangements for obtaining informed consent, or detailed justifications for not obtaining informed consent, to collect health information, and (3) appropriate safeguards for protecting the confidentiality of registry data, in addition to any other information required by the IRB on the risks and benefits of the research.[23]

As noted previously, for human subjects research conducted or supported by most Federal departments and agencies that have adopted the Common Rule, an FWA satisfies the requirement for an approved assurance of compliance. Some research organizations extend the application of their FWA to all research, regardless of the funding source. Under these circumstances, any patient information registry created and

maintained within the organization may be subject to the Common Rule. In addition, some research organizations have explicit institutional policies and procedures that require IRB review and approval of all human subjects research.

2.2.2 The Privacy Rule

In the United States, HIPAA and its implementing regulations[4] (here collectively called the Privacy Rule) created legal protections for the privacy of individually identifiable health information created and maintained by "covered entities" and their "business associates." "Individually identifiable health information" is information, including demographic data, created or received by a health care provider, health plan, employer, or health care clearinghouse, that identifies an individual or could be used to identify an individual, and relates to (1) an individual's past, present, or future physical or mental health condition; (2) the provision of health care to an individual; or (3) the past, present, or future payment for health care to an individual. With certain exceptions, "individually identifiable health information" is "protected health information" (PHI) under the Privacy Rule when it is transmitted or maintained by a covered entity or a business associate on behalf of a covered entity.[24] Because registries may exist over long periods of time, it is important to note one exception that provides that the individually identifiable information of persons who have been deceased for more than 50 years is not considered PHI.

Covered entities are health care providers that engage in certain standard financial or administrative health care transactions electronically, health plans, and health care clearinghouses.[25] Business associates generally are persons or organizations, other than a member of a covered entity's workforce, that perform certain functions or services (e.g., claims processing, data analysis, data aggregation, patient safety activities) on the covered entity's behalf that involve access to protected health information.[26] For the purposes of this chapter, the relevant entities are covered health care providers, which may include individual health care providers (e.g., a physician, pharmacist, or physical therapist), health care insurance plans, and their business associates. The discussion in this chapter assumes that the data sources for registries are HIPAA-covered entities or their business associates, which are subject to provisions of the Privacy Rule. In the event that a registry developer intends to collect and use data from sources that are not covered entities or their business associates under the Privacy Rule, such as personal health record vendors not working on behalf of a covered entity or business associate, these sources still may be subject to other laws, such as State laws, that may govern the protection of patient information.

Although data sources are assumed to be subject to the Privacy Rule, registry developers and the associated institutions where the registry will reside may not be. Notably, the Privacy Rule does not apply to registries that reside outside of a covered entity or business associate. Within academic medical centers, for example, registry developers may be associated with units that are outside of the institutional health care component to which the Privacy Rule applies, such as a biostatistics or economics department. But because many, if not virtually all, data sources for registries are covered entities or their business associates, registry developers are likely to find themselves deeply enmeshed in the Privacy Rule. In addition, the formal agreements required by the Privacy Rule in certain circumstances in order to access, process, manage, and use certain forms of patient information impose legally enforceable continuing conditions upon data recipients under contract law. Such conditions of use also may result in direct liability under HIPAA if the registry is considered a business associate of a data source that is a covered entity (see the discussion in Section 3.4.3 below of the HITECH Act, which extended direct liability for compliance with certain requirements of the HIPAA Rules to business associates of covered entities, where before business associates were required and liable to protect the information to which they had access only through their business associate contracts with covered entities). Therefore, registry developers should be cognizant of the patient privacy considerations confronting their likely data sources—as well as themselves, if they are performing functions or services on behalf of their data sources as business associates—and should consider following certain

Privacy Rule procedures, required or not, depending on their arrangements with those data sources.

In general, the Privacy Rule defines the circumstances under which covered entities (e.g., health plans and most health care providers) and their business associates may use and disclose patient information for a variety of purposes, including research. The Privacy Rule establishes a Federal baseline of protections, which do not pre-empt more stringent (more protective) State laws.[27] For example, the Privacy Rule requires covered entities to include certain information in patient authorizations for the use or disclosure of individually identifiable information, including an expiration date or event that can be many years in the future. The laws of the State of Maryland, however, specifically require that, absent certain exceptions, a patient's authorization may only be valid for a maximum period of 1 year.[28] In this case, a covered entity located in Maryland can and should satisfy both the Privacy Rule and State law requirements by complying with the State's one-year maximum expiration deadline on its patient authorization forms.

The Privacy Rule regulates the use of identifiable patient information within health care providers' organizations and health insurance plans (and within their business associates), and the disclosure of patient information to others outside of the entity.[29] As a result, the initial collection of registry data from covered entities or business associates is subject to Privacy Rule requirements, which may apply in different ways depending on the registry's purpose, whether the registry resides within a covered entity or outside of a covered entity, whether the registry is considered a business associate of the covered entity, and the extent to which the patient information identifies individuals. Health care providers or insurance plans, as well as their business associates, that create, use, and disclose patient information for clinical use or business purposes are subject to civil—and in some cases criminal—liability for violations of the Privacy Rule.

Registry developers should be sufficiently knowledgeable about the Privacy Rule to facilitate the necessary processes for their data sources. In developing a registry, they should expect to interact with clinicians, the privacy officer, the IRB or privacy board staff, health information system representatives, legal counsel, compliance officials, and contracting personnel. Registry developers should also maintain awareness of regulatory modifications or amendments to, or new guidance on how to comply with, the Privacy Rule, which can be expected as the use of electronic health information becomes more prevalent. For example, on January 25, 2013, HHS issued significant modifications to the Privacy Rule, many of which implemented HITECH Act requirements.[30] One of the most relevant modifications for registry developers, as mentioned above and discussed more fully below in Section 3.4.3, is the extension of certain Privacy Rule requirements and liability directly to business associates.[31]

Subsequent use and sharing of registry data may be affected by the regulatory conditions that apply to initial collection, as well as by other ethical concerns and legal issues. The Privacy Rule created multiple pathways by which registries can compile and use patient information. For instance, a registry within a covered entity may obtain a HIPAA authorization from each patient contributing identifiable information to a registry for a particular research project, such as the relationship between hypertension and Alzheimer's disease. If the registry subsequently seeks to use the data for another research purpose, it may do so if it obtains new authorizations or satisfies the conditions of another permission in the Privacy Rule—for example, by de-identifying the data to Privacy Rule standards. Alternatively, the registry can obtain authorizations for use and disclosure of individuals' health information for future research purposes at the same time that it obtains authorization to place the information in the registry, as long as the authorization for future research use or disclosure adequately describes the purposes of the future research such that it would be reasonable for the individual to expect that his or her information could be used or disclosed for the future research.

The authors recommend that registry developers establish a detailed tracking system, based on the

extent to which registry data remain identifiable for individual patients, for the collection, uses, and disclosures of registry data. The tracking system should produce comprehensive documentation of compliance with both Privacy Rule requirements and legally binding contractual obligations to data sources.

With regard to registries developed for research purposes, the Privacy Rule defines research as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[32] Commentary by HHS on the Privacy Rule explicitly includes within this definition of research the development (building and maintenance) of a repository or database for future research purposes.[33] The definition of research in the Privacy Rule partially restates the definition of research in the preexisting Common Rule for the protection of human subjects, adopted by HHS and certain other Federal agencies.[34] Some implications of this partial restatement of the definition of research are discussed later in this chapter.

The National Institutes of Health (NIH) has published guidance, in collaboration with the Office for Civil Rights and other HHS offices and agencies, on the impact of the Privacy Rule on health services research and research databases and repositories. The NIH guidance identifies the options available to investigators under the Privacy Rule to gain access to health information held by health care providers and insurance plans.[35] For example, in addition to provisions for the use or disclosure of identifiable patient information for research, the Privacy Rule permits health care providers and insurance plans (and business associates on their behalf) to use or disclose patient information for certain defined public health activities.[36] The Privacy Rule defines a public health authority as "an agency or authority of the United States, a State, a territory, a political subdivision of a State or territory, or an Indian tribe, or a person or entity acting under a grant of authority from or contract with such public agency… that is responsible for public health matters as part of its official mandate."[32] The Centers for Disease Control and Prevention and HHS have jointly published specific guidance on the Privacy Rule for public health activities.[37] Other Privacy Rule provisions permit the use or disclosure of patient health information as required by other laws.[38]

The protections for patient information created by the Privacy Rule that are generally relevant to registries developed for research purposes include explicit individual patient authorization for the use or disclosure of identifiable information,[39] legally binding data use agreements for the release of "limited data sets" between health information sources and users,[40] the removal of specified identifiers or statistical certification to achieve de-identification of health information,[41] an accounting of disclosures to be made available to patients at their request,[42] and notification in the event of a breach of unsecured protected health information to affected individuals who may be affected by the breach. In addition, if certain criteria required by the Privacy Rule are satisfied, an IRB or privacy board may grant a waiver of individual patient authorization for the use or disclosure of health information in research.[43]

2.2.3 FDA Regulations

U.S. Food and Drug Administration (FDA) regulatory requirements for research supporting an application for FDA approval of a product also include protections for human subjects, including specific criteria for protection of privacy and maintaining the confidentiality of research data.[44]

2.2.4 Applicability of Regulations to Research; Multiple-Purpose Registries

At many institutions, the IRB or the office that provides administrative support for the IRB is the final arbiter of the activities that constitute human subjects research, and thus may itself determine what activities require IRB review. A registry developer is strongly encouraged to consult his or her organization's IRB or a central IRB, as applicable, early in the registry planning process to avoid delays and lessen the need for multiple revisions of documentation submitted to the IRB. Distinctions between research and other activities that apply scientific methodologies are frequently unclear. Such other activities include both public

health practice[45] and quality-related investigations.[46] Both the ostensible primary and secondary purposes of an activity are factors considered in the determination of whether registry activities constitute research subject to the Common Rule. As interpreted by OHRP, if any secondary purpose of an activity is research, then the activity should be considered research.[47] This OHRP interpretation of research purpose differs from that of the Privacy Rule with respect to quality-related studies performed by health care providers and insurance plans. Under the Privacy Rule, only if the primary purpose of a quality-related activity is to obtain generalizable knowledge do the research provisions of the Privacy Rule apply; otherwise, the Privacy Rule defines the activity as a "health care operation."[48]

Registry developers should rely on their privacy officer's and IRB's experience and resources in defining research and other activities for their institutions and determining which activities require IRB review as research. In meeting accreditation standards, inpatient facilities typically maintain standing departmental (e.g., pediatrics) or service (e.g., pharmacy or nursing) committees to direct, review, and analyze quality-related activities. Some physician groups also establish and maintain quality-related programs, because good clinical practice includes ongoing evaluation of any substantive changes to the standard of care. These institutional quality committees can provide guidance on the activities that usually fall within their purview. Similarly, public health agencies typically maintain systematic review processes for identifying the activities that fit within their legal authority.

As mentioned previously, use of registry data for multiple research purposes may entail obtaining additional permissions from patients or satisfying different regulatory requirements for each research purpose. Standard confidentiality protections for registry data include requirements for physical, technical, and administrative safeguards to be incorporated into plans for a registry. In some instances, an IRB may not consider legally required protections for the research use of patient information sufficient to address relevant confidentiality concerns, including the Privacy Rule protections that may be applicable to registries created by or maintained within covered entities, such as health care providers and insurance plans, or business associates. For example, information about certain conditions (such as alcoholism or HIV-positive status) and certain populations (such as children) may be associated with a greater potential for harm from social stigma and discrimination. Under these circumstances, the IRB can make approval of a registry plan contingent on implementation of additional safeguards that it determines are necessary to minimize the risks to the individuals contributing health information to the registry.

3. Applicable Regulations

This section discusses the specific applicability of the Common Rule[49] and the Privacy Rule[50] to the creation and use of health information registries. Registry developers are strongly encouraged to consult with their organization's privacy officer and IRB or privacy board early in the planning process to clarify applicable regulatory requirements and the probable effect of those requirements on registry design and development.

This discussion assumes three general models for health information registries. One model is the creation of a registry containing the contact, demographic, and diagnostic or exposure information of potential research subjects who will be individually notified about projects in which they may be eligible to participate. The notification process permits the registry to shield registry participants from an inordinate number of invitations to participate in research projects, as well as to protect privacy and confidentiality. This model is particularly applicable to patients with unusual conditions, patients who constitute a vulnerable population,[51] or both (e.g., children with a rare condition). A second model is the creation of a registry and the conduct of all subsequent research using registry data by the same group of investigators. No disclosures of registry data will occur and all research activities have the same scientific purpose. This model

applies, in general, to quality improvement registries and other quality-related investigations of a clinical procedure or service. Note, however, that some quality improvement registries may involve confidential feedback to providers as well as public reporting of provider performance in a patient de-identified format. These activities may or may not constitute research as defined by the Common Rule. Under the Privacy Rule, these activities may be regulated as the health care operations of the covered entity that provides the data to the registry, rather than research, provided the obtaining of generalizable knowledge is not the primary purpose of the activities. A third model is the creation of a registry for an initial, specific purpose by a group of investigators with the express intent to use registry data themselves, as well as to disclose registry data to other investigators for additional related or unrelated scientific purposes. An example of this last model is a registry of health information from patients diagnosed with a condition that has multiple known comorbidities to which registry data can be applied. This third model is most directly applicable to industry-sponsored registries. The American College of Epidemiology encourages the data sharing contemplated in this last registry model.[52] Data sharing enhances the scientific utility of registry data and diminishes the costs of compilation.

The extent to which the regulations will apply to each of these registry models will depend on factors such as the registry developer, purpose of the registry, potential for individual patient identification, consent process, and inclusion of genetic information. These factors are discussed further below.

3.1 Public Health, FDA-Regulated Products, Health Oversight

When Federal, State, or municipal public health agencies create registries in the course of public health practice, specific legislation typically authorizes the creation of the registries and regulates data acquisition, maintenance, security, use, and disclosures of registry data for research. Ethical considerations and concerns about maintaining the confidentiality of patient information used by public health authorities are similar to those for research use, but they generally are explicitly balanced against potential social benefits during the legislative process. Nonetheless, if the registry supports human subjects research activities as well as its public health purposes, Common Rule requirements for IRB review may apply to the creation and maintenance of the registry.

Cancer registries performing public health surveillance activities mandated by State law are well-known exceptions to Common Rule regulation. However, secondary uses of public health registry data for research and the creation of registries funded by public health agencies, such as the Centers for Disease Control and Prevention and the Agency for Healthcare Research and Quality, may be subject to the Common Rule as sponsored research activities. The Common Rule's definitions of human subjects research[34] may encompass these activities, which are discussed in the next subsections of this chapter. Not all cancer registries support public health practice alone, even though the registries are the result of governmental programs. For example, the Surveillance Epidemiology and End Results (SEER) program, funded by the National Cancer Institute, operates and maintains a population-based cancer reporting system of multiple registries, including public use data sets with public domain software. SEER program data are used for many research purposes in addition to aiding public health practices. These latter research activities may be subject to the Common Rule.[53]

Disclosures of health information by health care providers and insurance plans and their business associates for certain defined public health activities are expressly permitted by the Privacy Rule without patient authorization.[36] An example of a public health activity is the practice of surveillance, in which the distributions and trends of designated risk factors, injuries, or diseases in populations are monitored and disseminated.[54] Health care providers or insurance plans are likely to insist upon documentation of public health authority for legal review before making any

disclosures of health information. Registry developers should obtain this documentation from the agency that funds or enters into a contract for the registry, and present it to the health care provider or insurance plan well in advance of data collection efforts.

The Privacy Rule characterizes responsibilities related to the quality, safety, or effectiveness of a product or activity regulated by FDA as public health activities. This public health exception allows uses and disclosures of patient information to a person subject to FDA jurisdiction with respect to FDA-regulated products or activities for which the person has responsibility, such as for adverse event reporting; product tracking; product recalls, repairs, replacement, or look-back; and postmarketing surveillance (e.g., as part of a risk management program that is a condition for approval of an FDA-regulated product).[55]

The Privacy Rule also permits uses and disclosures by health care providers and insurance plans (and their business associates on their behalf) for "health oversight activities" authorized by law.[56] These activities include audits and investigations necessary for oversight of the "health care system" and other entities subject to government regulatory programs for which health information is relevant to determining compliance with program standards.[57] The collection of patient information, such as occurrences of decubitus ulceration, from nursing homes that are operating under a compliance or corporate integrity agreement with a Federal or State health care program, is an example of a health oversight activity.

3.2 Research Purpose of a Registry

The Common Rule defines research, and its definition is partially restated in the Privacy Rule, as described above. These regulatory definitions affect how the regulatory requirements of each rule are applied to registry activities.[58]

In the Common Rule:

> Research means a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other purposes. For example, some demonstration and service programs may include research activities.[34]

OHRP interprets this Common Rule definition of research to include activities having any research purpose, no matter what the stated objectives of the activity may be. Compliance with Common Rule requirements depends on the nature of the organization where the registry resides. If an organization receives Federal funding for research, then it is likely that Common Rule requirements apply.

The Privacy Rule's definition of research[32] restates the first sentence of the Common Rule definition. However, the Privacy Rule distinguishes between research and quality I/A conducted by covered entities or their business associates,[59] which are defined as "health care operations."[32] Under the Privacy Rule, if the primary purpose of a quality-related registry maintained by a covered entity is to support a research activity (i.e., to create generalizable knowledge), Privacy Rule requirements for research apply to the use or disclosure of the patient information to create the registry and to subsequent research use of registry data. If, however, the primary purpose is other than to create generalizable knowledge, the study is considered a health care operation of the covered entity and is not subject to Privacy Rule requirements for research activities, including the requirement to obtain patient authorization or a waiver of authorization from an IRB or privacy board for the uses or disclosures.

As noted earlier, both public health practice and quality I/A can be difficult to distinguish from research activities.[60] The determination of whether a particular registry should be considered as or include a research activity depends on a number of different factors, including the nature of the organization where the registry will reside; the employment duties of the individuals performing the activities associated with the registry; the source of funding for the registry; the original,

intended purpose of the registry; the sources of registry data; whether subsequent uses or disclosures of registry data are likely; and other circumstances of registry development.

Quality I/A activities entail many of the same ethical concerns about protecting the confidentiality of health information as research activities do. Obtaining express patient consent to participate in quality I/A activities is not the usual practice; instead, the professional and cultural norms of health care providers, both individual and institutional, regulate these activities. Registry developers should consider whether the ethical concerns associated with a proposed quality I/A or patient safety registry require independent review and the use of special procedures such as notice to patients or providers. Registry advisory committee members, quality I/A and patient safety literature,[61] hospital ethics committees, IRB members, and clinical ethicists can make valuable contributions to these decisions.

To avoid surprises and delays, the decision about the nature of the activity that the registry is intended to support should be made prospectively, in consultation with appropriate officials of the funding agency and officials of the organization where the registry will reside. Some research institutions may have policies that either require IRB review for quality I/A, especially if publication of the activity is likely, or exclude them from IRB review. Frequently, IRBs make this determination on a case-by-case basis.

3.3 Potential for Individual Patient Identification

The specific regulatory requirements applicable to the use or disclosure of patient information for the creation of a registry to support human subjects research depend in part on the extent to which patient information received and maintained by the registry can be attributed to a particular person. Various categories of information, each with a variable potential for identifying individuals, are distinguished in the Privacy Rule: *individually identifiable* health information, *de-identified* information (all identifying elements removed), and a *limited data set* of information (specified direct identifiers removed).[62] The latter two categories of information may include certain codes linked to identifiers, provided certain conditions are met.

If applicable, Common Rule requirements affect all research involving patient information that is individually identifiable and obtained by the investigator conducting the research. The definition of "human subject" in the Common Rule is "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information."

This regulatory definition further explains that:

> Private information includes…information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.[63]

In short, the Common Rule definition of *human subject* makes all research use of identifiable patient information subject to its requirements; if the identity of the patients whose information is used for research purposes is not readily ascertainable to the investigator, the research is not human subjects research to which the Common Rule applies. Moreover, research involving the collection of information from existing records is exempt from the Common Rule if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through a coded link to identifiers. Registry developers should consult the IRB early in the process of selecting data elements to obtain guidance about whether registry activities constitute human subjects research or may be exempt from Common Rule requirements.

Also among the criteria specified by the Common Rule for IRB approval of research involving human subjects are provisions to protect the privacy of subjects and to maintain the confidentiality of data.[64] In addition, the consent process for research subjects should include explicit information about the confidentiality protections in place when records containing identifiers are going to be used.[65]

Data collection frequently requires patient identifiers, especially in prospective registries with ongoing data collection, revision, and updates. Secondary or subsequent research use by outside investigators (i.e., those not involved in the original data collection) of patient information containing direct identifiers is complicated, however, because ethical principles for the conduct of human subjects research require that risks, including risks to confidentiality of patient identifiable information, be minimized. In addition, the Privacy Rule requires patient authorization to describe the purpose of the use or disclosure of patient information. In order for a patient authorization to allow for secondary research uses of the patient information, the authorization must adequately describe the purpose of the future research, such that it would be reasonable for the patient to expect that his or her information could be used or disclosed for such purpose. Thus, unless the registry developer has anticipated and adequately described the purposes of the secondary research in the initial authorization received from a patient, that initial authorization may not constitute authorization for the use of identifiable registry data for secondary research purposes. In cases where there is no authorization for the secondary research, there may be other options under the Privacy Rule for secondary use of the data collected, such as de-identification of the information or the creation of a limited data set.[66] Chapter 16 provides a discussion of the technical and legal considerations related to linking registry data for secondary research purposes.

Direct identifiers, as specified by the Privacy Rule, include, for example, a patient's name, contact information, medical record number, and Social Security number. As stated in the Privacy Rule standard, a limited data set of patient information does not include specified direct identifiers of the patient, or the patient's relatives, employer, or household members.[67]

In an electronic environment, masking of individual identities is a complex task. Data suppression limits the utility of the information from the registry. Linkage or triangulation of information can enable the re-identification of individuals. A technical assessment of electronic records for their uniqueness within any data set is necessary to minimize the potential for re-identification. In aggregated published data, standard practice assumes that a subgroup size of less than six may also be identifiable, depending on the nature of the data. An evaluation for uniqueness should be performed to ensure that the electronic format does not produce a potential for identification greater than this standard practice, including when the information is triangulated within a record or linked with other data files.

If a registry for research, public health, or other purposes will use any of the categories of health information discussed below, a registry developer should consult the IRB, the privacy officer, and the institutional policies developed specifically in response to the Privacy Rule early in his or her planning. These consultations should establish the purpose of the registry, the applicability of the Common Rule requirements to registry activities, and the applicability of the Privacy Rule to the collection and use of registry data. In addition, the registry developer should consult a representative of the information technology or health information system office of each health care provider or insurance plan that will be a source of data for the registry, as well as a representative of the IRB or privacy board for each data source, so as to obtain feasibility estimates of data availability and formats.

3.3.1 De-Identified Patient Information

The Privacy Rule describes two methods for de-identifying health information.[68] One method requires the removal of certain data elements. The other method requires that a qualified expert certify that the potential for identifying an individual from the data elements is very small.

A qualified expert should have "appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable" in order to make this determination.[69] De-identified information may include a code permitting re-identification of the original record by the data source (covered entity), provided certain conditions are met.[70] The code may not be derived from information about an individual, including hash codes,[71] and should resist translation. In addition, the decoding key must remain solely with the health care provider or plan that is the source of the patient information, and the covered entity cannot use or disclose the code for any other purpose.[70]

Research using existing data in which individual patients cannot be identified directly or indirectly through linked identifiers does not involve human subjects as defined by the Common Rule, and thus is not subject to the requirements of the Rule.[63] Refer to the discussion later in this chapter.

As a prudent business practice, each health care provider or insurance plan or their business associate that is a source of de-identified information is likely to require an enforceable legal agreement with the registry developer. It should be signed by an appropriate institutional official on behalf of the registry developer. At a minimum, this agreement will likely contain the following terms, some of which may be negotiable: the identification of the content of the data and the medium for the data; a requirement that the data recipient, and perhaps the health care provider or insurance plan or their business associate providing the data, make no attempt to identify individual patients; the setting of fees for data processing and data use; limitations on disclosure or further use of the data, if any; and an allocation of the risks of legal liability for any improper use of the data.

3.3.2 Limited Data Sets of Health Information

De-identified health information may not suffice to carry out the purposes of a registry, especially if the registry is designed to receive followup information as a result of monitoring patients over time or information from multiple sources in order to compile information on a health event (e.g., cancer incidence). Dates of service and geographic location may be crucial to achieving the purposes of the registry or to the integrity and use of the data. Health information provided to the registry that is not fully de-identified but excludes direct identifiers may constitute a *limited data set* as defined by the Privacy Rule.[67] A health care provider or insurance plan (or business associate on behalf of a provider or plan if permitted by the terms of the business associate agreement) may disclose a limited data set of health information for research, public health, or health care operations purposes, provided it enters into a data use agreement (DUA) with the recipient. The terms of the DUA must satisfy specific Privacy Rule requirements.[72] Institutional officials for both the data source and the registry developer should sign the DUA so that a legal contract results. The DUA establishes the permitted uses of the limited data set by the registry developer (i.e., the creation of the registry and subsequent use of registry data for specified research purposes). The DUA may not authorize the registry developer to use or disclose the information in a way that would result in a violation of the Privacy Rule if done by the data source.[73] Furthermore, the DUA for a limited data set of health information must provide that the data recipient will appropriately safeguard the information and not attempt to identify individual patients or to contact those patients.[74] Certain other requirements also apply.

An investigator who works for a health care provider or insurance plan to which the Privacy Rule applies and that is the source of the health information for a registry may use a limited data set to develop a registry for its own research purpose. In these circumstances, the Privacy Rule still requires a DUA that satisfies the requirements of the Privacy Rule between the health care provider or insurance plan and the investigator. This agreement may be in the form of a written confidentiality agreement.[75]

A registry developer may assist a health care provider or insurance plan or their business associate by creating the limited data set.[67] In some situations, this assistance may be crucial to ensuring that data are accessible and available to the registry. In order for the registry developer to create a limited data set on behalf of a data source, the Privacy Rule requires that the data source (the covered entity or their business associate) and the registry developer (in this instance acting as a business associate) enter into a business associate agreement that satisfies certain regulatory criteria.[76] The business associate agreement is a binding legal arrangement that should be signed by appropriate institutional officials on behalf of the data source and registry developer. This agreement contains terms for managing health information as required by the Privacy Rule.[76] Most health care providers have developed a standard business associate agreement in response to the Privacy Rule and will likely insist on using it, although some modifications may need to be negotiated in order to produce registry data. A registry developer hired to create a limited data set must return or destroy the direct identifiers once the business associate relationship formed for purposes of creating the limited data set terminates and the registry developer now wants to use the limited data set (subject to the required data use agreement) for the registry purposes.

The registry populated with a limited data set may include a coded link that connects the data back to patient records, provided the link does not replicate part of a direct identifier.[67] The key to the code (e.g., encryption key) may allow health information obtained from patients over time to supplement existing registry data or allow the combination of information from multiple sources.

If the registry data obtained by investigators constitute a limited data set, then the research does not involve human subjects, as defined by HHS regulations at 45 Code of Federal Regulations (CFR) 46.102(f), and the Common Rule requirements would not apply to the registry.[77] An IRB or an institutional official knowledgeable about the Common Rule requirements should make the determination of whether a research registry involves human subjects; frequently, a special form for this purpose is available from the IRB. The IRB (or institutional official) should provide the registry developer with documentation of its decision.

3.3.3 Direct Identifiers: Authorization and Consent

As discussed above, the Privacy Rule permits the use or disclosure of patient information for research with a valid, written authorization from each patient whose information is used or disclosed.[78] The Privacy Rule specifies the content of this authorization, which gives permission for a specified use or disclosure of the health information.[79] Health care providers and insurance plans frequently insist on using the specific authorization forms that they have developed in order to avoid additional legal review and minimize any potential liability that they believe might be associated with use of other forms.

One exception to the requirement for an authorization occurs when a health care provider or insurance plan creates a registry to support its "health care operations."[80] The Privacy Rule's definition of "health care operations" explicitly includes quality I/A, outcomes evaluation, and the development of clinical guidelines, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities.[32] For example, a hospital registry created to track its patient outcomes against a recognized clinical care standard as a quality improvement initiative has a health care operations purpose. The hospital would not be required to obtain authorizations from its patients for use or disclosure of the health information it tracks in this registry.

Research use of health information containing identifiable information constitutes human subjects research as defined by the Common Rule.[63] In general, the Common Rule requires documented, legally effective, voluntary, and informed consent of each research subject.[81]

Documentation of the consent process required by the Common Rule may be combined with the authorization required by the Privacy Rule for

disclosure and use of health information.[82] However, registry developers should be aware that a health care provider or insurance plan may not immediately accept the combined form as a valid authorization and may insist on legal review of the combined form before disclosing any health information.

Authorizations for the use or disclosure of health information under the Privacy Rule and informed consent to participate in research under the Common Rule must be legally effective (i.e., obtained from a legally competent subject or the subject's legally authorized representative and documented in a manner consistent with the regulations and applicable laws of the jurisdiction). Adults, defined in most States as persons who are at least 18 years old, are generally presumed legally competent in the absence of a judicially approved guardianship. Minors commonly are defined as individuals less than 18 years old and are presumed legally incompetent; therefore, a biological, adoptive, or custodial parent or guardian generally must provide consent and authorization on the minor child's behalf, unless the child has been legally emancipated or another exception applies. Registry developers should consult legal counsel about situations in which these presumptions seem inapplicable, such as when a registry is created to investigate contraceptive drug and device use by adolescents, or other situations in which State or Federal law exceptions may apply.

In addition to being voluntary and legally effective, an individual's consent should be informed about the research, including what activities are involved, as well as the expected risks and potential benefits from participation. The Common Rule requires the consent process to include specific elements of information.[81] Registry developers should provide non–English-speaking patients with appropriate resources to ensure that the communication of these elements during the consent process is comprehensible. All written information for patients should be translated, or else arrangements should be made for qualified translators to assist in the consent process.

IRBs may approve waivers or alterations of both authorization (for use or disclosure of patient information for registry purposes) and consent (to registry participation), provided the research use of health information satisfies certain regulatory conditions. In addition, the Privacy Rule provides for the ability of privacy boards to approve waivers of authorization for the research use of health information where an organization does not have an IRB.[83] Waivers are discussed in detail below.

In certain limited circumstances, research subjects can consent to future unspecified research using their identifiable patient information. The Common Rule permits an IRB-approved consent process to be broader than a specific research project[81] and to include information about research that may be done in the future. In its review of such future research, an IRB can subsequently determine that the previously obtained consent (1) satisfies or (2) does not satisfy the regulatory requirements for informed consent. If the previously obtained consent is not satisfactory, an additional consent process may be required; alternatively, the IRB may grant a waiver of consent, provided the regulatory criteria for a waiver are satisfied.

As such, an IRB-approved consent process for the creation of a research registry should include a description of the specific types of research to be conducted using registry data. For any future research that involves private identifiable information maintained by the registry, the IRB may determine that the original consent process (for the creation of the research registry) satisfies the applicable regulatory requirements because the prospect of future research and future research projects were adequately described. The specific details of that future research using registry data may not have been known when data were collected to create the registry, but that research may have been sufficiently anticipated and described to satisfy the regulatory requirements for informed consent. For consent to be informed as demanded by the ethical principle of respect for persons, however, any description of the nature and purposes of the research should be as specific as possible.

If a registry developer anticipates subsequent research use of identifiable private registry data, he or she should request an assessment by the IRB of the description of the research that will be used in the consent process for potential subjects at the time the data are initially collected. Nonetheless, in its review of any subsequent research, an IRB may require an additional consent process for each research subject or may grant a waiver for obtaining further consent.

Historically, HHS rejected broadening the description of purpose in authorizations under the Privacy Rule to allow for future unspecified research.[84] As a result, an authorization for the use or disclosure of health information to create a research registry could not also authorize the future research uses of the information if the specific details of the future studies were not known when the authorization was obtained.[85] However, under the modified HIPAA Privacy Rule released on January 25, 2013, HHS modified its prior interpretation and guidance that research authorizations must be research study specific.[86] While this modification does not make any changes to the authorization requirements at 45 CFR § 164.508, HHS no longer interprets the "purpose" provision for authorizations as permitting only study-specific descriptions. This change now allows future research to be authorized provided the authorization adequately describes the purposes of any future research such that it would be reasonable for the individual to expect that his or her health information could be used or disclosed for such future research.[87] Where an authorization for the use or disclosure of registry data for the future research does not exist, a health care provider or health insurance plan maintaining the registry may need to obtain an additional authorization for the research from individuals or seek a waiver of authorization from an IRB or privacy board. Alternatively, the use or disclosure of a limited data set or de-identified registry data can occur, provided regulatory criteria are satisfied. Registries maintained by organizations to which the Privacy Rule does not apply (e.g., funding agencies for research that are not health care providers or insurance plans, professional societies, or non-health care components of hybrid entities such as in many universities) are not legally bound by the limited purpose of any original authorization that was obtained to permit data sources to disclose identifiable patient information to the registry. However, data sources or their business associates that are subject to the Privacy Rule are unlikely to be willing to provide patient information without a written agreement with the registry developer that includes legally enforceable protections against redisclosure of identifiable patient information. Regardless of whether such a written agreement is in place, a valid authorization must contain a warning to patients that their health information may not be protected by Privacy Rule protections once disclosed to recipient organizations.[88]

In addition to (or in lieu of) obtaining health information from covered entities or business associates, registry developers can request that patients obtain and share with the registry copies of their own records from their health care providers or insurance plans. This strategy can be useful for collecting data on mobile populations, such as elderly retirees who occupy different residences in winter and summer, and for collecting the health records of school children. A Federal privacy law[89] protects the health records of children that are held by schools from disclosure without explicit parental consent; thus, parents can often obtain copies of these records more easily than investigators. Alternatively, individuals can simply be asked to volunteer health information in response to an interview or survey. These information collection strategies do not require obtaining a Privacy Rule authorization from each subject; however, IRB review and other requirements of the Common Rule, including careful protections of the confidentiality of registry data still may apply to a registry project with a research purpose. Moreover, a registry developer may encounter Privacy Rule requirements for the use or disclosure of patient information by a health care provider or insurance plan for purposes of recruiting registry participants. For example, a patient authorization or waiver of authorization (discussed below) may be necessary for the disclosure of patient contact information by a health care provider or insurance plan (or their business associate) to a registry developer.

3.3.4 Certificates of Confidentiality and Other Privacy Protections

Certificates of confidentiality granted by the NIH permanently protect identifiable information about research subjects from legally compelled disclosure. For the purposes of certificates of confidentiality, identifiable information is broadly defined to include any item, or combination of items, in research data that could directly or indirectly lead to the identification of a research participant.[90] Federal law authorizes the Secretary of HHS (whose authority is delegated to NIH) to provide this privacy protection to subjects of biomedical, behavioral, clinical, and other research.[91] Federal funding for the research is not a precondition for obtaining a certificate of confidentiality.[92] An investigator whose research project has been granted a certificate of confidentiality may refuse to disclose identifying information collected for that research even though a valid subpoena demands that information for a civil, criminal, administrative, or legislative proceeding at the Federal, State, or local level. The protection provided by a certificate of confidentiality is intended to prevent the disclosure of personal information that could result in adverse effects on the social, economic, employment, or insurance status of a research subject.[93] Detailed information about certificates of confidentiality is available on the NIH Web site.[94]

The grant of a certificate of confidentiality to a research project, however, is not intended to affect State laws requiring health care and other professionals to report certain conditions to State officials; for example, designated communicable diseases, neglect and abuse of children and the elderly, or threats of violent harm. If investigators are mandatory reporters under State law, in general, they continue to have a legal obligation to make these reports.[95] In addition, other limitations to the privacy protection provided by certificates of confidentiality exist and may be relevant to particular research projects. Information on the NIH Web site describes some of these other legal limitations.[94]

Registry developers should also be aware that Federal law provides specific confidentiality protections for the identifiable information of patients in drug abuse and alcoholism treatment programs that receive Federal funding.[96] These programs may disclose identifiable information about their patients for research activities only with the documented approval of the program director or authorization of the patient.[97] The basis for the director's approval is receipt of written assurances about the qualifications of the investigator to conduct the research and the confidentiality safeguards incorporated into the research protocol, and an assurance that there will be no further disclosure of identifying information by the investigator. Moreover, an independent review of the research project should determine and verify in writing that the protocol provides adequate protection of the rights and welfare of the patients and that the benefits of the research outweigh any risks to patients.[97] Prior to submitting proposed consent documentation to an IRB, registry developers should consult legal counsel about the limitations of these confidentiality protections.

As a condition of approval, IRBs frequently require investigators to obtain a certificate of confidentiality for research involving information about substance abuse or other illegal activities (e.g., underage purchase of tobacco products), sexual attitudes and practices, and genetic information. Registry developers should consult legal counsel to determine if and how the limitations of a certificate of confidentiality may affect privacy protection planning for registry data. In all circumstances, the consent process should ensure that clear notice is given to research subjects about the extent of privacy protections they may expect for their health information when it is incorporated into a registry.

In the absence of a certificate of confidentiality, a valid subpoena or court order for registry data will usually compel disclosure of the data unless State law specifically protects the confidentiality of data. For example, Louisiana's laws specifically protect the collection of information related to tobacco use

from subpoena.[98] On the other hand, a subpoena or court order may supersede State law confidentiality protections. These legal instruments can be challenged in the court having jurisdiction for the underlying legal proceeding. In some circumstances, research institutions may be willing to pursue such a challenge. The remote yet definite possibility of this sort of disclosure should be clearly communicated to research subjects as a limitation on confidentiality protections, both during the consent process and in an authorization for use or disclosure of patient information.

State law may assure the confidentiality of certain quality I/A activities performed by health care providers as peer review activities.[99] When State law protects the confidentiality of peer review activities, generally, it is implementing public policy that encourages internal activities and initiatives by health care providers to improve health care services by reducing the risks of medical errors and systematic failures. Protection by peer review statutes may limit the use of data generated by quality I/A activities for any other purposes.

3.3.5 Waivers and Alterations of Authorization and Consent

As mentioned above, the Privacy Rule provides for the ability of privacy boards and IRBs to sometimes waive or alter authorizations by individual patients for the disclosure or use of health information for research purposes. (See Case Example 13.) In addition, the Common Rule authorizes IRBs to waive or alter the consent process. It is important for registry developers to keep distinct the terms "consent" and "authorization," as they are not interchangeable with respect to the Privacy Rule and Common Rule. As described above, authorization is the term used to describe individuals' written agreement to the use or disclosure of their health information under the Privacy Rule, while consent is the term used to describe research subjects' agreement to participate in research, as required by the Common Rule. There are separate requirements for each of these permissions.

The Privacy Rule and the Common Rule each specify the criteria under which waivers or alterations of authorization and the consent process are permitted.[100] There are potential risks to patients participating in the registry resulting from these waivers of permission. A waiver of authorization potentially imposes the risk of a loss of confidentiality and consequent invasion of privacy. A waiver of consent potentially imposes risks of harm from the loss of self-determination, dignity, and privacy expected under the ethical principles of respect for persons and beneficence. Acknowledging these potential risks, regulatory criteria for waiver and alterations require an IRB or privacy board to determine that risks are minimal, in addition to other criteria. This determination is a necessary condition for approval of an investigator's request for a waiver or alteration of these permissions.

The following discussion refers only to waivers; registry developers should note that privacy boards and IRBs may approve alterations to authorizations or the consent process, provided a requested alteration satisfies all of the criteria required for a waiver by the Privacy Rule or Common Rule. Alterations are generally preferable to waivers in an ethical analysis based on the principle of respect for persons, because they acknowledge the importance of self-determination. In requesting alterations to an authorization or to the consent process, registry developers should be prepared to justify each proposed change or elimination of required elements (such as description of alternative procedures, courses of treatment, or benefits). Plausible justifications include a registry to which a specific element does not apply or a registry in which one element contradicts other required information in the authorization or consent documentation. The justifications for alterations should relate as specifically and directly as possible to the regulatory criteria for IRB or privacy board approval of waivers and alterations.

The Privacy Rule permits a covered entity to obtain the approval of an IRB or privacy board for a waiver of authorization if the following criteria are met: (1) the use or disclosure involves no more than minimal risk to the privacy of individuals;

(2) the research cannot practicably be conducted without the waiver; and (3) the research cannot be practicably conducted without access to, and use of, the health information. The determination of minimal risk to privacy includes several elements: an adequate plan to protect identifiers from improper use or disclosure; an adequate plan to destroy identifiers at the earliest opportunity, unless a health or research justification exists to retain them; and adequate written assurances that the health information will not be reused or disclosed to others, except as required by law, as necessary for oversight of the research, or as permitted by the Privacy Rule for other research.[101] The registry developer should provide detailed documentation of the IRB or privacy board's decision to the health care provider or insurance plan (covered entity) that is the source of the health information for registry data.[43] The documentation should clearly communicate that each of the criteria for a waiver required by the Privacy Rule has been satisfied.[102] The privacy board or IRB documentation should also provide a description of the health information it determined to be necessary to the conduct of the research and the procedure it used to approve the waiver.[103] A health care provider or insurance plan might insist on legal review of this documentation before disclosing any health information.

The criteria for a waiver of consent in the Common Rule are similar to those for a waiver of authorization under the Privacy Rule. An IRB should determine that: (1) the research involves no more than minimal risk to subjects; (2) the waiver will not adversely affect the rights and welfare of subjects; (3) the research cannot practicably be carried out without a waiver; and (4) whenever appropriate, subjects will be provided with additional information after participation.[104] The criterion for additional information can be satisfied at least in part by public disclosure of the purposes, procedures, and operations of a registry, as discussed in Section 4.1.

Some IRBs produce guidance about what constitutes "not practicable" justifications and the circumstances in which justifications are applicable. For population-based research projects, registry developers may also present the scientific justification of avoiding selection bias. A waiver permits the registry to include the health information of all patients who are eligible. An IRB may also agree to consider requests for a limited waiver of consent that applies only to those individuals who decline use of their health information in a registry project. This limited waiver of consent most often permits the collection of de-identified and specified information sufficient to characterize this particular population.

An important difference between the Common Rule and FDA regulations for the protection of human subjects involves consent to research participation. The FDA regulations require consent, except for emergency treatment or research, and do not permit the waiver or alteration of informed consent.[105] If registry data are intended to support the labeling of an FDA-regulated product, a registry developer should plan to obtain the documented, legally effective, voluntary, and informed consent of each individual whose health information is included in the registry.

The Common Rule also permits an IRB to waive documentation of the consent process under two different sets of regulatory criteria. The first set of conditions for approval of this limited waiver requires that the only record linking an individual subject to the research is the consent document, and that the principal risk to subjects is the potential harm from a breach of confidentiality. Each subject individually determines whether his or her consent should be documented.[106] Alternatively, an IRB can waive documentation of consent if the research involves no more than minimal risk of harm to subjects and entails no procedures for which written consent is normally obtained outside of a research context.[107] For either set of regulatory criteria, the IRB may require the investigator to provide subjects with written information about the research activities in which they participate.[108] The written information may be as simple as a statement of research purposes and activities, or it may be more elaborate, such as a Web site for regularly updated information describing the progress of the research project.

The Privacy Rule creates a legal right for patients to receive, upon request, an accounting of certain disclosures of their health information that are made by health care providers, insurance plans, and their business associates.[109] The accounting must include disclosures that occur with a waiver of authorization approved by a privacy board or IRB. The Privacy Rule specifies the information that an accounting should contain[110] and requires it to cover a six-year period or any requested shorter period of time.[111] If multiple disclosures are made to the same recipient for a single purpose, including a research purpose, a summary of these disclosures may be made. In addition, because most waivers of authorization cover records of many individuals, and thus an individualized accounting in such circumstances may be burdensome or impossible, the Privacy Rule provides that if the covered entity has disclosed the records of 50 or more individuals for a particular research purpose, the covered entity may provide to the requestor a more general accounting, which lists the research protocols for which the requestor's information may have been disclosed, among other items.[112]

3.3.6 Patient Safety Organizations

The final rule (the "Patient Safety Rule") implementing the Patient Safety and Quality Improvement Act of 2005 (PSQIA) became effective on January 19, 2009.[113] The PSQIA statute was enacted in response to a 1999 report by the Institute of Medicine that identified medical errors as a leading cause of hospital deaths in the United States, with many such errors being preventable.[114] In accordance with the statute, the Patient Safety Rule allows health care providers to voluntarily report patient safety data, known as patient safety work product (PSWP), to independent patient safety organizations (PSOs). In general, PSWP falls into three general categories: (1) information collected or developed by a provider for reporting to a PSO and actually reported; (2) information developed by the PSO itself as part of patient safety activities; and (3) information that identifies or constitutes the deliberations or analysis of, or identifies the fact of reporting to, a patient safety evaluation system.[115] The Patient Safety Rule broadly defines PSWP to include any data, reports, records, memoranda, analyses, and statements that can improve patient safety, health care quality, or health care outcomes, provided that all such data must be developed for the purpose of reporting it to a PSO. Certain categories of information are expressly excluded from being PSWP. These include "a patient's medical record, billing and discharge information, or any other original patient or provider information...[and] information that is collected, maintained, or developed separately, or exists separately, from a patient safety evaluation system."[116]

Once PSWP is received by a PSO, it may be aggregated and analyzed by the PSO to assist a provider in determining, among other things, certain quality benchmarks and underlying causes of patient risks. Under the PSQIA, PSWP is considered privileged and confidential, and it may not be disclosed unless certain requirements are met. In addition, the Privacy Rule's limitations on uses and disclosures may apply where PSWP includes PHI. Civil money penalties may be imposed for violations.[117]

The Patient Safety Rule permits PSOs to disclose PSWP—that is, they may release, transfer, provide access to, or otherwise divulge PSWP to another person—in certain circumstances, including where disclosures are authorized by the identified health care providers, limited to nonidentifiable PSWP, or made to FDA, among other specified circumstances.[118] With respect to disclosure of PSWP for purposes of research, the regulations provide a narrow exception. The Rule allows for disclosure of identifiable PSWP to entities carrying out "research, evaluation, or demonstration projects authorized, funded, certified, or otherwise sanctioned by rule or other means by the Secretary [of Health and Human Services], for the purpose of conducting research."[119] Notably, the disclosure of PSWP for general research activities is not permitted under the PSQIA statute or the Patient Safety Rule.

An organization desiring to become a PSO must complete and submit a certification form to the Agency for Healthcare Research and Quality to become "listed" as a PSO.[120] A registry may choose to become listed as a PSO; however, the

registry should consider whether the obligations imposed on it in its capacities as a PSO would limit or otherwise restrict its attainment of its original objectives and whether it can fully meet the regulatory requirements that apply to a PSO.[121] In evaluating whether or not to be listed as a PSO, the registry developer should carefully review the registry's organizational structure and data collection processes to help ensure that there is a clear distinction between the collection of registry-related data and PSWP. For example, certain registries may publish certain information and results related to the data collected in the registry. As described above, if that registry is a PSO, then it must ensure that publication of such data does not constitute unauthorized disclosure for purposes of the PSQIA (as well as HIPAA, if applicable).

Finally, it is important for registry developers to know that, instead of becoming a PSO itself, a registry may elect to form a separate division or legal organization that it controls and segregate PSWP into that component. These types of PSOs are referred to as "Component PSOs." By keeping registry data and PSWP separate, the registry can reduce the possibility of an impermissible disclosure of PSWP.

3.4 Developments Affecting the HIPAA Privacy Rule

3.4.1 The Institute of Medicine Report

On February 4, 2009, the Institute of Medicine (IOM) published a report that examined how research was being conducted within the framework of the Privacy Rule. The IOM Report presented findings and recommendations of an IOM Committee tasked with assessing the impact of the HIPAA Privacy Rule on health research. This group had proposed recommendations to ensure that important health research might be conducted while maintaining or strengthening privacy protections for research subjects' health information.[122] The IOM Report stated that certain Privacy Rule requirements were difficult to reconcile with other regulations governing the conduct of research, including the Common Rule and the FDA regulations, and it noted a number of inconsistencies among applicable regulations related to the de-identification of data and the ability to obtain informed consent for future research studies, among other differences.

Citing more uniform regulations in other countries, the IOM Report affirmed that "a new direction is needed, with a more uniform approach to patient protections, including privacy, in health research."[123] As its primary recommendation, the IOM Committee held that Congress should authorize HHS and other Federal agencies to develop a new approach to protecting privacy that would apply uniformly to all health research and to exempt health research from the Privacy Rule when this new approach was implemented. Until such an overhaul could be accomplished, the IOM Committee called upon HHS to revise the Privacy Rule and associated guidance to address certain issues. HHS addressed some of these issues in the January 25, 2013, modifications to the Privacy Rule, such as by allowing HIPAA authorizations to encompass future research and removing prohibitions on combining certain HIPAA authorizations for multiple research studies, thereby harmonizing these HIPAA Privacy Rule requirements with the Common Rule. Nevertheless, registry operators should be aware that additional clarifications or modifications to the Privacy Rule as it relates to research activities may continue to be made in the future.

3.4.2 The Genetic Information Nondiscrimination Act of 2008

The Genetic Information Nondiscrimination Act of 2008 (GINA) was signed into law on May 21, 2008. In general, GINA prohibits discrimination in health insurance coverage (Title I) and employment (Title II) based on genetic information. GINA defines genetic information as, with respect to any individual, information about the individual's genetic tests, the genetic tests of the individual's family members, and the manifestation of a disease or disorder in the individual's family members (e.g., family health history). Title I of GINA took effect for most health insurance plans on May 22, 2009, and Title II became effective for employers on November 21, 2009. GINA also specifies that the definition of genetic information includes the genetic information of a fetus carried by a pregnant woman and an embryo legally held by an

individual or family member utilizing an assisted reproductive technology. Pursuant to GINA, health insurers are prohibited from using the genetic information of individuals for underwriting purposes (e.g., determining health insurance eligibility and coverage, or premium setting), and employers are prohibited from using genetic information in making employment-related decisions.

In addition to its nondiscrimination requirements, GINA also required related amendments to the Privacy Rule to clarify that genetic information is health information for purposes of the Privacy Rule, and to prohibit certain health plans from using or disclosing genetic information for underwriting purposes.[124]

3.4.3 The HITECH Act

The American Recovery and Reinvestment Act of 2009 (ARRA) was signed into law on February 17, 2009. Funds appropriated as a result of passage of ARRA are supporting new registries developed to study comparative effectiveness of treatments and protocols. It should be noted that there are no specific exceptions to regulatory or ethical requirements for such comparative effectiveness registries. Title XIII of division A and Title IV of division B of ARRA, the Health Information Technology for Economic and Clinical Health Act (HITECH Act) significantly modified the obligations of HIPAA covered entities and the business associates who perform certain services on behalf of covered entities.

Perhaps most significantly, the HITECH Act extends to business associates direct liability for compliance with many of the key privacy and security obligations contained in the HIPAA Rules, where before business associates were required to protect (and liable for failing to protect) the information to which they had access only through their business associate contracts with covered entities. Specifically, the HITECH Act imposed direct liability on business associates for compliance with the administrative, physical, and technical safeguard requirements for electronic protected health information under the HIPAA Security Rule, as well as for compliance with the use and disclosure provisions of the Privacy Rule and the business associate agreement. While many business associate agreements previously contained general safeguarding requirements (e.g., requiring the business associate to maintain appropriate technical safeguards), these agreements often had not imposed specific security requirements (e.g., a requirement that the business associate implement procedures to terminate an electronic session after a predetermined time of inactivity). The HITECH provisions now subject business associates to civil and criminal penalties once reserved only for covered entities under the HIPAA Rules. The HITECH Act obligations imposed on business associates were finalized through HHS rulemaking on January 25, 2013, with compliance required by September 23, 2013.[31]

The HITECH Act also created new breach notification requirements for covered entities and business associates. Following a breach of unsecured protected health information, covered entities must notify the affected individuals, HHS, and in some cases, the media. Notification must be provided without unreasonable delay but in no case later than 60 calendar days after the breach is discovered (except in cases of reports to HHS of breaches affecting less than 500 individuals, in which case, notification to HHS is required not later than 60 days after the end of the calendar year in which the breach was discovered). Depending on the circumstances, individual notification may include both direct, written notification to affected individuals via first-class mail or email, as well as substitute notice via conspicuous posting on the entity's Web site or in major print or broadcast media.

If a business associate experiences a breach of any unsecured protected health information it maintains, the business associate must provide notification to the applicable covered entity without unreasonable delay, and in no case later than 60 calendar days after the breach is discovered, so that the covered entity can provide the notifications described above with respect to the breach or delegate that responsibility to the business associate. Any notification by a business associate must include the identification of any individual(s) whose information was accessed,

acquired, or disclosed during the breach. The Department issued an interim final rule to implement the breach notification requirements on August 24, 2009, which became effective on September 23, 2009. On January 25, 2013, the Department published modifications to and made permanent these breach notification requirements.[125]

3.4.4 Summary of Regulatory Requirements

The use and disclosure of health information by health care providers and insurance plans for research purposes, including for registries, *are assumed* by the authors of this chapter to be subject to regulation under the Privacy Rule and *may* be subject to the Common Rule.

In general, the Privacy Rule permits a covered entity (or business associate on its behalf) to use or disclose patient information for registry purposes, subject to specific conditions, where the use or disclosure is: (1) for a registry supporting certain public health activities, including registries developed in connection with FDA-regulated products; (2) for a registry supporting the health care operations of a health care provider or insurance plan (covered entities), such as for quality I/A; (3) for a registry created by health oversight authorities for health system oversight activities authorized by law; (4) limited to de-identified health information; (5) limited to a "limited data set" of patient information that lacks specified direct identifiers, and a data use agreement is in place with the recipient; (6) pursuant to patient authorizations; or (7) consistent with a waiver or alteration of authorization by an IRB or privacy board.

The Common Rule will apply to the creation and use of registry data if (1) the registry is funded by HHS or the organization responsible for the registry has an FWA that encompasses the registry project, regardless of funding, *and* (2) the creation of the registry and subsequent research use of the registry data constitute nonexempt human subject research as defined by the Common Rule. As interpreted by OHRP, human subject research includes registry activities which have a research purpose, which may be in addition to the main purpose of the registry. Registry developers are strongly encouraged to consult the IRB, not only about the applicability of the Common Rule, but also about the selection of data elements, the content of the consent process or the regulatory criteria for waiver, and any anticipated future research involving identifiable registry data.

State laws regulate public health activities and may also apply in various ways to the research use of health information. NIH can issue certificates of confidentiality to particular research projects for the protection of identifiable personal information from most legally compelled disclosures. Federal law provides specific privacy protections to the health information of patients in substance abuse programs that receive Federal funding. The institutional policies of health care providers and insurance plans may also affect the use and disclosure of the health information of their patient or insured populations.

Legal requirements applying to use or disclosure of health information for research are evolving and can significantly influence the planning decisions of registry developers and investigators. It is prudent to obtain early and frequent consultation, as necessary, with institutional privacy officers, privacy board or IRB staff and members, information system representatives of health care providers and insurance plans, plus technology transfer representatives and legal counsel.

4. Registry Transparency, Oversight, and Data Ownership

4.1 Registry Transparency

Efforts to make registry operations transparent (i.e., to make information about registry operations public and readily accessible to anyone who is interested) are desirable. Such efforts may be crucial to realizing the potential benefits of research using health information. Registry transparency can also educate about scientific processes. Transparency contributes to public and professional confidence in the scientific integrity and validity of registry processes, and therefore in the conclusions reached as a result of registry activities. Public information about registry

operations may also increase the scientific utility of registry data by promoting inquiries from scientists with interests to which registry data may apply.

Registry developers can promote transparency by making the registry's scientific objectives, governance, eligibility criteria, sampling and recruitment strategies, general operating protocol, and sources of data available to anyone who is interested. Proprietary interests of funding agencies, contractual obligations, and licensing terms for the use of patient or claims information may limit, to some extent, the information available to the public about the registry. It is important to stress that, while transparency and access to information are to be encouraged, the intent is not to discourage or criticize investments in patient registries that produce proprietary information. Neither the funding source nor the generation of proprietary information from a registry determines whether a registry adheres to the good practices described in this handbook. Funding agencies, health care providers, and insurance plans do, however, have an important stake in maintaining public confidence in how health information is managed. The extent of registry transparency should be prospectively negotiated with these entities.

Creating a Web site of information about registry objectives and operations is one method of achieving transparency; ideally, registry information should be available in various media. An IRB may require registry transparency as a condition of approval to satisfy one of the regulatory criteria for granting a waiver of consent, which is to provide "additional pertinent information after participation."[126] For those interested, a useful example of registry transparency can currently be found on an international transplant registry Web site.[127]

4.2 Registry Oversight

Registry governance must reflect the nature and extent of registry operations. As described in Chapter 2, governing structures can vary widely, from one in which the registry developer is the sole decisionmaker to a system of governance by committee(s) comprised of representatives of all stakeholders in the registry, including investigators, the funding agency, patients, clinicians, biostatisticians, information technology specialists, and government agencies.

Registry developers should also consider appointing an independent advisory board to provide oversight of registry operations. An advisory board can assist registry operations in two important ways: (1) providing guidance for the technical aspects of the registry operations and (2) establishing the scientific independence of the registry. The latter function can be valuable when controversies arise, especially those related to patient safety and treatment, or resulting from actions by a regulatory agency. Advisory boards collectively should have relevant technical expertise, but should also include representatives of other registry stakeholders, including patients. Advisory board actions should be limited to making recommendations to the ultimate decisionmaker, whether an executive committee or the registry developer.

Registry developers may also appoint other types of oversight committees to resolve specific recurring problems, such as verifying diagnoses of patient conditions or adjudicating data inconsistencies.

4.3 Data Ownership

4.3.1 Health Information Ownership in General

The Privacy Rule was not intended to affect existing laws governing the ownership of health records.[128] However, multiple entities are often in a position to assert ownership claims to health information in various forms. For example, certain States have enacted laws that assign ownership of health records. At the current time, such claims of ownership are plausible, but none are known to be legally tested or recognized, with the exception of copyright claims. Entities that could claim ownership include health care providers and insurance plans, funding agencies for registry projects, research institutions, and government agencies. Notably, State law requires health care providers to maintain documentation of the

services they provide. This documentation is the medical-legal record compiled on each patient who receives health care services from an individual or institutional provider. Individuals, including patients (who may have a potential liberty interest in maintaining control of its use), registry developers, and investigators, may also assert ownership claims to health information. The basis for these claims is control of the tangible expression of and access to the health information.

There is no legal basis for assertions of ownership of facts or ideas; in fact, established public policy supports the free exchange of ideas and wide dissemination of facts as fundamental to innovation and social progress.[129] However, as a tangible expression of health information moves from its creation to various derived forms under the control of successive entities, rights of ownership may be transferred (assigned), shared, or maintained, with use of the information under a license (i.e., a limited transfer of rights for use under specific terms and conditions). Currently, in each of these transactions, the rights of ownership are negotiated on a case-by-case basis and formalized in written private agreements. The funding agency for a registry may also assert claims to ownership as a matter of contract law in their sponsorship agreements with research organizations.

Many health care providers are installing systems for electronic health records at great expense. Many also are contemplating an assertion of ownership in their health records, which may include ownership of copyright. The claim to ownership by health care providers may be an overture to commercialization of their health care information in aggregate form.[130] Public knowledge of and response to such assertions of ownership are uncertain at this time. A licensing program for the use of health information may enable health care providers to recoup some of their investment costs of electronic health records including the expenses associated with the technicians engaged to maintain them. In the near future, research use of health information for a registry may require licensing, in addition to the terms and conditions in data use agreements and, if necessary, in business associate agreements required by the Privacy Rule . Subsequent research use of the registry data will likely be based on the terms of the original license.

Among the changes ARRA has made in the regulation of health care information is a prohibition on its sale, subject to certain exceptions, including a limited one for the disclosure of health information for research purposes. This exception permits covered entities to recover a reasonable cost-based fee to cover the cost for the preparation and transmittal of the health information for the research purpose.[131]

For academic institutions, publication rights are an important component of intellectual property rights in data. Formal institutional policies may address publication rights resulting from faculty educational and research activities. Moreover, the social utility and benefit of any registry is evaluated on the basis of its publicly known findings and any conclusions based on them. The authors strongly encourage registry developers to maximize public communication of registry findings through the customary channels of scientific conferences and peer-reviewed journals. The goal of public communication for scientific findings and conclusions applies equally to registries operated outside of academic institutions (i.e., directly by industry or professional societies). For further discussion of developing data access and publication policies for registries, see Chapter 2.

The concept of ownership does not fit comfortably in the context of health information, because it largely fails to acknowledge individual patient privacy interests in health information. An inescapable personal nexus exists between individuals and information about their health. A recent failure that illustrates this relationship, with regard to patient interests in residual tissue from clinical procedures, resulted in widely publicized litigation to determine who owned the residual tissue and how it could be used for future research.[132] The legal concept of custody may be a useful alternative to that of ownership. Custodians have legal rights and responsibilities like those that a guardian has for a ward or parents have for their

children. Custody also has a protective function, consistent with public expectations of confidentiality practices that preserve the privacy and dignity of individual patients. Custody and its associated legal rights and responsibilities are transferable from one custodian to another. The concept of custody can support health care provider investments in information systems and the licensed use of health information for multiple, socially beneficial purposes without denying patient interests in their health information.

The sharing of registry data subsequent to their collection presents special ethical challenges and legal issues.[133] The criteria used to determine the conditions for shared use include applicable Federal or State law as well as the regulatory requirements in place when the health information was originally obtained. These legal and regulatory requirements, as well as processing and licensing fees, claims of property rights, and concerns about legal liability, are likely to result in formal written agreements for each use of registry data. Moreover, to educate patients and establish the scientific independence of the registry, registry developers should make known the criteria under which data is used.

Currently, there are no widely accepted social or legal standards that govern property rights in health information, with the possible exception of copyright, which is discussed below. At the time of this writing, health information sources and other users privately reach agreement to manage access and control. The Privacy Rule regulates the use and disclosure of health information by covered entities (including health plans and most health care providers), plus business associates working on behalf of covered entities, without regard to ownership interests over the data; the Privacy Rule does not affect existing laws, if any, regarding property rights in health information.[134]

4.3.2 Copyright Protection for Health Information Registries

In terms of copyright theory, a health information registry is likely to satisfy the statutory definition of a compilation[135] and reflect independent creativity by its developer.[136] Thus, copyright law may provide certain protections for a health information registry existing in any medium, including electronic digital media. The "facts" compiled in a health information registry, however, do not correlate closely to other compilations protected by copyright, such as telephone books or even genetic databases.[137] Instead, registry data constitute legally protected, confidential information about individual patients to which independent and varied legal protections apply. Copyright protections may marginally enhance, but do not diminish, other legal restrictions on access to and use of health information and registry data. For more information on copyright law, see Appendix B.

5. Conclusions

Ethical considerations arise in many of the essential aspects of planning and operating a registry. These considerations can affect the scientific, logistical, and regulatory components of registry development, as well as claims of property rights in health information. The guiding ethical principles for these considerations are *beneficence*, *justice*, and *respect for persons and avoidance of harm*.

At the most fundamental level, investigations that involve human subjects and that are not capable of achieving their scientific purpose are unethical. The risk-benefit ratio of such studies is unacceptable in an analysis based on the principle of *beneficence*, which obligates investigators to avoid harming subjects, as well as maximize the benefits and minimize the harms of research projects. Ethical scientific design must be robust, must be based on an important question, and must ensure sufficient statistical power, precise eligibility criteria, appropriately selected data elements, and adequately documented operating procedures and methodologies.

In addition, an ethical obligation to minimize harms requires planning for and establishing adequate protections to ensure the confidentiality of the health information disclosed to a registry. Such planning should include developing policies and procedures for the appropriate use and disclosure of registry data, and implementing physical, technical, and administrative safeguards

to limit access to and use of registry data accordingly. Reducing the potential harms associated with the use of health information in a registry is particularly important, because generally no directly offsetting benefit from participation in a registry accrues to individuals whose health information is used in the registry. According to an analysis applying the principle of *justice*, research activities that produce a significant imbalance of potential risks and benefits to participating individuals are unethical.

Protection of the confidentiality of the health information used to populate a registry reflects the ethical principle of *respect for persons and avoidance of harm*. Health information intimately engages the privacy and dignity of patients. Registry developers should acknowledge public expectations of protection for patient privacy and dignity with clear and consistent communications to patients about protections in place to prevent inappropriate access to and use of registry data.

The regulatory requirements of the Privacy Rule and Common Rule address past ethical concerns about research involving human subjects, as well as general social anxiety about potential loss of privacy associated with rapid advances in health information systems technology and communications and biomedical developments in human genetics. Compliance with these regulatory requirements not only is a cost of doing business for a registry project, but also demonstrates recognition of the ethical considerations accompanying use of health information for scientific purposes. Compliance efforts by registry developers also acknowledge the important public relations and liability concerns of health care providers and insurance plans, public health agencies, health oversight agencies, and research organizations. Regulatory compliance contributes to, and generally supports, the credibility of scientific research activities and research organizations, as well as that of particular projects.

These and other Federal and State privacy laws may affect registry development, especially registries created for public health purposes. Such laws express an explicit, legislatively determined balance of individual patient interests in health information against the potential social benefits from various uses of that information, including in research. Consultation with legal counsel is strongly recommended to determine the possible effect of these laws on a particular registry project.

Additional ethical considerations also affect the operational aspects of registries, including governance, transparency, and data ownership. Registry governance, discussed in Chapter 2, should reflect both appropriate expertise and representation of stakeholders, including patients. An independent advisory committee can provide useful guidance to registry developers and managers, especially on controversial issues. Transparency involves making information about registry governance and operations publicly available. Registry transparency improves the credibility of the scientific endeavors of a registry, the use of health information for scientific purposes, and the results based on analyses of registry data. In short, registry transparency promotes public trust.

Claims of "ownership" of health information and registries are plausible, but have not yet been legally tested. In addition, how the public would respond to such claims is uncertain, particularly as many such claims do not seem to acknowledge patient interests in health information. Nonetheless, in theory, copyright protections for compilations may be applied to the patient information held by health care providers and insurance plans, as well as to registries. In general, property rights related to health information are likely to be negotiated privately under the terms and conditions of formal agreements between registry developers, funding agencies, and health care providers or insurance plans. As a practical matter, "ownership" implies operational control of registry data and publication rights.

In summary, careful attention to the ethical considerations associated with the design and operation of a registry, and fulfillment of the applicable legal requirements, are critical to the success of registry projects and to the realization of their social and scientific benefits.

6. Summary of Privacy Rule and Common Rule Requirements

Table 7-1 summarizes Privacy Rule and Common Rule requirements. The table generally assumes that the Privacy Rule applies to the data source—i.e., that the data source is a "covered entity" or their "business associate." The exception is Category 8, registry developers that use data not subject to the Privacy Rule.

Note that the information in the table is a simplified summary of material; there may be other laws and individual institutional policies that also apply. Each research project is unique. Therefore, this table is not intended to provide answers to specific questions that arise in the context of a given project. This table is no substitute for consultation with institutional officials and others about the regulatory requirements that apply to a particular registry project.

Table 7-1. Summary of Privacy Rule and Common Rule requirements

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
1A. Federal or State public health agency: Registry for *public health practice within agency's legal authority* **not** *involving research.*	No requirements.	The Privacy Rule permits use or disclosure to a public health authority for public health activities. The Common Rule is not applicable.	The Privacy Rule permits use or disclosure to a public health authority for public health activities. The Common Rule is not applicable.	Waivers are not applicable.
1B. Federal or State public health agency: Registry is an agency *research project*.	No requirements.	The Privacy Rule permits the use or disclosure of a limited data set, provided the data source and registry developer enter into a DUA. The Common Rule may apply.**	The Privacy Rule permits use or disclosure with patient authorization or IRB or privacy board waiver of authorization. If the Common Rule applies,** IRB review and documented consent are required, unless an IRB grants a waiver of documentation or waiver for the consent process.	With respect to the Privacy Rule, privacy board or IRB approval of a waiver of authorization depends on satisfaction of specific regulatory criteria. If the Common Rule applies,** IRB approval of a waiver of consent documentation or process depends on satisfaction of specific regulatory criteria.
2. Registry producing evidence in support of labeling for an *FDA-regulated product.*	No requirements.	The Privacy Rule permits use or disclosure to a person responsible for an FDA-regulated product. The Common Rule may apply.**	The Privacy Rule permits use or disclosure to a person responsible for an FDA-regulated product. FDA regulations, and Common Rule, if applicable,** require IRB review, a documented consent process, and protection of confidentiality of research data.	Waivers are not applicable with respect to the Privacy Rule. If the Common Rule applies,** IRB approval of a waiver of consent documentation or process depends on satisfaction of specific regulatory criteria.

Table 7-1. Summary of Privacy Rule and Common Rule requirements (continued)

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
3. Health oversight agency registry to perform *a health oversight activity* **not** *involving research*.	No requirements.	The Privacy Rule permits use or disclosure for health oversight activities authorized by law. The Common Rule is not applicable.	The Privacy Rule permits use or disclosure for health oversight activities authorized by law. Institutional policy may apply the Common Rule or require IRB review.	Waiver of authorization is not applicable under the Privacy Rule. If institutional policy applies the Common Rule, IRB approval of a waiver of consent documentation or process depends on satisfaction of specific regulatory criteria.
4. *Registry required by law*; Common Rule may apply if registry involves research.	No requirements.	The Privacy Rule permits use or disclosure required by other law. If the Common Rule applies,** it permits an IRB grant of exemption if the data is existing or publicly available, unless a re-identification code is used. The Common Rule may apply, however the study may qualify for exemption.	The Privacy Rule permits use or disclosure required by other law. The Common Rule may apply, however the study may qualify for exemption. Institutional policy may apply the Common Rule or require IRB review whether or not a research purpose is involved.	Waiver of authorization is not applicable under the Privacy Rule. If the Common Rule applies,** IRB approval of a waiver of consent documentation or process depends on satisfaction of specific regulatory criteria.
5. *Quality I/A registry not involving research*.	No requirements.	The Privacy Rule permits the use or disclosure of a limited data set for health care operations, provided the data source and registry developer enter into a data use agreement. The Common Rule is not applicable.	The Privacy Rule permits use or disclosure for the "health care operations" of the data source and, in certain circumstances, of another covered entity. The Common Rule is not applicable.	Waivers are not applicable.

Table 7-1. Summary of Privacy Rule and Common Rule requirements (continued)

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
6. Research registry residing in organization to which Common Rule applies.**	No requirements.	The Privacy Rule permits the use or disclosure of a limited data set for research, provided the data source and registry developer enter into a DUA. The Common Rule permits an IRB grant of exemption from review if the data is existing or publicly available, unless a re-identification code is used.	The Privacy Rule permits use or disclosure for research with individual patient authorization or an IRB or privacy board waiver of authorization. The Common Rule requires IRB review and documented consent unless the IRB grants a waiver of documentation of consent or a waiver for the consent process.	IRB or privacy board approval depends on satisfaction of specific regulatory criteria.
7. *Research registry* developed by organization that is not a health care provider or insurance plan and is not subject to the Common Rule, using health information obtained from a health care provider or insurance plan.	No requirements.	The Privacy Rule permits the disclosure of a limited data set, provided the data source and registry developer enter into a DUA.	The Privacy Rule permits use or disclosure for research with individual patient authorization or waiver of authorization.	Privacy board approval of a waiver of authorization depends on satisfaction of specific regulatory criteria.

Table 7-1. Summary of Privacy Rule and Common Rule requirements (continued)

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
8. Research registry developed by organization that is not a health care provider or insurance plan and is not subject to the Common Rule, using health information collected from entities not subject to the Privacy Rule.	No requirements.	No requirements.	No requirements.	Waivers are not applicable.

DUA = data use agreement; FDA = U.S. Food and Drug Administration; IRB = institutional review board.
Note: Reference to this table is not a substitute for consultation with appropriate institutional officials about the regulatory requirements that may apply to a particular registry project.
*Information lacks the data elements specified in the Privacy Rule standard for de-identification.
**The Common Rule likely applies if: (1) Federal funding is involved with the registry project, (2) the organization within which the registry will reside has agreed in its Federalwide Assurance (FWA) to apply the Common Rule to all research activities conducted in its facilities or by its employees. Note that institutional policy may also apply the Common Rule.

Case Example for Chapter 7

Case Example 13. Obtaining a waiver of informed consent

Description	The STS/ACC TVT Registry™ tracks patient safety and real-world outcomes for patients undergoing a transcatheter aortic valve replacement (TAVR) procedure for treatment of aortic stenosis. The registry collects data on patient demographics, procedure details, and facility and physician information to support analyses of patient outcomes and clinical practice patterns. The Centers for Medicare & Medicaid Services (CMS) approved the registry as meeting the requirements outlined in the Medicare National Coverage Decision on TAVR. The participant agreement for the registry permits the registry sponsors to conduct cardiovascular research using a limited data set. The registry sponsors have signed a Federalwide Assurance that requires that all research be conducted consistent with the Common Rule.
Sponsor	The Society of Thoracic Surgeons (STS) and the American College of Cardiology (ACC)
Year Started	2012
Year Ended	Ongoing
No. of Sites	247 hospitals
No. of Patients	9,051 patients

Challenge

Aortic stenosis is the most common valvular abnormality in the United States, and the prevalence of this condition is expected to increase as the U.S. population ages. Until recently, surgical aortic valve replacement has been the only effective treatment for adults with severe symptoms. The TAVR procedure is a new option for patients who are considered to be inoperable for conventional aortic valve replacement surgery, but evidence is lacking on long-term patient outcomes.

In 2012, CMS issued a Medicare National Coverage Decision for TAVR. Under the decision, CMS permits Medicare coverage for TAVR only when (1) the procedure is performed with a device approved by the U.S. Food and Drug Administration (FDA) consistent with labeled indications and any other FDA requirement; (2) the procedure is performed in facilities meeting certain requirements; and (3) when all patients undergoing the TAVR procedure are included in a national TAVR registry or participate in an approved clinical study. The national TAVR registry must consecutively enroll TAVR patients, accept all manufactured devices, follow patients for at least one year, and comply with all relevant regulations related to the protection of human research subjects. The National Coverage Decision specifically requires that the registry collect data on the following outcomes: stroke, all-cause mortality, transient ischemic attacks, major vascular events, acute kidney injury, repeat aortic valve procedures, and quality of life.

The development of a national registry to meet these requirements was challenging, particularly due to the need to collect at least one year of followup and quality of life data. The registry was expected to enroll hundreds of sites and thousands of patients, making it time consuming, administratively cumbersome, and expensive to obtain local institutional review board (IRB) approval for each site and informed consent for each patient.

Proposed Solution

The registry developers determined that the national TAVR registry was most likely to be successful if it collected data that was already routinely documented as part of the standard of

Case Example 13. Obtaining a waiver of informed consent (continued)

Proposed Solution (continued)

care and was able to obtain a waiver of informed consent from a central IRB. To obtain a waiver of informed consent, the registry must meet all of the following four criteria, as documented in 45 CFR 46.116(d):

1. The research involves no more than minimal risk to the subjects;
2. The waiver or alteration will not adversely affect the rights and welfare of the subjects;
3. The research could not practicably be carried out without the waiver or alteration; and,
4. Whenever appropriate, the subjects will be provided with additional pertinent information after participation.

When designing the data collection instruments for the registry, the developers worked closely with surgeons and interventional cardiologists to understand which data are already collected. The developers were able to limit the registry data collection effort to data already collected routinely, thereby allowing registry data to be abstracted from the medical record with no data collected solely for the purposes of the registry. In particular, the registry developers carefully considered how to collect followup data and quality of life data without requiring the collection of information solely for the purposes of the registry.

Based on discussions with surgeons and interventional cardiologists, the developers determined that patients are seen for followup care routinely at 30 days and 1 year following the procedure. Published guidelines have established the use of the Kansas City Cardiomyopathy Questionnaire (KCCQ) to assess quality of life as a standard of care for TAVR patients at these follow-up visits. The registry was designed to use the data collected at these followup visits, including the KCCQ, to meet the requirements for collecting long-term outcomes and quality of life information.

Results

The registry began collecting data in 2012 and has been approved by CMS as meeting the requirements of the Medicare National Coverage Decision. The ACC and STS, the institutions operating the registry, are considered the only entities engaged in research, while the participating sites are considered to be providing data only. The registry was approved only by the single IRB designated under the ACC/STS's Federalwide Assurance. Based on the registry protocol, the IRB determined that the ACC/STS are engaged in research on human subjects and granted a waiver of informed consent. The waiver of informed consent was awarded primarily because the participating sites are collecting and submitting information that is already documented in the medical record as part of the standard of care. As the registry operators, the ACC and STS submit data, including patient identifiers, to CMS as required by the National Coverage Decision. However, the ACC and STS only conduct research on a limited data set, and any research studies not covered by the protocol are submitted to the IRB for review.

Because the ACC and STS have IRB approval and a waiver of informed consent, and because the data are already collected as part of the standard of care, the individual sites participating in the registry do not necessarily need to go through an IRB prior to enrolling in the registry. Some individual sites elect to submit the registry to their local IRB, in many cases because they intend to use the data they collect for the registry in additional research studies. The local IRBs generally have reached the same conclusion as the central IRB. However, a local IRB may reach a different conclusion, perhaps due to differences in the data collection process at an individual site. For example, the data collection process at an individual site may provide an opportunity to consent the patient, in which case the IRB may not grant a waiver of informed consent. In these cases, the individual site will follow the advice of the local IRB.

Case Example 13. Obtaining a waiver of informed consent (continued)

Key Point

Protecting the subjects whose data will be used is of the utmost importance when developing a registry. When developing a registry, sponsors should consider the planned data collection effort in the context of requirements for IRB review and informed consent. This approach may help the registry identify a strategy that protects patients' privacy without overburdening the participating sites.

For More Information

STS/ACC TVT Registry. https://www.ncdr.com/TVT/Home/Default.aspx.

References for Chapter 7

1. See, for example, Section III, Article 8, of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
2. Based to a certain extent on 45 CFR 160.103: definition of health information; and 45 CFR 46.102(f): definition of human subject.
3. 45 CFR Part 46.
4. Part C of Title XI of the Social Security Act, 42 USC §§ 1320d to 1320d-9 , and section 264 of the Health Insurance Portability and Accountability Act of 1996, 42 USC §1320d-2 note ; 45 CFR Parts 160 and 164.
5. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. April 18, 1979. http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html. Accessed March 4, 2014.
6. Public Law 93-348 (1974), Title II.
7. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Noted to be under revision. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014. See especially sections entitled General Ethical Principles and Informed Consent..
8. Grant RW, Sugarman J. Ethics in human subjects research: do incentives matter? J Med Philos. 2004 Dec;29(6):717-38. PMID: 15590518.
9. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Note 6, at paragraphs 11 and 12. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014.
10. U.S. Department of Health and Human Services. Office of the Inspector General. Recruiting Human Subjects: Sample Guidelines for Practice. OEI-01-97-00196. June 2000. p. 5. http://oig.hhs.gov/oei/reports/oei-01-97-00196.pdf.
11. U.S. Department of Health and Human Services. Office of the Inspector General. Recruiting Human Subjects: Sample Guidelines for Practice. Appendix A. OEI-01-97-00196. June 2000. http://oig.hhs.gov/oei/reports/oei-01-97-00196.pdf.
12. Patient Protection and Affordable Care Act. Section 6002, Transparency Reports and Reporting of Physician Ownership or Investment Interests. Public Law 111-148. March 23, 2010. http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf.
13. Massachusetts regulation 105 CMR 970.000 implementing Massachusetts General Law, Chapter 111N, Pharmaceutical and Medical Device Manufacturer Conduct, as enacted under Chapter 305 of the Acts of 2008, An Act To Promote Cost Containment, Transparency and Efficiency in the Delivery of Quality Health Care. http://www.lawlib.state.ma.us/source/mass/cmr/cmrtext/105CMR970.pdf.
14. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Note 6, at paragraphs 18–21. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014.

15. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Note 6, at paragraph 26. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014.

16. See generally Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Note 6, at paragraph 43. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014. See also the Patient Protection and Affordable Care Act of 2010, Sec. 1101. http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf.

17. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Note 6, at paragraph 40. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014.

18. See, for example, U.S. Department of Health and Human Services regulations at 45 CFR Part 46; 21 CFR Parts 50 and 56 for research conducted in support of products regulated by the U.S. Food and Drug Administration (FDA); and Williams ED. Federal Protection for Human Research Subjects: An Analysis of the Common Rule and Its Interactions with FDA Regulations and the HIPAA Privacy Rule. Congressional Research Service, Library of Congress. Updated June 2, 2005. http://www.fas.org/sgp/crs/misc/RL32909.pdf.

19. Regulations identical to 45 CFR 46 Subpart A apply to research funded or conducted by a total of 17 Federal agencies, some of which may also require additional legal protections for human subjects.

20. U.S. Department of Health and Human Services. Office for Human Research Protections. Terms of the Model Federalwide Assurance (FWA) for the Protection of Human Subjects. http://www.hhs.gov/ohrp/assurances/assurances/filasurt.html. Accessed March 4, 2014.

21. U.S. Department of Health and Human Services. Office for Human Research Protections. Guidance on Research Involving Coded Private Information or Biological Specimens. Oct 16, 2008. http://www.hhs.gov/ohrp/policy/cdebiol.html.

22. 45 CFR Part 46, Subpart A.

23. See, for example, on the essential elements of a protocol: Epstein M, International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiology Practices (GPP). Pharmacoepidemiol Drug Safety. 2005 August;14(8):589–95. PMID: 15918159.

24. See 45 CFR 160.103.

25. 45 CFR 160.102, Applicability, and 160.103, definitions of covered entity, health care provider, health plan, health care clearinghouse, and transaction.

26. 45 CFR 160.103, definition of business associate.

27. 45 CFR 160.203.

28. Maryland Health General Statute § 4–303(b)(4).

29. 45 CFR 160.103 defines both "disclosure" and "use" for the purposes of the HIPAA Rules.

30. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; Other Modifications to the HIPAA Rules; Final Rule, 78 Fed. Reg. 5566 (January 25, 2013) (codified at 45 CFR pts 160 and 164).

31. HITECH Act §13404(a); 45 CFR 164.104(b); 78 Fed. Reg. at 5591.

32. 45 CFR 164.501.

33. 67 Fed Reg 53231, August 14, 2002.

34. 45 CFR 46.102(d).

35. National Institutes of Health. Health Services Research and the HIPAA Privacy Rule. NIH Publication Number 05-5308. May 2005. http://privacyruleandresearch.nih.gov/healthservicesprivacy.asp. See also National Institutes of Health. Research Repositories, Databases, and the HIPAA Privacy Rule. NIH Publication Number 04-5489, January 2004. http://privacyruleandresearch.nih.gov/research_repositories.asp.

36. 45 CFR 164.512(b).

37. Centers for Disease Control and Prevention. HIPAA Privacy Rule and Public Health: Guidance from CDC and the U.S. Department of Health and Human Services. MMWR. 2003:52. PMID: 12741579.

38. 45 CFR 164.512(a).
39. 45 CFR 164.508(a).
40. 45 CFR 164.514(e).
41. 45 CFR 164.514(a)–(c).
42. 45 CFR 164.528.
43. 45 CFR 164.512(i)(1)(i).
44. 21 CFR 65.111(a)(7).
45. Centers for Disease Control and Prevention. Guidelines for Defining Public Health Research and Public Health Non-Research. Revised October 4, 1999. http://www.cdc.gov/od/science/integrity/docs/defining-public-health-research-non-research-1999.pdf. Accessed March 4, 2014. Gostin LO. Public Health Law: Power, Duty, Restraint. Berkeley and Los Angeles, CA: University of California Press; New York: The Milbank Memorial Fund; 2000. pp. 126–127. See also Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Note 6, Introduction, noting that epidemiological practice and research may overlap. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014.
46. Bellin E, Dubler NN. The quality improvement-research divide and the need for external oversight. Am J Public Health. 2001;91(9): 1512–1517. PMID: 11527790. Lindenauer PK, Benjamin EM, et al. The role of the institutional review board in quality improvement: a survey of quality officers, institutional review board chairs, and journal editors. Am J Med. 2002;113(7): 575–9. PMID: 12459404. Lo B, Groman M. Oversight of quality improvement: focusing on benefits and risks. Arch Intern Med. 2003;163(12):1481–6. PMID: 12824099.
47. U.S. Department of Health and Human Services. Office for Human Research Protections. http://www.hhs.gov/ohrp. Accessed March 4, 2014.
48. National Institutes of Health. Health Services Research and the HIPAA Privacy Rule. NIH Publication Number 05-5308, pp. 2–3. See also 45 CFR 164.501 for the definition of health care operations. http://privacyruleandresearch.nih.gov/healthservicesprivacy.asp.
49. A copy of the HHS version of the "Common Rule," 45 CFR Part 46, subpart A, and additional subparts B, C, and D regarding vulnerable populations may be obtained on the Web site of the Office for Human Research Protection (OHRP) in the U.S. Department of Health and Human Services. http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html. Accessed March 4, 2014.
50. A copy of the "Privacy Rule," 45 CFR Parts 160 and 164, may be obtained on the Web site of the Office for Civil Rights (OCR) in the U.S. Department of Health and Human Services. http://www.hhs.gov/ocr/hipaa/finalreg.html. Accessed March 4, 2014.
51. The Common Rule as adopted by the U.S. Department of Health and Human Services contains special protections for certain defined "vulnerable" populations, i.e., women, human fetuses, neonates, prisoners, and children. See 45 CFR Part 46, Subparts B, C, D.
52. American College of Epidemiology: Policy Statement on Sharing Data from Epidemiologic Studies. May 2002. http://www.acepidemiology.org/sites/default/files/DataSharing.pdf. Accessed March 4, 2014.
53. National Cancer Institute. Surveillance Epidemiology and End Results. http://seer.cancer.gov. Accessed August 27, 2012.
54. Gostin LO, Lillienfeld DE, Stolley PD Foundations of Epidemiology (revised). Oxford University Press; 1994. p. 104. See also Gostin LO. Public Health Law: Power, Duty, Restraint. Berkeley and Los Angeles, CA: University of California Press; New York: The Milbank Memorial Fund; 2000. Note 43, at 114, Table 5.
55. 45 CFR 164.512(b)(1)(iii).
56. 45 CFR 164.512(d).
57. 45 CFR 164.512(d)(1).
58. 45 CFR 46.102(d) and 45 CFR 164.501, respectively.
59. National Institutes of Health. Health Services Research and the HIPAA Privacy Rule. NIH Publication No. 05-5308. May 2005. pp. 2–3. http://privacyruleandresearch.nih.gov/healthservicesprivacy.asp. Accessed March 13, 2014.

60. Centers for Disease Control and Prevention. Guidelines for Defining Public Health Research and Public Health Non-Research. October 4, 1999. pp. 126-7. http://www.cdc.gov/od/science/integrity/docs/defining-public-health-research-non-research-1999.pdf. Accessed March 4, 2014. Gostin LO. Public Health Law: Power, Duty, Restraint. Berkeley and Los Angeles, CA: University of California Press; New York: The Milbank Memorial Fund; 2000. Note 43. See also Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. Note 6, Introduction, noting that epidemiological practice and research may overlap. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed March 4, 2014. And Bellin E, Dubler NN. The quality improvement-research divide and the need for external oversight. Am J Public Health. 2001;91(9): 1512–1517. Note 44. PMID: 11527790.
61. Bellin E, Dubler NN. The quality improvement-research divide and the need for external oversight. Am J Public Health. 2001;91(9): 1512–1517. Note 44. PMID: 11527790. Lindenauer PK, Benjamin EM, et al. The role of the institutional review board in quality improvement: a survey of quality officers, institutional review board chairs, and journal editors. Am J Med. 2002;113(7):575–9. PMID: 12459404. Lo B, Groman M Oversight of quality improvement: focusing on benefits and risks. Arch Intern Med. 2003;163(12):1481–6. PMID: 12824099.
62. See 45 CFR 160.103 for the definition of individually identifiable health information and 45 CFR 164.514(a)–(c) and (e) on the de-identification of health information and limited data sets, respectively.
63. 45 CFR 46.102(f).
64. 45 CFR 46.111(a)(7).
65. 45 CFR 46.116(a)(5).
66. See 45 CFR 164.514(a)–(c) and (e) on the deidentification of health information and limited data sets, respectively.
67. 45 CFR 164.514(e)(2).
68. 45 CFR 164.514(b).
69. 45 CFR 164.514(b)(1).
70. 45 CFR 164.514(c).
71. 67 FR 53182, 53233, August 14, 2002.
72. 45 CFR 164.514(e)(4).
73. 45 CFR 164.514(e)(4)(ii)(A).
74. 45 CFR 164.514(e)(4)(ii)(C)(2) and (5), respectively.
75. 67 Fed Reg 53182, 53236, August 14, 2002.
76. 45 CFR 164.504(e).
77. 45 CFR 46.101(b)(4).
78. 45 CFR 164.508.
79. 45 CFR 164.508(c).
80. 45 CFR 164.502(a)(1).
81. 45 CFR 46.116.
82. 45 CFR 164.508(b)(3).
83. 45 CFR 164.512(i)(1)(i)(B).
84. 67 Fed Reg 53182, 53226, August 14, 2002.
85. National Institutes of Health. Institutional Review Boards and the HIPAA Privacy Rule. NIH Publication Number 03-5428. August 2003. pp. 15–16. http://privacyruleandresearch.nih.gov/irbandprivacyrule.asp. Accessed March 4, 2014.
86. 45 CFR 164.508(b)(3); 78 Fed. Reg. at 5612.
87. 45 CFR 164.508(c) and 164.508(c)(1)(iv).
88. 45 CFR 164.508(c)(2)(iii).
89. Family Educational Rights and Privacy Act (FERPA), 20 USC 1232g, 34 CFR Part 99.
90. National Institutes of Health. Office of Extramural Research. Grants & Funding. "Frequently Asked Questions, Certificates of Confidentiality. Definitions – What does identifying characteristic mean?" http://grants1.nih.gov/grants/policy/coc/faqs.htm#369. Accessed March 4, 2014.
91. Public Health Services Act Section 301(d), 42 USC 241(d) as amended. See also 42 CFR Part 2a about research activities on mental health, including the use and effect of alcohol and other psychoactive drugs.
92. U.S. Department of Health and Human Services. Office for Human Research Protection. Guidance on Certificates of Confidentiality. Background. February 25, 2003. http://www.hhs.gov/ohrp/policy/certconf.pdf. Accessed March 4, 2014.

93. National Institutes of Health. "NIH Announces Statement on Certificates of Confidentiality." Notice NOTOD-02-037. March 15, 2002. http://grants1.nih.gov/grants/guide/notice-files/NOT-OD-02-037.html. Accessed March 4, 2014.
94. Information about obtaining a certificate of confidentiality is available at the "Certificates of Confidentiality Kiosk" on the National Institutes of Health Web site. http://grants.nih.gov/grants/policy/coc/index.htm. Accessed March 4, 2014.
95. National Institutes of Health. Office of Extramural Research. Certificates of Confidentiality: Background Information. Febrary 14, 2006. http://grants.nih.gov/grants/policy/coc/background.htm. Accessed March 4, 2014.
96. 42 USCS 290dd-2 and 290ee-3; 42 CFR Part 2.
97. 42 CFR 2.52(a).
98. In re Philip Morris, 706 So. 2d 665 (La.App. 4 Cir. Jan. 28, 1998) (holding that raw data from research on tobacco use is protected under Louisiana statutes that govern the confidentiality of public health data).
99. See, for example, Wis. Stat. 146.38.
100. See 45 CFR 164.512(i) and 46.116(d), respectively.
101. 45 CFR 164.512(i)(2)(ii).
102. 45 CFR 164.512(i)(2).
103. 45 CFR 164.512(i)(2)(iii) and (iv).
104. 45 CFR 46.116(d).
105. 21 CFR 50.20 and 50.23.
106. 45 CFR 46.117(c)(1).
107. 45 CFR 46.117(c)(2).
108. 45 CFR 46.117(c).
109. 45 CFR 164.528(b)(1).
110. 45 CFR 164.528(b).
111. 45 CFR 164.528(a)(1).
112. 45 CFR 164.528(b)(3) and (4).
113. 42 CFR 3.
114. Kohn LT, Corrigan JM, Donaldson MS, eds. To Err is Human: Building a Safer Health System. Washington, D.C.: National Academies Press; 1999.
115. See 73 Fed. Reg. 70,739.
116. 42 C.F.R. § 3.20.
117. 42 CFR 3.402.
118. 42 CFR 3.206.
119. 42 CFR 3.206(b)(6).
120. 42 CFR 3.102.
121. 42 USC 299b–22(i). 45 CFR 164.501; 78 Fed. Reg. at 5592.
122. Nass SJ, Levit LA, Gostin LO, eds. Committee on Health Research and the Privacy of Health Information. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. National Academies Press; The HIPAA Privacy Rule; Institute of Medicine. http://www.nap.edu/catalog/12458.html. Accessed March 4, 2014.
123. Nass SJ, Levit LA, Gostin LO, eds. Committee on Health Research and the Privacy of Health Information. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. National Academies Press; The HIPAA Privacy Rule; Institute of Medicine. http://www.nap.edu/catalog/12458.html. Accessed March 4, 2014. pp. 26.
124. 45 CFR § 160.103, definition of health information; 78 Fed. Reg. at 5658-61.
125. 74 FR 42740 (August 24, 2009); 78 Fed. Reg. at 5638-58 (January 25, 2013); 45 CFR 164.400-164.414.
126. 45 CFR 46.116(d)(4).
127. Center for International Blood & Marrow Transplant Research. http://www.cibmtr.org/pages/index.aspx. Accessed March 4, 2014.
128. 67 F.R. 53213, August 14, 2002.
129. Joyce C, Patry W, Leaffer M, et al. Introduction: the landscape of copyright. In: Copyright Law. 3rd ed., New York and San Francisco: Matthew Bender & Co., Inc; 1994 (reprinted 1997):chapter 1, pp. 1–41.

130. American Medical Association House of Delegates, Connecticut Delegation. Guiding Principles, Collection and Warehousing of Electronic Medical Record Information. Resolution No. 802. September 16, 2005. http://www.ama-assn.org/meetings/public/interim05/802i05.pdf. Accessed March 4, 2014. Bailey S. Your data for sale? Boston Globe. March 24, 2006. http://www.boston.com/business/healthcare/articles/2006/03/24/your_data_for_sale/. Accessed March 4, 2014.

131. American Recovery and Reinvestment Act, section 13405(d)(2)(B); 45 CFR § 164.502(a)(5)(ii).

132. Washington University v. Catalona, Case No. 4:03CV1065SNL (E.D. Mo., filed Mar. 31, 2006). Washington University v. William J. Catalona, M.D., No. 06-2286 and No. 06-2301 (8th Cir. June 20, 2007).

133. American College of Epidemiology. Policy Statement on Sharing Data from Epidemiologic Studies. May 2002. http://www.acepidemiology.org/sites/default/files/DataSharing.pdf. Accessed March 4, 2014. National Institutes of Health. Final NIH Statement on Sharing Research Data. Notice NOT-OD-03-032. February 26, 2003. http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html. Accessed March 4, 2014.

134. 78 FR 5566 at 5606 (January 25, 2013).

135. 17 USC § 101.

136. Feist Publications, Inc. v. Rural Telephone Service, Co., Inc., 499 U.S. 340, 345, 348 (1991).

137. Id., 340 et seq; Harris RK, Rosenfield SS. Copyright Protection for Genetic Databases, 2005. Jurimetrics J. 45:225–250.

Chapter 8. Informed Consent for Registries

1. Introduction

This chapter identifies the best practices for obtaining informed consent and permission for registry participation. It builds on some of the general ethical and legal principles discussed in Chapter 7, focusing specifically on the application of the regulations governing human subjects research and the requirements of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

The purpose of this chapter is to provide an ethical framework for obtaining informed consent and permission for registry participation and to distinguish registries from clinical research protocols with respect to these issues. It is not designed to provide specific legal guidance, nor can it substitute for Institutional Review Board (IRB) review. Moreover, legal discussion is limited to U.S. law, and more specifically, to Federal as opposed to State statutes and regulations. Some States have guidelines governing the conduct of research involving human subjects or statutes addressing privacy, and an exploration of either area is beyond the scope of this chapter. Likewise, analysis of the relevant international standards and laws is left to others. Case examples 14, 15, and 16 provide descriptions of practical issues that registries have encountered in this area.

2. Registries, Research, and Other Activities

The purpose of this volume is to provide guidance for registries used to evaluate patient outcomes, such as efforts to describe the natural history of disease, determine clinical and/or cost effectiveness, assess safety or harm, and measure or improve quality of care. As a result, the focus of this chapter is on informed consent and authorization issues that arise in registries used for research. Some registries now used for research may have been developed initially for clinical purposes (e.g., a registry of names and contact information of patients using a particular treatment to facilitate notifications or recalls). In fact, registries are increasingly being used for research purposes even when initially developed for clinical purposes, and thus it is suggested that in all cases, consideration should be given to the informed consent issues, as well as HIPAA privacy requirements, discussed in this chapter. The HIPAA Privacy Rule governs the use and disclosure of most individually identifiable health information (called protected health information or "PHI") held by covered entities (health plans, health care clearinghouses, and most health care providers).

The Federal research regulations promulgated by the U.S. Department of Health and Human Services (HHS) focus on research involving human subjects. The U.S. Food and Drug Administration (FDA) regulations apply to "all clinical investigations" regulated by the FDA—defined as "any experiment that involves a test article and one or more human subjects".[1] The HHS regulations apply only to "human subjects research," where "research" is defined as a "systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge" and "human subject" is defined as a living person about whom the investigator obtains either data through intervention or interaction, or identifiable private information.[2] Thus, investigations that involve non-living individuals, or that do not collect data through intervention/interaction with the individual, or do not collect identifiable private information are not governed by the HHS research regulations. Despite the apparent limitations of the regulatory language, institutions may choose to apply the frameworks more broadly (sometimes under an "assurance," i.e., an agreement with HHS that the institution will apply the regulations to all research at the institution regardless of funding source). Even when the activity in question meets the HHS definition of involving human subjects, a series of exemptions may apply.[3]

2.1 Registry Research vs. Clinical Research

Some differences between registry research and clinical research are worth noting. In particular, the use of a control group in a registry setting is often substantively different from the concept of a control group in a clinical research setting. Registry controls may be pulled from a general population—in some cases a population that may not have interacted with health professionals or a health institution. Unlike clinical controls, who may be exposed to placebos (and thus need to consent) or exposed to a standard treatment (and thus will already be involved in the treatment system), registry controls may be identified from an unaffected population. This raises ethical questions about the initial contact with an individual who may have no link to the registry topic area and who may view the contact as an unwelcome intrusion or perhaps even an incorrect indication of problematic health status.[4] Furthermore, since a clinical research trial may involve double-blind procedures, "controls" may agree to participate because of the potential for direct therapeutic benefits or even the indirect therapeutic benefits that come from better attendant care. In other situations, controls may participate because they hope to help others suffering from their ailments (altruistic reasoning) or perhaps because they seek monetary compensation. In contrast, controls in a registry trial have no similar potential therapeutic (direct or indirect) or monetary benefits. While altruism may play a role in this context, its effects may be less than ideal. There is a great concern about the potential for selection bias in the creation of a control group for registry trials—as there is also significant concern about the effect of bias in clinical trials. Those who may agree to participate in a registry may be qualitatively different from those who do not agree, and this disparity can threaten the external validity of research findings. Concerns about selection bias will be heightened for diseases with a low prevalence in the general population, since there will be a greater possibility that the bias will affect the data. Developing consent requirements in such a way as to avoid selection bias will be extremely important in this setting.

Questions about adapting the regulatory requirements to research that does not fit the typical clinical model are not unusual. There are two other areas that have raised questions about the how the Federal regulations apply and that are particularly relevant to registry evaluations: public health activities and quality improvement/assurance (QI/QA).

2.2 Public Health Activities

The HIPAA Privacy Rule expressly permits the disclosure of Protected Health Information (PHI) to a public health authority that is authorized by law to collect or receive such information for the purpose of preventing or controlling disease, injury, or disability, including for activities related to disease, injury, or vital event reporting. Thus, a covered entity may disclose PHI, without individual authorization, for a registry maintained by a public health authority (or by an entity acting under a grant of authority from or contract with such public agency) for authorized public health purposes; such as, for example, immunization registries, State cancer registries, birth and death registries, and general disease reporting (although this last type of reporting is often anonymous). The HIPAA Privacy Rule also allows the disclosure of PHI to a person subject to the jurisdiction of FDA for FDA-regulated–product reporting.

Public health activities may not be considered "human subjects research" under HHS or FDA research regulations. However, differentiating between public health practice and public health research activities can be challenging. According to the Belmont Report, the document on which the Federal research regulations are based, if any aspect of an activity constitutes "research," then the entire activity should undergo regulatory review. The Office for Human Research Protections (OHRP) interpretation of the HHS regulations implies that if any part of the activity falls under the regulations the entire activity is covered.[5] By contrast, the Centers for Disease Control and Prevention (CDC) only consider an activity research if the primary intent is to contribute to or generate generalizable knowledge.[6] The CDC, however, does not provide official interpretations of the HHS research regulations, as the authority for this rests with

OHRP. Local IRB policies in this area vary; some focus on whether the primary intent of an activity is to gain generalizable knowledge, and others categorically exclude normal public health department activities.

To address confusion regarding what is considered a public health activity versus a research activity, the Council of State and Territorial Epidemiologists (CSTE) issued a report clarifying that public health practice activities are those for which: there is a specific or general legal authorization to conduct (e.g., State statutory cancer registries, or reports of newborn hearing screening to the State health department); the specific intent of the authority conducting the activity is to promote the health of, or prevent harm to, the individuals or communities involved (as opposed to research where the intent is to generate generalizable knowledge); and there are, in fact, health benefits to the individuals involved or to the target community.[7] Moreover, public health activities, unlike research, are not likely to involve experimental procedures, to have one (or more) individuals, such as a principal investigator (PI), responsible for the development and conduct of the activity, or to entail individual randomization for access to interventions.

Alternatively, a public health activity may fit the definition of human subjects research, but fall into one of the various HHS exemptions. For example, there are exemptions for research involving surveys, interviews, or observations of public behavior, provided certain requirements are met.[8] There is also an exemption for the collection or examination of existing data, if publicly available and information is recorded "in such a manner that subjects cannot be identified, directly or through identifiers."[9]

2.3 Quality Improvement/Quality Assurance Activities

As with certain public health activities, the HIPAA Privacy Rule provides an explicit permission to use or disclose PHI for "health care operations," which are defined as certain activities of a covered entity, including "conducting quality assessment and improvement activities…, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities."[10] Individual authorization for use or disclosure of PHI in this context is not required, but a covered entity can choose to obtain individual consent for such uses and disclosures.

The Federal research regulations do not have an explicit exemption for QI/QA activities. However, many of the efforts in this area will: (a) not meet the regulatory definitions of "research," (b) not involve "human subjects," (c) fall under a delineated exemption, or (d) not be supported by HHS, involve an FDA regulated product, or otherwise be covered by an assurance of compliance. Some local IRBs appear to consider all QI/QA activities outside the scope of the regulations, as they would public health activities.[11]

The application of the human subjects research regulations does not rest on whether or not a procedure is considered "standard" or part of the "standard of care;" rather, it rests on the purpose of the activity. Intent to publish the results of a QI/QA activity is not determinative of whether the human subjects regulations apply. Registries developed within an institution to implement a practice to improve the quality of patient care or to collect data regarding the implementation of such a practice are not considered "research" under the regulations. Nor are registries designed to collect provider performance data for clinical, practical, or administrative uses. Registries that involve existing data that are not individually identifiable may entail "research," but do not involve "human subjects" as defined by the HHS research regulations; therefore the HHS research regulations do not apply. However, a QI/QA project that involves an untested clinical intervention (whether or not part of the standard of care) for purposes of gathering scientific evidence of efficacy (i.e., a systematic investigation designed to contribute to generalizable knowledge) would be governed by the regulations, although a specific exemption may apply (e.g., if it is part of the evaluation of a public benefit program).[12] Even if the regulations apply, waivers or alterations to the consent process may be approved as noted below.

3. Current Challenges for Registries

3.1 Electronic Health Records

The development of large-scale data registries raises a variety of regulatory questions, and this is nowhere more evident than in the discussions about electronic health records (EHRs). These issues are explored in detail in Chapter 15. This chapter focuses only on the relevant consent issues. There are currently few, if any, efforts to obtain individual consent for the creation of an EHR (or, for that matter, the creation of any health record). Yet, these databases have enormous research potential. For example, Kaiser Permanente, a leader in the use of health information technology, created and maintains one of largest private-sector EHR systems, collecting health information from more than 8.7 million Kaiser members nationwide. Moreover, there are a number of efforts to develop (sometimes via State legislation) multi-payer claims registries to support comparative effectiveness research (CER). Various steps have been discussed to ensure the privacy and confidentiality of the individual health information gathered into these registries (e.g., the use of coded identifiers). Application of traditional consent models for the secondary use of these databanks for research may prove inefficient and may result in selection bias, impacting the usefulness of downstream analyses.

As the development of EHRs, claims registries, health information exchanges, and linkages between innumerable health databases moves forward, keeping records private becomes more difficult to manage. Personal health information may be accessed and shared in ways patients never imagined, often for the purpose of secondary analysis and often without patient consent. Although studies consistently indicate that Americans are generally supportive of EHRs and even secondary uses of the data, they want to be informed about how and to what extent their information will be shared and disclosed to others.[13-16]

Despite apparent public unease with a system of open access to EHRs, the Institute of Medicine (IOM) in 2009 released a statement that informed consent for research using EHRs should not be required, with the justification that obtaining permission from patients is too burdensome for researchers and should be eliminated entirely.[17] This proposal generated widespread concern that its implementation would undermine the trust that forms the basis of the patient-physician relationship and also more broadly increased concerns about patients' privacy and confidentiality protections. Given the strong arguments on both sides, establishing consensus on the topic has been slow.

In the effort to resolve the debate, additional work in this area should focus on striking the appropriate balance between providing patients with control over information and facilitating necessary research. Commentators have suggested a variety of different approaches, including recognition of public ownership of large electronic databases[18] or the creation of licensed data centers that would control access to information without individual consent.[19] It is not clear from the empirical evidence that patients want full consent protections in this context. One study, for example, found that patients were more likely to be comfortable with the research uses of their EHR information when they were asked about the use by a specific entity (e.g., universities, hospitals, or disease foundations) rather than when asked in the abstract, and that they fully supported public health uses of their data.[20] Public education about the scope of research uses may alleviate some patient concerns about the use of EHR data without consent.[21] Similarly, addressing underlying fears about unauthorized access to identifiable data or discriminatory uses of the information can also be helpful in increasing support for this type of research. Given the vast potential for using EHRs to conduct large-scale observational studies, development of an alternative to specific individual consent may be useful. On July 26, 2011, HHS and the Office of Science and Technology Policy (OSTP) published an advance notice of proposed rulemaking

(ANPRM) entitled, "Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators."[22] A number of changes have been proposed to enhance protections of research participants, while facilitating valuable research and reducing the burden of research investigators. Included in the proposal are suggestions that specifically address EHRs and large-scale electronic databanks.

3.2 Biobanks

The increasing availability of biobanks, which are defined as facilities that store biological material (e.g., serum, genomic material, pathology specimens) from humans, raises additional concerns. In addition to the Federal regulations described below, there are also guidelines governing the creation of a data repository or biobank (see Chapter 7). In particular, IRBs are charged with reviewing protocols for obtaining, storing and sharing information; verifying informed consent; and protecting privacy and confidentiality.

The Secretary's Advisory Committee on Human Research Protection (SACHRP) advises the Secretary of HHS on issues related to the protection of human subjects. SACHRP developed Frequently Asked Questions (FAQs) to provide a framework for IRBs, institutions, and investigators to consider relating to the collection, use, and storage of biospecimens.[23] One of the FAQs states that generally consent is necessary before moving excess identifiable clinical specimens to a centralized databank. In rare circumstances, an IRB may determine that the conditions for a waiver of consent have been met. Relevant factors to consider include: governance and oversight of the bank; protections in place for confidentiality/privacy; policies regarding access to specimens; the nature of the research for which specimens will be used; the databank's ability to locate/contact subjects; the risk of introducing bias into the collection process; potential anxiety/confusion for subjects; number of subjects; length of time since the specimens were first collected; and the likelihood a subject would object to research use.[23] While these factors are designed to address the use of clinical specimens, the issues raised are also applicable to the use of clinical data. Similarly, SACHRP suggests that an IRB determine whether a transfer of specimens to a new bank or institution is permissible under the initial consent—a relevant point for information transfers as well.

As with EHRs, there have been a variety of challenges to the use of biobanks for research without specific individual consent. Many long-standing biobanks were established either for non-research purposes (e.g., newborn blood spot banks) or under a general consent allowing the use of leftover tissue in hospitals. While more recent banks and repositories have been set up with a variety of consent protections, it is unclear what to do with existing repositories created without these protections, or how to manage access to archived data within the repositories where initial consent was either silent on the matter or significantly limits future research. At least one author has suggested the creation of a new regulatory oversight framework that would substitute for the necessary individual informed consent for the use of existing data or tissue samples.[24] Another suggests using broad initial consents to cover a variety of future uses.[25] Litigation in Texas and Minnesota regarding the use of newborn blood spots has highlighted this issue in the national dialogue, and development of additional regulations at the State level is likely. The ANPRM cited above includes among its proposed changes mechanisms to improve informed consent, including consent for the secondary use of pre-existing biospecimens and data.

Key unresolved issues relevant to both biobanks and large information data repositories include: obligations to return individually relevant research results, future unforeseen research uses, the need to recontact participants (some of whom may not wish to be recontacted or who are deceased), the financial burdens of recontacting, the limits on withdrawal of the sample or information, whether the sample/information can be kept indefinitely, whether commercial uses of the bank should be treated differently from noncommercial uses, and the implications of large-scale database research for socially identifiable groups. Moreover, as

technology continues to progress, so will the ability to re-identify participants from data deposited into biobanks and large data repositories.

De-identification and aggregate reporting alone does not completely conceal identity.[26, 27] For example, there is a considerable push to make de-identified, aggregate-level data from Genome Wide Association Studies (GWAS) publicly available in large repositories so that the data can be combined with other studies for more powerful analysis. However, an individual can be re-identified by assessing the probability that an individual or relative participated in a GWAS through composite statistics across cohorts (such as allele frequency or genotype counts). BioVU, the Vanderbilt DNA Databank, has taken steps to diminish the risk of re-identification. BioVU is linked to a de-identified version of data extracted from an EHR in which all personal identifiers have been removed. Thus, there is no identifiable information attached to the records. The disadvantages or tradeoffs in such design are that it explicitly precludes both recontacting and linking with any information other than that contained within the original EHR. It also prevents the return of individual results—an issue that remains controversial even when the study design allows it.

The informed consent documents initially used for biobanking research either stated explicitly that no results would be returned to participants or remained silent on the issue. More recently, there is general agreement in the scientific community supporting the return of aggregate results to research participants. There is less agreement on return of individual results. Moreover, there is still debate regarding the most ethically appropriate mechanisms for returning results (e.g., when, how, and by whom—physician or investigator). In 2010, a National Heart, Lung, and Blood Institute (NHLBI) Working Group released revised recommendations providing guidance on many of these issues, but the issue is far from settled.[28]

3.3 Reconsidering the Ethical Framework Governing Research

Perhaps the most challenging part of the shift to large database research and the current regulatory structure is the potential reframing of the underlying ethical issues. The July 2011 HHS-issued ANPRM states, "Although the regulations have been amended over the years, they have not kept pace with the evolving human research enterprise, the proliferation of multi-site clinical trials and observational studies, the expansion of health services research, research in the social and behavioral sciences, and research involving databases, the Internet, and biological specimen repositories, and the use of advanced technologies, such as genomics."[22] The current Federal research regulations are based on the Belmont Report, which focused on the traditional clinical research context. The HIPAA Privacy Rule was put in place more recently to protect the privacy of individually identifiable health information and demonstrates the challenges involved with balancing individual privacy with the information needs of a comprehensive health system. The future focus on electronic data repositories and the potential for large-scale observational studies to replace some clinical trial data require consideration of whether the approaches used thus far should be adapted.

For example, in discussing the possible use of the FDA's Sentinel System as a pharmacoepidemiological research database, Professor Barbara Evans identified three "novel challenges in applying familiar ethical frameworks."[29] The first is the possibility that with the shorter time period between research results and clinical application, the historical conceptualization of research versus treatment (or even public health practice versus research) may be incorrect. Perhaps IRBs will need to consider both the potential direct medical benefits of an observational study, and potential participant health risks such as negative insurance coverage determinations or changes in physician prescribing patterns. Second, the creation of these massive databanks that span numerous States (and sometimes countries) raises issues about whether the "local context review" that forms the basis for the IRB system continues to be relevant. Although a detailed examination of State regulations is not part of this chapter, it is worth emphasizing the challenges faced by multistate registries, which may face different requirements for informed consent, different privacy protections, and even

different definitions of "human subjects research" from state to state. This can add enormous burden to the regulatory oversight system and significant complexity to these endeavors. Finally, this type of research raises questions about the meaning of vulnerability and susceptibility to harm, and who should be identified as a "vulnerable" population in need of additional protections. It may be that the groups traditionally considered vulnerable in the clinical research context are not especially vulnerable in this context. Conversely, there may be groups particularly vulnerable to re-identification, or for whom re-identification poses unique risks of psychosocial or economic harms, but which would not usually be considered vulnerable in clinical research. In fact, the need to understand potential group harms highlights the limitations of the traditional ethical framework that assumes the focus should be on the individual. More work is needed to consider the ethical framework that should guide large-scale observational studies, but such exploration is beyond the scope of this chapter. The challenges raised by these studies have implications for research more generally and may lead to broader regulatory changes such as those proposed in the ANPRM.

4. Regulatory Consent Requirements

While a number of issues remain unanswered in this area, there is some clear guidance for registries that fall under the Federal research regulations. There are two primary sets of Federal regulations governing the conduct of human subjects research. HHS regulates research supported by Federal money or covered under an institutional "assurance of compliance" (see Chapter 7). The FDA regulates research that will be used to support an FDA-regulated product. Both sets of regulations largely have the same consent requirements; relevant differences are indicated below. The HIPAA Privacy Rule also contains individual authorization requirements for uses and disclosures of individually identifiable health information for research purposes. Each of these Federal regulatory areas will be discussed in turn.

4.1 HHS and FDA General Consent Requirements

For activities covered by the HHS and FDA research regulations, eight basic elements of information must be provided to research participants:[30]

1. A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental.
2. A description of any foreseeable risks or discomforts to the subject.
3. A description of any benefits to the subject or to others which may reasonably be expected from the research.
4. A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.
5. A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained.
6. For research involving more than minimal risk, an explanation as to whether any compensation and an explanation as to whether any medical treatments are available if injury occurs and, if so, what they consist of, or whether further information may be obtained.
7. An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of a research-related injury to the subject.
8. A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

In addition, the FDA announced on January 4, 2011, that informed consent forms for applicable clinical trials must include a statement that the trial

information will be entered into the National Institutes of Health (NIH) clinical trial registry.[31] Both the HHS and the FDA regulations also require, where appropriate, additional elements of informed consent, including the following:[32]

1. A statement that the particular treatment or procedure may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant) which are currently unforeseeable
2. Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent.
3. Any additional costs to the subject that may result from participation in the research.
4. The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject.
5. A statement that significant new findings developed during the course of research which may related to the subject's willingness to continue participation will be provided to the subject.
6. The approximate number of subjects involved in the study.

The HHS regulations allow an IRB to approve a waiver or alteration of the consent requirements for minimal risk research where the waiver or alteration will not affect the rights of the subjects, the research cannot be carried out without the waiver, and, when appropriate, subjects will be provided information after participation.[33] The FDA regulations do not allow waivers or alterations under these circumstances, but do allow for waivers in life-threatening situations and allow Presidential waivers for some military research.[34] Both sets of regulations allow for waiver of consent requirements for research conducted in specific types of emergency situations.[35, 36]

4.2 Documentation and Format of Consent

There are varying requirements for documentation of the consent process. Both FDA and HHS regulations speak to the documentation of informed consent.[37] Unlike treatment consents, research consents are usually written and the consent form functions both as documentation of the consent process, and in some cases as an aspect of the consent itself (since in long form the document contains all of the necessary consent disclosures and participants may be given the form to read as part of the consent process). HHS allows an IRB to waive the written documentation requirement in whole or in part when (1) the only record linking the subject and the research would be the consent document and the principal risk of the study would be potential harm resulting from a breach of confidentiality, or (2) the research involves no more than minimal risk of harm and involves no procedures for which written consent is not normally obtained in a clinical context.[38] In either case, the participant may still be provided a written summary. Waivers may be used to allow oral consent procedures without written documentation. Such an approach is often used in research involving interviews conducted by telephone.

Under certain conditions, an IRB may approve the use of a short-form written consent.[39] In these cases, oral presentation of informed consent information is accompanied by a short-form written consent document (stating that the elements of consent have been presented orally) and a written summary of what has been presented orally. A witness to the oral presentation is required, and the participant (or representative) must be given copies of the short form document and the summary. The participant must sign the short form document and the witness must sign both the short form and the summary.

E-consents may be considered written documentation under either set of regulations and are within the scope of the IRBs' power to authorize as an alternate way of fulfilling written documentation requirements under the HHS regulations. HHS specifically allows electronic signatures on research consent documents, provided they are legally valid in the specific jurisdiction.[40] FDA also has provisions for e-signatures on electronic records, but does not speak directly to e-consent for research

participation.[41] While the Federal Electronic Signatures in Global and National Commerce Act (E-SIGN) attempts to provide some uniformity among State laws governing electronic transactions, there remain some variations.[42, 43] The primary goals of e-signature laws are authenticating the signature and ensuring privacy and confidentiality of electronic information. Although there have been some suggestions for standardized electronic consent procedures, there is little discussion focused specifically on the research area.[44]

4.3 Informed Consent Form Revisions and Re-consent

Changes to the informed consent document may require re-consent of patients; for example, re-consent may be necessary if there are changes in the scope of the registry, substantive changes to the protocol, addition of procedures not previously addressed in the consent, changes in reasonably foreseeable risks or potential benefit, changes in data sharing or reporting procedures, or identified errors or omissions in the original consent document. Such revisions must be reviewed and approved by an IRB prior to the revised consent being utilized, except when necessary to eliminate apparent immediate hazards to subjects. Re-consent may also be necessary if the participants were below the age of consent when initially enrolled but reach the age of majority when the registry is still active. The decision to change the informed consent form and subsequently obtain the re-consent of participants needs to be carefully considered due to possible challenges in obtaining the re-consent. Challenges may be particularly evident for registries that have been in place for several years. These difficulties are what prompted the interest in broad general initial consents. In situations where the initial consent does not cover the change, registries may seek IRB waiver of re-consent requirements.

For studies in which re-consent is sought, registry developers should consider the potential effects of selection bias and the implications for external validity. Re-consented participants may be systematically different from non-re-consented participants. For example, participants that have not re-consented may have died or been lost to followup for health-related reasons, leaving an overall healthier group of participants. Additionally, even among those who can be contacted, individuals who agree to continue participation may be different from those who refuse to provide consent. As a result, one important requirement for studies that undertake re-consent may be to evaluate characteristics of the original study population, as compared with the subset of patients that do re-consent, and consider the implications for research outcomes. The evaluation of whether re-consents are more common for particular populations should be done for any analyses that have comparative arms.

Minor changes to a consent document do not necessitate re-consent. Re-consent is necessary, however, where the terms of the study or the background preconditions have changed. In some long-term studies, re-informing participants, but not re-consent, may be necessary. Even where re-consent is needed, IRBs may waive requirements. Alternatively, data collection, sharing, and reporting for participants who cannot be re-consented could be maintained in accordance with the terms of the original consent. In those situations in which a re-consent process is implemented, participants should be told the reasons for the recontact and given a summary of consent form changes. Additionally, as with the original consent, documentation of the re-consent must be maintained as required by the registry, the IRB, and any relevant regulations.

4.4 Applying the Federal Research Regulations to Registries

Some of the regulatory requirements appear better suited to traditional clinical research trials than to registries. For example, of the eight basic elements listed earlier, requirements 4 (alternatives) and 6 (compensation/injury) are crafted to address issues raised in traditional clinical trials, rather than registries. Other elements have aspects that clearly encompass registry research (such as basic elements 1, 2, and 7), but other parts seem less applicable, since registries will not involve "experimental procedures" that must be identified, entail no physical "discomforts to the subject," and do not pose a risk of (physical) "research-related injury."

Other requirements may pose challenges for registries, such as basic element 8, which requires that subjects be informed about a right to withdraw. While registry participants may refuse to provide additional information about their medical status or care, withdrawing from a registry may undermine the data collection. In situations where the data have been anonymized, withdrawal will likely prove impossible. In many such cases, registry informed consents may contain language notifying subjects that in the event of withdrawal, data that was collected prior to the withdrawal may continue to be used and disclosed according to the consent in order to preserve the scientific integrity of the registry. However, even where data have not been anonymized, some argue that the registry must retain all records to be a valid information tool. The FDA explicitly requires the retention of identifiable data even after a subject withdraws from a study. HHS permits the retention of such data, but also permits the investigator to omit or destroy the data if retention is not required by FDA regulations for study integrity. OHRP suggests that IRBs provide guidance on documentation of participant withdrawal. Moreover, the OHRP guidance dated September 21, 2010, on this issue clarifies that once a subject withdraws, the investigator must stop interacting with the subject to obtain data, and stop collecting identifiable private information from other sources unless the subject specifically provides consent to the continued data collection.[45]

4.5 HIPAA

The HIPAA Privacy Rule may apply to uses or disclosures of health information into or out of a registry, or the use of such information to create a registry, or both. In addition, because the HIPAA Privacy Rule protects individually identifiable health information held by covered entities, the Privacy Rule requirements may apply even if the human subjects research regulations do not. Moreover, the Food and Drug Administration Amendments Act of 2007 (FDAAA) requires all qualified entities with which it contracts to provide analyses of drug safety data, regardless of whether they are a HIPAA-covered entity, to follow the minimal requirements of the Privacy Rule.[46] Chapter 7 describes the general Privacy Rule framework in this context and the specifics of coverage. The Privacy Rule requires a covered entity to obtain written authorization for the use or disclosure of an individual's PHI for research purposes unless the use or disclosure is permitted by another provision of the Rule (e.g., where an IRB waiver of authorization applies). A subject's informed consent to participate in research can be combined with a HIPAA authorization in one document. There are six core elements and three required statements for a HIPAA authorization:[47]

Core Elements

- A description of the PHI to be used or disclosed, identifying the information in a specific and meaningful manner.
- The names or other specific identification of the person or persons (or class of persons) authorized to make the requested use or disclosure.
- The names or other specific identification of the person or persons (or class of persons) to whom the covered entity may make the requested use or disclosure.
- A description of each purpose of the requested use or disclosure.
- Authorization expiration date or expiration event that relates to the individual or to the purpose of the use or disclosure. "End of the research study" or "none" are permissible for research, including for the creation and maintenance of a research database or repository.
- Signature of the individual and date. If the individual's legally authorized representative signs the Authorization, a description of the representative's authority to act for the individual must also be provided.

Required Statements

- A statement of the individual's right to revoke Authorization in writing, and either: (1) a description of how to do so, and the exceptions to the right to revoke authorization, or (2) reference to the corresponding section of the covered entity's notice of privacy practices.

- Whether treatment, payment, enrollment, or eligibility for benefits can be conditioned on the individual's signing the Authorization, including research-related treatment, and consequences of refusing to sign the Authorization, if applicable.
- A statement of the potential for the PHI to be re-disclosed by the recipient and no longer protected by the Privacy Rule. This may be a general statement that the Privacy Rule may no longer protect health information disclosed to the recipient.

Authorization is not needed for activities that are "preparatory to research," which may include scanning a patient database to determine feasibility for creating a registry. Before allowing an investigator access to PHI for such purposes, however, the covered entity must obtain from the researcher representations that: (1) the use or disclosure of PHI is sought solely for purposes preparatory to research, (2) no PHI will be removed from the covered entity during the review, and (3) access to the PHI is necessary for the research purposes.[48] These preparatory activities may aid investigators in the identification of potential research participants. Subsequent contact of potential research participants for purposes of obtaining authorization for the use or disclosure of the individual's PHI may be permitted under the Privacy Rule in a variety of ways depending on the relationship between the investigator and the covered entity. An investigator who is a workforce member of the covered entity is permitted to contact potential participants directly or through another person at the covered entity, such as a treating provider, to obtain authorization. Alternatively, a covered entity is permitted to hire a business associate—who may be an investigator—to contact patients to obtain authorization on behalf of the covered entity. Finally, a covered entity is permitted to provide contact information of potential research subjects to an investigator who is not part of the covered entity or a business associate, if the covered entity obtains documentation that an IRB or privacy board has waived the authorization requirement for the disclosure.

Additionally, uses or disclosures of decedents' PHI to a research registry or from a registry for research purposes do not require an authorization (as long as certain representations are provided to the covered entity that is providing the information).[49] Authorizations are also not required for uses or disclosures of de-identified data sets, provided the information has been de-identified in accordance with the Privacy Rule.[50] Nor are authorizations required for uses or disclosures of "limited data sets," as defined by the Rule (so long as a data use agreement is in place with the recipient of the limited data set).[51] See Chapter 7.

In addition, an IRB or privacy board may waive or alter aspects of the HIPAA authorization requirements. Like the requirements for a waiver or alteration under the human subjects research regulations described above, these are limited to situations in which the research could not practicably be carried out without the waiver or alteration and could not practicably be carried out without access to the PHI, and the use or disclosure information involves no more than minimal risk to privacy because there is: (a) an adequate plan to protect the identifiers from improper use or disclosure; (b) an adequate plan to destroy identifiers if possible; and (c) adequate written assurances that the PHI will not be reused or disclosed except as required by law, as needed for research oversight, or for other research in a way permitted by the Privacy Rule.[52]

Finally, if a subject was enrolled in a research protocol prior to the compliance date of the Privacy Rule (for most covered entities, April 14, 2003) and pursuant to a valid informed consent, an authorization may not be required unless after the compliance date another informed consent is sought from the subject.[53] This provision may be especially relevant to registries that were created prior to the application of the Privacy Rule.

The HIPAA Privacy Rule also speaks to the issue of withdrawal from a registry. The Privacy Rule explicitly gives individuals the right to revoke their authorization for the use and disclosure of protected health information (the revocation must be in writing), except to the extent that a covered

entity has already relied on the authorization. HHS guidance on the application of the Privacy Rule to research makes it clear that a covered entity that has disclosed PHI for research in reliance on an authorization is not required to retrieve information it disclosed prior to receiving the revocation, and may also continue to use and disclose PHI already obtained to the extent necessary to preserve the integrity of the study (e.g., as necessary to account for the subject's withdrawal). As noted above, FDA states that the data gathered as part of research under their regulatory authority is necessary and must be retained; but even for those registries outside the scope of FDA oversight, HIPAA permits the continued use of data as necessary to protect the integrity of the research.

In the past, concerns had been raised that the HIPAA authorization requirements were not sufficiently harmonized with the human subjects research informed consent requirements.[17, 54] This was due to the fact that, while a HIPAA authorization could be combined with a research informed consent document (and elements already present in the research consent were not required to be repeated in the authorization), there were some situations in which an additional separate authorization may have been necessary for a separate research activity or future research activity. In January 2013, HHS published changes to the HIPAA Rules that addressed these issues. In particular, the Privacy Rule now permits the use of a compound authorization to authorize the use or disclosure of an individual's PHI for a conditioned research activity (e.g., a clinical trial where treatment is conditioned on the individual's authorization) as well as an unconditioned research activity (e.g., the use or disclosure of PHI to create or contribute to a separate research database or repository), provided that the compound authorization clearly differentiates between the conditioned and unconditioned research components and provides the opportunity for the individual to opt in to the unconditioned research activities. The final rule also modifies HHS' interpretation of the Privacy Rule's research authorization requirements to permit authorization for future research uses and disclosures, as long as the future research purposes are adequately described in the authorization such that it would be reasonable for the individual to expect that her or his PHI could be used or disclosed for such future research.

4.6 Special Consent Issues: Incapacitated Adults and Children

In addition to the general requirements discussed above, there are also additional requirements for certain specific research populations. HHS has regulations that apply to pregnant women and fetuses, children, and prisoners. FDA has regulations that apply to children (which, for the most part, match the HHS regulations). Both also allow research to be conducted with adults lacking decisional capacity, although consent must be obtained by a "legally authorized representative," who may be a guardian, proxy, or surrogate decisionmaker (the terms are defined by State law). Likewise, HIPAA also allows for authorizations from "personal representatives" (again, generally defined by State law).

Of particular interest to registries are the research regulations pertaining to children. Unlike research involving adults, research involving children must fit into one of four categories: minimal risk,[55] greater than minimal risk/prospect of direct therapeutic benefit,[56] minor increase over minimal risk/likely to yield generalizable knowledge about subject's disorder or condition,[57] and research not otherwise approvable but authorized by the Secretary of HHS in consultation with an expert panel.[58] Most registry research is likely to fall into the minimal risk category. For these studies, permission must be obtained from at least one parent/guardian and assent obtained from the child, if the child is capable of assenting.

Waivers of both permission and assent are possible. Under HHS regulations, a waiver of parental permission is allowed under the same conditions that allow for a waiver of informed consent in adult populations;[59] or when parental permission is not a reasonable requirement to protect the subjects.[60] FDA regulations do not allow for waivers of parental permission. Both HHS and FDA regulations allow a waiver of assent when the research involves an intervention holding the potential for direct therapeutic benefit and is

not available except through participation; or when parental permission is waived in accord with section 46.116 of Subpart A.[61] Moreover, when some or all of the children involved are not capable of providing assent, an IRB can determine that assent in not necessary (for the child or children in question, or for all children if appropriate). Both sets of regulations allow an IRB to determine that permission is only required from one parent, even when required from both under 406 or 407, in limited circumstances.[62] Where authorization must be obtained for the use or disclosure of PHI, the HIPAA Privacy Rule requires authorization from only one personal representative of the individual, such as one parent of a minor child, and does not require assent of the child.

OHRP has indicated that when the research in question involves a treatment for which the child would have legal authority to consent, the child's consent may suffice and parental permission may be unnecessary. The HIPAA Privacy Rule also generally provides that when a minor has legal authority to consent to a particular health care service without the involvement of a parent, the minor and not the parent has authority to act as the individual with respect to the PHI pertaining to that health care service. State statutes granting decisionmaking authority to minors vary. Many address issues such as treatment for sexually transmitted infections (STIs), access to contraception, and some even allow consent for mental health or substance abuse treatments. Registries involving these areas may be able to rely on the minor's consent, rather than the parental permission/assent framework. However, more specific legal guidance on the particulars of State statutory interpretation may be warranted in these situations.

Another important consideration is what to do when a minor who is involved in a registry reaches the age of majority. OHRP interprets the continuing consent standard to require that legal consent be sought from the participant upon reaching the age of majority. An authorization under the Privacy Rule, including one signed by a parent as the personal representative of a minor, remains valid until it expires or is revoked, even if such time extends beyond the child's age of majority. If the authorization expires on the date the minor reaches the age of majority, a covered entity would be required to obtain a new authorization signed by the individual in order to further use or disclose PHI covered by the expired authorization (or ensure that the use or disclosure falls within another regulatory permission in the Rule). Registries that involve children that will retain identifiable information past the child's age of majority will need to take steps to gain the appropriate consent and, if necessary, authorization for continued use. Less clear is whether investigators should seek a child's assent to continued participation when the initial consent was provided by parents at a time when the child lacked the capacity to play any role in decisionmaking.

5. A Proposed Framework for Registry Consents

5.1 Current Practices and Problems

There are three current approaches to consent: opt-in, opt-out, and non-consent. An opt-in approach assumes that an individual will not be part of the registry until they have specifically consented to participation. An opt-out approach assumes that all individuals will be part of a registry, unless there is a specific refusal to participate. Finally, a non-consent model does not seek or require individual consent or refusal, but includes all relevant individuals in a registry. The labeling of the approaches may vary in the literature, but the general concepts remain consistent. Additionally, some registries involve a mix of one or more approaches or a combined consent mechanism, where an opt-in approach is used for one aspect (access to a particular treatment) and non-consent for the other (listing in the treatment registry). This may also be referred to as "conditional access."

5.1.1 The Opt-In Approach

An opt-in procedure may involve a consent process similar to that used for clinical research protocols. It may be used separately for a registry, or it may

be appended to a consent document used for a particular treatment (for example, individuals who consent to the use of a particular device may also be asked to participate in a registry for that device). While an opt-in approach has the benefit of assuring compliance with the Federal regulations, a number of the regulatory requirements are difficult to apply to registries (as discussed above). This has led many to suggest a modified opt-in approach—using elements of the clinical research framework but adjusting to fit the registry model. But, even with a modified model, there are concerns that the strict informed consent requirements of the clinical research consent will have negative effects on subject selection, resulting in biases that will undermine the validity and thus affect the usefulness of the registry. An analysis of the Canadian Stroke Network estimated that dealing with consent issues cost $500,000 over the first 2-3 years of the registry, and the requirement to obtain written informed consent introduced significant selection biases undermining the usefulness of the registry.[63] Alternative consent approaches may need to be considered for large-scale observational studies.

5.1.2 The Opt-Out Approach

An opt-out procedure shifts the presumption from one in which each individual must consent to participate, to one in which each individual must refuse to participate. There is a great deal of discussion about the usefulness of an opt-out model, particularly for registries (e.g., organ donation registries). To be a valid opt-out model, individuals must be fully informed about the existence of the registry and their rights to opt-out of participation. In many cases, the information requirements are the same as the information requirements for an opt-in procedure—the only difference is that instead of explicitly agreeing to participate, the person must take steps explicitly to refuse to participate. While the information requirements may not change, the psychological shift may be significant. If the expectation is that everyone will participate, people may be more inclined to acquiesce. There is evidence in other areas of decisionmaking that setting the default to participation results in greater inclusion than setting the default to non-participation, even when individuals are given an easy way to opt-in or opt-out.[64] While the Federal research regulations appear to assume an opt-in approach, in some circumstances an IRB could approve a modification that allowed a shift to an opt-out. In order for an IRB to approve an opt-out approach for non-exempt, HHS-supported human subjects research, the researchers must document that the waiver of informed consent is appropriate for the research. An opt-out approach may be especially useful for registries. Nonetheless, Privacy Rule requirements will preclude this approach unless the situation fits within one of the delineated permissible uses without an individual authorization (e.g., with a waiver of authorization for research or for public health activities).

IRBs could consider the opt-out approach for research that meets the four criteria for a waiver or alteration of consent under the HHS guidelines: (1) the research involves no more than minimal risk to participants; (2) the waiver or alteration will not adversely affect the rights and welfare of participants; (3) the research could not practicably be carried out without the waiver or alteration; (4) participants will be provided with additional pertinent information after participation. For example, the Vermont Diabetes Information System (VDIS) is a quality-improvement, registry-based decision support and reminder system targeted to primary care physicians and their patients with diabetes. With IRB approval, VDIS incorporated an opt-out consent process.[65] Patients are notified by mail of their eligibility and inclusion in the registry and given a mechanism to opt-out by calling a toll-free number.

5.1.3 The Non-Consent Approach

Non-consent is not really a consent mechanism and thus will not be addressed here in detail. Nonetheless, this approach may, and probably should, entail providing participants with information about the registry. The format and process of disclosure may vary. In some cases, general public notifications (perhaps listing on a Web site, or posting prominently in a place likely to be seen) will be sufficient. In other cases, individual notification may be appropriate. A non-consent approach is used currently for registries that fall outside the Federal research

regulations such as State-mandated public health reporting or quality improvement activities. One primary methodological advantage of the non-consent approach in no-risk and minimal risk studies is that it can function to reduce concerns about biases introduced by the consent process, such as those that occur when individuals who consent to participate in the registry systematically differ from those who do not or cannot consent. Besides debates about when the use of a non-consent approach is acceptable (based on the level of permissible risk), most of the focus in this area should be on the type and extent of required notifications.

5.2 Scope of Consent

Consents may be broad or narrow. A so-called "blanket consent" approach asks for consent to a wide category of uses and assumes that consent will cover all uses, unless one is specifically excluded. Blanket consent should be distinguished from broad or general consent that does not necessarily imply "blanket" consent to all uses. In agreement with legislation, broad consent refers to use in biomedical research, not to other kinds of uses, such as for forensic use or for use by immigration authorities. A blanket consent model has historically been relatively common and still exists in some contexts. For example, patients entering a health institution or agreeing to a procedure sometimes have a notation at the bottom of the general consent form allowing the use of leftover tissue in any way deemed appropriate by the institution. Extremely broad blanket consents are not generally viewed as valid exercises of autonomy and thus may not truly be considered to be "informed consent." At best, blanket consent may be viewed as a type of notification procedure, alerting individuals to the possible uses of their information. Neither the Federal human subjects research regulations nor the Privacy Rule permit extremely broad blanket consents. Some registries will have been created, with the use of a prior express legal permission from individuals, before the compliance date of the Privacy Rule, and may additionally fall under an exemption to the human subjects research regulations; in these circumstances previously obtained broad blanket consent may be deemed sufficient.

The real question related to the scope of consent is to what extent consent can and should authorize future unspecified uses. In other words, how broad a consent is permissible? The exercise of autonomy should include the ability to consent both to specific and to non-specific research participation. An individual who would like to give broad permission for the use of their data in any future registry (or for use in a particular registry, but include permission that the information may be shared with investigators for any future research query) is exercising a form of autonomy. In addition, part of the issue is in determining whether a broad consent was truly informed. In the absence of specific details about the future uses, decisionmaking is necessarily less informed than if every future use is spelled out clearly. However, the ethical doctrine of informed consent does not require this level of detail. Moreover, requiring multiple consent dialogues may respect autonomy less than permitting broad consent if the individual does, in fact, wish to give broad permission and does not want continued recontact. In some contexts, such as the donation of biological samples, broad consents are more acceptable (e.g., there is a long history of allowing unrestricted tissue donations). It has become common to provide a menu of options in a consent form for biological or genetic databanking. These allow participants to specify any constraints they would like to place on the of their samples, such as permitting use only for the specific study listed, or for all studies in a particular research areas (e.g., heart disease), or for any future study in any area. Details regarding whether and under what circumstances the participant would like to be recontacted may also be collected. By contrast, in other areas such as consent to participate in a clinical trial, broad consents (e.g., "I give consent to participate in any clinical trial") are insufficient on both ethical and regulatory grounds. For situations such as the use of medical information, the scope of a broad consent is less clear. The debates about the scope of consent are ongoing.

5.3 Oversight and Community Consultation

Consent is only one aspect of the protections in place for human research participants. The second

major part involves IRB review and oversight. Other chapters discuss the oversight roles for IRBs and registry governance boards.

Community consultation, a third concept that usually appears in the context of discussions on human subjects research consent, is not actually part of this system of protections. In fact, there is no simple community analog to individual consent. Consent requirements for research arise from the principle of autonomy, and there is no corresponding principle at the community level. Thus, concepts such as "community consent" or "community authorization" can be incoherent, in part because there is no unitary concept of a community. Communities may be defined on social, biological, religious, racial, cultural, or geographic grounds. Most people belong to multiple, sometimes overlapping, communities. Some of these communities may have a designated spokesperson, but this individual may not represent the interests of all members of the community. (Consider, for example, the complex relationship between the Pope and Catholics in the United States.) Other communities have no clearly identified spokesperson. It is inappropriate to consider community consultation as a replacement for individual consent. Rather than view community involvement as an aspect of consent, it should be considered as part of oversight (and an analog of IRB review).[66] Nevertheless, community involvement in the design and oversight of a registry may be particularly important when the registry involves socially identifiable groups that have been subject to historic discrimination or when it involves sensitive genetic information. In addition, in some cases, community involvement can enhance participant understanding of consent and thereby increase individual participation.

6. Consent Guidance

Although general agreement has been reached about the required elements of informed consent for clinical research, this model may not be entirely applicable to informed consent for the creation of and participation in registries. Risks to participants (and, when applicable, risks to groups and/or communities) should be balanced carefully with the public health benefits of registry development. The sensitive nature of information about participants and the potential for broad data distribution highlight the importance of the informed consent process. Moving forward, informed consent elements and guidelines specific to registry research should be developed.

6.1 Special Considerations

Given the nature of registry research, some elements of informed consent should be given special consideration, including: the scope of the use of registry data, potential for recontact, withdrawal, and information regarding the electronic data security and management to be employed.

6.1.1 Scope of Use of Registry Data

Registries constitute a valuable resource, since investigators often draw upon them to address questions extending far beyond those envisioned when the registries were first created. Therefore, informed consent for registry research that allows broad data sharing is optimal for promoting science. There may be instances, however, such as with respect to research on specific diseases (e.g., HIV/AIDS research), where more specific consent may be appropriate. Additional Federal-level guidance on the appropriate scope of broad consent for future uses will be important. In the meantime, registry developers should not only provide clear parameters regarding the scope of use of registry data when first creating the registry, but should also develop a mechanism to consider how future, possibly unanticipated, requests for data access will be evaluated. The registry governance board can play an important role in this situation.

6.1.2 Recontact

Individuals should be informed how their data/samples will be used at the point of entry into the registry. Whether and how participants will be recontacted should be established at the outset and included in the consent form. Exceptions should be considered specifically where data/samples were made irretrievably anonymous, since recontact would then be impossible. It is important to inform registry participants that the anonymization of their data will make withdrawal from the registry impossible.

6.1.3 Withdrawal

Many issues governing withdrawal from a registry have been discussed in this chapter. Consensus needs to be developed regarding whether withdrawal should be presented as an option to participants in the initial consent, and, if it is an option, how withdrawal will be managed. While withdrawal from a traditional research study is a basic subject right, withdrawal of collected data, even from clinical trials, may be restricted. It is extremely important that registry creators develop initial rules and procedures for withdrawal and fully inform participants of them.

6.1.4 Electronic Data Security

Given the public concerns about electronic data security, participants in the registry should be clearly informed as to the physical security of their data and/or biospecimens, including methods of coding and removal of identifiers, encryption techniques, potential for cloud computing, and quality assurance policies. As well, participants should be informed about the process of releasing and transferring data to future investigators as it relates to maintaining confidentiality. In some cases, this information will reassure participants, potentially increasing consent rates.

6.2 Proposed Consent Form Elements

The following is an outline of potential elements to consider when developing consent forms and engaging in consent dialogs for registry research. These elements were generated from the applicable HHS, FDA, and Privacy Rule requirements and include consent aspects particularly relevant to registry research. These are all issues that should be considered; there may be additional legal requirements (e.g., for a HIPAA authorization). The outline below should not be viewed as comprehensive or even applicable to all registries. Some registries will modify this outline to meet specific needs, while others will follow a consent procedure similar to one used for traditional clinical trials. However, this outline provides a starting place for understanding the scope of informed consent for registry participation. It is important to note that the responsibility for obtaining and assuring appropriate informed consent rests on multiple parties, including sponsors, investigators, Protocol Review Committees (PRCs), and IRBs. Moreover, despite the multitude of elements listed below, every effort should be made to keep consent forms as short as possible and at approximately a fifth to eighth grade reading level.

The following elements should be considered:

1. A statement that the individual is being asked to take part in a registry (or a research study, if applicable), including—
    a. The name of the specific registry for which consent is being obtained.
    b. An explanation of the purposes of the registry (why it was created, who will be included).
    c. The expected duration of participation.
    d. A description of the procedures entailed.
    e. The approximate number of subjects involved (if applicable).
2. A description of any foreseeable risks or inconveniences (specifically risks related to any potential breach of confidentiality related to the data being collected).
    a. When human genetic research is anticipated, information should include possible consequences of genetic testing (e.g., insurance risks, paternity determinations, potential risks to family and community) and other related confidentiality risks.
3. A description of the types of research that the repository will support, and any benefits to the subject or to others which may reasonably be expected, including—
    a. A statement about whether and how findings will be communicated to participants.
4. A statement describing the extent to which confidentiality of data/biospecimens identifying the subject will be maintained (including a description of the operations of the repository—how data/specimens will be stored and managed), including—

   a. If applicable, a statement about whether registry results will be published.

   b. A statement about the impact of participation on the subject's access to his/her medical records (e.g., that access may be limited until all work on the registry is completed).

5. The conditions and requirements under which data and/or specimens will be shared with recipient investigators, including—

   a. If applicable, a description that the data/specimens will be broadly shared and may be used for future research that is not yet identified.

   b. The fact that the data/specimens may be transferred to other institutions and explanation of a data transfer security plan.

6. A description of when recontact might be necessary, and how recontact will be handled.

7. A statement of whether there are any costs to participation and/or any payment for participation.

8. A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

   a. The consequences of a subject's decision to withdraw from the research, including the possibility that the previously collected data will continue to be used, and procedures for orderly termination of participation by the subject.

9. Details on who to contact for answers to pertinent questions about the research and research subjects' rights.

10. As appropriate, any State-specific addenda.

Case Examples for Chapter 8

Case Example 14. Issues with obtaining informed consent

Description	The Registry of the Canadian Stroke Network (RCSN), now known as the Ontario Stroke Registry, is a registry of stroke patients in Canada. The registry, currently in Phase V, is a non–consent-based registry that collects detailed clinical data on the acute stroke event from the onset of symptoms, including emergency medical service transport, emergency department care, and hospital discharge status. The purposes of the registry are to monitor stroke care delivery, to evaluate the Ontario Stroke System, and to provide a rich clinical database for research.
Sponsor	Canadian Stroke Network, Networks of Centres of Excellence, and Ministry of Health and Long Term Care of Ontario
Year Started	2001
Year Ended	Ongoing
No. of Sites	154
No. of Patients	More than 60,000

Challenge

The registry began in 2001 with Phase I, in which data were gathered from 21 hospitals in Canada. All patients admitted to the hospital or seen in the emergency department with symptoms of acute stroke within 14 days of onset or transient ischemic attack (TIA), as well as those with acute in-hospital stroke, were included in this phase. Research nurse coordinators identified eligible patients through daily reviews of emergency and admission patient lists and approached these patients for consent. Informed patient consent was required for full data collection, linkages to administrative data, and 6-month followup interviews.

Informed consent was required for full data collection. Unfortunately, consent was obtained for only 39 percent of eligible patients. Subsequent analyses showed that patients who consented to participate were not representative of the overall stroke population, as they were less likely to have severe or fatal stroke, and also less likely to have minor stroke or TIA.

Phase II of the registry began in 2002, with 21 hospitals and 4 Ontario Telestroke sites. In this phase, all patients admitted to the hospital or seen in the emergency department with symptoms of acute stroke within 14 days of onset or TIA were included. Patients with in-hospital stroke were no longer recruited. In order to standardize workload across the country, a random sample of eligible patients was selected to be approached for consent for full data collection. Consent was obtained from 50 percent of eligible patients.

After obtaining consent of only 39 percent and 50 percent of patients in Phases I and II, the team realized that obtaining written patient consent for participation in the registry on a representative sample of stroke patients was impractical and costly. Patient enrollment threatened the viability and generalizability of the stroke registry. The registry team published these findings in the New England Journal of Medicine in April 2004.

Proposed Solution

The registry team approached the Ontario Information and Privacy Commissioner to discuss a non-consent-based registry for Phase III. As a result of these discussions, the registry was "prescribed" by the Privacy Commissioner under the Personal Health Information Protection Act, 2004. This decision allowed the registry to collect data legally on stroke patients without written consent.

Case Example 14. Issues with obtaining informed consent (continued)

Results

Phase III of the registry included all patients presenting to emergency departments of the 11 "Stroke Centres" in Ontario and 1 center in Nova Scotia with a diagnosis of acute stroke or TIA within 14 days of onset. Nurse coordinators identified eligible patients through daily reviews of emergency and admission patient lists. Patients were identified prospectively, with retrospective chart review, without consent. No followup interviews were done. Because informed consent was not required, the data collected provided a representative sample of stroke patients seen at tertiary care centers in Canada, making the data more viable for use in research and in developing initiatives to improve quality of care. The registry has now expanded to include a population-based, province-wide audit of stroke care delivery on a 20-percent sample of patients from every acute care institution in Ontario.

Key Point

The impact of obtaining informed consent should be considered in developing a registry. Requiring that registries obtain the consent of patients with acute medical conditions such as stroke may result in limited selective participation, as it is not possible to obtain consent from all patients. For example, patients who die in the emergency department and patients who have brief hospital visits may be missed. Mechanisms such as obtaining a waiver of informed consent or using the approach outlined in this case may be alternatives.

For More Information

Tu JV, Willison DJ, Silver FL. et al. The impracticability of obtaining informed consent in the Registry of the Canadian Stroke Network. N Engl J Med. 2004;350:1414–21.

Case Example 15. Operationalizing informed consent for children

Description	TARGetKids! (Toronto Applied Research Group) is a prospective registry enrolling healthy children aged 0-5 years. The aim of the registry is to link early nutritional exposures to later health outcomes including obesity, micronutrient deficiency, and developmental outcomes.
Sponsor	University of Toronto
Year Started	2008
Year Ended	Ongoing
No. of Sites	7 primary care practices in Toronto, Canada
No. of Patients	4,287

Challenge

Research involving children faces unique challenges, including special requirements related to the informed consent process. In this pediatric patient registry, patients are recruited at their annual well-child visits and followed up for ten years during subsequent annual well-child visits. Participation involves completion of age-specific questionnaires related to the child (asking for nutritional, behavioural, and developmental information), collection of anthropometric measurements, and collection of the child's blood sample (4-7 mL) by a trained pediatric phebotomist.

Consent for the registry is provided by one or both parents. By signing consent, parents also authorize the collection of their child's health card number to allow researchers to access the child's health records. Registry staff anticipated challenges in obtaining informed consent for these activities, particularly given the infrequency of contact with patients and the fact that blood sample collection is not part of normal clinical care during these annual visits. After reviewing the registry protocol, a research ethics board recommended that steps be taken to minimize coercion when recruiting and obtaining consent from patients.

Case Example 15. Operationalizing informed consent for children

Proposed Solution

Two weeks before a scheduled well-child visit for an eligible patient, the physician's office mails a short informational letter to the child's home. The purpose of the letter is to provide information about the registry and prepare parents for their contact with registry staff during the visit. By providing this information in advance, the letter minimizes the possibility that parents will feel coerced to consent to registry participation.

On the day of the visit, the child's parents are approached by a registry research assistant to provide consent for participation of their child in the registry. The research assistant explains what participation entails (i.e., completing questionnaires, collection of anthropometric measurements, and collection of a blood sample). If the parent spontaneously expresses that they wish to participate in the registry but don't wish to participate in one of these activities, they are given the option to opt out of one portion of the registry (e.g., blood collection) while still consenting for others (e.g., questionnaires and anthropometric measurements).

Results

The registry is now following 4,287 children aged 0 to 5 years from seven primary care practices in Toronto, Canada. The participation rate for the registry is 49 percent of all eligible children in the targeted practices (defined as children aged 0-5 years with a well-child appointment scheduled). The primary reasons parents decline to participate in the registry include lack of time to answer the questionnaire, no legal guardian accompanying the child, the need to discuss participation with spouse, and their child not feeling well that day.

Of consenting parents, about 50 percent consent to the registry blood sample. One possible reason parents choose to consent to the blood sample is that they see value in the test results (i.e., for iron or vitamin D deficiency) which are not standard of care for children in Canada. Parents may perceive this as an added benefit, although, in an attempt to minimize coercion, the informed consent form and registry staff do not emphasize this fact.

Key Point

Providing patients (and parents of pediatric patients) with information about the registry in advance can give them time to prepare questions and thoughtfully consider whether they wish to consent to participation. A flexible consent structure that allows patients to opt out of activities of a sensitive nature can reduce barriers to consent and participation.

For More Information

http://targetkidssu.wordpress.com

Morinis J, Maguire J, Khovratovich M, et al. Paediatric obesity research in early childhood and the primary care setting: The TARGet Kids! Research Network. Int J Environ Res Public Health. 2012 Apr;9(4):1343-54. Epub 2012 Apr 16.

Birken CS, Maguire J, Mekky M, et al. Parental factors associated with screen time in preschool children in primary-care practice: a TARGet Kids! study. Public Health Nutr. 2011 Apr 5:1-5.

Maguire JL, Birken CS, Jacobson S, et al. Office-based intervention to reduce bottle use among toddlers: TARGet Kids! Pragmatic, randomized trial. Pediatrics. 2010 Aug;126(2):e343-50. Epub 2010 Jul 12.

Case Example 16. Using a patient-centered study design to collect informed consent, maximize recruitment and retention, and provide meaningful clinical data

Description	Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) is a prospective research registry tracking and studying long-term outcomes of elective total joint replacement (TJR) surgery, funded by an Agency for Healthcare Research and Quality (AHRQ) award to the University of Massachusetts Medical School. The registry seeks to understand patient-reported and clinical outcomes by collecting data on baseline patient attributes, procedure approach and technology, inpatient hospital stay, surgeon and institutional characteristics, longitudinal patient pain and function, and post-procedure complications and revisions. A diverse patient cohort allows the generation of aggregate severity-adjusted national and regional data against which participating surgeons can compare their own practice.
Sponsor	Agency for Healthcare Research and Quality
Year Started	2011
Year Ended	Ongoing
No. of Sites	36
No. of Patients	9,000 as of January 2013; 30,000 projected

Challenge

Total joint replacement (TJR) is a common procedure, with more than 700,000 primary hip and knee replacements performed in the United States each year. Although TJR can result in significant pain relief, physical function and activity levels can vary widely after surgery. FORCE-TJR collects data to track patient, provider, and site characteristics in order to evaluate their contributions to patient-reported and clinical outcomes of TJR over time.

TJR patients often have limited contact with their surgeons immediately after making the decision to have surgery, instead interacting with office and hospital staff to complete insurance or anesthesia pre-operative paperwork. Administrative site staff often do not have the time or training to effectively inform patients about the risks and benefits of participating in patient-centered studies. Further, clinical information that may contain important data about adverse events resides in various, disconnected points of care. Patients may return to the hospital in which TJR was performed or they may present at another hospital, urgent care center, or doctor's office. Often these disparate sites of care are not linked with the same electronic medical record, making data challenging to collect. Collecting informed consent, patient reported outcomes, and other followup data from TJR patients can therefore be challenging and requires an innovative approach.

Proposed Solution

Successful approaches to maximizing patient participation in research are based in creating a relationship with each patient and minimizing the burden on site staff. Patients who schedule a TJR are asked by administrative staff at the participating site to sign a short study contact form, giving permission for registry staff to contact them. Site staff give the patient an informational packet and fax the signed contact forms to the registry. To collect informed consent, registry research staff contact patients at their convenience via telephone to review the study procedures, informed consent form, and medical release forms in the informational packet. At this point, patients have the opportunity to ask questions of registry staff and discuss with them any concerns, facilitating a deep understanding of the registry and their role in its success. Patients return the signed informed consent and medical release forms to registry staff via U.S. mail.

Case Example 16. Using a patient-centered study design to collect informed consent, maximize recruitment and retention, and provide meaningful clinical data (continued)

Proposed Solution (continued)

Collecting clinical data that does not reside in a single medical record also relies upon patient engagement. Upon enrollment in the registry, patients are asked to authorize release of their medical records; at each contact following surgery, patients are asked if they sought medical care since their last contact with the registry. If registry staff determine the medical care could be related to the TJR, the related medical records are obtained and analyzed.

Results

The model described above uses registry staff to enroll patients, obtain informed consent, and deliver longitudinal information and motivation, enhancing participant enrollment and commitment over the long term. This procedure facilitates the longitudinal collection of patient-reported outcomes and medical records data, thus enabling more precise severity adjustment than relying on administrative data. Sites report high satisfaction with the model, contributing to an 80 percent overall patient recruitment ratio in the registry.

Key Point

Registries and other patient-centered research can benefit from a study design that engages patients at enrollment, thereby increasing their participation over the life of the study. For registries that require clinical data from patients who may not access all their care within one system, an approach that follows the patient across settings can be beneficial. Contacting patients at their convenience rather than in a health care setting can allow them more time to have their questions answered, increasing patient commitment.

For More Information

http://www.force-tjr.org/FORCE-index.html

References for Chapter 8

1. 21 CFR 50.3(c).
2. 46 CFR 102 (d) - (f).
3. 45 CFR 46.101(b).
4. Casarett D, Karlawish J, Andrews E, et al. Bioethical Issues in Pharmacoepidemiologic Research. In: Strom BL, ed. Pharmacoepidemiology. 4th ed. Sussex, England: John Wiley & Sons; 2005. p. 593.
5. U.S. Department of Health & Human Services. Text Version of OHRP Decision Charts. http://www.hhs.gov/ohrp/policy/decisioncharttext.html. Accessed August 15, 2012.
6. Centers for Disease Control and Prevention. Guidelines for Defining Public Health Research and Public Health Non-Research. October 4, 1999. http://www.cdc.gov/od/science/integrity/docs/defining-public-health-research-non-research-1999.pdf. Accessed August 27, 2012.
7. Council of State and Territorial Epidemiologists. Public Health Practice vs. Research. May 24, 2004. http://www.vdh.virginia.gov/OFHS/policy/documents/2012/irb/pdf/public%20Health%20Practice%20versus%20Research.pdf. Accessed August 15, 2012.
8. 46.101(b)(2) and (3).
9. 46.101(b)(4).
10. 45 CFR 164.501.
11. Johnson N, Vermeulen L, Smith KM. A survey of academic medical centers to distinguish between quality improvement and research activities. Qual Manag Health Care. 2006 Oct-Dec;15(4):215-20. PMID: 17047495.
12. U.S. Department of Health and Human Services. Office for Human Research Protections. Quality Improvement Activities – FAQs. http://answers.hhs.gov/ohrp/categories/1569. Accessed August 15, 2012.

13. National Committee on Vital and Health Statistics. Recommendations on Privacy and Confidentiality, 2006-2008. Section I: Privacy and Confidentiality in the Nationwide Health Information Network. May 2009. http://www.ncvhs.hhs.gov/privacyreport0608.pdf. Accessed August 27, 2012.

14. AF Westin. How the Public Sees Privacy and Health Research. February 2008. http://www.iom.edu/~/media/Files/Activity%20Files/Quality/VSRT/S4_1Westin.pdf. Accessed August 15, 2012.

15. Rothstein MA. Currents in contemporary ethics. Improving privacy in research by eliminating informed consent? IOM Report misses the mark. J Law Med Ethics. 2009 Fall;37(3):507-12. PMID: 19723261. PMCID: 3077820.

16. Schneider S, Kerwin J, Robins C, et al. Consumer engagement in developing electronic health information systems: Final report. (Prepared by Westat, Rockville, Maryland, under Contract No. 233-02-0087). AHRQ Publication Number 09-0081-EF. Rockville, MD: Agency for Healthcare Research and Quality. July 2009. http://healthit.ahrq.gov/sites/default/files/docs/citation/09-0081-EF.pdf. Accessed August 15, 2012.

17. Institute of Medicine. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. January 27, 2009. http://www.iom.edu/Reports/2009/Beyond-the-HIPAA-Privacy-Rule-Enhancing-Privacy-Improving-Health-Through-Research.aspx. Accessed August 15, 2012.

18. Rodwin MA. Patient data: property, privacy & the public interest. Am J Law Med. 2010;36(4):586-618. PMID: 21302847.

19. Care FH. Protecting Privacy in Health Research: The Limits of Individual Choice. Calif Law Rev. 2010 Dec;98:1765-804.

20. Willison DJ, Schwartz L, Abelson J, et al. Alternatives to project-specific consent for access to personal information for health research: what is the opinion of the Canadian public? J Am Med Inform Assoc. 2007 Nov-Dec;14(6):706-12. PMID: 17712084. PMCID: 2213476.

21. Miller FG. Research on medical records without informed consent. J Law Med Ethics. 2008 Fall;36(3):560-6. PMID: 18840249.

22. Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 143 (proposed July 26, 2011). http://www.gpo.gov/fdsys/pkg/FR-2011-07-26/html/2011-18792.htm. Accessed August 15, 2012.

23. Office for Human Research Protections. U.S. Department of Health and Human Services. Attachment A to January 24, 2011 SACHRP Letter to the Secretary: FAQs, Terms and Recommendations on Informed Consent and Research Use of Biospecimens. http://www.hhs.gov/ohrp/sachrp/20110124attachmentatosecletter.html. Accessed August 15, 2012.

24. Greely HT. Breaking the stalemate: a prospective regulatory framework for unforseen research uses of human tissue samples and health information. Wake Forest Law Rev. 1999 Fall;34(3):737-66. PMID: 12664910.

25. Wendler D. One-time general consent for research on biological samples: is it compatible with the health insurance portability and accountability act? Arch Intern Med. 2006 Jul 24;166(14):1449-52. PMID: 16864754.

26. Homer N, Szelinger S, Redman M, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 2008 Aug;4(8):e1000167. PMID: 18769715. PMCID: 2516199.

27. Jacobs KB, Yeager M, Wacholder S, et al. A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies. Nat Genet. 2009 Nov;41(11):1253-7. PMID: 19801980. PMCID: 2803072.

28. National Heart L, Blood Institute working g, Fabsitz RR, et al. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. Circ Cardiovasc Genet. 2010 Dec;3(6):574-80. PMID: 21156933. PMCID: 3090664.

29. Barbara J. Evans. Appropriate Human-Subject Protections for Research use of Sentinel System Data. FDA Sentinel System Meeting Series Issue Brief: Legal Issues in Active Medical Product Surveillance. March 2010. http://www.brookings.edu/~/media/events/2010/3/08%20fda%20legal%20issues/panel%203%20issue%20brief.pdf. Accessed August 15, 2012.

30. 45 CFR 46.116(a); 21 CFR 50.25(a).

31. Informed Consent Elements; Final rule. 76 Federal Register 2 (4 January 2011), pp. 256-270. http://www.gpo.gov/fdsys/pkg/FR-2011-01-04/pdf/2010-33193.pdf. Accessed August 15, 2012.

32. 45 CFR 46.116(b); 21 CFR 50.25(b).

33. 45 CFR 46.116(d).

34. 21 CFR 50.23.

35. U.S. Department of Health & Human Services. Informed Consent Requirements in Emergency Research. http://www.hhs.gov/ohrp/policy/hsdc97-01.html. Accessed July 18, 2013.

36. 21 CFR 50.24.

37. 45 CFR 46.117; 21 CFR 50.27.

38. 45 CFR 46.117(c).

39. 45 CFR 46.117(b); 21 CFR 50.27(b)(2).

40. U.S. Department of Health and Human Services. Office for Human Research Protections. Can an electronic signature be used to document consent or parental permission? http://answers.hhs.gov/ohrp/questions/7260. Accessed August 15, 2012.

41. Donawa M. New FDA guidance on electronic records and signatures. Med Device Technol. 2001 Nov;12(9):32-5. PMID: 12938536.

42. Buchanan Ingersoll & Rooney PC. Electronic Records and Signatures in Healthcare and the Interplay of E-Sign, HIPAA and UETA. 1-7, 1. http://corporate.findlaw.com/business-operations/electronic-records-andsignatures-in-healthcare-and-the-interplay.html.

43. Paterick TJ, Paterick BB, Paterick TE. Expanding Electronic Transmissions in the Practice of Medicine and the Role of Electronic Informed Consent. J Patient Saf. 2008 Dec;4(4):217-20.

44. Goldstein MM. Health information technology and the idea of informed consent. J Law Med Ethics. 2010 Spring;38(1):27-35. PMID: 20446981.

45. U.S. Department of Health & Human Services. Guidance on Withdrawal of Subjects from: Data Retention and Other Related Issues. www.hhs.gov/ohrp/policy/subjectwithdrawal.html. Accessed July 18, 2013.

46. Pub. L. No. 110-85, 121 Stat. 823 (2007) (Codified in 21 U.S.C.).

47. 45 CFR 164.508. For additional information on authorizations for research, see http://privacyruleandresearch.nih.gov/authorization.asp. Accessed August 15, 2012.

48. 45 CFR 164.512(i)(1)(ii).

49. 45 CFR 164.512(i)(1)(iii).

50. 45 CFR 164.514 (a)-(c).

51. 45 CFR 164.514(e).

52. 45 CFR 164.512(i)(2).

53. 45 CFR 164.532.

54. Mark Barnes & David Forster, SACHRP, Subcommittee on Harmonization (SOH) Update: July 20, 2011 (2011). http://www.hhs.gov/ohrp/sachrp/mtgings/mtg07-11/soh.pdf. Accessed August 15, 2012.

55. 404; 21 CFR 50.51.

56. 405; 50.52.

57. 406; 50.53.

58. 407; 50.54.

59. 45 CFR 46.116.

60. 45 CFR 46.408(c).

61. 46.408(a); 50.55.

62. 408b.

63. Tu JV, Willison DJ, Silver FL, et al. Impracticability of informed consent in the Registry of the Canadian Stroke Network. N Engl J Med. 2004 Apr 1;350(14):1414-21. PMID: 15070791.

64. Thaler RH, Sunstein CR. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale: Yale University Press; 2008.

65. Littenberg B, MacLean CD. Passive consent for clinical research in the age of HIPAA. J Gen Intern Med. 2006 Mar;21(3):207-11. PMID: 16637821. PMCID: 1828090.

66. Marshall PA, Berg JW. Protecting communities in biomedical research. Am J Bioeth. 2006 May-Jun;6(3):28-30; discussion W46-8. PMID: 16754445.

Chapter 9. Protecting Data: Confidentiality and Legal Concerns of Providers, Manufacturers, and Health Plans

1. Background

As the cost of care delivered in the United States continues to grow at an unsustainable rate without parallel improvements in the quality of care,[1-8] health care policy experts and lawmakers are paying increasing attention to initiatives that measure and publicly report information about the performance of physicians, hospitals, and other health care providers and about the services and procedures they are delivering. They believe this is an important step toward improving health care quality and controlling costs. For example, advancing quality improvement through greater access to and use of health information is a specific goal of the Patient Protection and Affordable Care Act,[9] which includes a number of provisions to incentivize quality measurement, improvement, and reporting and to enable more informed decisionmaking by consumers and other stakeholders.

Critical to the success of these initiatives is the availability and accessibility of relevant administrative and clinical data. Data registries (or repositories) are often used to collect, process, maintain, and release relevant data for these purposes. For example, many professional associations and societies organized around provider specialties or specific diseases and conditions administer their own registries. Typically, these registries are used for a variety of patient safety and quality improvement activities, including: matching patients with researchers, tracking the course of patients' care, tracking and identifying trends in medical errors or other patient safety issues, and tracking outcomes related to specific diseases or conditions or the effectiveness of specific treatments used to treat them. For example, the American College of Chest Physicians (ACCP) directs the ACCP Quality Improvement Registry, Evaluation and Education, or AQuIRE, which is intended to "assist the chest physician with meeting increasing demands placed upon them by the public, credentialing bodies, regulatory agencies, payers, and the institutions in which they practice."[10] Likewise, the American Orthopaedic Association spearheads the Own the Bone registry, designed to better coordinate patient care among a patient's providers, close the gaps associated with physician treatment recommendations, and alter patient and physician behaviors to reduce future incidence of bone fractures due to osteoporosis.[11] Quality improvement and clinical research registries are also organized by the American College of Rheumatology,[12] the American College of Radiology,[13] the Society for Thoracic Surgeons,[14] and the National Cardiovascular Data Registry.[15] Other registries, such as the American Joint Replacement Registry, are endorsed by surgeons but use a multi-stakeholder funding and sponsorship model. Manufacturers and health plans also often contribute data to or sponsor registries for quality improvement and clinical research, including identifying care delivery trends and determining whether or not a particular service, procedure, medical device, or pharmaceutical achieves the desired effect.

Administrative and clinical information submitted by or about providers, medical device and pharmaceutical manufacturers, and health plans and included in registries for research, quality measurement and improvement activities, and patient safety initiatives often includes sensitive patient, provider, and/or manufacturer or health plan–identifiable information. Release of this information in an identifiable or even nonidentifiable manner may compromise the privacy of individual patients and providers and compromise sensitive financial, commercial, or proprietary manufacturer or health plan (e.g., benefit design, reimbursement) information. As more and more registries are developed and used for a variety of research (including comparative effectiveness research), quality improvement, and patient safety programs, more and more

information about patient safety, quality of patient care, performance, and other details about providers, medical devices, pharmaceuticals, and health plans becomes available. This information is incredibly useful because it helps providers better understand and improve the care they are delivering, helps manufacturers refine and improve the devices and pharmaceuticals they are developing, and helps patients make more informed choices about their providers and treatment options. However, this information is also desirable for use in litigation or other judicial or administrative proceedings to demonstrate that a certain level of care was adhered to or not, or that a certain device or pharmaceutical works in a particular way.

Considerable attention has been directed towards ensuring the privacy and confidentiality of individually identifiable patient health information maintained in registries, particularly in regards to the requirements of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.[16] As such, the privacy and confidentiality of individually identifiable patient health information is well established and recognized by Federal and State agencies, courts of law, and others. (See Chapter 7.)

Significantly less attention has been directed towards the privacy and potentially proprietary nature of information about providers, medical device and pharmaceutical manufacturers, and health plans. Often providers, manufacturers, and health plans are the best source of information about the care they deliver and/or their products to support the effectiveness of a registry, whether it is used for research, quality improvement, patient safety initiatives, or other related activities. Even when they do not directly give a registry information about the care they deliver and/or their products, they may be included in information provided by other sources. (For example, a provider may submit information about Company Y's medical device related to a procedure performed on a patient.) Numerous policymakers and researchers have noted the "chilling effect" that the lack of protection for the information provided may have on the willingness of providers, manufacturers, and health plans to provide (or be included as a direct or indirect subject of) relevant information to support research, quality improvement, and patient safety initiatives conducted through registries. For example, commentary in the 2006 Journal of the American Medical Association reported that "wariness about liability exposure in the medical community may stymie public and private efforts."[17]

The 1999 Institute of Medicine report, *To Err is Human*, explicitly identified this "wariness" as a significant issue that hampers voluntary reporting and collaborative efforts, including the free exchange of information, to identify medical errors and prevent their repetition. To address this issue, the report recommended that Congress pass "legislation to extend peer review protections to data related to patient safety and quality improvement that are collected and analyzed by health care organizations for internal use or shared with others solely for purposes of improving safety and quality."[18] To date, however, no comprehensive Federal legislation has been passed. Thus, providers, manufacturers, and health plans must look to a variety of Federal and State laws that may offer protection from disclosure of information pursuant to a discovery request or other judicial or administrative proceedings.

2. Relevant Laws and Regulations: Variety of Sources, But Limited Protection

While no general Federal statutory privilege exists to protect information held in a registry submitted by or relating to providers, manufacturers, or health plans, a number of Federal laws may provide protection from discovery or disclosure in judicial or administrative proceedings. In addition, most States have specific peer review or quality assurance laws that may provide additional protection as well, but again in limited circumstances. This chapter will primarily focus on available Federal evidentiary protections, but will also address State-specific protections. It concludes with an overview of mechanisms that may be used to protect information included in a registry during judicial or administrative proceedings. While registries may collect

information from both within the United States and internationally, treatment of registries by international laws is outside the scope of this chapter.

2.1 Federal Laws

2.1.1 AHRQ's Confidentiality Statute

All identifiable research data obtained by the Agency for Health Research and Quality (AHRQ) is protected by the agency's confidentiality statute.[19] The statute requires that data collected by AHRQ-sponsored entities that identify individuals or establishments be used only for the purposes for which the data are supplied. Any effort to determine the identity of a person in an AHRQ database, or to use the information for any purpose other than for research, analysis, and aggregate statistical reporting, violates the AHRQ confidentiality statute. Recipients of a data set are also prohibited from releasing, disclosing, publishing, or presenting any individually identifying information. Specifically, the statute provides:

> No information, if an establishment or person supplying the information or described in it is identifiable, obtained in the course of activities undertaken or supported under this subchapter may be used for any purpose other than the purpose for which it was supplied unless such establishment or person has consented (as determined under regulations of the Director) to its use for such other purpose. Such information may not be published or released in other form if the person who supplied the information or who is described in it is identifiable unless such person has consented (as determined under regulations of the Director) to its publication or release in other form.[19]

Concerns have been raised that this protection may be vulnerable if the information is disclosed to an outside entity such as a registry. However, AHRQ has interpreted this provision to protect all AHRQ-funded research from discovery requests, including discovery requests in the course of litigation. A memorandum from senior AHRQ attorney Susan Merewitz noted that "if individuals inside a health care institution are gathering identifiable medical error information as part of AHRQ-supported grant or contract research, and it is conveyed outside the institution (e.g., for analysis in an AHRQ-supported central databank), even if the reporters lost their protection against being subpoenaed to testify under State law, the Federal statute would cover and protect the identifiable information they acquired pursuant to AHRQ's statutory research authority."[20] While this memorandum is not binding on any court of law and has yet to be introduced in a legal challenge, it clearly establishes AHRQ's protective position as it relates to any information collected under the auspices of an AHRQ-supported project. Registries participating in AHRQ-sponsored activities would certainly be able to avail themselves of this protection and, given Merewitz's comments, this protection may even extend to non-AHRQ activities in which an AHRQ-sponsored entity holds the data as a repository or intermediary.

Importantly, this protection is limited to AHRQ-sponsored registries. Therefore, registries maintained by professional associations or other organizations that are not sponsored by or otherwise participating in an AHRQ-sponsored project would not benefit from the protections the AHRQ confidentiality statute affords.

2.1.2 HHS Certificate of Confidentiality

The U.S. Department of Health and Human Services (HHS) may issue a "Certificate of Confidentiality" for any research project that collects personally identifiable, sensitive information and has been approved by an institutional review board. A Certificate protects an investigator, and others who have access to research records, from being required to disclose identifying information on research participants in any Federal or State judicial, administrative, or legislative proceeding. The Certificate may be used for biomedical, behavioral, clinical, or other types of research.

In the research arena, the National Institutes of Health (NIH) is the most common source for Certificates of Confidentiality. The NIH considers research to be sensitive if disclosing the

information could have adverse consequences for subjects or damage their financial standing, employability, insurability, or reputation. According to NIH, examples of studies that may be considered sensitive include those collecting genetic information; information on subjects' psychological well-being; information on sexually transmitted diseases or on subjects' sexual attitudes, preferences or practices; and information on substance abuse or other illegal conduct; as well as studies where subjects may be involved in litigation related to exposures under study (i.e., breast implants, environmental or occupational exposures).[21]

The specific statutory language provides that:

> The Secretary [of the U.S. Department of Health and Human Services] may authorize persons engaged in biomedical, behavioral, clinical, or other research (including research on mental health, including research on the use and effect of alcohol and other psychoactive drugs) to protect the privacy of individuals who are the subject of such research by withholding from all persons not connected with the conduct of such research the names or other identifying characteristics of such individuals. Persons so authorized to protect the privacy of such individuals may not be compelled in any Federal, State, or local civil, criminal, administrative, legislative, or other proceedings to identify such individuals.[22]

The application of Certificates of Confidentiality to registries has four inherent shortcomings.[23] First, the protections apply only to the identity of research subjects, or to data that would allow the possible identification of such individuals. Thus, de-identified patient safety data is still potentially discoverable. Second, there are questions as to whether a Certificate applies to patients who presumably have not consented to becoming research subjects. Third, individual patients may be able to waive the Certificate protections as to their own information, which they presumably would do if they were plaintiffs in a malpractice lawsuit. Fourth, the protections apply only to research information for which a Certificate of Confidentiality has been applied for and has been granted. Therefore, the protections afforded by a Certificate of Confidentiality are limited and do not provide meaningful opportunities for registries that are not engaged in research and that have not applied for and been granted a Certificate.

2.1.3 The Patient Safety and Quality Improvement Act of 2005

The Patient Safety and Quality Improvement Act of 2005 (PSQIA)[24] creates a Federal privilege from discovery in connection with Federal or State judicial or administrative proceedings for certain information identified as patient safety work product. To claim the privilege, providers must create the patient safety work product by collecting event information and reporting it to a formally recognized patient safety organization (PSO) for aggregation and analysis. The term "patient safety work product" encompasses any data, reports, records, memoranda, analyses, or written or oral statements that meet one of two criteria: (1) the materials "could improve patient safety, health care quality, or health care outcomes" and are gathered by a provider to be reported and are reported to a PSO or are developed by a PSO to conduct patient safety activities; or (2) the materials "identify or constitute the deliberations or analysis of, or fact of reporting to, a patient safety evaluation system."

Materials not gathered to be reported to the PSO and not actually transmitted to a PSO would not qualify for a privilege. The privilege specifically does not apply to medical records, billing and discharge information, or other records kept outside of a patient safety evaluation system. Furthermore, providers must comply with any State laws that require reporting of patient safety information and may not use PSO activities to shield them from reporting. Thus, if a patient safety investigation references medical records, the records themselves do not become part of the work product eligible for protection.

Documents created, maintained, or developed separately from a patient safety evaluation system are excluded from the definition of patient safety work product. Thus, individual patient medical records, billing and discharge information, and any original patient or provider records are not considered work product and are thus not

privileged. In addition, information collected to comply with external reporting requirements is not work product.

The preamble to the final regulations (44 FR 70,732) implementing the confidentiality protections of the PSQIA identifies several examples of information that must be reported and that does not merit protection as work product. These examples of required reportinginclude State incident reporting, adverse drug event information reporting, records for compliance with health oversight agency requirements, reporting physician disciplinary actions to the National Practitioner Data Bank, and disclosures required under Medicare's conditions of participation. Thus, a significant amount of data remains outside the patient safety work product definition. This includes registry data that is not maintained by a PSO and used for specific patient safety activities or is not identifiable. Therefore, the PSO statute and regulations provide no protection for registries acting outside the protected scope of the PSO arena.

2.1.4 Quality Improvement Organization Statute and Regulations

Quality Improvement Organizations (QIOs) are responsible for improving the effectiveness, efficiency, economy, and quality of services delivered to Medicare beneficiaries. The Centers for Medicare & Medicaid Services (CMS) contracts with one private, generally not-for-profit organization in every State, as well as the District of Columbia, Puerto Rico, and the U.S. Virgin Islands, to serve as that jurisdiction's QIO. QIO employees, consisting primarily of doctors and other health care professionals, are instructed to review medical care and assist beneficiaries with quality of care issues and complaints, as well as to implement improvements to the quality of providers' care.

The QIO statute states that any data or information acquired by a QIO in its course of duties must be kept confidential and may not be disclosed to any person, except as the information assists Federal and State agencies responsible for investigating fraud, abuse, and risks to the public health, or assists appropriate State agencies and national accreditation bodies responsible for licensing or certifying providers or practitioners.[25]

Furthermore, the statute explicitly states that "no patient record in the possession of" a QIO may be subject to subpoena or discovery proceedings in a civil action.[26] Additionally, no document or other information produced by a QIO in connection with its deliberations may be subject to subpoena or discovery in any administrative or civil proceeding. However, a QIO is required to provide, upon the request of a practitioner or other person adversely affected by such deliberations, a summary of the QIO's findings and conclusions.

Additionally, QIO regulations state that quality review study information with a patient identifier is not subject to subpoena or discovery in a civil action, including administrative, judicial, or arbitration proceedings.[27] This restriction, however, does not apply to HHS administrative subpoenas issued in the course of an audit or investigation of HHS programs, in the course of administrative hearings held under the Social Security Act, or to disclosures to the U.S. Government Accountability Office as necessary to carry out its statutory responsibilities.

Similar to the PSQIA, the QIO statute and regulations provide protection only to information that has been collected by a QIO under contract with CMS to perform specific statutory functions. To the extent a QIO is the owner and operator of a registry used to perform required functions, the information included in the registry would be protected. However, this does not apply to the vast majority of registries currently in existence today.

2.1.5 HIPAA Privacy Rule

The HIPAA Privacy Rule protects the privacy of individually identifiable health information.[28] The Privacy Rule applies to "covered entities," which include health plans, health care clearinghouses, and health care providers who conduct certain electronic health care transactions. Some of the Privacy Rule's provisions also apply to business associates of covered entities (e.g., contractors performing specific functions on their behalf).[29] The purpose of the Rule is to protect "individually identifiable health information" held or transmitted

by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. "Individually identifiable health information" is information, including demographic data, that relates to an individual's (1) past, present, or future physical or mental health condition; (2) receipt of health care; or (3) past, present, or future payment for health care. The information also must either directly identify the individual or be able to reasonably lead to identification of the individual. Common identifiers include an individual's name, address, birth date, and Social Security number.[29] The Privacy Rule refers to this information as "protected health information" (PHI).

The Privacy Rule permits the disclosure of PHI in the course of any judicial or administrative proceeding in response to an order of a court or administrative tribunal.[30] Absent a court order, a covered entity also may respond to a subpoena or discovery request from a party to the proceeding if the covered entity obtains either: (1) satisfactory assurances that reasonable efforts have been made to give the individual whose information has been requested notice of the request; or (2) satisfactory assurances that the party seeking such information has made reasonable efforts to secure a qualified protective order that will guard the confidentiality of the information.

In meeting the first test, a covered entity is considered to have received satisfactory assurances from the party seeking the information if the covered entity receives a written statement and documentation that the party has made a good-faith effort (such as by sending a notice to the individual's last known address) to provide written notice to the individual whose information is the subject of the request, that the written notice included sufficient information about the proceeding to permit the individual to raise an objection, and that the time for the individual to raise objections to the court or administrative tribunal has elapsed and no objections were filed, or any objections the individual filed have been resolved.

A "qualified protective order" means an order of a court or of an administrative tribunal or a stipulation that: (1) prohibits the parties from using or disclosing the protected health information for any purpose other than the litigation or proceeding for which the records are requested; and (2) requires the return to the covered entity or destruction of the protected health information (including all copies made) at the end of the litigation or proceeding. Satisfactory assurances of reasonable efforts to secure a qualified protective order are a statement and documentation that the parties to the dispute have agreed to a protective order and that it has been submitted to the court or administrative tribunal with jurisdiction, or that the party seeking the protected health information has requested a qualified protective order from such court or tribunal.

Importantly, the protections of the HIPAA Privacy Rule will apply only if a registry is considered a covered entity or the business associate of a covered entity. A registry may be considered a covered entity if the registry is maintained and administered by a HIPAA-covered health care provider. More commonly, the registry will be acting as the business associate of a covered entity (e.g., collecting and processing PHI on behalf of provider(s) and/or health plan(s)). Even if the registry is considered a covered entity or business associate, the HIPAA requirements that apply to disclosures pursuant to a court order, subpoena, or discovery request only apply to PHI. To the extent the requested information does not include PHI (e.g., the information is considered to be de-identified according to the requirements of the Privacy Rule and thus does not contain identifiable health information about individuals), HIPAA does not protect information about providers, manufacturers, or any other entities. Therefore, in the case of providers, manufacturers, and health plans seeking protection for information specific to them or their products but not part of PHI, HIPAA does not shield them from discovery requests in litigation or any other court proceedings.

2.1.6 Privacy Act of 1974

The Privacy Act of 1974[31] protects information about individuals, such as patients and providers, held or collected by the Federal Government that can be retrieved by personal identifiers such as name, Social Security number, or other identifying number or symbol. The Privacy Act authorizes a

Federal agency to release individually identifiable information to identified persons or to their designees with written consent or pursuant to 1 of 12 exemptions for disclosure. These exemptions include disclosure to Federal agency employees, the Census Bureau, the National Archives and Records Administration, other government entities for civil and criminal law enforcement purposes, the Comptroller General, Congress or its committees, and a consumer reporting agency.[32] Additional exemptions include disclosures for statistical research, disclosures required by Freedom of Information Act, disclosures in response to emergency circumstances, and importantly for purposes of this chapter, disclosures pursuant to a court order.

Unless the Federal Government maintains the registry, the Privacy Act of 1974 offers no protection from discovery for litigation or related court proceedings. Furthermore, even if the Federal Government maintains the registry, the Privacy Act specifically allows for the release of identifiable information about individuals without their written consent pursuant to a court order. This could include information not only about individual patients, but also individual providers (e.g., individual practitioners).

2.1.7 Freedom of Information Act

Enacted by Congress in 1966, and expanded in 1996 to cover electronic records,[33] the U.S. Freedom of Information Act (FOIA)[34] generally provides that any person has the right to obtain access to information contained in the records of Federal agencies, unless FOIA specifically protects the information from disclosure. With a goal of ensuring an informed citizenry, capable of holding the government accountable, FOIA effectively establishes a statutory right of public access to executive branch information, requiring that virtually every record held by a Federal agency be provided to individuals upon request.[35] Information that is subject to FOIA is likely to be disclosable pursuant to a discovery request or other court proceeding. However, FOIA does have limited exemptions and exclusions to the broad disclosure requirements. The most relevant exemptions for purposes of protection of registry information—in order of their importance—are Exemptions 4, 6, and 3.

Exemption 4 protects "trade secrets and commercial or financial information obtained from a person and privileged or confidential."[36] Importantly, however, it is not a mandatory bar to disclosure, but rather limits an agency's obligation to disclose specified information. "Trade secrets" are defined as commercially valuable plans or formulas for producing trade commodities to which has been invested substantial effort or innovation.[37] In the case of registries that contain information supplied by or about providers, medical device or pharmaceutical manufacturers, and health plans, it is likely that only the pharmaceutical and medical device manufacturers and health plans would have information that could be considered "trade secrets." Furthermore, information that manufacturers contribute to a registry is more likely to be considered "commercial" or "financial." For example, in *Public Citizen Health Research Group* v. *Food & Drug Administration*,[38] the court found that "because documentation of the health and safety experience of their products will be instrumental in gaining market approval for their products it seems clear that the manufacturers … have a commercial interest in the requested information."[39]

Exemption 6 states that Federal agencies can withhold from disclosure information about individuals in "personnel and medical files and similar files" when the disclosure of such information "would constitute a clearly unwarranted invasion of personal privacy."[40] In order to warrant protection under Exemption 6, the information at issue must first meet the threshold requirement of falling into one of three categories—personnel files, medical files, and similar files. The Supreme Court found that Congress intended these categories to be interpreted broadly and to protect information that "applies to a particular individual."[41] Once it has been established that the information meets this threshold, the focus shifts to whether the disclosure of such information would be an unwarranted invasion of privacy. This requires balancing the public's right to disclosure of the information

against the individual's right to privacy. After determining that a protectable privacy interest exists, the public's interest in disclosure of the information will be weighed against the individual's privacy interest in not disclosing the information.

The landmark Supreme Court decision in *United States Department of Justice* v. *Reporters Committee for Freedom of the Press*[42] governs how privacy interests under Exemption 6 are determined and balanced with public interest in the information. First, the Court clarified that a substantial privacy interest may exist in information that has already been released to the public at some point. Second, the Court held that the identity of the individual or party requesting the information may not be taken into consideration when determining if information should be released and "has no bearing on the merits of his or her FOIA request."[43] When considering the public interest in disclosing the information, the Court ruled that the determination should be based on the nature of the requested information and its relationship to the public interest generally, and not solely the purpose for which the request is made. Finally, the Court narrowed the scope of the public interest to the kind of interest to information that will "shed light on an agency's performance of its statutory duties."[44]

It is important to note, however, that Exemption 6 protects only information that identifies the individual in question. Thus, while patient records included in a registry are likely to be protected, if identifying information is removed, the information is no longer protected under Exemption 6. The most common types of information protected under Exemption 6 are age, home address, Social Security number, medical information about individuals participating in clinical research trials, claims files, and other personal information held by CMS.[45]

Exemption 3 protects information if it is "specifically exempted from disclosure by statute, provided that such statute (A) requires that the matters be withheld from the public in such a manner as to leave no discretion on the issue, or (B) establishes particular criteria for withholding or refers to particular types of matters to be withheld."[46] An example of a statute that may prevent disclosure or discovery of information contained in a registry under FOIA Exemption 3 would be the Patient Safety Quality Improvement Act of 2005 (discussed above), if the information met the PSQIA requirements.

2.1.8 Federal Trade Secrets Act

The Federal Trade Secrets Act[47] imposes fines or imprisonment on any Federal employee who discloses any information that relates to trade secrets, processes, operations, style of work, or apparatus, or to the identity, confidential statistical data, amount or source of any income, profits, losses or expenditures of any person or corporation. The Act applies only to public disclosures, and does not reach internal agency use of the data.

There is no private right of action under the statute; however, the Administrative Procedure Act may create a right of action to prevent a violation of the Trade Secrets Act or review a decision to disclose information.[48] Similar to Exemption 4 under FOIA, the Federal Trade Secrets Act may protect proprietary information, however, only to the extent that is held by the Federal Government and disclosed publicly (e.g., the Act may not reach information disclosed pursuant to a protection order as part of a discovery proceeding).

2.1.9 Federal Rules of Evidence and Civil Procedure

Federal legal rules of evidence and civil procedure may place limits on what information may be discoverable or otherwise used in a court proceeding. For example, Rule 401 of the Federal Rules of Evidence defines "relevant evidence" as evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence."[49] Rule 403 narrows the scope of Rule 401 stating "although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury…"[50] Often of the most relevance to information contained in registries is the prohibition in Rule 404 against "evidence of

other … acts … to prove the character of a person in order to show action in conformity therewith."[51] This may apply in instances where information is sought from a registry to show evidence of similar medical outcomes.

Turning to Civil Procedure, Rule 26(b)(2)(C)(iii) of the Federal Rules of Civil Procedure provides that a court may limit discovery if "the burden or expense of the proposed discovery outweighs its likely benefit, considering … the importance of the discovery in resolving the issues."[52] Rule 26(c) also allows parties from whom discovery is sought to move for a protective order.[53] Rule 45(c) protects individuals from unduly burdensome or expensive subpoenas. Specifically, 45(c) states, "a party or attorney responsible for issuing and serving a subpoena must take reasonable steps to avoid imposing undue burden or expense on a person subject to the subpoena. The issuing court must enforce this duty and impose an appropriate sanction—which may include lost earnings and reasonable attorney's fees—on a party or attorney who fails to comply."[54] It is left to the discretion of the court in each case to determine whether the facts and circumstances merit quashing a specific subpoena.

While these Federal rules of evidence and civil procedure may offer protection against discovery of registry information in certain situations, their application is left entirely to the discretion of the particular court in which the case is heard. Thus, the case law is mixed, with some courts allowing discovery and others not, depending on their application of a balancing test of the need for confidentiality versus hardship to the party seeking discovery. For example, in *Andrews* v. *Eli Lilly & Co., Inc.*, *E.R. Squibb & Sons* and *Rexall Drug Company*,[55] Squibb sought production of data from a University of Chicago registry that included information about a disease related to a products liability action against Squibb. The University of Chicago claimed that the contents of the registry were privileged and confidential. In balancing the privacy interests of the registry against the need for the information, the court stated, "the balance … tips in favor" of the registry.[56] "Squibb's need for the information is speculative and uncertain. Its essentially private interest in defending itself is outweighed by the compelling social interest in preventing harm to the Registry and the vital work it conducts."[56] The court, however, did allow discovery of information contained in the registry related to the plaintiffs. Squibb appealed, and in *Deitchman* v. *E.R. Squibb & Sons*,[57] the appellate court noted that the privilege was not "absolute" and remanded the case back to the lower court to determine the best way to provide information from the registry while preserving confidentiality.[58] In a secondhand-smoke case, *Wolpin* v. Phillip Morris, Inc.,[59] the court ordered disclosure of data from a statewide tumor registry from a study by the University of South California and California Department of Health, despite objections that the data was protected by State and Federal privacy laws. The court held that the need for the information in the lawsuit outweighed the confidentiality interest, but did require patient names to be removed prior to release.

2.1.10 Patient Protection and Affordable Care Act: Release of Medicare Claims Data for Provider Performance Measurement and Reporting

Section 10332 of the Affordable Care Act requires the Secretary to make Medicare claims data available to qualified entities for the evaluation of provider performance on measures of quality, efficiency, effectiveness, and resource use.[60] The data are standardized extracts of claims data under Medicare Parts A, B, and D, for items and services furnished for one or more specified geographic areas and time periods requested by a qualified entity. The qualified entities will be required to pay a fee to obtain the data and must submit to the Secretary of HHS a description of the methodologies that will be used to evaluate the performance of providers and suppliers.

All subsequent reports a qualifying entity publishes must include an understandable description of the measures, risk adjustment methods, physician attribution methods, other applicable methods, data specifications and limitations, and the sponsors, so that consumers, providers of services and suppliers, health plans, researchers, and other stakeholders can assess such reports. In addition, the reports must be made available to any provider or supplier identified in

the report, with an opportunity to appeal and correct any errors. Finally, the reports may include information on a provider or supplier only in an aggregate form as the Secretary determines appropriate.

The Secretary may not make claims data available to a qualified entity unless the entity agrees to release the information on the evaluation of performance of providers of services and suppliers. Section 10332 requires that data released to a qualified entity shall not be subject to discovery or admission as evidence in judicial or administrative proceedings without consent of the applicable provider or supplier.[61]

While limited in application to Medicare claims data provided to qualified entities, this provision is a significant recognition by Congress of concerns related to the "chilling effect" the fear of disclosure has on provider, supplier (including manufacturer), and health plan participation in quality measurement, improvement, and reporting programs. By shielding this information from discovery or admission as evidence without consent, Congress has explicitly protected and incentivized the activities of qualified entities, including the development of registries to support their performance measurement and reporting efforts.

2.2 State Laws

2.2.1 State Surveillance Laws

To encourage practitioner participation in State surveillance registries, many States have passed legislation providing immunity from civil and criminal penalties that may arise in conjunction with such reports. Most States protect providers from any liability that may result from a disease report unless the provider acted with some level of negligence or malicious intent.[62] Fewer States provide complete immunity for reporting disease cases to a registry, with no distinction made for negligent or intentionally malicious reports.[63] Additionally, other States provide immunity only for certain causes of action related to the information reported, or protect from civil liability only.[64] However, case law implicating such State harbors is sparse, with most cases affording immunity for health care professionals who report sexually transmitted diseases discovered in minors in potential child abuse cases.[65]

2.2.2 State Peer Review and Quality Assurance Laws

Presently, all 50 States have enacted statutes to protect the confidentiality of the peer review process. Most statutes offer a blanket protection for all accounts, records, and conclusions of the review process from being introduced into evidence during any court proceeding.[66] Without such protection, providers and hospitals may be less inclined to truthfully monitor their peers, evaluate the quality of care that is provided to patients, or adequately prevent or correct for adverse events.

A few States also have passed legislation specifically protecting information collected for quality improvement and other related purposes. These "safe harbor" laws protect a broader set of information beyond the traditional peer review process, including collection by organizations outside the scope of an internal peer review board or committee. For example, in Minnesota, information relating to patient care a nonprofit organization collects for purposes of "evaluating and improving the quality of health care" or "reviewing the safety, quality or cost of health care services" provided to health plan enrollees "shall not be disclosed to anyone … and shall not be subject to subpoena or discovery."[67] The Virginia Patient Safety Act protects the collection of patient safety data by "statewide or local associations" representing licensed health care providers. It treats the information collected as "privileged communications which may not be disclosed or obtained by legal discovery proceedings unless a circuit court, after a hearing and for good cause arising from extraordinary circumstances being shown, orders the disclosure of such proceedings, minutes, records, reports, or communications."[68] Similarly, the Illinois Medical Studies Act protects all information "used in the course of internal quality control or of medical study for the purpose of reducing morbidity or mortality, or for improving patient care or increasing organ and tissue donation." Such information is "privileged,

strictly confidential, and shall be used only for medical research, increasing organ and tissue donation, [or] the evaluation and improvement of quality care."[69] It is not "admissible as evidence, nor discoverable in any action of any kind in any court or before any tribunal, board, agency or person."[69]

However, it is important to note that the peer review privilege and other State-based protections are often not recognized in Federal cases outside the jurisdiction of State law. Federal courts have been resistant to the establishment of a Federal peer review or other privilege, and often subordinate the State peer review privileges in favor of other interests. For instance, the 11th Circuit Court of Appeals recently declined to recognize such a privilege during a Federal civil rights discrimination case, ordering the discovery of peer review documents.[70] Similar outcomes have been reached in many other Federal courts, including cases deciding Federal antitrust[71] and wrongful termination claims.[72]

2.3 Practical Considerations

As described above, Federal and State law currently do not provide any consistent or comprehensive protection from disclosure, pursuant to discovery or other judicial or administrative proceedings, of information submitted to registries by (or related to) providers, medical device or pharmaceutical manufacturers, or health plans. This leaves registries that are operating outside of the government-sponsored programs described above, their participants and subjects, vulnerable to discovery requests ranging from preliminary fact-finding requests to court orders. As the Institute of Medicine noted, this can have a "chilling effect" on willingness of providers to participate. The same is also true for manufacturers and health plans that similarly may be both a source of information as well as the subject of information included in a registry. Beyond concern for the privacy and confidentiality of the information, the costs and burden associated with discovery or other requests can be substantial, as the litigation process often takes months or years to unfold. These costs may include not only costs related to challenging the request, but also data production, including costs for redaction (particularly where patient identifiable information is involved), and the costs of legal representation.

Registries and their participants can take several steps to reduce their vulnerability to disclosure requests. As described above, several Federal programs protect information collected for specific patient safety and quality improvement purposes. (See Case Example 18.) If registries qualify for or are funded through these programs, the information collected and maintained would automatically be entitled to protection from discovery or other judicial or administrative proceedings. While not always possible or practical given their goals or priorities, registries should consider whether participation in any of these programs would be appropriate.

To the extent participation in one of these Federal programs is not possible, registries and their participants should consider forming in a State that provides broader protection, beyond the peer review process, for information collected (e.g., Virginia, Minnesota, or Illinois). Furthermore, registries and their participants should clearly articulate their roles and responsibilities, including how discovery or other requests will be handled. Registries should develop specific policies and procedures to guide their response to such requests and should ensure that all participants are familiar with the policies and procedures. (See Case Examples 17 and 19.) For example, registries might stipulate that they will direct all disclosure requests to the original source of the information where possible. Where information held within a registry has been aggregated and analyzed such that it is significantly modified from its original state, the registry will notify the original data sources prior to compliance with any discovery request and give them the opportunity to object.

In the event a registry is compelled to release information pursuant to a court order or other judicial or administrative order, registries may request that certain information be redacted or a protective order issued. A court protective order can stipulate who can see the information, who has access to the information, and how the data should be returned or destroyed. Similarly, a registry may request that the court "seal" information so that it

is not made public. These types of actions have historically been used to protect patient-identifiable information held in registries; however, they may be similarly applied to confidential or proprietary information related to providers, manufacturers, or health plans.

3. Summary

As more attention is focused on the development and implementation of quality improvement activities, including those tied to new payment models, availability and accessibility of underlying clinical and administrative data that registries can provide to support these efforts will be increasingly important. This emphasis will be further strengthened as the new Patient-Centered Outcomes Research Institute, which is authorized to support efforts to generate comparative effectiveness research, begins its work. Given this heightened interest in registries as data sources, the issue of protection of registry data from disclosure pursuant to a discovery request or other judicial or administrative proceedings will be increasingly important. Registry sponsors may be able to address concerns from potential participants about data protection by considering these issues during the registry development stage. In particular, providers, manufacturers, and health plans that are developing registries or considering participation in registries should look to the Federal and State laws described here that may offer protection and should consider the practical steps outlined above to reduce their vulnerability to disclosure requests.

Case Examples for Chapter 9

Case Example 17. Handling discovery requests for registry data

Description	The ICD Registry captures the characteristics, treatments, and outcomes of patients receiving implantable cardioverter defibrillators (ICDs). Participation in the registry is required by a coverage decision of the Centers for Medicare & Medicaid Services (CMS). In addition to CMS reporting, participating hospitals may use their data to monitor and improve the outcomes and management of ICD patients through implementation of evidence-based clinical guidelines.
Sponsor	American College of Cardiology Foundation (ACCF)
Year Started	2005
Year Ended	Ongoing
No. of Sites	Over 1,600 laboratories
No. of Patients	Over 800,000 procedures

Challenge

Before its launch, the ICD Registry received significant attention from the public and researchers because of the CMS coverage decision. The registry sponsors anticipated that the registry would generate interest from outside entities seeking to discover registry data, particularly those investigating potential fraud (e.g., the Office of the Inspector General), those desiring to use registry data in malpractice lawsuits (e.g., litigation attorneys), and those seeking to corroborate information published in the peer-reviewed literature (e.g., media representatives). Driven by these concerns, the sponsor and registry staff sought to address this anticipated issue proactively.

Proposed Solution

Registry staff implemented procedures to assist all parties involved in handling a discovery request. In coordination with general counsel and outside counsel, internal policies were drafted and provisions were written into contracts with participating sites that explained the way registry staff would handle discovery requests, the process of responding to a discovery request from an

Case Example 17. Handling discovery requests for registry data (continued)

Proposed Solution (continued)

attorney, the process of cooperating with the Office of Inspector General during an audit or investigation, and best practices for protecting registry data from discovery. This language prepared sites for these procedures should they ever occur, made them aware of their options in these situations (i.e., cooperate with or dispute the request), and clarified the level of protection that the registry offered for their data.

Standard operating procedures were implemented to train staff to recognize a discovery request and subpoena and to describe the actions that should be taken to appropriately triage and respond to discovery requests for registry data (e.g., that site support staff direct such requests to the registry compliance office). More in-depth staff training was provided, including role-play scenarios in which staff tested their skills and confidence through simulated discovery requests.

Results

Since 2005, the registry has received five different requests for registry data, from sources as varied as attorneys, the Office of the Inspector General, and members of the press. The requests are managed in a consistent, documented way, regardless of whether the registry first receives these requests via site support staff or other venues. The registry maintains a relationship with the ACCF general counsel so that questions about future requests can be promptly resolved.

Key Point

Registries can take proactive steps to manage discovery requests for their data. Appropriate steps may include confidentiality provisions in contracts with sites and targeted training for all levels of registry staff.

For More Information

https://www.ncdr.com/webncdr/ICD/Default.aspx

Case Example 18. Meeting the confidentiality and quality improvement needs of providers through a patient safety organization

Description	The Pediatric Peri-Operative Cardiac Arrest (POCA) Registry investigated the incidence, causes, and outcomes of cardiac arrest among children undergoing anesthesia. The Wake Up Safe (WuS) Initiative is a quality improvement initiative and registered patient safety organization (PSO) that aims to improve processes of care and outcomes for children undergoing anesthesia.
Sponsor	POCA: American Society of Anesthesiologists and the University of Washington; WuS: Society for Pediatric Anesthesia
Year Started	1994 (POCA) and 2008 (WuS)
Year Ended	2005 (POCA) and ongoing (WuS)
No. of Sites	58–79 (POCA) and 18 (WuS)
No. of Patients	374 events from >3.5 million anesthetics (POCA) and 518,000 anesthetic records (WuS)

Challenge

Children undergoing anesthesia have an increased risk of cardiac arrest compared with adults. Although factors associated with this increased risk had been identified, the causes of these arrests and their outcomes were not well understood. The American Society of Anesthesiologists and the University of Washington formed the POCA Registry in 1994 to collect detailed case reports of cardiac arrests during anesthesia in pediatric patients. Institutions submitted these case reports anonymously to protect the participating institutions and the registry from legal risks of disclosure.

In 2000, the registry analyzed the first 4 years of data (1994–1997) and found that the incidence of anesthesia-related cardiac arrest was 1.4 per 10,000 anesthetics, with a mortality rate of 26 percent. The most common causes were medication related and cardiovascular, with cardiovascular depression from halothane (a commonly used anesthetic in children) accounting for two-thirds of medication-related cardiac arrests. This suggested a target for preventive strategies, including avoidance of halothane when a new agent (sevoflurane) was available. In 2007, the registry published an update on its findings, comparing cardiac arrests from 1998 through 2004 with the cardiac arrests in the initial report. This report found fewer medication-related cardiac arrests (associated with a decline in use of halothane) and more arrests with undetermined causes.

Despite these promising findings, the pediatric anesthesiology community sought a more comprehensive approach to quality care improvement. The registry collected only data on pediatric cardiac arrest, and not on any other outcomes, processes, or quality of care indicators. In addition, because data was submitted anonymously to the registry, benchmarking at the institution level was not possible. There were also concerns about the protection of registry data. The registry was housed in the State of Washington, which had legal protections in place to protect quality improvement data from legal discovery. However, not all States had these same protections in place.

Proposed Solution

In 2005, the POCA Steering Committee decided to halt case collection, as compliance with reporting had declined. While the registry had contributed much to understanding of anesthesia-related cardiac arrest in children, it was felt that further data collection using the same methodology was unlikely to produce additional insights.

In 2007, the Society for Pediatric Anesthesia (SPA) began garnering support to form the Wake Up Safe Initiative, a new quality improvement initiative for pediatric anesthesiology. As institutions joined, they signed participation agreements with the SPA that were intended to

Case Example 18. Meeting the confidentiality and quality improvement needs of providers through a patient safety organization (continued)

Proposed Solution (continued)

address intellectual property issues, but evolved as each hospital had its own concerns about privacy, protection, and anonymity in reporting. By 2008, 10 institutions had signed on to participate in the initiative. The same year, the Patient Safety and Quality Improvement Act of 2005 came into effect, which provides Federal legal protections to information reported by providers to a patient safety organization (PSO). WuS applied for and was granted PSO status in 2008.

Results

The WuS initiative captures a broad range of outcomes, and individual institutions are not identifiable. Anesthetic records for pediatric patients are extracted from the administrative billing systems of member hospitals and provided to the registry on a quarterly basis. This information provides background data on which to base incidence calculations. Individual event case reports are provided to the registry on an ad hoc basis, as they occur. To date, 18 member hospitals have contributed 518,000 anesthetic records and 450 event case reports to the registry.

As a PSO, the registry provides more complete protections for the providers contributing data. These include limits on the use of registry data in civil, administrative, and some criminal proceedings, and provisions for monetary penalties for violations of confidentiality or privilege protections.

Key Point

Registries, particularly those that collect sensitive information on provider performance, should consider taking advantage of the legal protections that are available to patient registries. Official designation as a patient safety organization (PSO) offers broader Federal protections that individual States may not be able to offer. This may provide an incentive for participation, especially for registries that collect sensitive information on performance or quality of care.

For More Information

http://www.wakeupsafe.org/

Morray JP, Geiduschek JM, Ramamoorthy C, et al. Anesthesia-related cardiac arrest in children: initial findings of the Pediatric Perioperative Cardiac Arrest (POCA) Registry. Anesthesiology. 2000 Jul;93(1):6-14.

Posner KL, Geiduschek J, Haberkern CM, et al. Unexpected cardiac arrest among children during surgery, a North American registry to elucidate the incidence and causes of anesthesia related cardiac arrest. Qual Saf Health Care. 2002 Sep; 11(3):252-7.

Bhananker SM, Ramammoorthy C, Geiduschek JM, et al. Anesthesia-related cardiac arrest in children: update from the Pediatric Perioperative Cardiac Arrest Registry. Anesth Analg. 2007 Aug; 105(2):344-50.

Ramamoorthy C, Haberkern CM, Bhananker SM, et al. Anesthesia-related cardiac arrest in children with heart disease: data from the Pediatric Perioperative Cardiac Arrest (POCA) registry. Anesth Analg. 2010 May 1;110(5):1376-82.

Case Example 19. Protections available to registry data from institutional review boards and academic institutions

Description	The Postoperative Visual Loss (POVL) Registry consists of anonymous case reports of blindness after non-ophthalmologic surgery. Its goal is to identify patient and clinical factors associated with this complication.
Sponsor	American Society of Anesthesiologists
Year Started	1999
Year Ended	Ongoing
No. of Sites	Not applicable
No. of Patients	191

Challenge

Blindness after non-eye surgery is a rare but devastating complication. Blindness due to ischemic optic neuropathy after spine surgery appeared to be increasing in the 1990s, and its causation was unknown. Its rarity created difficulty in studying causative factors.

Proposed Solution

A registry was created to collect detailed case reports of blindness after non-ophthalmologic surgery, and included data pertinent to all known theories of causation of postoperative visual loss. Due to the potential for malpractice litigation when postoperative blindness occurs, case reports were submitted without patient, provider, or institutional identifiers. It was hoped that anonymity of case reports would protect the registry from legal discovery and encourage case report submission by health care providers. The registry was housed at the University of Washington, and its institutional review board (IRB) approved these confidentiality procedures.

Results

In spite of these procedures to protect confidentiality of case reports, plaintiff attorneys submitted numerous requests for release of registry data. Following university policy, all requests were referred to the office of public information. All such public information requests were denied based on the institutional review board–approved confidentiality procedures.

One public information request for registry data was appealed through the court system. A registry investigator was serving as a defense expert in a malpractice lawsuit, basing her testimony on published registry results. The plaintiff requested raw registry data. When the university denied this request, the plaintiff sought to strike the investigator's testimony because she would not produce the raw registry data underlying the publication that formed the basis of her testimony. The trial court determined that the raw registry data was discoverable because the defense expert considered such data in forming her opinion. The university supported an appeal to the State Supreme Court in order to support institutional protections of research data. The State Supreme Court ruled that the author could testify based on the published results and that release of the underlying data was not required.

Key Point

It is critical to consider protection of sensitive registry data from legal discovery. When developing and implementing registries, consider protections that may be available through IRB study approval and academic institutions. Seek guidance from IRBs or university counsel, as needed.

For More Information

http://www.cobar.org/opinions/opinion.cfm?opinionid=7861&courtid=2

Lee LA, Roth S, Posner, KL, et al. The American Society of Anesthesiologists Postoperative Visual Loss Registry: Analysis of 93 spine surgery cases with postoperative visual loss. Anesthesiology. 105:652-9, 2006

The Postoperative Visual Loss Study Group. Risk factors associated with ischemic optic neuropathy after spinal fusion surgery. Anesthesiology. 116:15-24, 2012.

References for Chapter 9

1. Fisher E, Goodman D, Skinner J, et al. Health Care Spending, Quality and Outcomes: More Isn't Always Better. Dartmouth Atlas Project Topics Brief. February 27, 2009. http://www.dartmouthatlas.org/downloads/reports/Spending_Brief_022709.pdf. Accessed August 15, 2012.
2. Fisher ES, Wennberg DE, Stukel TA, et al. The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. Ann Intern Med. 2003 Feb 18;138(4):273-87. PMID: 12585825.
3. Fisher ES, Wennberg DE, Stukel TA, et al. The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care. Ann Intern Med. 2003 Feb 18;138(4):288-98. PMID: 12585826.
4. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003 Jun 26;348(26):2635-45. PMID: 12826639.
5. Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. Health Aff (Millwood). 2002 Jul-Dec;Suppl Web Exclusives:W96-114. PMID: 12703563.
6. Wennberg JE, Cooper MM, Bubolz TA, et al. The Dartmouth Atlas of Health Care in the United States: American Hospital Publishing, Inc.; 1996. http://www.dartmouthatlas.org/downloads/atlases/96Atlas.pdf. Accessed August 15, 2012.
7. Wennberg JE, Brownlee S, Fischer ES, et al. An Agenda for Change: Improving Quality and Curbing Health Care Spending: Opportunities for the Congress and the Obama Administration. The Dartmouth Institute for Health Policy & Clinical Practice. December 2008. http://www.dartmouthatlas.org/downloads/reports/agenda_for_change.pdf. Accessed April 30, 2012.
8. de Brantes F, Rosenthal MB, Painter M. Building a bridge from fragmentation to accountability--the Prometheus Payment model. N Engl J Med. 2009 Sep 10;361(11):1033-6. PMID: 19692682.
9. Patient Protection and Affordable Care Act of 2010, Public Law No. 111-148.
10. American College of Chest Physicians. AQuIRE Registry. http://www.chestnet.org/Guidelines-and-Resources/Payment-Practice-and-Quality/Practice-Resources/AQuIRE. Accessed August 7, 2013.
11. American Orthopaedic Association. http://www.ownthebone.org/. Accessed August 15, 2012.
12. American College of Rheumatology. ACR Rheumatology Clinical Registry. http://www.rheumatology.org/Practice/Clinical/Rcr/Rheumatology_Clinical_Registry/. Accessed August 7, 2013.
13. American College of Radiology. National Radiology Data Registry. http://www.acr.org/Quality-Safety/National-Radiology-Data-Registry. Accessed August 7, 2013.
14. Society for Thoracic Surgeons National Database. http://www.sts.org/national-database. Accessed August 15, 2012.
15. National Cardiovascular Data Registry. https://www.ncdr.com/webncdr/. Accessed August 7, 2013.
16. 45 CFR pts. 160, 164 (2010).
17. Kesselheim AS, Ferris TG, Studdert DM. Will physician-level measures of clinical performance be used in medical malpractice litigation? JAMA. 2006 Apr 19;295(15):1831-4. PMID: 16622145.
18. Kohn LT, Corrigan JM, Donaldson MS, eds. To Err is Human: Building a Safer Health System. Washington, D.C.: National Academies Press; 1999.
19. 42 U.S.C. § 299c-3(c).
20. Memorandum from Susan Merewitz, Senior Attorney, AHRQ, on Statutory Confidentiality Protection of Research Data, to Nancy Foster, Coordinator for Quality Activities, AHRQ. April 16, 2001. http://www.ahrq.gov/fund/datamemo.htm. Accessed August 15, 2012.
21. See, National Institutes of Health, Certificates of Confidentiality: Background Information. http://grants.nih.gov/grants/policy/coc/background.htm. Accessed August 15, 2012.
22. 42 U.S.C. § 241(d).

23. Rothman K, Greenland S, Poole C, et al. Causation and Causal Inference. In: Rothman K, Greenland S, Lash TL, eds. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008, pp. 5-31.
24. 42 U.S.C. §§ 299b-21-26.
25. 42 U.S.C. § 1320c-9(b).
26. 42 U.S.C. § 1320c-9(d).
27. 42 CFR § 480.140.
28. Health Insurance Portability and Accountability Act of 1996, Public Law No. 104-191, 110 Stat. 139 (1996) (codified as amended in scattered sections of 42 U.S.C.). HIPAA Privacy Rule Regulations codified at 45 CFR pts. 160 & 164 (2010).
29. 45 CFR § 160.103 (2010).
30. 45 CFR § 164.512(e).
31. Privacy Act of 1974, Public Law No. 93-579, § 3, 88 Stat. 1896, 1896 (codified as amended at 5 U.S.C. § 552a (2006)).
32. Id. at 5 U.S.C. § 552a(b).
33. Electronic Freedom of Information Amendment Acts of 1996, Public Law No. 104-231, 110 Stat. 3048 (1996) (codified as amended at 5 U.S.C. § 552).
34. Freedom of Information Act, 5 U.S.C. § 552 (2006), amended by OPEN Government Act of 2007, Public Law No. 117-175.
35. U.S. Department of Justice. Freedom of Information Act Guide (2004). http://www.justice.gov/oip/introduc.htm#N_2_. Accessed August 15, 2012.
36. 5 U.S.C. § 552(b)(4).
37. 45 CFR § 5.65(a).
38. *Pub. Citizen Health Research Group* v. *Food & Drug Admin.*, 704 F.2d 1280 (D.C. Cir. 1983).
39. 704 F.2d at 1290.
40. 5 U.S.C. § 552(b)(6).
41. See *U.S. Dept of State* v. *Washington Post Co.*, 456 U.S. 595, 602 (1982).
42. *U.S. Dept of Justice* v. *Reporters Comm. for Freedom of the Press*, 489 U.S. 749 (1989).
43. 489 U.S. at 771.
44. 489 U.S. at 773.
45. 45 CFR § 5.67(c).
46. 5 U.S.C. § 552(b)(3).
47. 18 U.S.C. § 1905.
48. 76 C.J.S. Records §111 (2011).
49. Fed. R. Evid. 401.
50. Fed. R. Evid. 403.
51. Fed. R. Evid. 404(b).
52. Fed. R. Civ. P. 26(b)(2)(C)(iii).
53. Fed. R. Civ. P. 26(c).
54. Fed. R. Civ. P. 45(c).
55. *Andrews* v. *Eli Lily & Co., Inc.*, 97 F.R.D. 494 (1983).
56. 97 F.R.D. at 502.
57. *Deitchman* v. *E.R. Squibb & Sons, Inc.*, 740 F.2d 556 (7th Cir. 1984).
58. 740 F.2d at 561.
59. *Wolpin* v. *Phillip Morris, Inc.*, 189 F.RD. 418 (C.D. Cal. 1999).
60. Patient Protection and Affordable Care Act, Public Law 111-148, §10332 (2010).
61. Patient Protection and Affordable Care Act, Public Law 111-148, §10332(e)(4)(D) (2010).
62. See, e.g. Va. Code Ann. §32.1-38 (2011) (immunity from civil liability or criminal penalty unless such person acted with "gross negligence or malicious intent"); Ariz. Rev. Stat. Ann. §36-666 (2010) (immunity from civil or criminal liability if the provider acted in "good faith and without malice"); Iowa Code §139A.3 (2010) (immunity from any liability, civil or criminal, for any person "acting reasonably and in good faith" while reporting); Kan. Stat. Ann. §65-118 (2010) (immunity from any liability, civil or criminal, if provider reported "in good faith and without malice").
63. See, e.g. N.C. Gen Stat. §130A-142 (2010) (a provider who reports "shall be immune from any civil or criminal liability that might otherwise be incurred or imposed as a result of making that report").
64. See, e.g. Neb. Rev. Stat. §71-503.01 (2010) (immunity for providers from suits for slander, libel, or breach of privileged communication); Nev. Rev. Stat. §629.069 (2010) (immunity for providers from civil liability only).

65. See, e.g., *KB* v. *Mills*, 639 N.W.2d 261 (Mich. Ct. App. 2002); State v. Superior Court, 930 P.2d 488 (Ariz. Ct. App. 1997); *Alicia T.* v. *Cnty of L.A.*, 222 Cal App. 3d 869 (Cal. Ct. App. 1990); *Criswell* v. *Brentwood Hospital*, 551 N.E.2d 1315 (Ohio Ct. App. 1989).

66. See e.g., Idaho Code § 39-1392 (2011) ("all peer review records shall be confidential and privileged, and shall not be directly or indirectly subject to subpoena or discovery proceedings or be admitted as evidence, nor shall testimony relating thereto be admitted in evidence, or in any action of any kind in any court or before any administrative body, agency or person for any purpose whatsoever"); Ohio Rev. Code Ann. § 2305.252 (2011) ("proceedings and records within the scope of a peer review committee of a health care entity shall be held in confidence and shall not be subject to discovery or introduction in evidence in any civil action against a health care entity or health care provider…).

67. Minn. Code § 154.61 (a, g), § 145.64 (2010).

68. Va. Code Ann. § 8.01-581.17 (2011).

69. 735 Ill. Comp. Stat. 5/8-2101 (2010).

70. *Adkins* v. *Christie*, 488 F.3d 1324 (11th Cir. 2007).

71. *Marshall* v. *Spectrum Medical Group*, 198 F.R.D. 1 (D. Me. 2000).

72. *Price* v. *Howard County Gen. Hosps.*, 950 F. Supp. 141 (D. Md. 1996).

Section III

Operating Registries

Chapter 10. Recruiting and Retaining Participants in the Registry

1. Introduction

Recruitment and retention of participants are essential elements in the design and operation of a registry. Registries are often intended to be representative of a certain population of patients and reflective of the practices of certain providers and geographic areas. The problems commonly associated with clinical studies—such as difficulties with patient enrollment, losses to followup, and certain sites contributing the majority of patients—can also have profound consequences on validity of registry data. When registry patients are not representative of the target population, the value of the results is diminished. For example, in regard to policy determinations, the enrolled sites or providers must be representative of the types of sites and providers to which the policy determination would apply in order for the results of the registry to be generalizable. Differences in how effectively sites enroll or follow patients can skew results and overly reflect the sites with the most data. This oversampling within a particular site or location must also be considered in sample size calculations. If the sample size of a key unit of analysis (patient, provider, or institution) is not sufficient to detect a clinically important difference, the validity of the entire registry is weakened. (See Chapters 3 and 13.)

Well-planned strategies for enrollment and retention are critical to avoiding these biases that may threaten registry validity. Because registries typically operate with limited resources and with voluntary rather than mandatory participation, it is particularly important to balance the burdens and rewards of participation in the registry. The term "voluntary" in this context is intended to mean that participation in the registry by either providers or patients is not mandated (e.g., by the U.S. Food and Drug Administration), nor is participation required as a necessary condition for a patient to gain access to a health care product or for a provider to be eligible for payment for a health care service.

Registries that are not voluntary have different drivers for participation. In general, the burden of participation should be kept as low as possible, while the relative rewards, particularly nonmonetary rewards, should be maximized. As described in Chapters 2 and 4, minimizing burden typically starts with focusing on the key goals of the registry.

Building participation incentives into a registry should also be included in the planning phase. A broad range of incentives—spanning a spectrum from participation in a community of researchers, to access to useful data or quality improvement benefits, to continuing medical education, to public recognition or certification, to payments or access to patients—have been used in registries. The ability to offer certain incentives (e.g., linking payment for a service to participation in a registry or access to patients) may be available only to certain registry developers (e.g., payers, licensing entities). Many registries incorporate multiple types of incentives, even when they pay for participation. Monetary incentives (e.g., from payers or sponsors) are very helpful in recruiting sites. However, because the payments should not exceed fair market value for work performed, registries cannot solely rely on these incentives. A number of nonmandated registries have achieved success in recruitment and retention by providing a combination of ethical incentives that are tailored to and aligned with the specific groups of sites, providers, and patients that are asked to participate. (See Case Examples 20, 21, 22, and 23.)

2. Recruitment

Depending on the purpose of a registry, recruitment may occur at any of three levels: facility (e.g., hospital, practice, and pharmacy), provider, or patient. While recruitment at these levels is frequently part of a design to accrue a sufficient number of patients for sample size purposes, such as for a safety registry, the individual levels may also constitute potential units

of analysis (and as such, may further affect sample size, as discussed in Chapter 3, Section 8). As an example, a registry focused on systems of care that is examining both hospital system processes and patient outcomes might need to consider characteristics of the individual patients, the providers, and/or the places where they practice (i.e., clusters). If the question is about the practices of orthopedic surgeons in the United States, the registry will be strengthened by describing the number and characteristics (e.g., age, gender, and geographic distribution) of U.S. orthopedic surgeons, perhaps by citing membership data from the American Academy of Orthopedic Surgeons. This will allow documentation of the similarities and differences in the characteristics of the surgeons participating in the registry compared with the target population. (See Chapter 3, Section 7.)

2.1 Hospital Recruitment

A hospital or health system may choose to participate in a patient registry for many reasons, including the research interest of a particular investigator or champion, the ability of the hospital to achieve other goals through the registry (such as requirements for reimbursement, certification, or recognition), or the general interest of the particular institution in the disease area (e.g., specialty hospitals). Increasingly, external mandates to document compliance with practice standards provide an incentive for hospitals to participate in registries that collect and report mandatory hospital performance or quality-of-care data. For example, a number of registries allow hospitals to document their performance to meet the Joint Commission requirements for hospital accreditation.[1] Hospitals in the United States must submit these data to maintain accreditation. Therefore, hospital administrators may be willing to supply the staff time to collect these data without additional financial incentives from the registry sponsor, provided that registry participation allows the hospital to meet external quality-of-care mandates. In other cases, participation in a quality monitoring or health system surveillance registry may be required by payers or governments for reimbursement, differential payments, or patient referrals under various programs, ranging from the Centers for Medicare & Medicaid Services (CMS) public reporting initiative, to centers of excellence programs, to pay-for-performance programs. One particular example, CMS's Coverage With Evidence Development programs,[2] which may require participation in a registry for the center or provider to qualify for payment for a procedure, can have a dramatic impact on registry participation. Registry participation requirements have existed for implantable cardioverter defibrillators for preventing sudden cardiac death in heart failure, for bariatric surgery, for positron emission tomography scan use in cancer, and for other procedures and devices. These requirements have rapidly resulted in high participation rates for registries meeting them.

The presence of quality assurance departments in U.S. hospitals provides an infrastructure for participation in many hospital-based registries, and these departments are therefore a natural target for recruiting. However, hospital size, service line (e.g., disease-specific centers), and competing activities may limit institutional interest. The American Hospital Association database provides a valuable resource for identifying hospitals by key characteristics, including hospital ownership, number of beds, and the presence of an intensive care unit.

Table 10–1 summarizes the key factors for successful hospital recruitment, and lists specific methods that might be used to recruit hospitals. While programs need not incorporate all of these characteristics or use all of these methods, successful programs typically incorporate several.

Table 10–1. Hospital recruitment	
Keys to hospital recruitment	• The condition being studied satisfies one of the hospital's quality assurance mandates. Sufficient funds, data, or other benefits will be realized to justify the effort required to participate. • The confidentiality of the hospital's performance data is ensured, except to the extent that the hospital elects to report it. • Clinically relevant, credible, timely, and actionable self-assessment data—ideally, data that are risk adjusted and benchmarked—are provided back to the hospital to help it identify opportunities for enhancing patient care outcomes. • High-profile hospitals (regional or national) are participating in the registry. • Burden is minimized. • Participation assists the hospital in meeting coverage and reimbursement mandates, gaining recognition as a center of excellence, or meeting requirements for pay-for-performance initiatives.
Methods of hospital recruitment	• Identify eligible hospitals from the American Hospital Association database. • Use stakeholder representatives to identify potentially interested hospitals. • Enroll hospitals through physicians who work there and are interested in the registry. • Use invitation letters or calls to directors of quality assurance or the chief of the clinical department responsible for the condition targeted by the registry. • Ask physician members of an advisory board (if applicable) to network with their colleagues in other hospitals. • Reach out to physicians or hospital administrators through relevant professional societies or hospital associations. • Leverage mandates by external stakeholders, including third-party payers, health plans, or government agencies.

2.2 Physician Recruitment

A physician practice may or may not choose to participate in a voluntary registry for many reasons. As with hospitals, these reasons can include the research interests of the physician and the ability of the practice to achieve other goals through the registry (such as reimbursement or recognition). When deciding to participate, physicians often focus on several concerns:

- *Relevance*: Does the registry have meaning for the practice and patients?
- *Trust*: Are the registry leaders credible? Are the goals clearly stated?
- *Risks*: Will confidentiality be maintained? Are patient records secure?
- *Effort*: Will the amount of effort expended be fairly compensated?
- *Disruption*: Will participation disrupt workflow of the staff?
- *Value*: What benefits will be derived from participation? Will it improve the care provided? Will it enhance the evidence base for future practice?

Physicians who manage only a few patients per year with the disease that is the subject of the registry are less likely to be interested in enrolling their patients than physicians who see many such patients—unless the disease is rare or extremely rare, in which case the registry may be of great interest.

Because most registries are voluntary and physicians in nonacademic practice settings may have less infrastructure and staff available to enroll their patients, recruitment of representative physicians is a major challenge for registries that aim to compare physician practices across a full

spectrum of practice settings. In general, community-based physicians are less well equipped than hospital-based or academic physicians to collect data for research studies because they work in busy practices geared to routine clinical care rather than research. To increase recruitment of nonacademic physicians, it can be helpful to clearly explain the purpose and objectives of the registry; how registry data will be used; and, specifically, that individual results will not be shared (except at the direction of the physician) or published, and that registry outcomes data will be released only in large aggregates that protect the identities of individual hospitals, physicians, and patients. In addition, any incentives should be clearly articulated.

Table 10–2 describes the key factors for successful physician recruitment and lists several methods that might be used for recruiting physicians.

Table 10–2. Physician recruitment

Keys to physician recruitment	• The condition being studied is part of the physician's specialty. • The registry is a valuable scientific endeavor. • The registry is led by respected physician opinion leaders. • The registry is endorsed by leading medical, government, or patient advocacy organization(s). • The effort needed to recruit patients and collect and submit data is perceived as reasonable. • Useful practice pattern and/or outcome data are provided. • The registry meets other physician data needs, such as maintenance of certification requirements, credentialing requirements, or quality-based, differential, reimbursement payment programs (pay-for-performance).
Methods of physician recruitment	• Purchase mailing lists from physician specialty organizations. • Ask opinion leaders in the field to suggest interested colleagues. • Partner with local and national medical societies or large physician hospital organizations. • Use stakeholder representatives to identify interested physicians. • Recruit and raise awareness at conferences. • Advertise using email and the Web. • Register in the Registry of Patient Registries (RoPR) to increase awareness. • Leverage practice-based research networks.

2.3 Vetting Potential Hospital and Physician Participants

Once potential hospital or physician participants have been identified, it is important to vet them to ensure that the registry is gathering the appropriate mix of data. Issues to consider when vetting potential participants include—

- Representativeness
- Hospital characteristics (e.g., bed size, geographic location)
- Physician characteristics (e.g., specialty training)
- Practice setting (health maintenance organization [HMO], private practice)
- Ability to recruit patients
- Volume of target cases
- Internal resources
- Availability of a study coordinator
- Availability of Internet connectivity for studies with electronic data capture
- Prior performance, including reliability and accuracy of data entry

2.4 Patient Recruitment

Patients may be recruited based on the judgment of the physician who provides their care; the diagnosis of a disease; receipt of a procedure, operation, device, or pharmaceutical; membership in a health insurance plan; or membership in a group of individuals who have a particular exposure. Recruitment of patients by the physician who is providing their care is one of the most successful strategies. The direct involvement in and support of the registry by their personal physicians is an important factor for patients.

Since registries should not modify the usual care that physicians provide to their patients, there should be little or no conflict between their role of physician and that of participant in the registry. (See Chapter 7.) In addition, patients may see participation in the registry as an opportunity to increase their communication with their clinician. Another incentive for many patients is the feeling that they are contributing to the knowledge base of sometimes poorly understood and undertreated conditions.

Recruitment of patients presents different challenges, depending on the nature of the condition being studied. In general, patient recruitment plans should address the following questions:

- Does the plan understand the needs and interests of potential participants?
- Does the plan address patient recruitment issues and procedural challenges, including informed consent and explanation of risks?
- What are the patient retention goals? What is a reasonable followup period? What is a reasonable followup rate? When does reduced retention compromise validity?
- What, if any, patient incentives are offered, including different types of incentives and the ethical, legal, or study validity issues to be considered with patient incentives?
- What are the costs of patient recruitment and retention?

Table 10–3 summarizes the key factors for patient recruitment and lists several specific methods that might be used for recruiting patients, grouped by the basic categories of patients at the time of recruitment.

Table 10–3. Patient recruitment	
Keys to patient recruitment	• Recruit through a physician who is caring for the patient. • Communicate to the patient that registry participation may help to improve care for all future patients with the target condition. • Write all patient materials (brochures, consent forms) in a manner that is easily understandable by the lay public. • Keep the survey forms short and simple. • Provide incentives. These can be nonmonetary, such as functions relevant to the patient's care (reports) or community (newsletters, portals). In some cases, monetary incentives can be offered if approved by the institutional review board. • Actively plan how to include minorities or other populations of interest.
Methods of patient recruitment	• Noninstitutionalized residents of the general U.S. population: – Recruit via letter survey, telephone, or email. – Recruit during well-patient visits to outpatient clinics. – Recruit via patient advocacy and support groups, health information Web sites, etc. – Register in the Registry of Patient Registries (RoPR) to increase awareness. • Outpatients attending the clinic of a physician who is participating in the registry: – Recruit through the patient's physician. – Recruit via brochures placed in physician's office. • Hospital inpatients who are hospitalized for treatment of a condition that is the subject of the registry: – Recruit through the patient's physician. – Recruit through hospitalists or consultant specialists. – Recruit through a hospital research coordinator. • Residents of nursing homes and similar long-term care facilities: – Establish a relationship with the nursing home and staff.

2.5 Partnerships To Facilitate Recruitment

Many agencies/organizations can assist in the recruitment of physicians and patients. These partners may have access to patients or their families and physicians who treat the condition, and they may lend credibility to the effort. These agencies/organizations include—

- Government agencies
- Physician professional associations or State medical associations
- Certifying boards (e.g., American Board of Neurological Surgeons)
- Patient advocacy groups (e.g., Muscular Dystrophy Association)
- Nonprofit foundations (e.g., Robert Wood Johnson Foundation)
- Industry (e.g., pharmaceutical companies)
- HMOs and other third-party insurance providers

2.6 Procedural Considerations Related To Recruitment

When developing a recruitment plan for a registry, consideration should be given to the procedural concerns that may be factored into potential participants' decisions. These concerns include the roles and responsibilities of each party, the need and process for obtaining institutional review board (IRB) approval, and the management of patient and provider confidentiality.

The contract between registry sites and the sponsor or coordinating center should clearly state the roles and responsibilities of the participants, the registry-coordinating center, and the sponsor. If

remuneration is being offered, the data-entry requirements that need to be fulfilled before payments are made should be stated. It is often helpful to explain to sites the concept of fair market value. Where the Health Insurance Portability and Accountability Act (HIPAA) applies, consideration should also be given to new requirements in the HIPAA Privacy Rule that prohibit a covered entity from disclosing protected health information in exchange for remuneration from the recipient of that information without the individual's authorization, subject to a number of exceptions (see 45 CFR 164.502(a)(5)(ii)). There is no specific formula (such as whether to separate startup payments from per-patient payments), but total remuneration must reflect work effort for the specific registry. Some individual factors, ranging from location to specialty, may have a bearing on fair market value. It is also important to spell out which entity will have ownership of the data and how the data will be used.

The contract should clearly explain the registry policy regarding any necessary approvals. If review by an IRB is required, generic templates can be offered to participants to assist them in obtaining ethical and IRB approval. Because the costs of obtaining IRB approval are often substantial, it is essential that the contract with the participants clearly indicate which party is responsible for bearing this cost. If the registry developer believes that IRB or privacy board review or approval is not required or may be waived, then a clear rationale should be provided to the prospective participants (see Case Example 57). As discussed in Chapter 7, the research purpose of the registry, the type of entity that creates and maintains it, whether the Common Rule applies to the particular site, and the extent to which the data are individually identifiable largely determine which regulatory requirements apply. For example, for registries limited to certain purposes, such as quality improvement, institutions may not need IRB approval.[3]

Patient privacy and participant confidentiality should be addressed in the registry materials. Methods of ensuring patient privacy need to be clearly elucidated in all registry-related documentation. Case report forms and patient logs must be designed to minimize patient identification (such as by transmitting limited data sets rather than more identifiable information, if such information is not required to meet a registry objective).

The intended management of the confidentiality of participating providers should be explained in the contract. Mechanisms for protecting provider confidentiality, including Certificates of Confidentiality and Patient Safety Organizations, are discussed in Chapters 7 and 9. If third-party or public reporting is an intended component of the registry, the specific data to be shared, the level of the disclosure (e.g., hospital and/or physician level), and the permitted receiving entities need to be articulated and the control mechanisms explained.

3. Retention

3.1 Providers

Once hospitals and physicians are recruited to participate in a registry, retaining them becomes a key to success. All of the factors identified as important for recruitment are important for retention as well. A critical factor in retention is delivery on promises made during recruitment (e.g., that the burden of participation is low). By carefully pilot testing all aspects of the registry prior to full recruitment, there is less likelihood that problems will arise that threaten the registry's reputation. Registries with an advisory board or steering committee can use this resource to help with retention. A visible and independent advisory board adds transparency and credibility, sets appropriate expectations among its peers on what to expect from a registry (e.g., compared with a clinical trial), ensures that the burden of the registry is minimized (or at least never outweighs its value to participants), and maintains the relevance and currency of the registry for the investigators. Ideally, advisory board members serve as ambassadors for the program. The level of credibility, engagement, practicality, and enthusiasm of the advisory board can significantly affect provider recruitment and retention. For

example, an advisory board whose clinical members are not themselves participating in the registry will have greater difficulty than a board with participating members in addressing the concerns of participating practices that invariably arise over the course of the registry. Including patients or patient advocacy organization representatives on the advisory board also can support patient retention efforts. These representatives can provide feedback to the board on patient issues or concerns about the registry, and they can facilitate communication about the registry's purpose and value to their peers or members.

Throughout the registry's duration, communication from the data coordinating center and the advisors, as well as community building, are important for strong retention. Early and continued engagement of the site champions or principal investigators is very important. Some registries use periodic face-to-face meetings of principal investigators from participating sites. When this approach is not economically feasible, well-planned online meetings can serve the same purpose.

Visibility of the registry at relevant national meetings can help maintain clinician awareness and sense of community, and regular demonstration of its value through presentations and publications reinforces the credibility of the registry to its participants. As the data set grows, so too does the value of the registry for all participants, and regular updates on the registry growth can be important. Finally, enhancing site value through nonfinancial rewards can be particularly useful in retention, and the registry should continually seek to bring value to the participants in creative and useful ways.

Participation retention tools include—

- Web sites
- Newsletters
- Telephone helplines
- Instruction manuals
- Training meetings
- Site audit/retraining visits
- Customer satisfaction/opinion surveys
- Regular data reports to stakeholders
- Presentations at conferences
- Regular reports to registry participants on registry growth and publications
- Ability of participating physicians to publish based on registry data (depending on the data access policy of the registry)

3.2 Patients

Retaining patients as active participants in registries with longitudinal followup is an ongoing challenge. Many factors need to be considered in developing a retention plan, including how long the patient is likely to return to the enrolling site. Patients enrolled in a primary care practice for a chronic illness can likely be followed in that practice for some time, although there should be a plan for how the registry will (or will not) address the issue of patients who transfer to unenrolled practices. Different solutions are needed for patients enrolled in a hospital at discharge or through a specialist who does not follow the patient long term. Several options exist. They include enlisting site staff to reach out to patients beyond their standard interactions,[4] following patients directly through a central patient management center,[5] and linking to other data sources (e.g., National Death Index, claims data) to obtain key long-term outcomes data on patients who are lost to followup. Retention plans, including contingencies, should be considered during registry planning, as they may require additional permissions (e.g., for direct contact) or data elements (e.g., for linkage). Maintaining ethical incentives for patient participation (ranging from newsletters to payments) is important for some registries (e.g., those that collect patient-reported outcomes data). Beyond planning for how to retain patients in a registry, it is important to track actual versus expected followup rates over time and to respond if rates are not meeting expectations. The resources available for patient retention efforts should also be clear. Followup rates can often be improved with more efforts, such as more attempts to contact the patient, but these efforts add costs and, at some level, will yield diminishing returns.

4. Pitfalls in Recruitment and Retention

Pitfalls abound in recruitment and retention. The most important of these pitfalls is the risk of selection bias. Targeting hospital-based or academic physicians to the exclusion of community-based physicians is tempting because the former are often more accessible and are frequently more open to involvement in, and more experienced in, research projects. Similarly, targeting high-volume practices or centers will improve efficiency of patient enrollment, but may not yield an adequately representative sample of care practices. If an advisory board or committee is used to help design the registry and aid in recruitment, there may be a tendency for advisors to recruit known colleagues or to target disease experts, when a wider range of participants may be necessary to provide the appropriate data to meet the research goals. Including representatives from the range of anticipated site types on the advisory board can be helpful.

Even with an appropriate mix of physician participants in a registry, biases in patient recruitment may still occur. For example, older and more seriously ill patients may be excluded because of challenges in enrollment and followup or poorer outcomes. From the outset, physicians involved in recruitment efforts need to be aware of the potential for bias, and they must understand the importance of adhering to well-delineated inclusion and exclusion criteria. They must also adhere to the registry's enrollment strategy, which is typically designed to reduce this bias (e.g., consecutive or randomized enrollment). In addition, overly demanding data collection requirements can affect retention. The schedule should be designed to obtain relevant data in a timely fashion without overtaxing the resources of patients and providers. It is also important to consider approaches that will distinguish patients who are lost to followup from those who have missing data for other reasons (such as a patient who missed a visit but is still in the registry).

Another major pitfall is confusing terminology. This can be a major problem when the registry is international. When designing training materials, instruction manuals, and questionnaires, it is critical that the language and terminology be clear and concise. Materials that are translated into other languages must undergo strict quality assurance measures to ensure that terms are translated properly (e.g., back translation).

5. International Considerations

While many general principles are similar for participant enrollment and retention in other parts of the world, there are many different customs or regulations regarding contract language, requirements for ethics committee or other submissions, informed consent, and allowable approaches to patient retention. Registries that extend to other countries should consult national and local regulations in those countries.

Case Examples for Chapter 10

Case Example 20. Building value as a means to recruit hospitals

Description	Get With The Guidelines® is the flagship program for in-hospital quality improvement of the American Heart Association (AHA) and American Stroke Association (ASA). The program uses the experience of the AHA and ASA to ensure that the care that hospitals provide for heart failure, stroke, and resuscitation is aligned with the latest evidence-based guidelines.
Sponsor	American Heart Association and American Stroke Association
Year Started	2000
Year Ended	Ongoing
No. of Sites	3,150
No. of Patients	3,174,462 patients and 4,253,461 patient records in Get With The Guidelines-Inpatient (Stroke, HF, and Resuscitation)

Challenge

Recruiting hospitals for registries or quality improvement (QI) programs can be arduous. Human and financial capital is constrained. Accreditation and reimbursement programs, such as those of The Joint Commission (formerly the Joint Commission on Accreditation of Healthcare Organizations, or JCAHO) and Centers for Medicare & Medicaid Services (CMS), contend for the same valuable human and financial capital. As a result, in the absence of specific benefits, many hospitals defer the data collection and report utilization required for successful QI execution.

Like most registries and QI programs, the sponsor's program faced barriers to data entry. Unlike other registries, Get With The Guidelines offered no reimbursements for data entry and entered a market characterized by significant competition. The registry team wanted to motivate resource-strapped hospitals to consistently and proactively enter data and analyze improvement.

Proposed Solution

The registry team began by listening to the hospitals through in-depth interviews designed to understand the motivations and deterrents underlying behavior. Interviews were conducted with hospital decisionmakers at all levels (nurses, QI professionals, administrators/chief executive officers, and physicians).

Based on the research findings, the team developed strategies that differentiated and built value for the program. Some of the more noteworthy strategies included the following:

- Systems were designed to allow data transmission from and to Joint Commission and CMS vendors, enabling hospitals to reduce the burden of duplicate data entry while still participating in other programs.
- A new tagline, Turning Guidelines into LifelinesSM, linked the brand's value proposition to the brand name and logo. Key messages for each target audience were included in marketing communications.
- A newly designed national recognition program motivated participation and advancement, and received the attention of hospital decisionmakers.
- Return-on-investment studies for the program demonstrated the value of participation.

Product innovations/enhancements created additional incentives to participate. Immediate point-of-care flags highlighted variances from guidelines. Benchmarking filters/reports empowered decisionmakers to benchmark performance with national averages and data from similar institutions. Customizable notes explaining diseases, tests, and medications can be sent to both the referring physician and the patient.

Case Example 20. Building value as a means to recruit hospitals (continued)

Results

By providing a mix of innovative nonfinancial incentives, the program increased both enrollment and advancement by about one-third in 12 months. Currently, 3,150 hospitals participate in the program. The database includes 4,253,461 patient records and is considered by many to be the most robust database for heart failure, stroke, and resuscitation. In 2004, the program received the Innovation in Prevention Award from the Department of Health and Human Services.

Key Point

Nonfinancial incentives that meet the needs of decisionmakers can assist in recruitment of sites. When creating such incentives, consider both tangible and nontangible benefits.

For More Information

http://www.heart.org/quality

Case Example 21. Using registry tools to recruit sites

Description	The objective of the OPTIMIZE-HF (Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure) registry was to improve quality of care and promote evidence-based therapies in heart failure. The registry provided a comprehensive process-of-care improvement program and gathered data that allowed hospitals to track their improvement over time.
Sponsor	GlaxoSmithKline
Year Started	2003
Year Ended	2005
No. of Sites	270
No. of Patients	More than 50,000

Challenge

The registry was designed to help hospitals improve care for patients hospitalized with heart failure. The objective was to accelerate the adoption of evidence-based guidelines and increase the use of the guideline-recommended therapies, thereby improving both short-term and long-term clinical outcomes for heart failure patients.

Proposed Solution

To increase compliance with guidelines, the registry team promoted the implementation of a process-of-care improvement component and the use of comprehensive patient education materials. They combined these materials into a hospital toolkit, which included evidence-based practice algorithms, critical pathways, standardized orders, discharge checklists, pocket cards, and chart stickers. The toolkit also included algorithms and dosing guides for the guideline-recommended therapies and a comprehensive set of patient education materials. The team engaged the steering committee in designing the toolkit to ensure that the materials reflected both the guideline-recommended interventions and the practical aspects of hospital processes.

In addition to the toolkit, the registry offered point-of-care tools, such as referral notes and patient letters, that could be customized for each patient based on data entered into the registry. The registry also included real-time performance reports that hospitals could use to assess their improvement on a set of standardized measures based on the guidelines.

Results

The hospital toolkit was a key component of the registry's marketing campaign. Hospitals could view the toolkit at recruitment meetings, but they did not receive their own copy until they joined the program. The toolkit gained credibility among

Case Example 21. Using registry tools to recruit sites (continued)

Results (continued)

hospitals because its creators included some of the most prominent members of the heart failure research and treatment community. Hospitals also actively used the reports to track their improvement over time and identify areas for additional work. Overall, the registry recruited 270 hospitals and met its patient accrual goal six months ahead of schedule.

Key Point

Nonfinancial incentives, such as patient education materials, toolkits, and reports, can encourage sites to join a registry. Incentives that also add value for the site by improving their processes or providing materials that they use frequently can aid retention.

For More Information

Fonarow GG, Abraham WT, Albert NM, et al. Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF): rationale and design. Am Heart J. 2004;148(1):43–51.

Gheorghiade M, She L, Abraham WT. et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. JAMA. 2006;296:2217–26.

Fonarow GC, Abraham WT, Albert NM. et al. Association between performance measures and clinical outcomes for patients hospitalized with heart failure. JAMA. 2007;297:61–70.

Case Example 22. Using a scientific advisory board to support investigator research projects

Description	The National LymphoCare Study is a large, prospective, disease-based registry in the area of follicular lymphoma in the United States. There are a number of open clinical questions about follicular lymphoma treatment, including whether anthracyclines should be used early in the course of disease, and whether there is a group of patients for whom observation (as opposed to active treatment) is the best choice, given the indolent nature of the disease. The registry follows patients for up to 10 years, and specific outcomes of interest include overall response rate, progression-free survival, time to subsequent therapy, and overall survival for common frontline and subsequent therapeutic strategies.
Sponsor	Genentech, Inc., and Biogen Idec, Inc.
Year Started	2004
Year Ended	Ongoing
No. of Sites	250 community and academic sites
No. of Patients	Over 2,700

Challenge

The National LymphoCare Study includes a large number of community-based sites in addition to many academic sites. Many of the principal investigators at the community-based sites are interested in using the registry data to answer clinical questions, but they do not have sufficient research experience to design a research question, conduct data analysis, and share the results with the scientific community. One aim of the registry sponsors and scientific advisory board (SAB) is to facilitate research among the community investigators, both to increase interest in the registry and to increase the scope of research questions addressed using registry data.

Proposed Solution

The registry sponsors and the SAB developed a plan to allow investigators at enrolling sites to propose a question of interest; work with an SAB member, clinical scientists, epidemiologists, and biostatisticians to develop an analysis plan to answer the question; and present findings at scientific meetings. The plan was implemented in 2007, when the registry issued a call for research proposals to all participating investigators. The proposal outlined the types of data that were available at that point (e.g., descriptive data on demographics, initial treatments, etc.). Several community-based investigators sent in proposals, which the SAB then reviewed. The SAB selected the proposals that it felt were most appropriate for the available data and that answered the most valuable questions from a clinical standpoint.

The community investigator for each selected proposal was then paired with a member of the SAB to further develop the research question. This process included conference calls and emails to refine the question and the high-level analytic plan. Once the plan was ready, the investigator and the SAB member submitted the proposal and analytic plan to the registry sponsor. The sponsor provided support for analytic design and biostatistics. The investigator, in consultation with the SAB member, developed an abstract based on the results. Abstracts were reviewed by the full SAB before being submitted for presentation.

Results

In 2007, a community-based investigator project developed through this process was accepted for abstract presentation at the annual American Society of Hematology (ASH) meeting. In 2009, a community-based investigator and a fellow at an academic institution developed abstracts that have been submitted for presentation at the annual ASH meeting.

Case Example 22. Using a scientific advisory board to support investigator research projects (continued)

Results (continued)

With outcomes data now available in the registry, registry sponsors plan to issue calls for proposals twice per year, with the goal of generating abstracts for the annual ASH meeting and the annual American Society of Clinical Oncology meeting. To date, the research program has been well received by community-based investigators, who have the opportunity to author their own research projects with mentoring from an experienced advisor. The SAB has also been enthusiastic about working with community-based physicians on research methodology and adding to the scientific knowledge about this disease.

Key Point

Community-based investigators who participate in a registry may be interested in pursuing research opportunities but may not have all of the necessary resources or expertise. By utilizing an engaged advisory board, a registry can provide investigators with research opportunities, resulting in more publications and presentations based on registry data, and potentially more engaged investigators.

For More Information

Friedberg JW, Taylor MD, Cerhan JR. et al. Follicular lymphoma in the United States: first report of the National LymphoCare Study. J Clin Oncol. 2009;27:1202–8.

Friedberg JW, Wong EK, Taylor MD, et al. Characteristics of patients with stage I follicular lymphoma (FL) selected for watchful waiting (WW) in the US: report from the National LymphoCare Study (NLCS). American Society of Hematology; 2007. Abstract 3315.

Link BK, Taylor MD, Brooks JM, et al. Correlates of treatment intensity for initial management of follicular lymphoma (FL) in the United States: report from the National LymphoCare Study (NLCS). American Society of Hematology; 2007. Abstract 2612.

Matasar MJ, Saxena R, Wong EK, et al. Practice patterns in the diagnosis of follicular lymphoma (FL): report from the National LymphoCare Study (NLCS). American Society of Hematology; 2007. Abstract 2613.

Nabhan C, Morawa E, Bitran JD, et al. Patterns of care in follicular lymphoma (FL): are minorities being treated differently? Report from the National LymphoCare Study (NLCS). American Society of Hematology; 2007. Abstract 367.

Case Example 23. Identifying and addressing recruitment and retention barriers in an ongoing registry

Description	Tracking Operations and Outcomes for Plastic Surgeons (TOPS) is a national registry of plastic surgery procedures and outcomes used to track and assess 30-day post-operative outcomes.
Sponsor	American Society of Plastic Surgeons (ASPS)
Year Started	2002
Year Ended	Ongoing
No. of Sites	425–450 annually
No. of Patients	611,682 complete cases and 1,094,268 plastic surgery procedures

Challenge

TOPS was initially launched in 2002 as a program designed to provide ASPS member plastic surgeons with a mechanism to track demographic, procedural, and outcomes information to help physicians benchmark and evaluate their practice. The registry uses an electronic data capture interface to collect common demographic, risk factor, procedural, and 30-day outcome data elements, which allow registry users to evaluate outcomes based on patient comorbidities and risk factors and track the rate and type of surgical incidences that could occur postoperatively. The registry has become an integral part of ASPS efforts and is used in many of the society's key initiatives, including developing evidence-based practice parameters, monitoring clinical outcomes and emerging trends, supporting research and educational programs, and compilation of the National Clearinghouse Plastic Surgery Statistics.

A majority of the registry participants contribute data annually. However, due to the voluntary nature of the registry, the registry also includes "one-time" users. Since 2002, more than 1,600 ASPS members have participated in the registry at any given time, leaving a large number of ASPS members choosing not to participate at all, and some ASPS members who choose to stop participating or participate sporadically. In order to encourage broad and continued participation from society members, the society sought to fully understand the needs of its members and develop strategies to retain and recruit participants for the ongoing registry.

Proposed Solution

To improve the program's value to member surgeons, the ASPS regularly evaluates member and organization needs to make upgrades and improve the registry. The ASPS surveyed its society membership in 2008 and 2012 to better understand reasons for both continued participation and lack of use. Survey results indicated that earning credits towards continuing medical education, society-awarded patient safety credits, and contributing to important scientific endeavors were strong reasons for continued participation. In addition, individual physician practices perceived value in the registry's ability to track patients' outcomes over time and benchmark practice data against other practices in the registry. Many nonusers reported their willingness to participate if the registry supported integration with their practice's electronic medical records (EMR) system. Other reasons for nonparticipation included perceived limited usefulness of data collection to physician practice and site-based barriers such as insufficient resources.

Results

Based on the survey feedback, the ASPS elected to increase functionality to make the registry more relevant to members. Updates to the registry were made in 2007, 2009, and 2011, and included the introduction of credentialing reports, enhanced benchmarking reports, collection of patient-reported outcomes (PRO), and enhanced data entry and querying reporting functions. The credentialing reports allow registry users to review their individual patient data by facility, with options to filter by medical record number, name, date range, procedure type, and outcome. The existing benchmarking reports were enhanced, allowing users to benchmark their

Case Example 23. Identifying and addressing recruitment and retention barriers in an ongoing registry (continued)

Results (continued)

individual data against aggregate registry data; customizations were also added, allowing users to filter by time period and procedure type. The registry also began electronically collecting PRO data from breast augmentation and reconstruction patients using the BREAST-Q© questionnaire; the PRO data supports metrics for documenting clinical performance appraisal and improvement based on the patients' responses. Users can develop PRO reports and create dynamic customized graphs and charts in order to assess patient satisfaction across PRO domains, such as satisfaction with outcome, psychosocial well-being, and satisfaction with care. In response to member requests, the upgrades also included new functionality to support data transfer from EMR and practice management programs to ease data entry burden on sites.

Key Point

Within an ongoing, voluntary registry, retaining and recruiting participants requires maintaining relevance to users. Surveys or other methods of collecting feedback from registry participants and potential participants can be useful tools for discovering recruitment or retention barriers and identifying potential improvements to maintain relevance.

For More Information

http://www.plasticsurgery.org/for-medical-professionals/surgeon-community/tops.html

References for Chapter 10

1. Scios. "Scios Announces ADHERE Heart Failure Registry Meets JCAHO Standards for Hospital Accreditation." http://www.prnewswire.com/news-releases/scios-announces-adhere-heart-failure-registry-meets-jcaho-standards-for-hospital-accreditation-77976192.html. Accessed August 15, 2012.
2. Centers for Medicare & Medicaid Services. Medicare Coverage Database. http://www.cms.gov/medicare-coverage-database/overview-and-quick-search.aspx?id=8. Accessed August 15, 2012.
3. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. Am Heart J. 2009 Jun;157(6):971-82. PMID: 19464406.
4. Gheorghiade M, Abraham WT, Albert NM, et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. JAMA. 2006 Nov 8;296(18):2217-26. PMID: 17090768.
5. Spertus JA, Peterson E, Rumsfeld JS, et al. The Prospective Registry Evaluating Myocardial Infarction: Events and Recovery (PREMIER)--evaluating the impact of myocardial infarction on patient outcomes. Am Heart J. 2006 Mar;151(3):589-97. PMID: 16504619.

Chapter 11. Data Collection and Quality Assurance

1. Introduction

This chapter focuses on data collection procedures and quality assurance principles for patient registries. Data management—the integrated system for collecting, cleaning, storing, monitoring, reviewing, and reporting on registry data—determines the utility of the data for meeting the goals of the registry. Quality assurance, on the other hand, aims to assure that the data were, in fact, collected in accordance with these procedures and that the data stored in the registry database meet the requisite standards of quality, which are generally defined based on the intended purposes. In this chapter, the term registry coordinating activities refers to the centralized procedures performed for a registry, and the term registry coordinating center refers to the entity or entities performing these procedures and overseeing the registry activities at the site and patient levels.

Because the range of registry purposes can be broad, a similar range of data collection procedures may be acceptable, but only certain methodologies may be suitable for particular purposes. Furthermore, certain end users of the data may require that data collection or validation be performed in accordance with their own guidelines or standards. For example, a registry that collects data electronically and intends for those data to be used by the U.S. Food and Drug Administration (FDA) should meet the systems validation requirements of that end user of the data, such as Title 21 of the Code of Federal Regulations Part 11 (21 CFR Part 11). Such requirements may have a substantial effect on the registry procedures. Similarly, registries may be subject to specific processes depending on the type of data collected, the types of authorization obtained, and the applicable governmental regulations.

Requirements for data collection and quality assurance should be defined during the registry inception and creation phases. Certain requirements may have significant cost implications, and these should be assessed on a cost-to-benefit basis in the context of the intended purposes of the registry. This chapter describes a broad range of centralized and distributed data collection and quality assurance activities currently in use or expected to become more commonly used in patient registries.

2. Data Collection

2.1 Database Requirements and Case Report Forms

Chapter 1 defined key characteristics of patient registries for evaluating patient outcomes. They include specific and consistent data definitions for collecting data elements in a uniform manner for every patient. As in randomized controlled trials, the case report form (CRF) is the paradigm for the data structure of the registry. A CRF is a formatted listing of data elements that can be presented in paper or electronic formats. Those data elements and data entry options in a CRF are represented in the database schema of the registry by patient-level variables. Defining the registry CRFs and corresponding database schema are the first steps in data collection for a registry. Chapter 4 describes the selection of data elements for a registry.

Two related documents should also be considered part of the database specification: the data dictionary (including data definitions and parameters) and the data validation rules, also known as queries or edit checks. The data dictionary and definitions describe both the data elements and how those data elements are interpreted. The data dictionary contains a detailed description of each variable used by the registry, including the source of the variable, coding information if used, and normal ranges if relevant. For example, the term “current smoker” should be defined as to whether “smoker” refers to tobacco or other substances and whether “current” refers to active or within a recent time period. Several cardiovascular registries, such as the Get With The Guidelines® Coronary Artery Disease[1] program define “current smoker” as someone who smoked tobacco within the last year.

Data validation rules refer to the logical checks on data entered into the database against predefined rules for either value ranges (e.g., systolic blood pressure less than 300 mmHg) or logical consistency with respect to other data fields for the same patient; these are described more fully in Section 2.5, "Cleaning Data," below. While neither registry database structures nor database requirements are standardized, the Clinical Data Interchange Standards Consortium[2] is actively working on representative models of data interchange and portability using standardized concepts and formats. Chapter 4 further discusses these models, which are applicable to registries as well as clinical trials.

2.2 Procedures, Personnel, and Data Sources

Data collection procedures need to be carefully considered in planning the operations of a registry. Successful registries depend on a sustainable workflow model that can be integrated into the day-to-day clinical practice of active physicians, nurses, pharmacists, and patients, with minimal disruption. (See Chapter 10.) Programs can benefit tremendously from preliminary input from the health care workers or study coordinators who are likely to be participants.

2.2.1 Pilot Testing

One method of gathering input from likely participants before the full launch of a registry is pilot testing. Whereas feasibility testing, which is discussed in Chapter 2, Section 2.4, focuses on whether a registry should be implemented, pilot testing focuses on how it should be implemented. Piloting can range from testing a subset of the procedures, CRFs, or data capture systems, to a full launch of the registry at a limited subset of sites with a limited number of patients.

The key to effective pilot testing is to conduct it at a point where the results of the pilot can still be used to modify the registry implementation. Through pilot testing, one can assess comprehension, acceptance, feasibility, and other factors that influence how readily the patient registry processes will fit into patient lifestyles and the normal practices of the health care provider. For example, some data sources may or may not be available for all patients. Chapter 4, Section 5 discusses pilot testing in more detail.

2.2.2 Documentation of Procedures

The data collection procedures for each registry should be clearly defined and described in a detailed manual. The term manual here refers to the reference information in any appropriate form, including hard copy, electronic, or via interactive Web or software-based systems. Although the detail of this manual may vary from registry to registry depending on the intended purpose, the required information generally includes protocols, policies, and procedures; the data collection instrument; and a listing of all the data elements and their full definitions. If the registry has optional fields (i.e., fields that do not have to be completed on every patient), these should be clearly specified.

In addition to patient inclusion and exclusion criteria, the screening process should be specified, as should any documentation to be retained at the site level and any plans for monitoring or auditing of screening practices. If sampling is to be performed, the method or systems used should be explained, and tools should be provided to simplify this process for the sites. The manual should clearly explain how patient identification numbers are created or assigned and how duplicate records should be prevented. Any required training for data collectors should also be described.

If paper CRFs are used, the manual should describe specifically how they are used and which parts of the forms (e.g., two-part or three-part no-carbon-required forms) should be retained, copied, submitted, or archived. If electronic CRFs are used, clear user manuals and instructions should be available. These procedures are an important resource for all personnel involved in the registry (and for external auditors who might be asked to assure the quality of the registry).

The importance of standardizing procedures to ensure that the registry uses uniform and systematic methods for collecting data cannot be overstated. At the same time, some level of customization of data entry methods may be required or permitted to enable the participation of

particular sites or subgroups of patients within some practices. As discussed in Chapter 10, if the registry provides payments to sites for participation, then the specific requirements for site payments should be clearly documented, and this information should be provided with the registry documents.

2.2.3 Personnel

All personnel involved in data collection should be identified, and their job descriptions and respective roles in data collection and processing should be described. Examples of such "roles" include patient, physician, data entry personnel, site coordinator, help desk, data manager, and monitor. The necessary documentation or qualification required for any role should be specified in the registry documentation. As an example, some registries require personnel documentation such as a curriculum vitae, protocol signoff, attestation of intent to follow registry procedures, or confirmation of completion of specified training.

2.2.4 Data Sources

The sources of data for a registry may include new information collected from the patient, new or existing information reported by or derived from the clinician and the medical record, and ancillary stores of patient information, such as laboratory results. Since registries for evaluating patient outcomes should employ uniform and systematic methods of data collection, all data-related procedures—including the permitted sources of data; the data elements and their definitions; and the validity, reliability, or other quality requirements for the data collected from each source—should be predetermined and defined for all collectors of data. As described in Section 3, "Quality Assurance," below, data quality is dependent on the entire chain of data collection and processing. Therefore, the validity and quality of the registry data as a whole ultimately derive from the least, not the most, rigorous link.

In Chapter 6, data sources are classified as primary or secondary, based on the relationship of the data to the registry purpose and protocol. Primary data sources incorporate data collected for direct purposes of the registry (i.e., primarily for the registry). Secondary data sources consist of data originally collected for purposes other than the registry (e.g., standard medical care, insurance claims processing). The sections below incorporate and expand on these definitions.

2.2.5 Patient-Reported Data

Patient-reported data are data specifically collected from the patient for the purposes of the registry rather than interpreted through a clinician or an indirect data source (e.g., laboratory value, pharmacy records). Such data may range from basic demographic information to validated scales of patient-reported outcomes (PROs). From an operational perspective, a wide range of issues should be considered in obtaining data directly from patients. These range from presentation (e.g., font size, language, reading level) to technologies (e.g., paper-and-pencil questionnaires, computer inputs, telephone or voice inputs, or hand-held patient diaries). Mistakes at this level can inadvertently bias patient selection, invalidate certain outcomes, or significantly affect cost. Limiting the access for patient reporting to particular languages or technologies may limit participation. Patients with specific diagnoses may have difficulties with specific technologies (e.g., small font size for visually impaired, paper and pencil for those with rheumatoid arthritis). Other choices, such as providing a PRO instrument in a format or method of delivery that differs from how it was validated (e.g., questionnaire rather than interview), may invalidate the results. For more information on patient-reported outcome development and use, see Chapter 5.

2.2.6 Clinician-Reported Data

Clinician-reported or -derived data can also be divided into primary and secondary subcategories. As an example, specific clinician rating scales (e.g., the National Institutes of Health Stroke Scale)[3] may be required for the registry but not routinely captured in clinical encounters. Some variables might be collected directly by the clinician for the registry or obtained from the medical record. Data elements that the clinician must collect directly (e.g., because of a particular definition or need to assess a specific comorbidity that may or may not be routinely present in the

medical record) should be specified. These designations are important because they determine who can collect the data for a particular registry or what changes must be made in the procedures the clinician follows in recording a medical record for a patient in a registry. Furthermore, the types of error that arise in registries (discussed in Section 3, "Quality Assurance") will differ by the degree of use of primary and secondary sources, as well as other factors. As an example, registries that use medical chart abstracters, as discussed in Section 2.2.7 below, may be subject to more interpretive errors.[4]

2.2.7 Data Abstraction

Data abstraction is the process by which a data collector other than the clinician interacting with the patient extracts clinician-reported data. While physical examination findings, such as height and weight, or laboratory findings, such as white blood cell counts, are straightforward, abstraction usually involves varying degrees of judgment and interpretation.

Clarity of description and standardization of definitions are essential to the assurance of data quality and to the prevention of interpretive errors when using data abstraction. Knowledgeable registry personnel should be designated as resources for the data collectors in the field, and processes should be put in place to allow the data collectors in the field continuous access to these designated registry personnel for questions on specific definitions and clinical situations. Registries that span long periods, such as those intended for surveillance, might be well served by a structure that permits the review of definitions on a periodic basis to ensure the timeliness and completeness of data elements and definitions, and to add new data elements and definitions. A new product or procedure introduced after the start of a registry is a common reason for such an update.

Abstracting data from unformatted hard copy (e.g., a hospital chart) is often an arduous and tedious process, especially if free text is involved, and it usually requires a human reader. The reader, whose qualifications may range from a trained "medical record analyst" or other health professional to an untrained research assistant, may need to decipher illegible handwriting, translate obscure abbreviations and acronyms, and understand the clinical content to sufficiently extract the desired information. Registry personnel should develop formal chart abstraction guidelines, documentation of processes and practical definitions of terms, and coding forms for the analysts and reviewers to use.

Generally, the guidelines include instructions to search for particular types of data that will go into the registry (e.g., specific diagnoses or laboratory results). Often the analyst will be asked to code the data, using either standardized codes from a codebook (e.g., the ICD-9 [International Classification of Diseases, 9th Revision] code) corresponding to a text diagnosis in a chart, or codes that may be unique to the registry (e.g., a severity scale of 1 to 5).

All abstraction and coding instructions must be carefully documented and incorporated into a data dictionary for the registry. Because of the "noise" in unstructured, hard-copy documents (e.g., spurious marks or illegible writing) and the lack of precision in natural language, the clinical data abstracted by different abstracters from the same documents may differ. This is a potential source of error in a registry.

To reduce the potential for this source of error, registries should ensure proper training on the registry protocol and procedures, condition(s), data sources, data collection systems, and most importantly, data definitions and their interpretation. While training should be provided for all registry personnel, it is particularly important for nonclinician data abstracters. Training time depends on the nature of the source (charts or CRFs), complexity of the data, and number of data items. A variety of training methods, from live meetings to online meetings to interactive multimedia recordings, have all been used with success.[5] Training often includes test abstractions using sample charts. For some purposes, it is best practice to train abstracters using standardized test charts. Such standardized tests can be further used both to obtain data on the inter-rater reliability of the CRFs, definitions, and coding instructions and to determine whether individual abstracters can perform up to a defined

minimum standard for the registry. Registries that rely on medical chart abstraction should consider reporting on the performance characteristics associated with abstraction, such as inter-rater reliability.[6] Examining and reporting on intra-rater reliability may also be useful. Some key considerations in standardizing medical chart abstractions are—

- Standardized materials (e.g., definitions, instructions)
- Standardized training
- Testing with standardized charts
- Reporting of inter-rater reliability

2.2.8 Electronic Medical Record

An electronic medical record (EMR) is an electronic record of health-related information on an individual that can be created, gathered, managed, and consulted by authorized clinicians and staff within one health care organization. More complete than an EMR, an electronic health record (EHR) is an electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards and that can be created, managed, and consulted by authorized clinicians and staff across more than one health care organization.[7] For the purposes of this discussion, we will refer to the more limited capabilities of the EMR.

The EMR (and EHR) will play an increasingly important role as a source of clinical data for registries. The medical community is currently in a transition period in which the primary repository of a patient's medical record is changing from the traditional hard-copy chart to the EMR. The main function of the EMR is to aggregate all clinical electronic data about a patient into one database, in the same way that a hard-copy medical chart aggregates paper records from various personnel and departments responsible for the care of the patient. Depending on the extent of implementation, the EMR may include patient demographics, diagnoses, procedures, progress notes, orders, flow sheets, medications, and allergies. The primary sources of data for the EMR are the health care providers. Data may be entered into the EMR through keyboards or touch screens in medical offices or at the bedside. In addition, the EMR system is usually interfaced with ancillary systems (discussed below), such as laboratory, pharmacy, radiology, and pathology systems. Ancillary systems, which usually have their own databases, export relevant patient data to the EMR system, which imports the data into its database.

Since EMRs include the majority of clinical data available about a patient, they can be a major source of patient information for a registry. What an EMR usually does not include is registry-specific (primary source) data that are collected separately from hard-copy or electronic forms. In the next several years, suitable EMR system interfaces may be able to present data needed by registries in accordance with registry-specified requirements, either within the EMR (which then populates the registry) or in an electronic data capture system (which then populates the EMR). EMRs already serve as secondary data sources in some registries, and this practice will continue to grow as EMRs become more widely used. In these situations, data may be extracted from the EMR, transformed into registry format, and loaded into the registry, where they will reside in the registry database together with registry-specific data imported from other sources. In a sense, this is similar to medical chart abstraction except that it is performed electronically.

Electronic capture differs from manual medical chart abstraction in two key respects. First, the data are "abstracted" once for all records. In this context, abstraction refers to the mapping and other decisionmaking needed to bring the EMR data into the registry database. It does not eliminate the potential for interpretive errors, as described later in this chapter, but it centralizes that process, making the rules clear and easily reviewed. Second, the data are uploaded electronically, eliminating duplicative data entry, potential errors associated with data reentry, and the related cost of this redundant effort.

When the EMR is used as a data source for a registry, a significant problem occurs when the information needed by the registry is stored in the EMR as free text, rather than codified or structured data. Examples of structured data include ICD-9 diagnoses and laboratory results. In contrast,

physician progress notes, consultations, radiology reports, et cetera, are usually dictated and transcribed as narrative free text. While data abstraction of free text derived from an EMR can be done by a medical record analyst, with the increasing use of EMRs, automated methods of data abstraction from free text have been developed. *Natural language processing* (NLP) is the term for this technology. It allows computers to process and extract information from human language. The goal of NLP is to parse free text into meaningful components based on a set of rules and a vocabulary that enable the software to recognize key words, understand grammatical constructions, and resolve word ambiguities. Those components can be extracted and delivered to the registry along with structured data extracted from the EMR, and both can be stored as structured data in the registry database.

An increasing number of NLP software packages are available (e.g., caTIES from the National Cancer Institute,[8] i2b2 (Informatics for Integrating Biology and the Bedside),[9] and a number of commercial products). However, NLP is still in an early phase of development and cannot yet be used for all-purpose chart abstraction. In general, NLP software operates in specific clinical domains (e.g., radiology, pathology), whose vocabularies have been included in the NLP software's database. Nevertheless, NLP has been used successfully to extract diagnoses and drug names from free text in various clinical settings.

It is anticipated that EMR/EHR use will grow significantly with the incentives provided under the American Recovery and Reinvestment Act of 2009 health information technology provisions. Currently, only a minority of U.S. patients have their data stored in systems that are capable of retrieval at the level of a data element. Furthermore, only a small number of these systems currently store data in structured formats with standardized data definitions for those data elements that are common across different vendors. A significant amount of attention is currently focused on interchange formats between clinical and research systems (e.g., from Health Level Seven [HL-7][10] to Clinical Data Interchange Standards Consortium[2] models). Attention is also focused on problems of data syntax and semantics. The adoption of common database structures and open interoperability standards will be critical for future interchange between EHRs and registries. This topic is discussed in depth in Chapter 15.

2.2.9 Other Data Sources

Some of the clinical data used to populate registries may be derived from repositories other than EMRs. Examples of other data sources include billing systems, laboratory databases, and other registries. Chapter 6 discusses the potential uses of other data sources in more detail.

2.3 Data Entry Systems

Once the primary and any secondary data sources for a registry have been identified, the registry team can determine how data will be entered into the registry database. Many techniques and technologies exist for entering or moving data into the registry database, including paper CRFs, direct data entry, facsimile or scanning systems, interactive voice response systems, and electronic CRFs. There are also different models for how quickly those data reach a central repository for cleaning, reviewing, monitoring, or reporting. Each approach has advantages and limitations, and each registry must balance flexibility (the number of options available) with data availability (when the central repository is populated), data validity (whether all methods are equally able to produce clean data), and cost. Appropriate decisions depend on many factors, including the number of data elements, number of sites, location (local preferences that vary by country, language differences, and availability of different technologies), registry duration, followup frequency, and available resources.

2.3.1 Paper CRFs

With paper CRFs, the clinician enters clinical data on the paper form at the time of the clinical encounter, or other data collectors abstract the data from medical records after the clinical encounter. CRFs may include a wide variety of clinical data on each patient gathered from different sources (e.g., medical chart, laboratory, pharmacy) and from multiple patient encounters. Before the data on formatted paper forms are entered into a

computer, the forms should be reviewed for completeness, accuracy, and validity. Paper CRFs can be entered into the database by either direct data entry or computerized data entry via scanning systems.

With direct data entry, a computer keyboard is used to enter data into a database. Key entry has a variable error rate depending on personnel, so an assessment of error rate is usually desirable, particularly when a high volume of data entry is performed. Double data entry is a method of increasing the accuracy of manually entered data by quantifying error rates as discrepancies between two different data entry personnel; data accuracy is improved by having up to two individuals enter the data and a third person review and manage discrepancies. With upfront data validation checks on direct data entry, the likelihood of data entry errors significantly decreases. Therefore, the choice of single versus double data entry should be driven by the requirements of the registry for a particular maximal error rate and the ability of each method to achieve that rate in key measures in the particular circumstance. Double data entry, while a standard of practice for registrational trials, may add significant cost. Its use should be guided by the need to reduce an error rate in key measures and the likelihood of accomplishing that by double data entry as opposed to other approaches. In some situations, assessing the data entry error rates by re-entering a sample of the data is sufficient for reporting purposes.

With hard-copy structured forms, entering data using a scanner and special software to extract the data from the scanned image is possible. If data are recorded on a form as marks in checkboxes, the scanning software enables the user to map the location of each checkbox to the value of a variable represented by the text item associated with the checkbox, and to determine whether the box is marked. The presence of a mark in a box is converted by the software to its corresponding value, which can then be transmitted to a database for storage. If the form contains hand-printed or typed text or numbers, optical character recognition software is often effective in extracting the printed data from the scanned image. However, the print font must be of high quality to avoid translation errors, and spurious marks on the page can cause errors. Error checking is based on automated parameters specified by the operator of the system for exception handling. The comments on assessing error rates in the section above are applicable for scanning systems as well.

2.3.2 Electronic CRFs

An electronic CRF (eCRF) is defined as an auditable electronic form designed to record information required by the clinical trial protocol to be reported to the sponsor on each trial subject.[11] An eCRF allows clinician-reported data to be entered directly into the electronic system by the data collector (the clinician or other data collector). Site personnel in many registries still commonly complete an intermediate hard-copy worksheet representing the CRF and subsequently enter the data into the eCRF. While this approach increases work effort and error rates, it is still in use because it is not yet practical for all electronic data entry to be performed at the bedside, during the clinical encounter, or in the midst of a busy clinical day.

An eCRF may originate on local systems (including those on an individual computer, a local area network server, or a hand-held device) or directly from a central database server via an Internet-based connection or a private network. For registries that exist beyond a single site, the data from the local system must subsequently communicate with a central data system. An eCRF may be presented visually (e.g., computer screen) or aurally (e.g., telephonic data entry, such as interactive voice response systems). Specific circumstances will favor different presentations. For example, in one clozapine patient registry that is otherwise similar to Case Example 24, both pharmacists and physicians can obtain and enter data via a telephone-based interactive voice response system as well as a Web-based system. The option is successful in this scenario because telephone access is ubiquitous in pharmacies and the eCRF is very brief.

A common method of electronic data entry is to use Web-based data entry forms. Such forms may be used by patients, providers, and interviewers to enter data into a local repository. The forms reside

on servers, which may be located at the site of the registry or co-located anywhere on the Internet. To access a data entry form, a user on a remote computer with an Internet connection opens a browser window and enters the address of the Web server. Typically, a login screen is displayed and the user enters a user identification and password, provided by personnel responsible for the Web site or repository. Once the server authenticates the user, the data entry form is displayed, and the user can begin entering data. As described in "Cleaning Data" (Section 2.5), many electronic systems can perform data validation checks or edits at the time of data entry. When data entry is complete, the user submits the form, which is sent over the Internet to the Web server.

Smart phones or other mobile devices may also be used to submit data to a server to the extent such transmissions can be done with appropriate information security controls. Mobility has recently become an important attribute for clinical data collection. Software has been developed that enables wireless devices to collect data and transmit them over the Internet to database servers in fixed locations. As wireless technology continues to evolve and data transmission rates increase, these will become more essential data entry devices for patients and clinicians.

2.4 Advantages and Disadvantages of Data Collection Technologies

When the medical record or ancillary data are in electronic format, they may be abstracted to the CRF by a data collector or, in some cases, uploaded electronically to the registry database. The ease of extracting data from electronic systems for use in a registry depends on the design of the interfaces of ancillary and registry systems, and the ability of the EMR or ancillary system software to make the requested data accessible. However, as system vendors increasingly adopt open standards for interoperability, transferring data from one system to another will likely become easier. Many organizations are actively working toward improved standards, including HL7,[10] the National eHealth Collaborative,[12] the National Institute of Standards and Technology,[13] and others. Chapter 15 describes standards and certifications specific to EHR systems.

Electronic interfaces are necessary to move data from one computer to another. If clinical data are entered into a local repository from an eCRF form or entered into an EMR, the data must be extracted from the source dataset in the local repository, transformed into the format required by the registry, and loaded into the registry database for permanent storage. This is called an "extract, transform, and load" process. Unless the local repository is designed to be consistent with the registry database in terms of the names of variables and their values, data mapping and transformation can be a complex task. In some cases, manual transfer of the data may be more efficient and less time-consuming than the effort to develop an electronic interface. Emerging open standards can enable data to be transferred from an EHR directly into the registry. This topic is discussed in more detail in Chapter 15.

If an interface between a local electronic system and registry system is developed, it is still necessary to communicate to the ancillary system the criteria for retrieval and transmission of a patient record. Typically, the ancillary data are maintained in a relational database, and the system needs to run an SQL (Structured Query Language) query against the database to retrieve the specified information. An SQL query may specify individual patients by an identifier (e.g., a medical record number) or by values or ranges of specific variables (e.g., all patients with hemoglobin A1c over 8 mg/dl). The results of the query are usually stored as a file (e.g., XML, CSV, CDISC ODM) that can be transformed and transferred to the registry system across the interface. A variety of interface protocols may be used to transfer the data.

Because data definitions and formats are not yet nationally standardized, transfer of data from an EMR or ancillary system to a registry database is prone to error. Careful evaluation of the transfer specifications for interpretive or mapping errors is a critical step that the registry coordinating center should verify. Furthermore, a series of test transfers and validation procedures should be performed and documented. Finally, error checking must be part of the transfer process because new formats or other errors not in the test databases may be introduced during actual

practice, and these need to be identified and isolated from the registry itself. Even though each piece of data may be accurately transferred, the data may have different representations on the different systems (e.g., value discrepancies such as the meaning of "0" vs. "1," fixed vs. floating point numbers, date format, integer length, and missing values). In summary, any system used to extract EMR records into registry databases should be validated and should include an interval sampling of transfers to ensure that uploading of this information is consistent over time.

The ancillary system must also notify the registry when an error correction occurs in a record already transferred to the registry. Registry software must be able to receive that notification, flag the erroneous value as invalid, and insert the new, corrected value into its database. Finally, it is important to recognize that the use of an electronic-to-electronic interchange requires not only testing but also validation of the integrity and quality of the data transferred. Few ancillary systems or EMR systems are currently validated to a defined standard. For registries that intend to report data to FDA or to other sponsors or data recipients with similar requirements, including electronic signatures, audit trails, and rigorous system validation, the ways in which the registry interacts with these other systems must be carefully considered.

2.5 Cleaning Data

Data cleaning refers to the correction or amelioration of data problems, including missing values, incorrect or out-of-range values, responses that are logically inconsistent with other responses in the database, and duplicate patient records. While all registries strive for "clean data," in reality, this is a relative term. How and to what level the data will be cleaned should be addressed upfront in a data management manual that identifies the data elements that are intended to be cleaned, describes the data validation rules or logical checks for out-of-range values, explains how missing values and values that are logically inconsistent will be handled, and discusses how duplicate patient records will be identified and managed.

2.5.1 Data Management Manual

Data managers should develop formal data review guidelines for the reviewers and data entry personnel to use. The guidelines should include information on how to handle missing data; invalid entries (e.g., multiple selections in a single-choice field, alphabetic data in a numeric field); erroneous entries (e.g., patients of the wrong gender answering gender-based questions); and inconsistent data (e.g., an answer to one question contradicting the answer to another one). The guidelines should also include procedures to attempt to remediate these data problems. For example, with a data error on an interview form, it may be necessary to query the interviewer or the patient, or to refer to other data sources that may be able to resolve the problem. Documentation of any data review activity and remediation efforts, including dates, times, and results of the query, should be maintained.

2.5.2 Automated Data Cleaning

Ideally, automated data checks are preprogrammed into the database for presentation at the time of data entry. These data checks are particularly useful for cleaning data at the site level while the patient or medical record is readily accessible. Even relatively simple edit checks, such as range values for laboratories, can have a significant effect on improving the quality of data. Many systems allow for the implementation of more complex data edit checks, and these checks can substantially reduce the amount of subsequent manual data cleaning. A variation of this method is to use data cleaning rules to deactivate certain data fields so that erroneous entries cannot even be made. A combination of these approaches can also be used. For paper-based entry methods, automated data checks are not available at the time the paper CRF is being completed but can be incorporated when the data are later entered into the database.

2.5.3 Manual Data Cleaning

Data managers perform manual data checks or queries to review data for unexpected discrepancies. This is the standard approach to cleaning data that are not entered into the database at the site (e.g., for paper CRFs entered via data

entry or scanning). By carefully reviewing the data using both data extracts analyzed by algorithms and hand review, data managers identify discrepancies and generate "queries" to send to the sites to resolve. Even eCRF-based data entry with data validation rules may not be fully adequate to ensure data cleaning for certain purposes. Anticipating all potential data discrepancies at the time that the data management manual and edit checks are developed is very difficult. Therefore, even with the use of automated data validation parameters, some manual cleaning is often still performed.

2.5.4 Query Reports

The registry coordinating center should generate, on a periodic basis, query reports that relate to the quality of the data received, based on the data management manual and, for some purposes, additional concurrent review by a data manager. The content of these reports will differ depending on what type of data cleaning is required for the registry purpose and how much automated data cleaning has already been performed. Query reports may include missing data, "out-of-range" data, or data that appear to be inconsistent (e.g., positive pregnancy test for a male patient). They may also identify abnormal trends in data, such as sudden increases or decreases in laboratory tests compared with patient historical averages or clinically established normal ranges. Qualified registry personnel should be responsible for reviewing the abnormal trends with designated site personnel. The most effective approach is for sites to provide one contact representative for purposes of queries or concerns by registry personnel. Depending on the availability of the records and resources at the site to review and respond to queries, resolving all queries can sometimes be a challenge. Creating systematic approaches to maximizing site responsiveness is recommended.

2.5.5 Data Tracking

For most registry purposes, tracking of data received (paper CRFs), data entered, data cleaned, and other parameters is an important component of active registry management. By comparing indicators, such as expected to observed rates of patient enrollment, CRF completion, and query rates, the registry coordinating center can identify problems and potentially take corrective action—either at individual sites or across the registry as a whole.

2.5.6 Coding Data

As further described in Chapter 4, the use of standardized coding dictionaries is an increasingly important tool in the ability to aggregate registry data with other databases. As the health information community adopts standards, registries should routinely apply them unless there are specific reasons not to use such standard codes. While such codes should be implemented in the data dictionaries during registry planning, including all codes in the interface is not always possible. Some free text may be entered as a result. When free text data are entered into a registry, recoding these data using standardized dictionaries (e.g., MedDRA, WHODRUG, SNOMED®) may be worthwhile. There is cost associated with recoding, and in general, it should be limited to data elements that will be used in analysis or that need to be combined or reconciled with other datasets, such as when a common safety database is maintained across multiple registries and studies.

2.5.7 Storing and Securing Data

When data on a form are entered into a computer for inclusion in a registry, the form itself, as well as a log of the data entered, should be maintained for the regulatory archival period. Data errors may be discovered long after the data have been stored in the registry. The error may have been made by the patient or interviewer on the original form or during the data entry process. Examination of the original form and the data entry log should reveal the source of the error. If the error is on the form, correcting it may require reinterviewing the patient. If the error occurred during data entry, the corrected data should be entered and the registry updated. By then, the erroneous registry data may have been used to generate reports or create cohorts for population studies. Therefore, instead of simply replacing erroneous data with corrected data, the registry system should have the ability to flag data as erroneous without deleting them and to insert the corrected data for subsequent use.

Once data are entered into the registry, the registry must be backed up on a regular basis. There are two basic types of backup, and both types should be considered for use as best practice by the registry coordinating center. The first type is real-time disk backup, which is done by the disk storage hardware used by the registry server. The second is a regular (e.g., daily) backup of the registry to removable media (e.g., tape, CD-ROM, DVD). In the first case, as data are stored on disk in the registry server, they are automatically replicated to two or more physical hard drives. In the simplest example, called "mirroring," registry data are stored on a primary disk and an exact replica is stored on the mirrored disk. If either disk fails, data continue to be stored on the mirrored disk until the failed disk is replaced. This failure can be completely transparent to the user, who may continue entering and retrieving data from the registry database during the failure. More complex disk backup configurations exist, in which arrays of disks are used to provide protection from single disk failures.

The second type of periodic backup is needed for disaster recovery. Ideally, a daily backup copy of the registry database stored on removable media should be maintained off site. In case of failure of the registry server or disaster that closes the data center, the backup copy can be brought to a functioning server and the registry database restored, with the only potential loss of data being for the interval between the regularly scheduled backups. The lost data can usually be reloaded from local data repositories or re-entered from hard copy. Other advanced and widely available database solutions and disaster recovery techniques may support a "standby" database that can be located at a remote data center. In case of a failure at the primary data center, the standby database can be used, minimizing downtime and preventing data loss.

2.6 Managing Change

As with all other registry processes, the extent of change management will depend on the types of data being collected, the source(s) of the data, and the overall timeframe of the registry. There are two major drivers behind the need for change during the conduct of a registry: internally driven change to refine or improve the registry or the quality of data collected, and externally driven change that comes as a result of changes in the environment in which the registry is being conducted.

Internally driven change is generally focused on changes to data elements or data validation parameters that arise from site feedback, queries, and query trends that may point to a question, definition, or CRF field that was poorly designed or missing. If this is the case, the registry can use the information coming back from sites or data managers to add, delete, or modify the database requirements, CRFs, definitions, or data management manual as required. At times, more substantive changes, such as the addition of new forms or changes to the registry workflow, may be desirable to examine new conditions or outcomes. Externally driven change generally arises in multiyear registries as new information about the disease and/or product under study becomes available, or as new therapies or products are introduced into clinical practice. Change and turnover in registry personnel is another type of change, and one that can be highly disruptive if procedures are not standardized and documented.

A more extensive form of change may occur when a registry either significantly changes its CRFs or changes the underlying database. Longstanding registries address this issue from time to time as information regarding the condition or procedure evolves and data collection forms and definitions require updating. Chapter 14 discusses in more detail the process for making significant modifications to a registry.

Proper management of change is crucial to the maintenance of the registry. A consistent approach to change management, including decisionmaking, documentation, data mapping, and validation, is an important aspect of maintaining the quality of the registry and the validity of the data. While the specific change management processes might depend on the type and nature of the registry, change management in registries that are designed to evaluate patient outcomes requires, at the very least, the following structures and processes:

- *Detailed manual of procedures*: As described earlier, a detailed manual that is updated on a regular basis—containing all the registry policies, procedures, and protocols, as well as a complete data dictionary listing all the data elements and their definitions—is vital for the functioning of a registry. The manual is also a crucial component for managing and documenting change management in a registry.
- *Governing body*: As described in Chapter 2, Section 6, registries require oversight and advisory bodies for a number of purposes. One of the most important is to manage change on a regular basis. Keeping the registry manual and data definitions up to date is one of the primary responsibilities of this governing body. Large prospective registries, such as the National Surgical Quality Improvement Program, have found it necessary to delegate the updating of data elements and definitions to a special definitions committee.
- *Infrastructure for ongoing training*: As mentioned above, change in personnel is a common issue for registries. Specific processes and an infrastructure for training should be available at all times to account for any unanticipated changes and turnover of registry personnel or providers who regularly enter data into the registry.
- *Method to communicate change*: Since registries frequently undergo change, there should be a standard approach and timeline for communicating to sites when changes will take place.

In addition to instituting these structures, registries should also plan for change from a budget perspective (Chapter 2) and from an analysis perspective (Chapter 13).

2.7 Using Data for Care Delivery, Coordination, and Quality Improvement

2.7.1 Improving Care

As registries increasingly collect data in electronic format, the time between care delivery and data collection is reduced. This shorter timeframe offers significant opportunities to use registry functionalities to improve care delivery at the patient and population levels. These functionalities (Table 11–1) include the generation of outputs that promote care delivery and coordination at the individual patient level (e.g., decision support, patient reports, reminders, notifications, lists for proactive care, educational content) and the provision of tools that assist with population management, quality improvement, and quality reporting (e.g., risk adjustment, population views, benchmarks, quality report transmissions). A number of registries are designed primarily for these purposes. Several large national registries[1, 14-16] have shown large changes in performance during the course of hospital or practice participation in the registry. For example, in one head-to-head study that used hospital data from Hospital Compare, an online database created by the Centers for Medicare & Medicaid Services, patients in hospitals enrolled in the American Heart Association's Get With The Guidelines® Coronary Artery Disease registry, which includes evidence-based reminders and real-time performance measurement reports, fared significantly better in measures of guidelines compliance than those in hospitals not enrolled in the registry.[17]

Table 11–1. Registry functionalities

Inputs: Obtaining data	• Identify/enroll representative patients (e.g., sampling). • Collect data from multiple sources and settings (providers, patients, labs, pharmacies) at key points. • Use uniform data elements and definitions (risk factors, treatments, and outcomes). • Check and correct data (validity, coding, etc.). • Link data from different sources at patient level (manage patient identifiers). • Maintain security and privacy (e.g., access control, audit trail).
Outputs: Care delivery and coordination	• Provide real-time feedback with decision support (evidence/guidelines). • Generate patient-level reports and reminders (longitudinal reports, care gaps, summary lists/plans, health status). • Send relevant notifications to providers and patients (care gaps, prevention support, self-management). • Share information with patients and other providers. • List patients/subgroups for proactive care. • Link to relevant patient education.
Outputs: Population measurement and quality improvement	• Provide population-level reports. – Real-time/rapid cycle – Risk adjusted – Including standardized measures – Including benchmarks – Enabling different reports for different levels of users • Enable ad hoc reports for exploration. • Provide tools to manage populations or subgroups. • Generate dashboards that facilitate action. • Facilitate third-party quality reporting (transmission).

2.7.2 Special Case: Performance-Linked Access System

A performance-linked access system (PLAS), also known as a restricted access or limited distribution system, is another application of a registry to serve more than an observational goal. Unlike a disease and exposure registry, a PLAS is part of a detailed risk-minimization action plan that sponsors develop as a commitment to enhance the risk-benefit balance of a product when approved for the market. The purpose of a PLAS is to mitigate a certain known drug-associated risk by ensuring that product access is linked to a specific performance measure. Examples include systems that monitor laboratory values, such as white blood cell counts during clozapine administration to prevent severe leukopenia, or routine pregnancy testing during thalidomide administration to prevent in utero exposure to this known teratogenic compound. Additional information on PLAS can be found in FDA's *Guidance for Industry: Development and Use of Risk Minimization Action Plans*.[18]

3. Quality Assurance

In determining the utility of a registry for decisionmaking, it is critical to understand the quality of the procedures used to obtain the data and the quality of the data stored in the database. As patient registries that meet sufficient quality criteria (discussed in Chapters 1 and 25) are

increasingly being seen as important means to generate evidence regarding effectiveness, safety, and quality of care, the quality of data within the registry must be understood in order to evaluate its suitability for use in decisionmaking. Registry planners should consider how to ensure quality to a level sufficient for the intended purposes (as described below) and should also consider how to develop appropriate quality assurance plans for their registries. Those conducting the registry should assess and report on those quality assurance activities.

Methods of quality assurance will vary depending on the intended purpose of the registry. A registry intended to serve as key evidence for decisionmaking[19] (e.g., coverage determinations, product safety evaluations, or performance-based payment) will require higher levels of quality assurance than a registry describing the natural history of a disease. Quality assurance activities generally fall under three main categories: (1) quality assurance of data, (2) quality assurance of registry procedures, and (3) quality assurance of computerized systems. Since many registries are large, the level of quality assurance that can be obtained may be limited by budgetary constraints.

To balance the need for sufficient quality assurance with reasonable resource expenditure for a particular purpose, a risk-based approach to quality assurance is highly recommended. A risk-based approach focuses on the most important sources of error or procedural lapses from the perspective of the registry's purpose. Such sources of error should be defined during inception and design phases. As described below, registries with different purposes may be at risk for different sources of error and focus on different practices and levels of assessment. Standardization of methods for particular purposes (e.g., national performance measurement) will likely become more common in the future if results are to be combined or compared between registries.

3.1 Assurance of Data Quality

Structures, processes, policies, and procedures need to be put in place to ascertain the quality of the data in the registry and to insure against several types of errors, including:

- *Errors in interpretation or coding*: An example of this type of error would be two abstracters looking for the same data element in a patient's medical record but extracting different data from the same chart. Variations in coding of specific conditions or procedures also fall under the category of interpretive errors. Avoidance or detection of interpretive error includes adequate training on definitions, testing against standard charts, testing and reporting on inter-rater reliability, and re-abstraction.
- *Errors in data entry, transfer, or transformation accuracy*: These occur when data are entered into the registry inaccurately—for example, a laboratory value of 2.0 is entered as 20. Avoidance or detection of accuracy errors can be achieved through upfront data quality checks (such as ranges and data validation checks), reentering samples of data to assess for accuracy (with the percent of data to be sampled depending on the study purpose), and rigorous attention to data cleaning.
- *Errors of intention*: Examples of intentional distortion of data (often referred to as "gaming") are inflated reporting of preoperative patient risk in registries that compare risk-adjusted outcomes of surgery, or selecting only cases with good outcomes to report ("cherry-picking"). Avoidance or detection of intentional error can be challenging. Some approaches include checking for consistency of data between sites, assessing screening log information against other sources (e.g., billing data), and performing onsite audits (including monitoring of source records) either at random or "for cause."

Steps for assuring data quality include:

- *Training*: Educate data collectors/abstracters in a structured manner.
- *Data completeness*: When possible, provide sites with immediate feedback on issues such as missing or out-of-range values and logical inconsistencies.
- *Data consistency*: Compare across sites and over time.

- *Onsite audits for a sample of sites*: Review screening logs and procedures and/or samples of data.
- *For-cause audits*: Use both predetermined and data-informed methods to identify potential sites at higher suspicion for inaccuracy or intentional errors, such as discrepancies between enrollment and screening logs, narrow data ranges, and overly high or low enrollment.

To further minimize or identify these errors and to ensure the overall quality of the data, the following should be considered.

3.1.1 A Designated Individual Accountable for Data Quality at Each Site

Sites submitting data to a registry should have at least one person who is accountable for the quality of these data, irrespective of whether the person is collecting the data as well. The site coordinator should be fully knowledgeable of all protocols, policies, procedures, and definitions in a registry. The site coordinator should ensure that all site personnel involved in the registry are knowledgeable and that all data transmitted to registry coordinating centers are valid and accurate.

3.1.2 Assessment of Training and Maintenance of Competency of Personnel

Thorough training and documentation of maintenance of competency, for both site and registry personnel, are imperative to the quality of the registry. A detailed and comprehensive operations manual, as described earlier, is crucial for the proper training of all personnel involved in the registry. Routine cognitive testing (surveys) of health care provider knowledge of patient registry requirements and appropriate product use should be performed to monitor maintenance of the knowledge base and compliance with patient registry requirements. Retraining programs should be initiated when survey results provide evidence of lack of knowledge maintenance. All registry training programs should provide means by which the knowledge of the data collectors about their registries and their competence in data collection can be assessed on a regular basis, particularly when changes in procedures or definitions are implemented.

3.1.3 Data Quality Audits

As described above, the level to which registry data will be cleaned is influenced by the objectives of the registry, the type of data being collected (e.g., clinical data vs. economic data), the sources of the data (e.g., primary vs. secondary), and the timeframe of the registry (e.g., 3-month followup vs. 10-year followup). These registry characteristics often affect the types and number of data queries that are generated, both electronically and manually. In addition to identifying missing values, incorrect or out-of-range values, or responses that are logically inconsistent with other responses in the database, specifically trained registry personnel can review the data queries to identify possible error trends and to determine whether additional site training is required. For example, such personnel may identify a specific patient outcome question or eCRF field that is generating a larger than average proportion of queries, either from one site or across all registry sites. Using this information, the registry personnel can conduct targeted followup with the sites to retrain them on the correct interpretation of the outcome question or eCRF field, with the goal of reducing the future query rate on that particular question or field. These types of "training tips" can also be addressed in a registry newsletter as a way to maintain frequent but unobtrusive communication with the registry sites.

If the registry purpose requires more stringent verification of the data being entered into the database by registry participants, registry planners may decide to conduct audits of the registry sites. Like queries discussed above, the audit plan for a specific registry will be influenced by the purpose of the registry, the type of data being collected, the source of the data, and the overall timeframe of the registry. In addition, registry developers must find the appropriate balance between the extensiveness of an audit and the impact on overall registry costs. Based on the objectives of the registry, a registry developer can define specific data fields (e.g., key effectiveness variables or adverse event data) on which the audit can be focused.

The term *audit* may describe examination or verification, may take place onsite (sometimes called monitoring) or offsite, and may be extensive

or very limited. The audit can be conducted on a random sample of participating sites (e.g., 5 to 20 percent of registry sites); "for cause" (meaning only when there is an indication of a problem, such as one site being an outlier compared with most others); on a random sample of patients; or using sampling techniques based on geography, practice setting (academic center vs. community hospital), patient enrollment rate, or query rate ("risk-based" audit strategy).

The approach to auditing the quality of the data should reflect the most significant sources of error with respect to the purpose of the registry. For example, registries used for performance measurement may have a higher risk of exclusion of higher risk patients ("cherry-picking"), and the focus of an audit might be on external sources of data to verify screening log information (e.g., billing data) in addition to data accuracy. (See Case Example 25.) Finally, the timeframe of the registry may help determine the audit plan. A registry with a short followup period (e.g., 3 months) may require only one round of audits at the end of the study, prior to database lock and data analysis. For example, in the OPTIMIZE-HF registry, a data quality audit was performed, based on predetermined criteria, on a 5-percent random sample of the first 10,000 patient records verified against source documents.[20] For registries with multiyear followup, registry personnel may conduct site audits every 1 or 2 years for the duration of the registry.

In addition to the site characteristics mentioned above, sites that have undergone significant staffing changes during a multiyear registry should be considered prime audit targets to help confirm adequate training of new personnel and to quickly address possible inter-rater variability. To minimize any impact on the observational nature of the registry, the audit plan should be documented in the registry manual.

Registries that are designed for the evaluation of patient outcomes and the generation of scientific information, and that use medical chart abstracters, should assess inter-rater reliability in data collection with sufficient scientific rigor for their intended purpose(s). For example, in one registry that uses abstractions extensively, a detailed system of assessing inter-rater reliability has been devised and published; in addition to requiring that abstracters achieve a certain level of proficiency, a proportion of charts are scheduled for re-abstraction on the basis of predefined criteria. Statistical measures of reliability from such re-abstractions are maintained and reported (e.g., kappa statistic).[21]

Subsequent to audits (onsite or remote), communication of findings with site personnel should be conducted face to face, along with followup written communication of findings and opportunities for improvement. As appropriate to meet registry objectives, the sponsor may request corrective actions from the site. Site compliance may also be enhanced with routine communication of data generated from the patient registry system to the site for reconciliation.

3.2 Registry Procedures and Systems

3.2.1 External Audits of Registry Procedures

If registry developers determine that external audits are necessary to ensure the level of quality for the specific purpose(s) of the registry, these audits should be conducted in accordance with pre-established criteria. Pre-established criteria could include monitoring of sites with high patient enrollment or with prior audit history of findings that require attention, or monitoring could be based on level of site experience, rate of serious adverse event reporting, or identified problems. The registry coordinating center may perform monitoring of a sample of sites, which could be focused on one or several areas. This approach could range from reviewing procedures and interviewing site personnel, to checking screening logs, to monitoring individual case records.

The importance of having a complete and detailed registry manual that describes policies, structures, and procedures cannot be overemphasized in the context of quality assurance of registry procedures. Such a manual serves both as a basis for conducting the audits and as a means of documenting changes emanating from these audits. As with data quality audits, feedback of the findings of registry procedure audits should be communicated to all stakeholders and documented in the registry manual.

3.2.2 Assurance of System Integrity and Security

All aspects of data management processes should fall under a rigorous life-cycle approach to system development and quality management. Each process is clearly defined and documented. The concepts described below are consistent across many software industry standards and health care industry standards (e.g., 21 CFR Part 11, legal security standards), although some specifics may vary. An internal quality assurance function at the registry coordinating center should regularly audit the processes and procedures described. When third parties other than the registry coordinating center perform activities that interact with the registry systems and data, they are typically assessed for risk and are subject to regular audits by the registry coordinating center.

3.2.3 System Development and Validation

All software systems used for patient registries should follow the standard principles of software development, including following one of the standard software development life-cycle (SDLC) models that are well described in the software industry.

In parallel, quality assurance of system development uses approved specifications to create a validation plan for each project. Test cases are created by trained personnel and systematically executed, with results recorded and reviewed. Depending on regulatory requirements, a final validation report is often written and approved. Unresolved product and process issues are maintained and tracked in an issue tracking or CAPA (Corrective Action/Preventive Action) system.

Processes for development and validation should be similarly documented and periodically audited. The information from these audits is captured, summarized, and reviewed with the applicable group, with the aim of ongoing process improvement and quality improvement.

3.3 Security

All registries maintain health information, and therefore security is an important issue. The HIPAA (Health Insurance Portability and Accountability Act of 1996) Security Rule establishes the standards for security for electronic protected health information that must be implemented by health plans, health care clearinghouses, and most health care providers (collectively, "covered entities"), as well as their business associates.[22] Therefore, covered entities and business associates that maintain registries with individually identifiable health information in electronic form must implement the technical, administrative, and physical safeguards specified and required by the HIPAA Security Rule with respect to the registry data. In addition, other Federal and State security laws may apply to registry data, depending on who maintains the registry, the type of data maintained, and other circumstances.

Aside from what may be required by applicable laws, this section generally discusses some of the components of a security program. Security is achieved not simply through technology but by clear processes and procedures. Overall responsibility for security is typically assigned. Security procedures are well documented and posted. The documentation is also used to train staff. Some registries may also maintain personal information, such as information needed to contact patients to remind them to gather or submit patient-reported outcome information.

3.3.1 System Security Plan

A system security plan consists of documented policies and standard operating procedures defining the rules of systems, including administrative procedures, physical safeguards, technical security services, technical security mechanisms, electronic signatures, and audit trails, as applicable. The rules delineate roles and responsibilities. Included in the rules are the policies specifying individual accountability for actions, access rights based on the principle of least privilege, and the need for separation of duties. These principles and the accompanying security practices provide the foundation for the confidentiality and integrity of registry data. The rules also detail the consequences associated with noncompliance.

3.3.2 Security Assessment

Clinical data maintained in a registry can be assessed for the appropriate level of security. Standard criteria exist for such assessments and are based on the type of data being collected. Part of the validation process is a security assessment of the systems and operating procedures. One of the goals of such an assessment is effective risk management, based on determining possible threats to the system or data and identifying potential vulnerabilities.

3.3.3 Education and Training

All staff members of the registry coordinating center should trained periodically on aspects of the overall systems, security requirements, and any special requirements of specific patient registries. Individuals should receive training relating to their specific job responsibilities and document that appropriate training has been received.

3.3.4 Access Rights

Access to systems and data should be based on the principles of least privilege and separation of duties. No individual should be assigned access privileges that exceed job requirements, and no individual should be in a role that includes access rights that would allow circumvention of controls or the repudiation of actions within the system. In all cases, access should be limited to authorized individuals.

3.3.5 Access Controls

Access controls provide the basis for authentication and logical access to critical systems and data. Since the authenticity, integrity, and auditability of data stored in electronic systems depend on accurate individual authentication, management of electronic signatures (discussed below) is an important topic.

Logical access to systems and computerized data should be controlled in a way that permits only authorized individuals to gain access to the system. This is normally done through a unique access code, such as a unique user ID and password combination that is assigned to the individual whose identity has been verified and whose job responsibilities require such access. The system should require the user to change the password periodically and should detect possible unauthorized access attempts, such as multiple failed logins, and automatically deauthorize the user account if they occur. The identification code can also be an encrypted digital certificate stored on a password-protected device or a biometric identifier that is designed so that it can be used only by the designated individual.

Rules should be established for situations in which access credentials are compromised. New password information should be sent to the individual by a secure method.

Intrusion detection and firewalls should be employed on sites accessible to the Internet, with appropriate controls and rules in place to limit access to authorized users. Desktop systems should be equipped with antivirus software, and servers should run the most recent security patches. System security should be reviewed throughout the course of the registry to ensure that management, operational, personnel, and technical controls are functioning properly.

3.3.6 Data Enclaves

With the growth of clinical data and demands for increasing amounts of clinical data by multiple parties and researchers, new approaches to access are evolving. Data enclaves are secure, remote-access systems that allow researchers to share respondents' information in a controlled and confidential manner.[23] The data enclave uses statistical, technical, and operational controls at different levels chosen for the specific viewer. This can be useful both for enhancing protection of the data and for enabling certain organizations to access data in compliance with their own organization or agency requirements. Data enclaves also can be used to allow other researchers to access a registry's data in a controlled manner. With the growth of registries and their utility for a number of stakeholders, data enclaves will become increasingly important.

3.3.7 Electronic Signatures

Electronic signatures provide one of the foundations of individual accountability, helping to ensure an accurate change history when used in conjunction with secure, computer-generated,

time-stamped audit trails. Most systems use an electronic signature. For registries that report data to FDA, such signatures must meet criteria specified in 21 CFR Part 11 for general signature composition, use, and control (sections 11.100, 11.200, and 11.300). However, even registries that do not have such requirements should view these as reasonable standards. Before an individual is assigned an electronic signature, it is important to verify the person's identity and train the individual in the significance of the electronic signature. In cases where a signature consists of a user ID and a password, both management and technical means should be used to ensure uniqueness and compliance with password construction rules. Password length, character composition, uniqueness, and validity life cycle should be based on industry best practices and guidelines published by the National Institute of Standards and Technology. Passwords used in electronic signatures should abide by the same security and aging constraints as those listed for system access controls.

3.3.8 Validation

Systems that store electronic records (or depend on electronic or handwritten signatures of those records) that are required to be acceptable to FDA must be validated according to the requirements set forth in the 21 CFR Part 11 Final Rule,[24] dated March 20, 1997. The rule describes the requirements and controls for electronic systems that are used to fulfill records requirements set forth in agency regulations (often called "predicate rules") and for any electronic records submitted to the agency. FDA publishes nonbinding guidance documents from time to time that outline its current thinking regarding the scope and application of the regulation. The current guidance document is Guidance for Industry: Part 11, Electronic Records; Electronic Signatures – Scope and Application,[25] dated August 2003. Other documents that are useful for determining validation requirements of electronic systems are Guidance for Industry: Computerized Systems Used in Clinical Investigations,[26] dated May 2007, and General Principles of Software Validation; Final Guidance for Industry and FDA Staff,[27] dated January 11, 2002.

4. Resource Considerations

Costs for registries can be highly variable, depending on the overall goals. Costs are also associated with the total number of sites, the total number of patients, and the geographical reach of the registry program. Each of the elements described in this chapter has an associated cost. Table 11–2 provides a list of some of the activities of the registry coordinating center as an example. Not all registries will require or can afford all of the functions, options, or quality assurance techniques described in this chapter. Registry planners must evaluate benefit versus available resources to determine the most appropriate approach to achieve their goals.

Table 11–2. Data activities performed during registry coordination

Data Management	• Defines all in-process data quality control steps, procedures, and metrics. • Defines the types of edit checks that are run against the data. • Defines required file-format specifications for electronic files, as well as schedules and processes for transfers of data. • Defines quality acceptance criteria for electronic data, as well as procedures for handling exceptions. • Develops guidelines for data entry. • Identifies areas of manual review where electronic checks are not effective. • Develops and maintains process for reviewing, coding, and reporting adverse event data. • Develops and maintains archiving process. • Develops and documents the process for change management. • Develops and maintains process for query tracking and creates standard reports to efficiently identify outstanding queries, query types per site, etc. • Relates queries to processes and activities (e.g., CRF design) requiring process improvements. • Follows up on query responses and errors identified in data cleaning by performing accurate database updates. • Defines registry-specific dictionaries and code lists. • Performs database audits as applicable. • Conducts user testing of systems and applications per written specifications. • Establishes quality criteria and quality error rate acceptance limits. • Evaluates data points that should be audited and identifies potential sources of data errors for audits. • Identifies root cause of errors in order to recommend change in process/technology to ensure the error does not occur again (continuous improvement). • Ensures that sampling audit techniques are valid and support decisions made about data. • Outlines all other data flow, including external data sources.
Documentation	• Documents the process, procedures, standards, and checklist(s) and provides training. • Documents and maintains process and standards for identifying signals and trends in data. • Documents database quality control actions performed.
Reporting	• Generates standard reports of missing data from the patient database. • Creates tools to track and inventory CRFs, and reports anticipated vs. actual CRF receipts.

CRF = case report form.

Case Examples for Chapter 11

Case Example 24. Developing a performance-linked access system

Description	The Teva Clozapine Patient Registry is one of several national patient registries for patients taking clozapine. The registry is designed as a performance-linked access system (PLAS) mandated by the U.S. Food and Drug Administration (FDA) to comply with a Risk Evaluation Mitigation Strategy. The goal is to prevent clozapine rechallenge in patients at risk for developing clozapine- induced agranulocytosis by monitoring lab data for signs of leukopenia or granulocytopenia.
Sponsor	Teva Pharmaceuticals USA
Year Started	1997
Year Ended	Ongoing
No. of Sites	Over 50,000 active physicians and pharmacies
No. of Patients	57,000 active patients

Challenge

Clozapine is classified as an atypical antipsychotic and is indicated for patients with severe schizophrenia who fail standard therapy, and for reducing the risk of recurrent suicidal behavior in schizophrenia or schizoaffective disorder. However, clozapine is known to be associated with a risk of developing agranulocytosis, a potentially life-threatening condition. The primary goal of the registry is to prevent clozapine-induced agranulocytosis. Patients at risk of developing clozapine-induced agranulocytosis are those who have a history of severe leukopenia or granulocytopenia (white blood cell [WBC] <2,000/mm^3 or absolute neutrophil [ANC] <1,000/mm^3). Because of the potential serious side effects, FDA requires manufacturers of clozapine to maintain a patient monitoring system. Designed as a PLAS, the registry needs to ensure the eligibility of patients, pharmacies, and physicians; monitor white blood cell (WBC) and absolute neutrophil (ANC) reports for low counts; ensure compliance with laboratory report submission timelines; and respond to inquiries and reports of adverse events.

Proposed Solution

The risk of developing agranulocytosis is mitigated by regular hematological monitoring and is a condition of access to the drug, also known as the "no-blood/no drug" requirement. Since there are multiple manufacturers of clozapine, FDA requires each company to share information with the single national non-rechallenge master file (NNRMF). The Teva Clozapine Patient Registry was developed to meet these goals. The core components of the system are a call center, a Web site, and a reminder system. Patients must be enrolled prior to receiving clozapine, and they must be assigned to a dispensing pharmacy and treating physician. After the patient has initiated therapy, a current and acceptable WBC count and ANC value are required prior to dispensing clozapine. Once a patient is enrolled and eligibility is confirmed, a 1-, 2-, or 4-week supply of clozapine can be dispensed, depending on patient experience and the physician's prescription.

Health care professionals are required to submit laboratory reports to the registry based on the patients' monitoring frequency. Patients are monitored weekly for the first 6 months. If there are no low counts, the patient can be monitored every 2 weeks for an additional 6 months. Afterward, if no low counts are detected after continuous therapy, the patient may qualify for monitoring every 4 weeks (depending on the physician's prescription). The registry provides reminders if laboratory data are not submitted according to the schedule. If a low count is

Case Example 24. Developing a performance-linked access system (continued)

Proposed Solution (continued)

identified, registry staff inform the health care providers to make sure that they are aware of the event and appropriate action is taken. If severe leukopenia or granulocytopenia is detected, the patient is posted to the NNRMF to prevent future exposure to the drug.

Results

Results indicate that the registry is achieving its goal of reducing the risk of agranulocytosis associated with the use of clozapine by serving as an early warning system. By linking access to clozapine to a strict schedule of laboratory data submissions, the sponsor can ensure that only eligible patients are taking the drug. The sponsor is also able to detect low counts, prevent inappropriate rechallenge (or re-exposure) in at-risk patients, and monitor the patient population for any adverse events. This system provides the sponsor with data on the frequency and severity of adverse events while ensuring that only the proper patient population receives the drug.

Key Point

A PLAS can ensure that only appropriate patients receive treatment. A secure, fully functional Web site allows health care professionals to manage their patients electronically. A reminder system permits rapid notification to providers to ensure that appropriate actions are taken when low counts are detected or if laboratory reports are not submitted in a timely manner. A call center with after-hours service ensures 24/7 availability, and data sharing with the NNRMF prevents rechallenge regardless of manufacturer. These systems can also help sponsors monitor the patient population to learn more about adherence, compliance, and the frequency of adverse events.

For More Information

Clozapine Package Insert (2012). https://www.clozapineregistry.com/insert.pdf.ashx

Honigfeld G. The Clozapine National Registry System: forty years of risk management. J Clin Psychiatry Monograph. 1996;14(2):29–32.

Karukin M, Conner J, Lage M. Incidence of leukopenia and agranulocytosis with the use of clozapine: Evidence from the Teva Clozapine Patient Registry. Poster Presented at the 23rd Annual U.S. Psychiatric and Mental Health Congress, Poster #219, November 20, 2010.

Peck CC. "FDA's position on the clozaril patient management system." Hosp Community Psychiatry. 1990 Aug;41(8): 876-7.

Case Example 25. Using audits to monitor data quality

Description	The Vascular Study Group of New England (VSGNE) is a voluntary, cooperative group of clinicians, hospital administrators, and research personnel, organized to improve the care of patients with vascular disease. The purpose of the registry is to collect and exchange information to support continuous improvements in the quality, safety, effectiveness, and cost of caring for patients with vascular disease.
Sponsor	Funded by participating institutions. Initial funding was provided by the Centers for Medicare & Medicaid Services.
Year Started	2002
Year Ended	Ongoing
No. of Sites	30 hospitals in Connecticut, Rhode Island, Massachusetts, Maine, New Hampshire, and Vermont
No. of Patients	Over 25,000

Challenge

VSGNE established a registry in 2002 as part of an effort to improve quality of care for patients undergoing carotid endarterectomy, carotid stenting, lower extremity arterial bypass, and open and endovascular repair of abdominal aortic aneurysms. The registry collects more than 120 patient, process, and outcome variables for each procedure at the time of hospitalization, and 1-year results are collected during a followup visit at the treating physician's office. All patients receiving one of the procedures of interest at a participating hospital are eligible for enrollment in the registry.

In considering the areas of greatest risk in evaluating the quality of this registry, the registry developers determined that incomplete enrollment of eligible patients was one major potential area for bias. It was determined that an audit of participating sites, focusing on included versus eligible patients, could reasonably address whether this was a significant issue. However, the group needed to overcome two logistical challenges: (1) the audit had to review thousands of eligible patients at participating hospitals in a timely, cost-effective manner; and (2) the audit could not overburden the hospitals, as they participate in the study voluntarily.

Proposed Solution

The registry team developed a plan to conduct the audit using electronic claims data files from the hospitals. Each hospital was asked to send claims data files for the appropriate time periods and procedures of interest to the registry. The registry team at Dartmouth-Hitchcock Medical Center then matched the claims data to the registry enrollment using ICD-9 (International Classification of Diseases, 9th Revision) codes with manual review of some patient files that did not match using a computer-matching process.

Results

The first audit performed in 2003 found that approximately 7 percent of eligible patients had not been enrolled in the registry. Because of concerns that the missing patients may have had different outcomes than the patients who had been enrolled in the registry, the registry team asked participating hospitals to complete registry forms for all missing patients. This effort increased the percentage of eligible patients enrolled in the registry to over 99 percent. The team also compared the discharge status of the missing patients and the enrolled patients, and found no significant differences in outcomes. The team concluded that the patients had been missed at random and that there were no systematic enrollment issues. Discussions with the hospitals identified the reasons for not enrolling patients as confusion about eligibility requirements, training issues, and questions about informed consent requirements.

Subsequent audits in 2006 and 2008 had similar outcomes, but considerable time was required to

Case Example 25. Using audits to monitor data quality (continued)

Results (continued)

clarify ICD-9 coding differences with procedures in the registry, since ICD-9 codes are not granular for vascular procedures. In 2011, the VSGNE model for regional vascular quality improvement was adopted by the Society for Vascular Surgery as the Vascular Quality Initiative, now a national network of regional quality groups like VSGNE, organized under the umbrella of the Society for Vascular Surgery's patient safety organization. In 2012, the now nationwide audit mechanism for data completeness switched from using ICD-9 codes to physician current procedural terminology (CPT®) claims data, since CPT codes are more precise for specific vascular procedures. Preliminary results in 2012 show more precise matching with registry data using CPT-based claims.

Key Point

For many registries, audits of participating sites are an important tool for ensuring that the data are reliable and valid. However, registries that rely on voluntary site participation must be cautious to avoid overburdening sites during the audit process. A remote audit using readily available electronic files, such as claims files, provided a reasonable assessment of the percentage of eligible patients enrolled in the registry without requiring large amounts of time or resources from participating sites.

For More Information

Cronenwett JL, Likosky DS, Russell MT, et al. A regional registry for quality assurance and improvement: the Vascular Study Group of Northern New England (VSGNNE). J Vasc Surg. 2007;46:1093–1102.

Cronenwett JL, Kraiss LW, Cambria RP. The Society for Vascular Surgery Vascular Quality Initiative. J Vasc Surg. 2012;55:1529-37.

References for Chapter 11

1. American Heart Association. Get With The Guidelines. http://www.heart.org/HEARTORG/HealthcareResearch/GetWithTheGuidelinesHFStroke/Get-With-The-Guidelines---HFStroke_UCM_001099_SubHomePage.jsp. Accessed May 1, 2013.
2. Clinical Data Interchange Standards Consortium. http://www.cdisc.org. Accessed May 1, 2013.
3. National Institutes of Health. National Institutes of Health Stroke Scale. http://www.ninds.nih.gov/doctors/NIH_Stroke_Scale.pdf. Accessed May 1, 2013.
4. Luck J, Peabody JW, Dresselhaus TR, et al. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. Am J Med. 2000 Jun 1;108(8):642-9. PMID: 10856412.
5. Reisch LM, Fosse JS, Beverly K, et al. Training, quality assurance, and assessment of medical record abstraction in a multisite study. Am J Epidemiol. 2003 Mar 15;157(6):546-51. PMID: 12631545.
6. Neale R, Rokkas P, McClure RJ. Interrater reliability of injury coding in the Queensland Trauma Registry. Emerg Med (Fremantle). 2003 Feb;15(1):38-41. PMID: 12656785.
7. A Healthcare IT Primer. http://geekdoctor.blogspot.com/2009/03/healthcare-it-primer.html. Accessed May 1, 2013.
8. cancer Text Information Extraction System (caTIES). http://caties.cabig.upmc.edu/. Accessed August 7, 2013.
9. Informatics for Integrating Biology and the Bedside. https://www.i2b2.org/. Accessed May 1, 2013.
10. Health Level Seven. http://www.hl7.org. Accessed May 1, 2013.
11. U.S. Food and Drug Administration. Guidance for Industry. E6 Good Clinical Practice: Consolidated Guidance. April 1996. http://www.fda.gov/downloads/Drugs/Guidances/ucm073122.pdf. Accessed August 7, 2013.
12. National eHealth Collaborative. http://www.nationalehealth.org. Accessed May 1, 2013.

13. National Institute of Standards and Technology. http://www.nist.gov/index.html. Accessed May 1, 2013.

14. OPTIMIZE-HF. Organized Program to Initiate LifeSaving Treatment in Hospitalized Patients with Heart Failure. http://www.optimize-hf.org/. Accessed May 1, 2013.

15. Fonarow GC, Heywood JT, Heidenreich PA, et al. Temporal trends in clinical characteristics, treatments, and outcomes for heart failure hospitalizations, 2002 to 2004: findings from Acute Decompensated Heart Failure National Registry (ADHERE). Am Heart J. 2007 Jun;153(6):1021-8. PMID: 17540205.

16. IMPROVE-HF. http://newsroom.medtronic.com/phoenix.zhtml?c=251324&p=irol-newsArticle&ID=1771401&highlight=. Accessed May 1, 2013.

17. Lewis WR, Peterson ED, Cannon CP, et al. An organized approach to improvement in guideline adherence for acute myocardial infarction: results with the Get With The Guidelines quality improvement program. Arch Intern Med. 2008 Sep 8;168(16):1813-9. PMID: 18779470. PMCID: 3086550.

18. U.S. Food and Drug Administration. Guidance for Industry: Development and Use of Risk Minimization Action Plans. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126830.pdf. Accessed May 1, 2013.

19. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006 Jan 26;354(4):353-65. PMID: 16436767.

20. Gheorghiade M, Abraham WT, Albert NM, et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. JAMA. 2006 Nov 8;296(18):2217-26. PMID: 17090768.

21. Fink AS, Campbell DA, Jr., Mentzer RM, Jr., et al. The National Surgical Quality Improvement Program in non-veterans administration hospitals: initial demonstration of feasibility. Ann Surg. 2002 Sep;236(3):344-53; discussion 53-4. PMID: 12192321. PMCID: 1422588.

22. The HIPAA Security Rule: Health Insurance Reform: Security Standards, February 20, 2003. 68 FR 8334.

23. National Institutes of Health. NIH Data Sharing Policy and Implementation Guidance. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave. Accessed May 1, 2013.

24. U.S. Food and Drug Administration. CFR - Code of Federal Regulations Title 21, Part 11: Electronic Records; Electronic Signatures. http://www.accessdata.fda.gov/SCRIPTs/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11&showFR=1. Accessed May 1, 2013.

25. U.S. Food and Drug Administration. Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm125125.pdf. Accessed May 1, 2013.

26. U.S. Food and Drug Administration. Guidance for Industry Computerized Systems Used in Clinical Investigations. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070266.pdf. Accessed May 1, 2013.

27. U.S. Food and Drug Administration. General Principles of Software Validation; Final Guidance for Industry and FDA Staff. http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM085371.pdf. Accessed May 1, 2013.

Chapter 12. Adverse Event Detection, Processing, and Reporting

1. Introduction

Registries that collect information on specific drugs and medical devices need to anticipate the need for adverse event (AE) detection, processing, and reporting. This chapter addresses the identification, processing, and reporting of AEs detected in situations in which a registry has contact with individual patients. This document is not a formal regulatory or legal document; therefore, any information or suggestions presented herein do not supersede, replace, or otherwise interpret Federal guidance documents that touch on these subjects. Registry sponsors are encouraged to discuss plans for AE collection and processing with local health authorities when planning a registry.

This chapter primarily focuses on AEs related to pharmaceutical products. Medical devices are significantly different from pharmaceutical products in the manner in which AEs and product problems (complaints) present themselves, in the etiology of their occurrence, and in the regulation governing the defining and reporting of these occurrences, as well as postapproval study requirements. Other sources provide more information about defining and reporting device-related AEs and product problems, and about postmarketing studies (including those involving registries).[1-3]

2. Identifying and Reporting Adverse Drug Events

The U.S. Food and Drug Administration (FDA) defines an adverse drug experience as any AE associated with the use of a drug in humans, whether or not considered drug related,[4] while the International Conference on Harmonisation (ICH) guideline ICH E2A similarly defines an AE as an untoward medical occurrence in a patient administered a pharmaceutical product, whether or not the occurrence is related to or considered to have a causal relationship with the treatment.[5]

For marketed products regulated by FDA, AEs are categorized for reporting purposes according to the seriousness and expectedness of the event (i.e., whether the event was previously observed and included in local product labeling), as it is presumed that all spontaneously reported events are potentially related to the product for the purposes of FDA reporting. Prior to marketing approval, relatedness is an additional determinant for reporting events occurring during clinical trials or preclinical studies associated with investigational new drugs and biologics. For AEs occurring in postapproval studies and reported during planned contacts and active solicitation of information from patients, as when registries collect data regarding one or more FDA-approved products,[6, 7] the requirements for mandatory reporting also include whether there is a reasonable possibility that the drug caused the adverse experience.[4] For registries that do not actively solicit AEs, incidentally reported events (e.g., those reported during clinician or consumer contact for another purpose) should typically be handled and evaluated as spontaneously reported events.

The medical device reporting regulations differ from those for drugs and biologics in that reportable events include both AEs and problems with the device itself.[8] Medical device reporting is required for incidents in which the device may have caused or contributed to a death or serious injury, or may have malfunctioned and would likely cause or contribute to death or serious injury if the malfunction were to recur.[9]

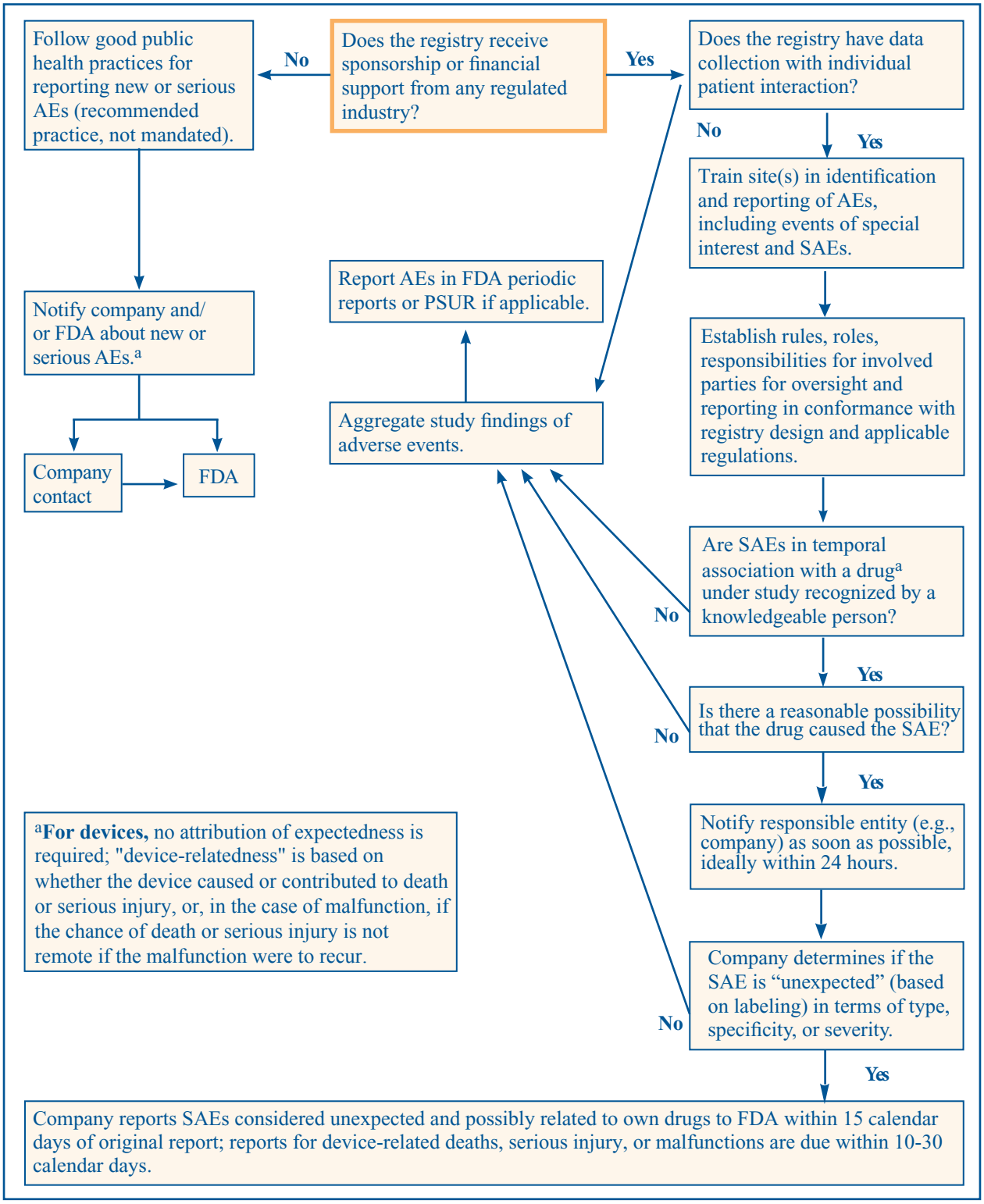
Most registries have the opportunity to identify and capture information on AEs for biopharmaceutical products and/or medical devices. With the passing of the FDA Amendments Act in September 2007 and the increased emphasis on ongoing monitoring of safety profiles, evaluation of risks unknown at the time of product approval, and proactive detection of potential safety issues, registries increasingly continue to be used to fulfill safety-related objectives.[10] Although no regulations in the United States specifically

compel registries to capture and process AE reports (aside from reporting requirements for registries that are sponsored by regulated industries), there is an implicit requirement from the perspective of systematic data collection and promoting public health: any individual who believes a serious risk may be associated with exposure to a medical product should be encouraged to report this AE either to the product sponsor or directly to FDA. The FDA maintains MedWatch, a Web-based reporting system that allows consumers and health professionals to voluntarily report serious adverse events and other serious problems that they suspect are associated with the use of an FDA-regulated product.[11]

The minimum dataset required to consider information as a reportable AE is indeed minimal, namely (1) an identifiable patient, (2) an identifiable reporter, (3) product exposure, and (4) an event. However, in addition to direct data collection, AEs can be detected through retrospective analysis of a population database, where direct patient or health care provider contact does not occur. Patient interactions include clinical interactions and data collection by phone, Internet, or other means; perusal of electronic medical records or insurance claims data would not be considered direct patient interaction. Reporting is rarely required for individual AEs observed in aggregate population data, since there is no direct patient interaction where an association might be suggested or inferred. Nevertheless, if aggregate or epidemiologic analyses suggest that an AE is associated with exposure to a drug or medical product, it is desirable that the minimum dataset information be forwarded to the manufacturer of the product, who will determine any need for, and timing of, reporting of study results to the relevant regulatory authorities.

Figure 12–1 provides a broad overview of the reporting requirements for AEs and shows how the reporting differs according to whether the registry has direct patient interaction, and whether it receives sponsorship and/or financial support from a regulated industry.[12] These industries may include entities with products subject to FDA regulation, including products with FDA approval, an FDA-granted license, and investigational products; and other entities such as manufacturers, user facilities, and distributors.

Figure 12–1. Best practices for adverse event reporting to FDA by registries of postmarket products



AE = adverse event; SAE = serious adverse event; FDA = U.S. Food and Drug Administration; PSUR = periodic safety update report.

All AE reporting begins with a suspicion by the physician (or responsible person who obtains or receives information) that a patient exposed to a medicinal product has experienced some AE and that the event has a reasonable possibility of being causally related to the product being used; this is referred to as the "becoming aware" principle. Some registries also collect and record AEs reported directly by the patients or their caregivers. It is important to develop a plan for detecting, processing, and reporting AEs for any registry that has direct patient contact. If the registry receives sponsorship in whole or part from a regulated industry (for drugs or devices), the sponsor has mandated reporting requirements, including stringent timelines. AE reporting requirements for registry sponsors are discussed later in this chapter.

Prior to registry launch, the process for detecting and reporting AEs should be established in collaboration with the sponsor and any oversight committees. (See Chapter 2.) Once the plans have been developed, the registry operator or sponsor should provide training to the physicians or other responsible parties (referred to as "sites" hereafter) on how to identify AEs and to whom they should be reported. AE reporting is based on categorization of the AE according to the seriousness of the event, its expectedness based on product labeling, and presumed causality or possible association with use of the product, as follows:

- *Seriousness*: Serious AEs (SAEs) include events that result in death, are life threatening (an event in which the patient was at risk of death at the time of the event), require or prolong inpatient hospitalization, result in persistent or significant disability or incapacity, or result in a congenital anomaly. Important medical events may also be considered serious when, based on medical judgment, they may jeopardize the person exposed and may require medical or surgical intervention to prevent one of the outcomes listed above (e.g., death or prolonged hospitalization).
- *Expectedness*: All AEs that are previously unobserved or undocumented are referred to as "unexpected," in that their nature and severity are not consistent with information provided in the relevant product information (e.g., approved professional package insert or product label). Determination of expectedness is made by the sponsor on a case-by-case basis. Expected events typically do not require expedited reporting to the regulatory authorities.
- *Relatedness*: Relatedness is a term intended to indicate that a determination has been made that the event had a reasonable possibility of being related to exposure to the product. This assessment of causality may be based on factors such as biological plausibility, prior experience with the product, and temporal relationship between product exposure and onset of the event, as well as dechallenge (discontinuation of the product to determine if the AE resolves) and rechallenge (reintroduction of the product to determine if the AE recurs). Many terms and scales are used to describe the degree of causality, including terms such as certainly, definitely, probably, possibly, or likely related or not related, but there is no standard nomenclature.[13] All spontaneous reports have an implied causal relationship as per regulatory guidance, regardless of the reporter's assessment.

The registry may use forms such as a structured questionnaire or an AE case report form to collect the information from providers or patients. When solicitation of AEs is not prespecified in the registry's operating plans, the registry may permit AE detection by asking general questions to solicit events, such as "Have you had any problems since your last visit or since we last spoke?" and then following up any such reports with probes as to what happened, diagnoses, and other documentation. This practice is not required.

3. Collecting AE Data in a Registry

There are two key considerations regarding AE collection as part of a registry: (1) what data need to be collected to meet the registry's safety-related objectives, and (2) what processes need to be in place to ensure that the registry is in compliance with regulations regarding expedited and periodic

AE event reporting, if applicable. The data fields needed for the purpose of analysis by the registry may be minimal (e.g., event and onset date), whereas a complete SAE form for a subset of events reported to the registry may be sought to fulfill the sponsor's reporting requirements. Due to the nature of registries, the goal of collecting enough data to meet the registry's objectives must constantly be balanced with the burden on sites. To this end, the processes for AE reporting should be streamlined as much as possible.

The collection of AE data by a registry is generally either intentionally solicited (meaning that the data are part of the uniform collection of information in the registry) or unsolicited (meaning that the AE information is volunteered or noted in an unsolicited manner and not as a required data element through a case report form). As described further below, it is good practice for a registry to specify when and how AE information (and any other events of special interest) should and should not be solicited from patients by a site and, if that information has been obtained, how and when the site should inform the appropriate persons.

While an AE may be reported to the manufacturer, to FDA (e.g., via MedWatch), or to the registry itself (and then from the registry to the manufacturer), it is strongly encouraged that the protocol describe the procedures that should be followed, and that the sites be trained in these procedures as well as in their general obligations and the relevant public health considerations. A separate safety reporting plan that fully identifies the responsible parties and describes the operational considerations may also be considered to ensure that potentially reportable information is evaluated in an appropriate timeframe, and, for manufacturer-sponsored registries, in accordance with any applicable standard operating procedures. This type of plan also should describe how deviations or systemic failures in detection and reporting processes will be identified, addressed, and considered for corrective action.

Determining whether a registry should use a case report form to collect AEs should be based on the principles described in Chapter 4, which refer to the scientific importance of the information for evaluating the specified outcomes of interest. This may mean that all, some, or no AEs are collected on the case report forms. However, if some AEs are collected in an intentional, solicited manner (e.g., routine collection of a primary or secondary outcome via an AE case report form) and others come to the registry's attention in an unsolicited, "spontaneous" way (e.g., when an AE is reported in the course of a registry contact, such as a call to the sponsor or to registry support staff), then from a practical perspective it is even more important to have a clear process, so that AEs that require reporting are identified. In this scenario, one best practice that is often used in electronic registry studies is to have a notification sent promptly to the sponsor's safety group when a case report form is submitted that contains specific or potential information indicating that a serious AE has occurred. This process allows for rapid followup by the sponsor, as needed.

4. AE Reporting by the Registry

Once suspicion has been aroused that an unexpected serious event has a reasonable possibility of being causally related to a drug, the AE should be reported to FDA through MedWatch, to the company that manufactures the product, or to the registry coordinating center. (See Chapter 11.) A system should be developed such that all appropriate events are captured and duplicate reporting is avoided to the extent possible. Generally, AE reports are submitted directly by the site or by the registry to the manufacturer, since they are often most efficient at evaluating, processing, and reporting for regulatory purposes within the required time periods. Alternatively, sites could be instructed to report AEs directly to FDA according to their normal practices for marketed products; however, this often means that the companies are not notified of the AE and are not able to follow up or evaluate the event in the context of their safety database. In fact, companies are not necessarily notified by FDA if an AE report comes directly to FDA, since only certain reports are shared with industry, and reporters have an option to request that the information not be shared directly with the company.[14] When sites report AEs directly to

FDA, this process can also lead to inadvertent duplication of information for events recorded both by the registry and the company.

Systematic collection of all AEs provides a unique resource of consistent and contemporaneously collected comparison information that can be used at a later date to conduct epidemiologic assessments. Ideally, the practice for handling AEs and SAEs should be applied to all treatments (including comparators) recorded in the registry, so that all subjects are treated similarly. In fact, a strong advantage of registries with systematic data collection and internal comparators is that they provide both numerators and denominators for safety events; thus, reporting of comparative known AE rates in the context of a safety evaluation provides valuable information on real-world performance. The contrast with comparators helps promote clarity about whether the observed effects are unique to the product, unique to a class, or common to the condition being treated. Reporting AEs without denominator information is less useful from a surveillance perspective since events rates cannot be calculated without both numerators and denominators. The reliability of the denominator should always be judged, however, by considering the likelihood that all events were reported appropriately.

For postapproval registries not financially supported by pharmaceutical companies, health care providers at registry sites should be instructed that if they suspect or otherwise become aware of a serious AE that has a reasonable possibility of being causally related to a drug or product, they should report the event directly to the product manufacturer (who must then report to FDA under regulation) or to FDA's MedWatch program (or local health authority if the study is conducted outside of the United States). Reporting can be facilitated by providing the MedWatch Form 3500,[15] information regarding the process for submission, and MedWatch contact information.

For registries that are sponsored or financially supported in full or in part by a regulated industry and that study a single product, the most efficient monitoring system to avoid duplicate reporting is one in which all physicians participating in the registry report all AEs (or SAEs only) directly to the sponsor or centralized designated responsible personnel, who then reports to the regulatory authorities. However, when products other than those exclusively manufactured by the sponsor are involved, including other treatments, sponsors will need to determine how to process AE reports received for these other products. Since sponsors are not obligated to report AEs for their competitors, it is good practice from a public health perspective to specify how the site should address those AEs (e.g., whether to report directly to the other product's manufacturer or to FDA). Options for the sponsor include (1) recommending that sites report the AEs of comparators directly to the manufacturer or to FDA; (2) collecting all AEs and forwarding the AE report directly to the comparator's manufacturer (who would then, in turn, report to FDA); and (3) actually reporting the AE for the comparator product directly to FDA. As standard practice in pharmacovigilance, many sponsors report events potentially associated with another manufacturer's drug to that manufacturer's safety department as a courtesy, rather than report events directly to FDA, and choose to continue that practice when conducting a registry or other observational study.

Some disease registries are not focused on a specific product, but rather on conducting natural history studies or evaluating treatment patterns and outcomes in a particular patient population prior to marketing approval of the sponsor's product. In these situations, it is recommended that sites follow their own standard practices for spontaneous AE reporting, including reporting any events associated with a product known to be manufactured by the sponsor.

In most circumstances in which a serious drug-associated AE is suspected, sites are encouraged to submit supportive data to sponsors, such as laboratory values, vital signs, and examination results, along with the SAE report form. If the event is determined to be an AE, the sponsor will include it in the safety database, evaluate it internally, and transfer the AE report to the regulatory authorities if required. It should be noted that the regulations represent minimum requirements for compliance; special circumstances for a particular product may result

in additional events being reportable (e.g., expected events of particular interest to regulators). It should not be expected that registry participants be aware of all the reporting nuances associated with a particular product. To the extent possible, guidance on reporting events of special interest should be provided in the protocol and in any safety training.

If an external party manages a registry, SAEs should be submitted to the sponsors as quickly as possible after the registry becomes aware of the event. In this situation, the registry is an agent of the sponsor, and FDA's 15-calendar-day reporting requirement starts as soon as the event has come to the attention of the registry. (See Section 7 below.) This submission can be accomplished by phone or fax, or by means of automated rules built into the vehicle used for data collection (such as automatic triggers that can be designed into electronic data capture programs). For direct regulatory submissions, the MedWatch Form 3500A[16] should be used for postapproval reporting for drugs and therapeutic biologics unless other means of submission are agreed upon. For vaccines, the Vaccine Adverse Event Reporting System should be consulted.[17] Foreign events may be submitted on a CIOMS form (the World Health Organization's Council for International Organizations of Medical Sciences),[8, 18, 19] or a letter can be generated that includes the relevant information in narrative format.

5. Coding

Coding AEs into a standard nomenclature should be done by trained experts to ensure accuracy and consistency. Reporters, patients, health care providers, and registry personnel should do their best to capture the primary data clearly, completely, and in as "natural" clinical language as possible. Since reporters may use different verbatim terms to describe the same event, it is recommended that sponsors apply coding conventions to code the verbatim terms. The Medical Dictionary for Regulatory Activities (MedDRA®) is customarily used throughout the product development cycle and as part of pharmacovigilance; however, other coding systems are also used. For example, SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) is used instead of MedDRA in some electronic health records. Coding the different verbatim language to preferred terms allows similar events to be appropriately grouped, creates consistency among the terms for evaluation, and maximizes the likelihood that safety signals will be detected.

Sponsors or their designees should review the accuracy of the coding of verbatim AEs into appropriate terms. If coding is performed by someone other than the sponsor, any applicable coding conventions associated with the underlying condition or product should be shared. Review of the coding process should focus on terms that do not accurately communicate the severity or magnitude of the AE or possibly mischaracterize the AE. Review of the coded terms compared with reported verbatim terms should be performed in order to ensure consistency and accuracy of the AE reporting and to minimize variability of coding of similar AE terms. Attention to consistency is especially important, as many different individuals may code AEs over time, and this situation contributes to variability in the coding process. In addition to monitoring AEs individually for complete clinical evaluation of the safety data, sponsors should consider grouping and analyzing clinically relevant coded terms that could represent similar toxicities or syndromes. Combining terms may provide a method of detecting less common and serious events that would otherwise be obscured. However, sponsors should be careful when combining related terms to avoid amplifying a weak signal or obscuring important overall findings when grouping is overly broad. In addition to monitoring individual AEs, sites and registry personnel should be attentive to toxicities that may cluster into syndromes.

6. Adverse Event Management

In some cases, such as when a safety registry is created as a condition of regulatory approval, a data safety monitoring board (DSMB), data monitoring committee (DMC), or adjudication committee may be established with the primary

role of periodically reviewing the data as they are generated by the registry. Such activities are generally discussed directly with the regulatory authorities, such as FDA. These authorities are typically involved in the design and critique of protocols for postapproval studies. Ultimately, registry planning and the registry protocol should anticipate and clearly delineate the roles, responsibilities, processes, forms, and lines of communication for AE reporting for sites, registry personnel, the DSMB, DMC, or adjudication committee if one exists, and the sponsoring organization. Documentation should be provided for definitions and approaches to determining what is considered unexpected and possibly related to drug or device exposure. The management of AE reporting should be clearly specified in the registry protocol, including explanations of the roles, responsibilities, processes, and methods for handling AE reports by the various parties conducting the registry, and for performing followup activities with the site to ensure that complete information is obtained. Sponsors who are stakeholders in a registry should have a representative of their internal drug safety or pharmacovigilance group participate in the design and review of the registry protocol and have a role in the data collection and reporting process (discussed in Chapter 2) to facilitate appropriate and timely reporting and communication.

For postapproval studies financially sponsored by manufacturers, the overall company AE monitoring systems are usually operated by personnel experienced in drug safety (also referred to as pharmacovigilance, regulatory safety, product safety, and safety and risk management). If sites need to report or discuss an AE, they can call the contact number provided for the registry, and are then prompted to press a number if reporting an AE. This number then transfers them to drug safety surveillance so that they can interact directly with personnel in this division and bypass the registry coordinating group. These calls may or may not be tracked by the registry. Alternatively, the registry system can provide instructions to the site on how to report AEs directly to the sponsor's drug safety surveillance division. By this method, the sponsor provides a separate contact number for AE reporting (independent of the registry support staff) that places the site in direct contact with drug safety personnel. This process minimizes the possibility of duplicate AE reports and the potentially complicated reconciliation of two different systems collecting AE information. Use of this process is critical when dealing with products that are available via a registry system as well as outside of a registry system, and it allows sites to have one designated drug safety representative for interaction.

Sponsors of registries designed specifically for surveillance of product safety are strongly encouraged to hold discussions with the regulatory authorities when considering the design of the AE monitoring system. These discussions should be focused on the purpose of the registry, the "best fit" model for AE monitoring, and the timing of routine registry updates. With respect to internal operations chosen by the sponsor to support the requirements of an AE monitoring system, anecdotal feedback suggests that health authorities expect compliance with the agreed-upon requirements. Details regarding implementation are the responsibility of the sponsor.

It should also be noted that FDA's Proposed Rule for Safety Reporting Requirements for Human Drug and Biologics Products (68 FR 12406, March 14, 2003) suggests that the responsible point of contact for FDA should be provided for all expedited and periodic AE reports, and preferably, this individual should be a licensed physician. Although this proposed rule has never been finalized, the principle is similar to the Qualified Person for Pharmacovigilance (QPPV) in Europe, whereby a specific, qualified individual is identified to provide responses to health authorities, upon request, including those regarding AEs reported via the registry system. Updated pharmacovigilance regulations issued by the European Medicines Agency are expected to be implemented in July 2012.[20, 21]

7. Adverse Event Required Reporting for Registry Sponsors

The reporting requirements of the sponsor directly affect how registries should collect and report AEs. Sponsors that are regulated industries are subject to the requirements shown in Table 12–1. ICH guidelines describe standards for expedited reporting[5, 22] and provide recommendations for periodic safety update reports[23] that are generally accepted globally.

Table 12–1. Overview of serious adverse event reporting requirements for marketed products

Type of Requirement	Drug and Biologics	Devices
U.S. postmarketing regulations	Primary: 21 CFR 314.80 (drugs), 21 CFR 600.80 (biologics) Other: 21 CFR 310.305, 21 CFR 314.98	21 CFR 803.20
Required reporting source	Regulated industries	Manufacturer, importer, user facility
Required reports	Serious, unexpected, and with a reasonable possibility of being related to drug exposure (with some exceptions)	Death or serious injury; device malfunction
Alternative reports	Not applicable	Summary reports (periodic line-listing of reports of well-known events)
Timeframe for reporting	15 calendar days for expedited reports	5 workdays, 10 workdays, or 30 calendar days, depending on the source and action required
Standard reporting form	MedWatch 3500A (for mandatory reporting required of a regulated industry) MedWatch 3500 (for voluntary reporting by health care professionals, consumers, and patients)	
Web sites	http://www.fda.gov/Safety/MedWatch/default.htm	http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm#1

Requirements for regulated industries that sponsor or financially support a registry include expedited reporting of serious and unexpected AEs made known to them via spontaneous reports. For registries, the 15-calendar-day notification applies if the regulated industry believes there is a reasonable possibility that the unexpected SAE was causally related to product exposure. Best practices for international reporting are that all "affiliates" of a sponsor report serious, unexpected, and possibly related events to the sponsor in a timely fashion, ideally within 2 calendar days; this allows the sponsor, in turn, to complete notification to the responsible regulatory authority within a total of 15 calendar days. Events that do not meet the requirements of expedited reporting (such as nonserious events or serious events considered expected or not related) may require submission through inclusion in an appropriate safety update, such as the New Drug Application or Biologic Licensing Application Annual Report, Periodic Report, or Periodic Safety Update Report, as applicable.[4] In many cases, sponsors are also required to provide registry safety updates to the health authority. Thus, sponsors may coordinate registry safety updates (i.e., determining the date

for creating the dataset—the data cutoff date) with the timing of the New Drug Application Annual Report, Periodic Report, Periodic Safety Update Report, or other agreed-upon periodic reporting format. Devices, however, have different reporting requirements (see http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm). In any event, sponsors should discuss safety reporting requirements for their specific registries with the applicable health authorities (such as FDA and European Medicines Agency) before finalizing their registry protocol.

In some cases, a registry sponsor may encourage sites to systematically report all potential SAEs to the sponsor. Given the potential for various assessments by different sites of the seriousness and relatedness of a particular AE—and therefore, inconsistency across sites in the evaluation of a particular AE—this method has certain advantages. In addition, expectedness is not always a straightforward assessment, and the expectedness of events can have significant variability depending on the local approved product labeling. For this reason, it is important that this determination be made by the sponsor and not the reporter of the event. Although this approach may result in substantially greater demands on the sponsor to evaluate all reports, it helps ensure compliance and avoid underreporting. Furthermore, sponsors must make their own assessments regarding the causality of individual solicited events. This requirement typically does not affect the need for reporting, but allows the sponsor to provide its own evaluation in the full context of the safety database. For these reasons, planning for high-quality and consistent training in AE reporting requirements across sites is the preferred approach for a patient registry.

Regardless of who assesses presumed relatedness, sponsors should be prepared to manage the increased volume of AE reports, and sponsors' registry staff should be trained to understand company policy and regulations on AE reporting in order to ensure compliance with local regulations. This training includes the ability to identify and evaluate the attributes of each AE and determine whether the AE should be reported to the health authority in keeping with local regulation. Sponsors are encouraged to appoint a health care practitioner to this role in order to ensure appropriate assessment of the characteristics of an AE.

When biopharmaceutical or device companies are not sponsoring, financially supporting, or participating in a registry in any way, AE reporting is dependent upon the "become aware" principle. If any agent or employee of the company receives information regarding an AE report, the agent or employee must document receipt and comply with internal company policy and regulatory requirements regarding AE reporting, to ensure compliance with applicable drug and device regulations.

8. Special Case: Risk Evaluation and Mitigation Strategies (REMS)

Under the FDA Amendments Act (2007), FDA established a legally enforceable new framework for risk management of products with known safety concerns, called Risk Evaluation and Mitigation Strategies (REMS).[6, 10, 24] The purpose of REMS is to ensure that the benefits of a particular drug outweigh the risks. New REMS programs can be imposed by FDA during clinical development, as part of the approval process, or at any time postapproval, should a new safety signal be identified. Although each REMS is customized depending on the product and associated safety issues, potential components include some combination of—

- A medication guide and/or patient package insert. Medication guides are informational packets distributed with some prescription drugs, which provide important information to patients about possible side effects and drug-drug interactions. The FDA has indicated the situations in which a medication guide is required to be available and distributed to patients.[25] A medication guide alone can and frequently does constitute a REMS.
- A communication plan that specifies targeted education and outreach for physicians, pharmacists, and patients.

- Elements to assure safe use (ETASU), in some cases. ETASU may include restriction of prescribing to health care providers with particular training, experience, and certification; dispensing of the drug in restricted settings; documentation of safe use conditions (such as laboratory results or specific patient monitoring); and registries.[24]

Unlike the less structured disease or exposure registries discussed above, a restricted-access system associated with an ETASU is designed for approved products that have particular risk-benefit profiles that require more careful controls. The purpose of ETASU is to mitigate a certain known drug-associated risk by ensuring that product access is tightly linked to some preventive and/or monitoring measure. Examples include systems that monitor laboratory values, such as white blood cell counts during clozapine administration to prevent severe leucopenia, or routine pregnancy testing during thalidomide administration to prevent in utero exposure of this known teratogenic compound. When these programs include registries, the registries often prospectively collect a battery of information using standardized instruments.

Data collection under ETASU may carry special AE reporting requirements, and as a result of the extensive contact with a variety of potential sources of safety information (e.g., pharmacists and patients), care should be taken to identify all possible routes of reporting. If special requirements exist, they should be made explicit in the registry protocol, with clear definitions of roles, responsibilities, and processes. Training of involved health care providers, such as physicians, nurses, and pharmacists, can be undertaken with written instructions, via telephone or with face-to-face counseling. Training of these health care providers should also extend beyond AE reporting to the specific requirements of the program in question. Such training may include the intended use and associated risk of the product, appropriate patient enrollment, and specific patient monitoring requirements, including guidelines for product discontinuation and management of AEs, as well as topics to cover during comprehensive counseling of patients. The objectives of the ETASU system and overall REMS should be clearly stated (e.g., prevention of in utero exposure during therapy via routine pregnancy testing), and registration forms that document the physician's and pharmacist's attestation of their commitment to requirements of the patient registry system should be completed prior to prescribing or dispensing the product.

9. Reporting Breaches of Confidentiality or Other Risks

In addition to addressing regulatory responsibilities for reporting adverse events, registries must also understand regulatory and ethical requirements and expectations regarding breaches of confidentiality or the reporting of other risks to patients that may arise during the course of a registry. The Health Information Technology for Economic and Clinical Health Act (HITECH Act) requires HIPAA-covered entities (entities covered by the Health Insurance Portability and Accountability Act of 1996) and their business associates to provide notification following a breach of unsecured protected health information.[26] See Chapter 7 for a detailed discussion of the HITECH Act. State breach notification laws may also apply to registry data.

Beyond these legal requirements, registries should establish clear notification procedures for breaches of confidentiality or other risks that become known during the course of the registry, whether or not they are governed by HIPAA or subject to State laws.

References for Chapter 12

1. Baim DS, Mehran R, Kereiakes DJ, et al. Postmarket surveillance for drug-eluting coronary stents: a comprehensive approach. Circulation. 2006 Feb 14;113(6):891-7. PMID: 16476863.
2. U.S. Food and Drug Administration. Medical Device Reporting. http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm. Accessed August 15, 2012.

3. Gross TP, Witten CM, Uldriks C, et al. A view from the US Food and Drug Administration. In: Frank E. Johnson KSV, ed. The Bionic Human: Health Promotion for People with Implanted Prosthetic Devices. Totowa, New Jersey: Humana Press, Inc.; 2006:5. p. 61-87.
4. 21 CFR § 314.80 (2008). Revised April 1, 2011.
5. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH E2A: Clinical safety data management: definitions and standards for expedited reporting. 27 October 1994. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf. Accessed September 4, 2013.
6. U.S. Food and Drug Administration. Guidance for Industry: Establishing Pregnancy Exposure Registries. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071639.pdf. Accessed August 15, 2012.
7. U.S. Food and Drug Administration. Postmarketing Adverse Experience Reporting for Human Drug and Licensed Biological Products: Clarification of What to Report. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071981.pdf. Accessed August 15, 2012.
8. 21 CFR § 312.32 (2008). Revised April 1, 2011.
9. 21 CFR § 803 (2008).
10. Food and Drug Administration Amendments Act of 2007., Pub. L. No. 110-85 (2007), Title VI. Sect. 603. Critical Path Public-Private Partnerships.
11. U.S. Food and Drug Administration. Reporting Serious Problems to the FDA. http://www.fda.gov/Safety/MedWatch/HowToReport/default.htm. Accessed August 15, 2012.
12. Dreyer NA, Sheth N, Trontell A, et al. Good practices for handling adverse events detected through patient registries. Drug Information Association Journal. 2008 2008;5(42):421-8.
13. U.S. Food and Drug Administration. Guidance for Industry. Good pharmacovigilance practices and pharmacoepidemiologic assessment. March 2005. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126834.pdf. Accessed August 14, 2012.
14. U.S. Food and Drug Administration. General Instructions for Completing the Internet MedWatch Form. https://www.accessdata.fda.gov/scripts/medwatch/medwatch-online.htm. Accessed August 15, 2012.
15. U.S. Food and Drug Administration. MedWatch Form FDA 3500 - Voluntary Reporting. http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Forms/UCM163919.pdf. Accessed September 4, 2013.
16. U.S. Food and Drug Administration. Guidance for Industry: MedWatch Form FDA 3500A: Mandatory Reporting of Adverse Reactions Related to Human Cells, Tissues, and Cellular and Tissue-Based Products (HCT/Ps). http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Tissue/ucm074000.htm. Accessed August 15, 2012.
17. Vaccine Adverse Event Reporting System. http://vaers.hhs.gov. Accessed August 15, 2012.
18. Council for International Organizations of Medical Sciences. CIOMS Form. http://www.cioms.ch/index.php/cioms-form-i. Accessed September 4, 2013.
19. 21 CFR § 314.80(f)(1) (2008).
20. EU Directive 2010/84/EU. 15 December 2010. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:348:0074:0099:EN:PDF. Accessed August 15, 2012.
21. EU Regulation No 1235/2010. 15 December 2010. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:348:0001:0016:EN:PDF. Accessed August 15, 2012.
22. ICH Topic E2D. Post Approval Safety Data Management. European Medicines Agency; May2004. CPMP/ICH/3945/03. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2D/Step4/E2D_Guideline.pdf. Accessed August 15, 2012.
23. ICH E2C R1: Clinical Safety Data Management: Periodic Updated Safety Reports for Marketed Drugs. European Medicines Agency; Jun1997. CPMP/ICH/288/95. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002780.pdf. Accessed September 4, 2013.

24. U.S. Food and Drug Administration. DRAFT Guidance for Industry: Format and content of proposed risk evaluation and mitigation strategies (REMS), REMS assessments, and proposed REMS modifications. September 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM184128.pdf. Accessed August 15, 2012.

25. U.S. Food and Drug Administration. Guidance: Medication Guides – Distribution Requirements and Inclusion in Risk Evaluation and Mitigation Strategies (REMS). http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM244570.pdf. Accessed August 15, 2012.

26. 74 F.R. 42740 (August 24, 2009).

Chapter 13. Analysis, Interpretation, and Reporting of Registry Data To Evaluate Outcomes

1. Introduction

Registries have the potential to produce databases that are an important source of information regarding health care patterns, decisionmaking, and delivery, as well as the subsequent association of these factors with patient outcomes. Registries, for example, can provide valuable insight into the safety and/or effectiveness of an intervention or the efficiency, timeliness, quality, and patient centeredness of a health care system. The utility and applicability of registry data rely heavily on the quality of the data analysis plan and its users' ability to interpret the results. Analysis and interpretation of registry data begin with a series of core questions:

- *Study purpose*: Were the objectives/hypotheses predefined or post hoc?
- *Patient population*: Who was studied?
- *Data quality*: How were the data collected, reviewed, and verified?
- *Data completeness*: How were missing data handled?
- *Data analysis*: How were the analyses chosen and performed?

While registry data present many opportunities for meaningful analysis, there are inherent challenges to making appropriate inferences. A principal concern with registries is that of making inferences without regard to the quality of data, since quality standards have not been previously well established or consistently reported. In some registries, comparison groups may not be robustly defined, and information provided about the external validity of a registry sample is often limited. These factors must be considered when making inferences based on analyses of registry data.[1]

This chapter explains how analysis plans are constructed for registries, how they differ depending on the registry's purpose, and how registry design and conduct can affect analysis and interpretation. The analytic techniques generally used for registry data are presented, addressing how conclusions may be drawn from the data and what caveats are appropriate. The chapter also describes how timelines for data analysis can be built in at registry inception and how to determine when the registry data are complete enough to begin analysis.

2. Hypotheses and Purposes of the Registry

While it may be relatively straightforward to develop hypotheses for registries intended to evaluate safety and effectiveness, not all registries have specific, testable, or simple hypotheses. Disease registries commonly have aims that are primarily descriptive, such as describing the typical clinical features of individuals with a disease, variations in phenotype, and the clinical progression of the disease over time (i.e., natural history). These registries play a particularly important role in the study of rare diseases.

In the case of registries where the aim is to study the associations between specific exposures and outcomes, prespecification of the study methodology and presence or absence of a priori hypotheses or research questions may affect the acceptance of results of studies derived from registry data. The many possible scenarios are well illustrated by examples at the theoretical extremes.

On one extreme, a study may evolve out of a clear and explicit prespecified research question and hypothesis. In such a study, there may have been preliminary scientific work that laid the conceptual foundation and plausibility for the proposed study. The investigators fully articulate the objectives and analytic plan before embarking on any analysis. The outcome is clearly defined and the statistical approach documented. Secondary analyses are identified and may be highlighted as hypothesis generating. The investigators have no prior knowledge of analyses in this database that would bias them in the formulation of their study objective. The study is conducted and published

regardless of the result. The paper states clearly that the objective and hypothesis were prespecified. For registries intended to support national coverage determinations with data collection as a condition of coverage, the specific coverage decision question may be specified a priori as the research question in lieu of a formal hypothesis.

At the other extreme, a study may evolve out of an unexpected observation in a database in the course of doing analyses for another purpose. A study could also evolve from a concerted effort to discover associations—for example, as part of a large effort to understand disease causation. In such a study, the foundation for the study is developed post hoc, or after making the observation. Because of the way in which the observation was found, the rationale for the study is developed retrospectively. The paper publishing this study's results does not clearly state that the objective and hypothesis were not prespecified.

Of course, many examples fall between these extremes. An investigator may suspect an association for many variables but find the relationship for only one of them. The investigator decides to pursue only the positive finding and develop a rationale for a study or grant. The association was sought, but it was sought along with associations for many other variables and outcomes.

Thus, while there is substantial debate about the importance of prespecified hypotheses,[2, 3] there is general agreement that it is informative to reveal how the study was developed. Transparency in methods is needed so that readers may know whether these studies are the result of hypotheses developed independently of the study database, or whether the question and analyses evolved from experience with the database and multiple iterations of exploratory analyses. Both types of studies have value.

3. Patient Population

The purpose of a registry is to provide information about a specific patient population to which all study results are meant to apply. To determine how well the study results apply to the target population, five populations, each of which is a subset of the preceding population, need to be considered, along with how well each population represents the preceding population. These five subpopulations are shown in Figure 13–1.

The *target population* is defined by the study's purpose. To assess the appropriateness of the target population, one must ask the question, "Is this really the population that we need to know about?" For example, the target population for a registry of oral contraceptive users would include women of childbearing age who could become pregnant and are seeking to prevent pregnancy. Studies often miss important segments of the population in an effort to make the study population more homogeneous. For example, a study to assess a medical device used to treat patients for cardiac arrhythmias that defines only men as its target population would be less informative than it could be, because the device is designed for use in both men and women.

The *accessible population* is defined using inclusion criteria and exclusion criteria. The inclusion criteria define the population that will be used for the study and generally include geographic (e.g., hospitals or clinics in the New England region), demographic, disease-specific, and temporal (e.g., specification of the included dates of hospital or clinic admission), as well as other criteria. Conversely, the exclusion criteria seek to eliminate specific patients from study and may be driven by an effort to assure an adequate-sized population of interest for analysis. The same goals may be said of inclusion criteria, since it is difficult to separate inclusion from exclusion criteria (e.g., inclusion of adults aged 18 and older vs. exclusion of children younger than 18).

The accessible population may lose representativeness to the extent that convenience plays a part in its determination, because people who are easy to enroll in the registry may differ in some critical respects from the population at large. Similarly, to the extent that homogeneity plays a part in determining the accessible population, it is less likely to be representative of the entire population because certain population subgroups will be excluded.

Factors to be considered in assessing the accessible population's representativeness of the target population include all the inclusion and exclusion criteria mentioned above. One method of evaluating representativeness is to describe the demographics and other key descriptors of the registry study population and to contrast its composition with patients with similar characteristics who are identified from an external database, such as might be obtained from health insurers, health maintenance organizations, or the U.S. Surveillance Epidemiology and End Results (SEER) cancer registries.[4]

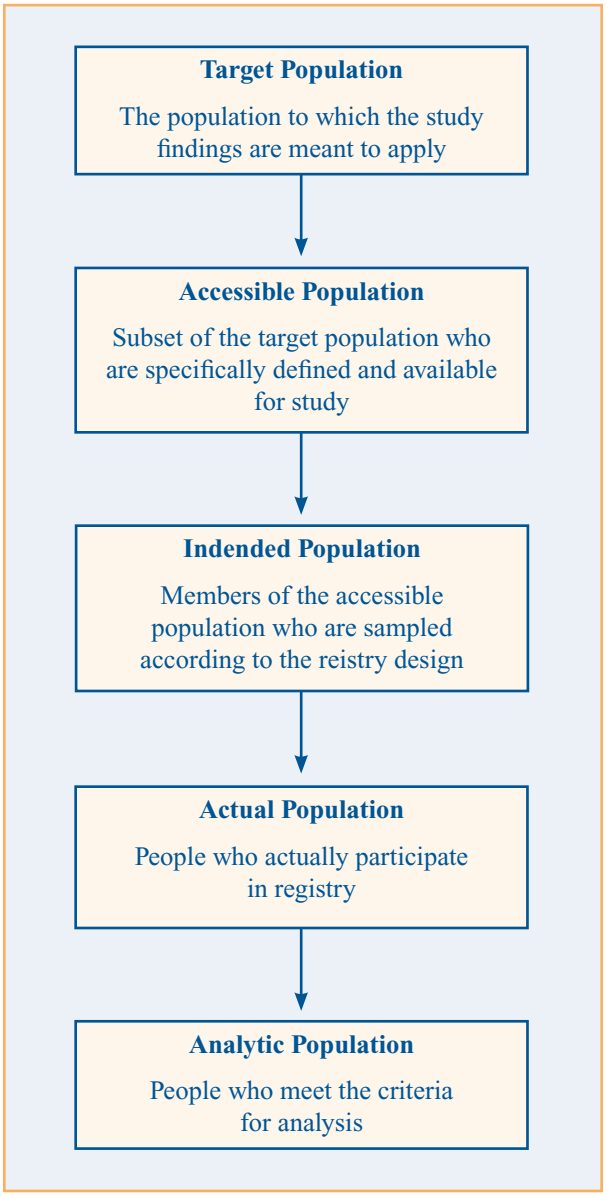
However, simple numerical/statistical representativeness is not the main issue. Representativeness should be evaluated in the context of the purpose of the study—that is, whether the study results can reasonably be generalized or extrapolated to other populations of interest outside of those included in the accessible population. (See Case Example 26.) For example, suppose that the purpose of the study is to assess the effectiveness of a drug in U.S. residents with diabetes. If the accessible population includes no children, then the study results may not apply to children, since children often metabolize drugs very differently from adults.

On the other hand, consider the possibility that the accessible population is generally drawn from a geographically isolated region, whereas the target population may be the entire United States or the world. In that case, the accessible population is not geographically representative of the target population, but that circumstance would have little or no impact on the representativeness of the study findings to the target population if the action of the drug (or its delivery) does not vary geographically (which we would generally expect to be the case, unless pertinent racial/genetic or dietary factors were involved). Therefore, in this example, the lack of geographical representativeness would not affect interpretation of results.

The reason for using an *intended population* rather than the whole accessible population for the study is simply a matter of convenience and practicality. The issues to consider in assessing how well the intended population represents the accessible population are similar to those for assessing how well the accessible population represents the target population. The main difference is that the intended population may be specified by a sampling scheme, which often tries to strike a balance among representativeness, convenience, and budget. If the intended population is a random sample of the accessible population, it may be reasonably assumed that it will represent the accessible population; however, for many, if not most, registries, a complete roster of the accessible population does not exist. More commonly, the intended population is compared with the accessible population in terms of pertinent variables.

To the extent that convenience or other design (e.g., stratified random sample) is used to choose the intended population, one must consider the extent to which the sampling of the accessible population—by means other than random sampling—has decreased the representativeness of the intended population. For example, suppose that, for the sake of convenience, only patients who attend clinic on Mondays are included in the study. If patients who attend clinic on Mondays are similar in every relevant respect to other patients, that may not constitute a limitation. But if Monday patients are substantially different from patients who attend clinic on other days of the week (e.g., well-baby clinics are held on Mondays) and if those differences affect the outcome that is being studied (e.g., proportion of baby visits for "well babies"), then that sampling strategy would substantially alter the interpretations from the registry and would be considered a meaningful limitation.

Figure 13–1. Patient populations



The extent to which the *actual population* is not fully representative of the intended population is generally a matter of real-world issues that prevent the initial inclusion of study subjects or adequate followup. In assessing representativeness, one must consider the likely underlying factors that caused those subjects not to be included in the analysis of study results and how that might affect the interpretations from the registry. For example, consider a study of a newly introduced medication, such as an anti-inflammatory drug that is thought to be as effective as other products and to have fewer side effects but that is more costly. Inclusion in the actual population may be influenced by prescribing practices governed by a health insurer. For example, if a new drug is approved for reimbursement only for patients who have "failed" treatment with other anti-inflammatory products, the resulting actual population will be systematically different from the target population of potential anti-inflammatory drug users. The actual population may be refractory to treatment or may have more comorbidities (e.g., gastrointestinal problems), and may be specifically selected for treatment beyond the intention of the study-specified inclusion criteria. In fact, registries of newly introduced drugs and devices may often include patients who are different from the ultimate target population.

Finally, the *analytic population* includes all those patients who meet the criteria for analysis. In some cases, it becomes apparent that there are too few cases of a particular type, or too few patients with certain attributes, such that these subgroups do not contribute enough information for meaningful analysis. Patients may also be excluded from the analysis population because their conditions are so rare that to include them could be considered a breach of patient confidentiality. Analytic populations are also created to meet specific needs. For example, an investigator may request a data set that will be used to analyze a subset of the registry population, such as those who had a specific treatment or condition.

A related issue is that of *early adopters*,[5] in which practitioners who are quick to use a novel health care intervention or therapy differ from those who use it only once it is well established. For example, a registry of the use of a new surgical technique may initially enroll largely academic physicians and only much later enroll community-based surgeons. If the outcomes of the technique differ between the academic surgeons (early adopters) and community-based surgeons (later adopters), then the initial results of the registry may not reflect the true effectiveness of the technique in widespread use. Patients selected for treatment with a novel therapy may also differ with regard to factors such as severity or duration of disease and prior treatment history, including treatment

failures. For example, patients with more severe or late-stage disease who have failed other treatments might be more likely to use a newly approved product that has shown efficacy in treating their condition. Later on, patients with less severe disease may start using the product.

Finally, patients who are included in the analytic population for a given analysis of registry data may also be subject to selection or inclusion criteria (admissibility criteria), and these may affect interpretation of the resulting analyses. (See Chapter 18.) For example, if only patients who remain enrolled and attend followup visits through 2 years after study initiation are included in analysis of adherence to therapy, it is possible or likely that adherence among those who remain enrolled in the study and have multiple followup visits will be different from adherence among those who do not. Differential loss to followup, whereby patients who are lost may be more likely to experience adverse outcomes, such as mortality, than those who remain under observation, is a related issue that may lead to biased results. (See Chapter 3.)

4. Data Quality Issues

In addition to a full understanding of study design and methodology, analysis of registry events and outcomes requires an assessment of data quality. One must consider whether most or all important covariates were collected, whether the data were complete, and whether the problem of missing data was handled appropriately, as well as whether the data are accurate.

4.1 Collection of All Important Covariates

While registries are generally constructed for a particular purpose or purposes, registry information may be collected for one purpose (e.g., provider performance feedback) and then used for another (e.g., addressing a specific clinical research question). When using an available database for additional purposes, one needs to be sure that all the information necessary to address a specific research question was collected in a manner that is sufficient to answer the question.

For example, suppose the research question addresses the comparative effectiveness of two treatments for a given disease using an existing registry. To be meaningful, the registry should have accurate, well-defined, and complete information, including potential confounding and effect-modifying factors; population characteristics of those with the specified disease; exposures (whether patients received treatment A or B); and patient outcomes of interest. Confounding factors are variables that influence both the exposure (treatment selection) and the outcome in the analyses. These factors can include patient factors (age, gender, race, socioeconomic factors, disease severity, or comorbid illness); provider factors (experience, skills); and system factors (type of care setting, quality of care, or regional effects). While it is not possible to identify all confounding factors in planning a registry, it is desirable to give serious thought to what will be important and how the necessary data can be collected. While effect modification is not a threat to validity, it is important to consider potential effect modifiers for data collection and analysis in order to evaluate whether an association varies within specific subgroups.[6] Analysis of registries requires information about such variables so that the confounding covariates can be accounted for, using one of several analytic techniques covered in upcoming sections of this chapter. In addition, as described in Chapter 3, eligibility for entry into the registry may be restricted to individuals within a certain range of values for potential confounding factors in order to reduce the effects of these factors. Such restrictions may also affect the generalizability of the registry.

4.2 Data Completeness

Assuming that a registry has the necessary data elements, the next step is to ensure that the data are complete. Missing data can be a challenge for any registry-based analysis. Missing data include situations in which a variable is directly reported as missing or unavailable, a variable is "nonreported" (i.e., the observation is blank), the reported data may not be interpretable, or the value must be imputed to be missing because of data inconsistency or out-of-range results. Before analyzing a registry database, the database should

be "cleaned" (discussed in Chapter 11, Section 2.5.), and attempts should be made to obtain as much missing data as realistically possible from source documents. Inconsistent data (e.g., a "yes" answer to a question at one point and "no" to the same question at another) and out-of-range data (e.g., a 500-year-old patient) should be corrected when possible. Finally, the degree of data completeness should be summarized for the researcher and eventual consumer of analyses from the registry. Detailed examples of sources of incomplete data are described in Chapter 18.

4.3 Missing Data

The intent of any analysis is to make valid inferences from the data. Missing data can threaten this goal both by reducing the information yield of the study and, in many cases, by introducing bias. A thorough review of types of missing data with examples can be found in Chapter 18. Briefly, the first step is to understand which data are missing. The second step is to understand why the data are missing (e.g., missing item-response or right censoring). Finally, missing data fall into three classic categories of randomness:[7]

- *Missing completely at random (MCAR)*: Instances where there are no differences between subjects with missing data and those with complete data. In such random instances, missing data only reduce study power without introducing bias.
- *Missing at random (MAR)*: Instances where missing data depend on known or observed values but not unmeasured data. In such cases, accounting for these known factors in the analysis will produce unbiased results.
- *Missing not at random (MNAR)*: Here, missing data depend on events or factors not measured by the researcher and thus potentially introduce bias.

To gain insight into which of the three categories of missing data are in play, one can compare the distribution of observed variables for patients with specific missing data to the distribution of those variables for patients for whom those same data are present.

While pragmatically it is difficult to determine whether data are MCAR or MAR, there are, nonetheless, several means of managing missing data within an analysis. For example, a *complete case* strategy limits the analysis to patients with complete information for all variables. This is the default strategy used in many standard analytic packages (e.g., SAS, Cary, NC). A simple deletion of all incomplete observations, however, is not appropriate or efficient in all circumstances, and it may introduce significant bias if the deleted cases are substantively different from the retained, complete cases (i.e., not MCAR). In observational studies with prospective, structured data collection, missing data are not uncommon, and the complete case strategy is inefficient and not generally used. For example, patients with diabetes who were hospitalized because of inadequate glucose control might not return for a scheduled followup visit at which HbA1c was to be measured. Those missing values for HbA1c would probably differ from the measured values because of the reason for which they were missing, and they would be categorized as MNAR. In an example of MAR, the availability of the results of certain tests or measurements may depend on what is covered by patients' health insurance (a known value), since registries do not typically pay for testing. Patients without this particular measurement may still contribute meaningfully to the analysis. In order to include patients with missing data, one of several *imputation* techniques may be used to estimate the missing data.

Imputation is a common strategy in which average values are substituted for missing data using strategies such as *unconditional* and *conditional mean*, *multiple hot-deck*, and *expectation maximum*, among others.[7, 8] For data that are captured at multiple time points or repeated measures, investigators often "carry forward" a last observation. However, such a technique can be problematic if early dropouts occur and a response variable is expected to change over time or when the effect of the exposure (or treatment) is intermittent. *Worst-case* imputation is another means of substitution in which investigators test the sensitivity of a finding by substituting a worst-case value for all missing results. While this

is conservative, it offers a *lower bound* to an association rather than an accurate assessment. One particular imputation method that has received significant attention in recent analyses has been termed *multiple imputation*. Rubin first proposed the idea to impute more than one value for a missing variable as a means of reflecting the uncertainty around this value.[9] The general strategy is to replace a missing value with multiple values from an approximate distribution for missing values. This produces multiple complete data sets for analysis from which a single summary finding is estimated.

There are several issues concerning how prognostic models for decisionmaking can be influenced by data completeness and missing data.[10] Burton and Altman reviewed 100 multivariable cancer prognostic models published in seven leading cancer journals in 2002. They found that the proportion of complete cases was reported in only 39 studies, while the percentage missing for important prognostic variables was reported in 52 studies. Comparison of complete cases with incomplete cases was provided in 10 studies, and the methods used to handle missing data were summarized in 32 studies. The most common techniques used for handling missing data in this review article were (a) complete case analysis (12), (b) dropping variables with high numbers of missing cases from model consideration (6), and (c) using some simple author imputation rule (6). Only one study reported using multiple imputation. The reviewers concluded that there was room for improvement in the reporting and handling of missing data within registry studies

Readers interested in learning more about methods for handling missing data and the potential for bias are directed to other useful resources by Greenland and Finkle,[11] Hernán and colleagues,[12] and Lash, Fox, and Fink.[13]

It is important to keep in mind that the impact of data completeness will differ, depending on the extent of missing data and the intended use of the registry. It may be less problematic with regard to descriptive research than research intended to support decisionmaking. For all registries, it is important to have a strategy for how to identify and handle missing data as well as how to explicitly report on data completeness to facilitate interpretation of study results. For more information on other specific types of missing data issues (e.g., left truncation), please see Chapter 18.

4.4 Data Accuracy and Validation

While observational registry studies are usually not required to meet U.S. Food and Drug Administration and International Conference on Harmonisation standards of Good Clinical Practice developed for clinical trials, sponsors and contract research organizations that conduct registry studies are responsible for ensuring the accuracy of study data to the extent possible. Detailed plans for site monitoring, quality assurance, and data verification should be developed at the beginning of a study and adhered to throughout its lifespan. Chapter 11 discusses in detail approaches to data collection and quality assurance, including data management, site monitoring, and source data verification.

Ensuring the accuracy and validity of data and programming at the analysis stage requires additional consideration. The Office of Surveillance and Epidemiology (OSE) of the Food and Drug Administration's Center for Drug Evaluation and Research uses the manual *Standards of Data Management and Analytic Process in the Office of Surveillance and Epidemiology* for analyses of databases conducted within OSE; the manual addresses many of these issues and may be consulted for further elaboration on these topics.[14] Topics addressed that pertain to ensuring the accuracy of data just before and during analysis include developing a clear understanding of the data at the structural level of the database and variable attributes; creating analytic programs with careful documentation and an approach to variable creation and naming conventions that is straightforward and, when possible, consistent with the Clinical Data Interchange Standards Consortium initiative; and complete or partial verification of programming and analytic data set creation by a second analyst.

For more detail about validation substudies, please see Chapter 18.

5. Data Analysis

This section provides an overview of practical considerations for analysis of data from a registry. As the name suggests, a descriptive study focuses on describing frequency and patterns of various elements of a patient population, whereas an analytical study focuses on examining associations between patients or treatment characteristics and health outcomes of interest (e.g., comparative effectiveness).

Statistical methods commonly used for descriptive purposes include those that summarize information from continuous variables (e.g., mean, median) or from categorical variables (e.g., proportions, rates). Registries may describe a population using incidence (the proportion of the population that develops the condition over a specified time interval) and prevalence (the proportion of the population that has the condition at a specific point in time). Another summary estimate that is often used is an incidence rate. The incidence rate (also known as absolute risk) takes into account both the number of people in a population who develop the outcome of interest and the person-time at risk, or the length of time contributed by all people during the period when they were in the population and the events were counted.

For studies that include patient followup, an important part of the description of study conduct is to characterize how many patients are "lost," or drop out, during the course of the registry, at what point they are lost, and if they return. Lasagna plots are one convenient method to visually assess missing data over time when conducting a longitudinal analysis.[15] Figure 13–2 illustrates key points of information that provide a useful description of losses to followup and study dropouts.

For analytical studies, the association between a risk factor and outcome may be expressed as attributable risk, relative risk, odds ratio, or hazard ratio, depending on the nature of the data collected, the duration of the study, and the frequency of the outcome. Attributable risk, a concept developed in the field of public health and preventive medicine, is defined as the proportion of disease incidence that can be attributed to a specific exposure, and it may be used to indicate the impact of a particular exposure at a population level. The standard textbooks cited here have detailed discussions regarding epidemiologic and statistical methods commonly used for the various analyses supported by registries.[6, 16, 17,18, 19]

For analytical studies of data derived from observational studies such as registries, it is important to consider the role of confounding. Although those planning a study try to collect as much data as possible to address known confounders, there is always the chance that unknown confounders will affect the interpretation of analyses derived from observational studies. It is important to consider the extent to which bias (systematic error stemming from factors that are related to both the decision to treat and the outcomes of interest [confounders]) could have distorted the results. For example, selective prescribing (confounding by indication) results when people with more severe disease or those who have failed other treatments are more likely to receive newer treatments; these patients are systematically different from other patients who may be treated with the product under study. Misclassification in treatment can result from the patient's incorrect recall of dose, or poor adherence or treatment compliance. Other types of bias include detection bias[20] (e.g., when comparison groups are assessed at different points in time or by different methods), selective loss to followup in which patients with the outcomes of most interest (e.g., sickest) may be more likely to drop out of one treatment group than another, and performance bias (e.g., systematic differences in care other than the intervention under study, such as a public health initiative promoting healthy lifestyles directed at patients who receive a particular class of treatment).

Confounding may be evaluated using stratified analysis, multivariable analysis, sensitivity analyses, and simple or quantitative bias analysis.[12] Appropriate methods should be used to adjust for confounding. For example, if an exposure or treatment varies over time and the confounding variable also varies over time, traditional adjustment using conventional

multivariable modeling will introduce selection bias. Marginal structural models use inverse probability weighting to account for time-dependent confounding without introducing selection bias.[21] The extensive information and large sample sizes available in some registries also support use of more advanced modeling techniques for addressing confounding by indication, such as the use of propensity scores to create matched comparison groups, or for stratification or inclusion in multivariable risk modeling.[22-25] New methods also include the high-dimensional propensity score (hd-PS) for adjustment using administrative data.[26] The uptake of these approaches in the medical literature in recent years has been extremely rapid, and their application to analyses of registry data has also been broad. Examples are too numerous for a few selections to be fully representative, but registries in nearly every therapeutic area, including cancer,[27] cardiac devices,[28] organ transplantation,[29] and rare diseases,[30] have published the results of analyses incorporating approaches based on propensity scores. As noted in Chapter 3, instrumental variable methods present opportunities for assessing and reducing the impact of confounding by indication,[31-33] but verification of the assumptions are important to ensure that an instrument is valid.[34] Violations in the instrumental variable assumptions or the use of a weak instrument will lead to results more biased than those from conventional methods.[35] While a variety of methods have been developed to address confounding, particularly confounding by indication, residual confounding may still be present even after adjustment; therefore, these methods may not fully control for unmeasured confounding.[35] For specific examples of the application of these methods, please see Chapter 18. Information bias, such as misclassification, and selection bias are also threats to the validity of our findings and examples can be found in Chapter 18. For further information on how to quantify bias, please see Lash, Fox, and Fink.[13]

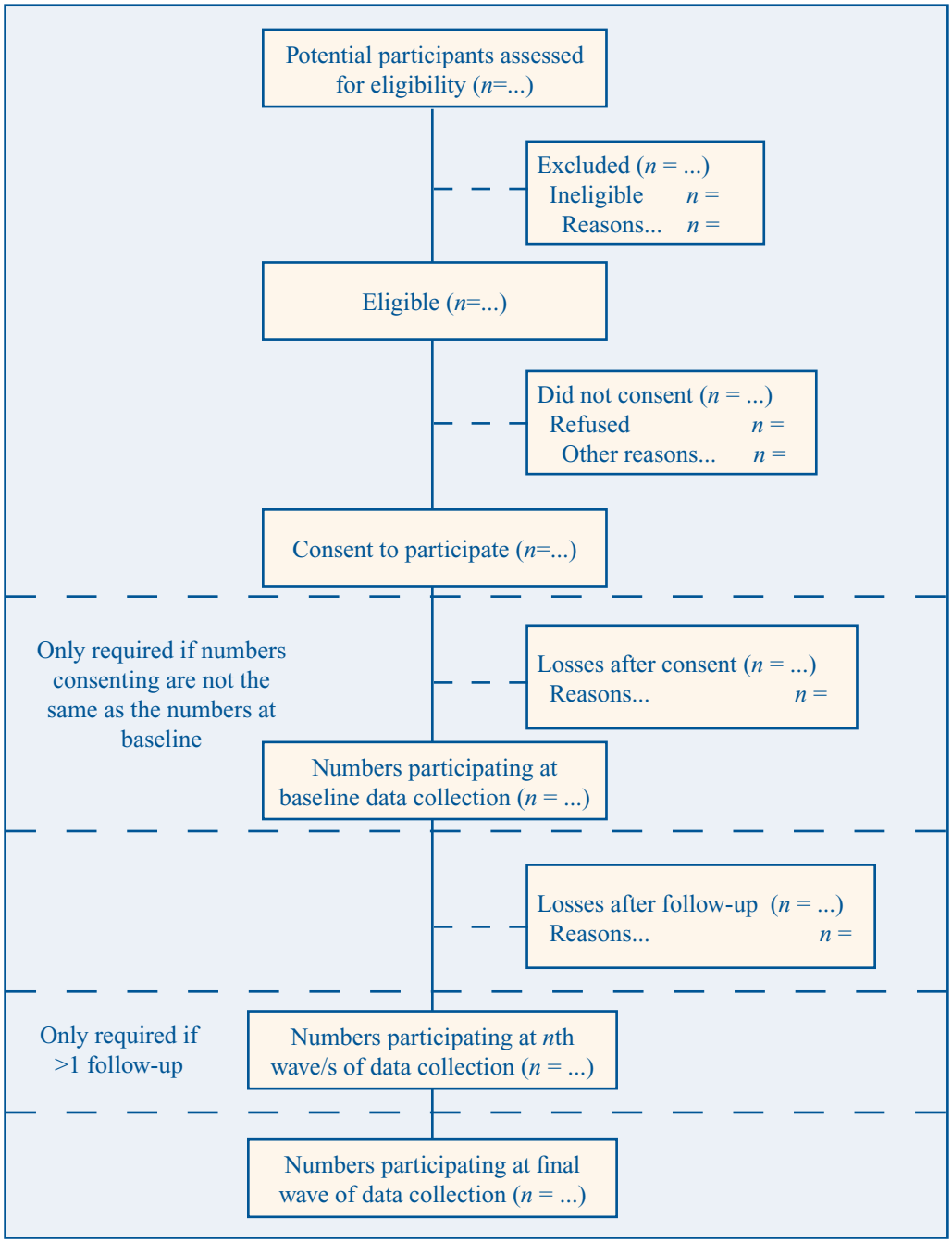
Groupings within a study population, such as patients seen by a single clinician or practice, residents of a neighborhood, or other "clusters," may themselves impact or predict health outcomes of interest. Such groupings may be accounted for in analysis through use of analytic methods including analysis of variance (ANOVA), and hierarchical or multilevel modeling.[36-39]

Heterogeneity of treatment effect is also an important consideration for comparative effectiveness research as the effect of a treatment may vary within subgroups of heterogeneous patients.[40] Stratification on the propensity score has been used to identify heterogeneity of treatment effect and may identify clinically meaningful differences between subgroups.

For economic analyses, the analytic approaches often encountered are cost-effectiveness analyses and cost-utility studies. To examine cost-effectiveness, costs are compared with clinical outcomes measured in units such as life expectancy or years of disease avoided.[41] Cost-utility analysis, a closely related technique, compares costs with outcomes adjusted for quality of life (utility) using measures known as quality-adjusted life years. Since most new interventions are more effective but also more expensive, another analytic approach examines the incremental cost-effectiveness ratio and contrasts that to the willingness to pay. (Willingness-to-pay analyses are generally conducted on a country-by-country basis, since various factors relating to national health insurance practices and cultural issues affect willingness to pay.) The use of registries for cost-effectiveness evaluations is a fairly recent development, and consequently, the methods are evolving rapidly. More information about economic analyses can be found in standard textbooks.[42-47]

It is important to emphasize that cost-effectiveness analyses, much like safety and clinical effectiveness analyses, require collection of specific data elements suited to the purpose. Although cost-effectiveness-type analyses are becoming more important and registries can play a key role in such analyses, registries traditionally have not collected much information on quality of life or resource use that can be linked to cost data.[48] To be used for cost-effectiveness analysis, registries must be developed with that purpose in mind.

Figure 13–2. The flow of participants into an analysis



Tooth L, Ware R, Bain C. Quality of reporting of observational longitudinal research. Am J Epidemiol 2005; 161(3):280–8. Reprinted with permission. Copyright restrictions apply. By permission of Oxford University Press on behalf of The Johns Hopkins Bloomberg School of Public Health.

5.1 Developing a Statistical Analysis Plan

5.1.1 Need for a Statistical Analysis Plan

It is important to develop a statistical analysis plan (SAP) that describes the analytical principles and statistical techniques to be employed in order to address the primary and secondary objectives, as specified in the study protocol or plan. Generally, the SAP for a registry study intended to support decisionmaking, such as a safety registry, is likely to be more detailed than the SAP for a descriptive study or health economics study. A registry may require a primary "master SAP" as well as subsequent, supplemental SAPs. Supplemental SAPs might be triggered by new research questions emerging after the initial master SAP was developed or might be needed because the registry has evolved over time (e.g., additional data collected, data elements revised). Although the evolving nature of data collection practices in some registries poses challenges for data analysis and interpretation, it is important to keep in mind that the ability to answer questions emerging during the course of the study is one of the advantages (and challenges) of a registry. In the specific case of long-term rare-disease registries, many of the relevant research questions of interest cannot be defined a priori but arise over time as disease knowledge and treatment experience accrue. Supplemental SAPs can be developed only when enough data become available to analyze a particular research question. At times, the method of statistical analysis may have to be modified to accommodate the amount and quality of data available. To the extent that the research question and SAP are formulated *before* the data analyses are conducted and results are used to answer specific questions or hypotheses, such supplemental analysis retains much of the intent of prespecification rather than being wide-ranging exploratory analyses (sometimes referred to as "fishing expeditions"). The key to success is to provide sufficient details in the SAP that, together with the study protocol and the case report forms, describe the overall process of the data analysis and reporting.

5.1.2 Preliminary Descriptive Analysis To Assist SAP Development

During SAP development, one particular aspect of a registry that is somewhat different from a randomized controlled study is the necessity to understand the "shape" of the data collected in the study by conducting a simple stratified analysis.[15] This may be crucial for a number of reasons.

Given the broad inclusion criteria that most registries tend to propose, there might be a wide distribution of patients, treatment, and/or outcome characteristics. The distribution of age, for example, may help to determine if more detailed analyses should be conducted in the "oldest old" age group (80 years and older) to help understand health outcomes in this subgroup that might be different from those of their younger counterparts.

Unless a registry is designed to limit data collection to a fixed number of regimens, the study population may experience many "regimens," considering the combination of various dose levels, drug names, frequency and timing of medication use (e.g., acute, chronic, intermittent), and sequencing of therapies. The scope and complexity of these variations constitute one of the most challenging aspects of analyzing a registry, since treatment is given at each individual physician's discretion. Grouping of treatment into regimens for analysis should be done carefully, guided by clinical experts in that therapeutic area. The full picture of treatment patterns may become clear only after a sizable number of patients have been enrolled. Consequently, the treatment definition in an SAP may be refined during the course of a study. Furthermore, there may be occasions where a particular therapeutic regimen is used in a much smaller number of patients than anticipated, so that specific study objectives focusing on this group of patients might become unfeasible. Also, the registry might have enrolled many patients who would normally be excluded from a clinical trial because of significant contraindications related to comorbidity or concomitant medication use. In this case, the SAP may need to define how these patients will be analyzed (either as a separate group or as part of the overall study population) and how these

different approaches might affect the interpretation of the study results.

There is a need to evaluate the presence of potential sources of bias and, to the extent feasible, use appropriate statistical measures to address such biases. For example, the bias known as *confounding by indication*[49] results from the fact that physicians do not prescribe medicine at random: the reason a patient is put on a particular regimen is often associated with their underlying disease severity and may, in turn, affect treatment outcome. (See Chapter 18 for more detailed discussion and examples.) To detect such a bias, the distribution of various prognostic factors at baseline is compared for patients who receive a treatment of interest and those who do not. A related concept is *channeling bias*, in which drugs with similar therapeutic indications are prescribed to groups of patients who may differ with regard to factors influencing prognosis.[50] To detect such a bias, registry developers and users must document the characteristics of the treated and untreated participants and either demonstrate their comparability or use statistical techniques to adjust for differences where possible. (Additional information about biases often found in registries is detailed in Chapter 3, Section 10.) In addition to such biases, analyses need to account for factors that are interrelated, also known as *effect modifiers*.[15] The presence of effect modification may also be identified after the data are collected. All of these issues should be taken into account in an SAP, based on understanding of the patient population in the registry.

5.2 Timing of Analyses during the Study

Unlike a typical clinical trial, registries, especially those that take several years to complete, may conduct intermediate analyses before all patients have been enrolled and/or all data collection has been completed. Such midcourse analyses may be undertaken for several reasons. First, many of these registries focus on serious safety outcomes. For such safety studies, it is important for all parties involved to actively monitor the frequency of such events at regular predefined intervals so that further risk assessment or risk management can be considered. The timing of such analyses may be influenced by regulatory requirements. Second, it may be of interest to examine treatment practices or health outcomes during the study to capture any emerging trends. Finally, it may also be important to provide intermediate or periodic analysis to document progress, often as a requirement for continued funding.

While it is useful to conduct such periodic analysis, careful planning should be given to the process and timing. The first questions are whether a sufficient number of patients have been enrolled and whether a sufficient number of events have occurred. Answers to both questions can be estimated based on the speed of enrollment and rate of patient retention, as well as the expected incidence rate of the event of interest. The second issue is whether sufficient time has elapsed after the initial treatment with a product so that it is biologically plausible for events to have occurred. (For example, some events, such as site reactions to injections, can be observed after a relatively short duration, compared with events like cancers, which may have a long induction or latency.) If there are too few patients or insufficient time has elapsed, premature analyses may lead to the inappropriate conclusion that there is no occurrence of a particular event. Similarly, uncommon events, occurring by random chance in a limited sample, may be incorrectly construed as a safety signal. However, it is inappropriate to delay analysis so long that an opportunity might be missed to observe emerging safety outcomes. Investigators should use sound clinical and epidemiological judgment when planning an intermediate analysis and, whenever possible, use data from previous studies to help to determine the feasibility and utility of such an analysis.

When planning the timing of the analysis, it may be helpful to consider substudies if emerging questions require data not initially collected. Substudies often involve data collection based on biological specimens or specific laboratory procedures. They may, for example, take the form of nested case-control studies. In other situations, a research question may be applicable only to a subset of patients, such as those who become pregnant while in the study. It may also be

desirable to conduct substudies among patients in a selected site or patient group to confirm the validity of study measurement. In such instances, a supplemental SAP may be a useful tool to describe the statistical principles and methods.

5.3 Factors To Be Considered in the Analysis

Registry results are most interpretable when they are specific to well-defined endpoints or outcomes in a specific patient population with a specific treatment status. Registry analyses may be more meaningful if variations of study results across patient groups, treatment methods, or subgroups of endpoints are reported. In other words, analysis of a registry should explicitly provide the following information:

- *Patient*: What are the characteristics of the patient population in terms of demographics, such as age, gender, race/ethnicity, insurance status, and clinical and treatment characteristics (e.g., past history of significant medical conditions, disease status at baseline, and prior treatment history)?
- *Exposure (or treatment)*: Exposure could be therapeutic treatment such as medication or surgery; a diagnostic or screening tool; behavioral factors such as alcohol, smoking habits, and diet; or other factors such as genetic predisposition or environmental factors. What are the distributions of the exposure in the population? Is the study objective specific to any one form of treatment? Is a new user design being used?[51] Does the exposure definition (index and reference group) and analysis avoid immortal-time bias?[52] Are there repeated measures or is the exposure intermittent?
- *Endpoints (or outcomes)*: Outcomes of interest may encompass effectiveness or comparative effectiveness, the benefits of a health care intervention under real-world circumstances,[53] and safety—the risks or harms that may be associated with an intervention. Examples of effectiveness outcomes include survival, disease recurrence, symptom severity, quality of life, and cost-effectiveness. Safety outcomes may include infection, sensitivity reactions, cancer, organ rejection, and mortality. Endpoints must be precisely defined at the data collection and analysis stages. Are the study data on all-cause mortality or cause-specific mortality? Is information available on pathogen-specific infection (e.g., bacterial vs. viral)? (See Case Example 27.) Are there competing risks?[54]
- *Covariates*: As with all observational studies, comparative effectiveness research requires careful consideration, collection, and analysis of important confounding and effect modifying variables. For medication exposures, are dose, duration, and calendar time under consideration? Directed acyclic graphs (DAGs) can be useful tools to illustrate how the exposure (or treatment), outcome and covariates are related.[55, 56]
- *Time*: For valid analysis of risk or benefit that occurs over a period of time following therapy, detailed accounting for time factors is required. For exposures, dates of starting and stopping a treatment or switching therapies should be recorded. For outcomes, the dates when followup visits occur, and whether or not they lead to a diagnosis of an outcome of interest, are required in order to take into account how long and how frequently patients were followed. Dates of diagnosis of outcomes of interest, or dates when patients complete a screening tool or survey, should be recorded. At the analysis stage, results must also be described in a time-appropriate fashion. For example, is an observed risk consistent over time (in relation to initiation of treatment) in a long-term study? If not, what time-related risk measures should be reported in addition to or instead of cumulative risk? When exposure status changes frequently, what is the method of capturing the population at risk? Many observational studies of intermittent exposures (e.g., use of nonsteroidal antiinflammatory drugs or pain medications) use time windows of analysis, looking at events following first use of a drug after a prescribed interval (e.g., 2 weeks) without drug use. Different analytic

approaches may be required to address issues of patients enrolling in a registry at different times and/or having different lengths of observation during the study period.

- *Potential for bias*: Successful analysis of observational studies also depends to a large extent on the ability to measure and analytically address the potential for bias. Refer to Chapter 3, Section 10 for a description of potential sources of bias. Directed acyclic graphs can also be useful for understanding and identifying the source of bias.[55, 56] Details and examples of quantification of bias can be found in Chapter 18. For details on how to quantify potential bias, see the textbook by Lash, Fox, and Fink.[13]

5.3.1. Choice of Comparator

An example of a troublesome source of bias is the choice of comparator. When participants in a cohort are classified into two or more groups according to certain study characteristics (such as treatment status, with the "standard of care" group as the comparator), the registry is said to have an *internal* or *concurrent* comparator. The advantage of an internal comparator design is that patients are likely to be more similar to each other, except for their treatment status, than patients in comparisons between registry subjects and external groups of subjects. When defining the comparator group, it is important not to introduce immortal time bias.[52] In addition, consistency in measurement of specific variables and in data collection methods make the comparison more valid. Internal comparators are particularly useful for treatment practices that change over time. Comparative effectiveness studies may often necessitate use of an internal comparator in order to maximize the comparability of patients receiving different treatments within a given study, and to ensure that variables required for multivariable analysis are available and measured in an equivalent manner for all patients to be analyzed.

Unfortunately, it is not always possible to have or sustain a valid internal comparator. For example, there may be significant medical differences between patients who receive a particularly effective therapy and those who do not (e.g., underlying disease severity or contraindications), or it may not be feasible to maintain a long-term cohort of patients who are not treated with such a medication. It is known that external information about treatment practices (such as scientific publications or presentations) can result in physicians changing their practice, such that they no longer prescribe the previously accepted standard of care. There may be a systematic difference between physicians who are early adopters and those who start using the drug or device after its effectiveness has been more widely accepted. Early adopters may also share other practices that differentiate them from their later-adopting colleagues.[5]

In the absence of a good internal comparator, one may have to leverage external comparators to provide critical context to help interpret data revealed by a registry. An external or historical comparison may involve another study or another database that has disease or treatment characteristics similar to those of registry subjects. Such data may be viewed as a context for anticipating the rate of an event. One widely used comparator is the U.S. SEER cancer registry data, because SEER provides detailed annual incidence rates of cancer stratified by cancer site, age group, gender, and tumor staging at diagnosis. SEER represents 28 percent of the U.S. population.[57] A procedure for formalizing comparisons with external data is known as *standardized incidence rate or ratio*;[15] when used appropriately, it can be interpreted as a proxy measure of risk or relative risk.

Use of an external comparator, however, may present significant challenges. For example, SEER and a given registry population may differ from each other for a number of reasons. The SEER data cover the general population and have no exclusion criteria pertaining to history of smoking or cancer screening, for example. On the other hand, a given registry may consist of patients who have an inherently different risk of cancer than the general population, resulting from the registry's having excluded smokers and others known to be

at high risk of developing a particular cancer. Such a registry would be expected to have a lower overall incidence rate of cancer, which, if SEER incidence rates are used as a comparator, may complicate or confound assessments of the impact of treatment on cancer incidence in the registry.

Regardless of the choice of comparator, similarity between the groups under comparison should not be assumed without careful examination of the study patients. Different comparator groups may result in very different inferences for safety and effectiveness evaluations; therefore, analysis of registry findings using different comparator groups may be used in sensitivity analyses or bias analyses to determine the robustness of a registry's findings. Sensitivity analysis refers to a procedure used to determine how robust the study result is to alterations of various parameters. If a small parameter alteration leads to a relatively large change in the results, the results are said to be sensitive to that parameter. Sensitivity and bias analyses may be used to determine how the final study results might change when taking into account those lost to followup. A simple hypothetical example is presented in Table 13–1.

Table 13–1. Hypothetical simple sensitivity analysis

Impact of Loss to Followup on Incidence Rates per 1,000 in a Study of 1,000 Patients in a Registry		
Various Assumptions of the Observed Incidence Rate	**Assuming a 10% Loss to Followup**	**Assuming a 30% Loss to Followup**
Assuming that the incidence of patients lost to followup is X times the rate of incidence estimated in those who stayed in the registry:		
X=0.5	106	94
X=1	111	110
X=2	122	143
X=3	156	242

Table 13–1 illustrates the extent of change in the incidence rate of a hypothetical outcome assuming varying degrees of loss to followup, and differences in incidence between those for whom there is information and those for whom there is no information due to loss to followup. In the first example, where 10 percent of the patients are lost to followup, the estimated incidence rate of 111/1,000 people is reasonably stable; it does not change too much when the (unknown) incidence in those lost to followup changes from 0.5 times the observed to 5 times the observed, with the corresponding incidence rate that would have been observed ranging from 106 to 156 per 1,000. On the other hand, when the loss to followup increases to 30 percent, the corresponding incidence rates that would have been observed range from 94 to 242. This procedure could be extended to a study that has more than one cohort of patients, with one being exposed and the other being nonexposed. In that case, the impact of loss to followup on the relative risk could be estimated by using sensitivity analysis. More examples are included in Chapter 18.

5.3.2 Patient Censoring

At the time of a registry analysis, events may not have occurred for all patients. For these patients, the data are said to be *censored*, indicating that the observation period of the registry was stopped before all events occurred (e.g., mortality). In these situations, it is unclear when the event will occur, if at all. In addition, a registry may enroll patients until a set stop date, and patients entered into the registry earlier will have a greater probability of having an event than those entered more recently because of the longer followup. An important assumption, and one that needs to be

assessed in a registry, is how patient prognosis varies with the time of entrance into the registry. This issue may be particularly problematic in registries that assess innovative (and changing) therapies. Patients and outcomes initially observed in the registry may differ from patients and outcomes observed later in the registry timeframe, either because of true differences in treatment options available at different points in time, or because of the shorter followup for people who entered later. Patients with censored data, however, contribute important information to the registry analysis. When possible, analyses should be planned so as to include all subjects, including those censored before the end of the followup period or the occurrence of an event. One method of analyzing censored data to estimate the conditional probability of the event occurring is to use the Kaplan-Meier method.[58] In this method, for each time period, the probability is calculated that those who have not experienced an event before the beginning of the period will still not have experienced it by the end of the period. The probability of an event occurring at any given time is then calculated from the product of the conditional probabilities of each time interval.

For information about right censoring and left truncation, please see Chapter 18.

6. Summary of Analytic Considerations

In summary, a meaningful analysis requires careful consideration of study design features and the nature of the data collected. Most typical epidemiological study analytical methods can be applied, and there is no one-size-fits-all approach. Efforts should be made to carefully evaluate the presence of biases and to control for identified potential biases during data analysis. This requires close collaboration among clinicians, epidemiologists, statisticians, study coordinators, and others involved in the design, conduct, and interpretation of the registry.

A number of biostatistics and epidemiology textbooks cover in depth the issues raised in this section and the appropriate analytic approaches for addressing them—for example, "time-to-event" or survival analyses[59] and issues of recurrent outcomes and repeated measures, with or without missing data,[60] in longitudinal cohort studies. Other texts address a range of regression and nonregression approaches to analysis of case-control and cohort study designs[61] that may be applied to registries.

7. Interpretation of Registry Data

Interpretation of registry data is needed so that the lessons from the registry can be applied to the target population and used to change future health care and improve patient outcomes. Proper interpretation of registry data allows users to understand the precision of the observed risk or incidence estimates, to evaluate the hypotheses tested in the current registry, and often also to generate new hypotheses to be examined in future registries or in randomized controlled trials. If the purpose of the registry is explicit, the actual population studied is reasonably representative of the target population, the data quality monitored, and the analyses performed so as to reduce potential biases, then the interpretation of the registry data should allow a realistic picture of the quality of medical care, the natural history of the disease studied, or the safety, effectiveness, or value of a clinical evaluation. Each of these topics needs to be discussed in the interpretation of the registry data, and potential shortcomings should be explored. Assumptions or biases that could have influenced the outcomes of the analyses should be highlighted and separated from those that do not affect the interpretation of the registry results. The use of a comparator of the highest reasonably possible quality is integral to the proper interpretation of the analysis.

Interpretation of registry results may also be aided by comparisons with external information. Examples include rates, or prevalence, of the outcomes of interest in other studies and different data sources (taking into account reasons why they may be similar or different). Such comparisons can put the findings of registry analyses within the context of previous study results and other pertinent clinical and biological considerations as

to the validity and generalizability of the results.

Once analyzed, registries provide important feedback to several groups. First analysis and interpretation of the registry will demonstrate strengths and limitations of the original registry design and will allow the registry developers to make needed design changes for future versions of the registry. Another group consists of the study's sponsors and related oversight/governance groups, such as the scientific committee and data monitoring committee. (Refer to Chapter 2, Section 2.6 for more information on registry governance and oversight.) Interpretation of the analyses allows the oversight committees to offer recommendations concerning continued use and/or adaptation of the registry and to evaluate patient safety. The final group consists of the end users of the registry output, such as patients or other health care consumers, health services researchers, health care providers, and policymakers. These are the people for whom the data were collected and who may use the results to choose a treatment or intervention, to determine the need for additional research programs to change clinical practice, to develop clinical practice guidelines, or to determine policy. Ideally, all three user groups work toward the ultimate goal of each registry—improving patient outcomes.

Case Examples for Chapter 13

Case Example 26. Using registry data to evaluate outcomes by practice

Description	The Epidemiologic Study of Cystic Fibrosis (ESCF) Registry was a multicenter, encounter-based, observational, postmarketing study designed to monitor product safety, define clinical practice patterns, explore risks for pulmonary function decline, and facilitate quality improvement for cystic fibrosis (CF) patients. The registry collected comprehensive data on pulmonary function, microbiology, growth, pulmonary exacerbations, CF-associated medical conditions, and chronic and acute treatments for children and adult CF patients at each visit to the clinical site.
Sponsor	Genentech, Inc.
Year Started	1993
Year Ended	Patient enrollment completed in 2005; followup complete.
No. of Sites	215 sites over the life of the registry
No. of Patients	32,414 patients and 832,705 encounters recorded

Challenge

Although guidelines for managing cystic fibrosis patients have been widely available for many years, little is known about variations in practice patterns among care sites and their associated outcomes. To determine whether differences in lung health existed between groups of patients attending different CF care sites, and to determine whether these differences were associated with differences in monitoring and intervention, data on a large number of CF patients from a wide variety of CF sites were necessary.

As a large, observational, prospective registry, ESCF collected data on a large number of patients from a range of participating sites. At the time of the outcomes study, the registry was estimated to have data on over 80 percent of CF patients in the United States, and it collected data from more than 90 percent of the sites accredited by the U.S. Cystic Fibrosis Foundation. Because the registry contained a representative population of CF patients, the registry database offered strong potential for analyzing the association between practice patterns and outcomes.

Proposed Solution

In designing the study, the team decided to compare CF sites using lung function (i.e., FEV1 [forced expiratory volume in 1 second] values), a common surrogate outcome for respiratory studies. Data from 18,411 patients followed in 194 care sites were reviewed, and 8,125 patients from 132 sites (minimum of 50 patients per site) were included. Only sites with at least 10 patients in a specified age group (ages 6–12, 13–17, and 18 or older) were included for evaluation of that age group. For each age group, sites were ranked in quartiles based on the median FEV1 value at each site. The frequency of patient monitoring and use of therapeutic interventions were compared between upper and lower quartile sites after stratification for disease severity.

Results

Substantial differences in lung health across different CF care sites were observed. Within-site rankings tended to be consistent across the three age groups. Patients who were cared for at higher-ranking sites had more frequent monitoring of their clinical status, measurements of lung function, and cultures for respiratory pathogens. These patients also received more interventions, particularly intravenous antibiotics for pulmonary exacerbations. The study concluded that frequent monitoring and increased use of appropriate medications in the management of CF are associated with improved outcomes.

Case Example 26. Using registry data to evaluate outcomes by practice (continued)

Key Point

Stratifying patients by quartile of lung function, age, and disease severity allowed comparison of practices among sites and revealed practice patterns that were associated with better clinical status. The large numbers of patients and sites allowed for sufficient information to create meaningful and informative stratification, and resulted in sufficient information within those strata to reveal meaningful differences in site practices.

For More Information

Johnson C, Butler SM, Konstan MW. et al. Factors influencing outcomes in cystic fibrosis: a center-based analysis. Chest. 2003;123:20–7.

Padman R, McColley SA, Miller DP. et al. Infant care patterns at Epidemiologic Study of Cystic Fibrosis sites that achieve superior childhood lung function. Pediatrics. 2007;119:E531–7.

Case Example 27. Using registry data to study patterns of use and outcomes

Description	The Palivizumab Outcomes Registry was designed to characterize the population of infants receiving prophylaxis for respiratory syncytial virus (RSV) disease, to describe the patterns and scope of the use of palivizumab, and to gather data on hospitalization outcomes.
Sponsor	MedImmune, LLC
Year Started	2000
Year Ended	2004
No. of Sites	256
No. of Patients	19,548 infants

Challenge

RSV is the leading cause of serious lower respiratory tract disease in infants and children and the leading cause of hospitalizations nationwide for infants under 1 year of age. Palivizumab was approved by the U.S. Food and Drug Administration (FDA) in 1998 and is indicated for the prevention of serious lower respiratory tract disease caused by RSV in pediatric patients at high risk of RSV disease. Two additional large retrospective surveys conducted after FDA approval studied the effectiveness of palivizumab in infants, again showing that it reduces the rate of RSV hospitalizations. To capture postlicensure patient demographic outcome information, the manufacturer wanted to create a prospective study that identified infants receiving palivizumab. The objectives of the study were to better understand the population receiving the prophylaxis for RSV disease and to study the patterns of use and hospitalization outcomes.

Proposed Solution

A multicenter registry study was created to collect data on infants receiving palivizumab injections. No control group was included. The registry was initiated during the 2000–2001 RSV season. Over 4 consecutive years, 256 sites across the United States enrolled infants who had received palivizumab for RSV under their care, provided that the infant's parent or legally authorized representative gave informed consent for participation in the registry. Data were collected by the primary health care provider in the office or clinic setting. The registry was limited to data collection related to subjects' usual medical care. Infants were enrolled at the time of their first injection, and data were obtained on palivizumab injections, demographics, and risk factors, as well as on medical and family history.

Followup forms were used to collect data on subsequent palivizumab injections, including dates and doses, during the RSV season. Compliance with the prescribed injection

Case Example 27. Using registry data to study patterns of use and outcomes (continued)

Proposed Solution (continued)

schedule was determined by comparing the number of injections actually received with the number of expected doses, based on the month that the first injection was administered. Infants who received their first injection in November were expected to receive five injections, whereas infants receiving their first injection in February would be expected to receive only two doses through March. Data were also collected for all enrolled infants hospitalized for RSV and were directly reported to an onsite registry coordinator. Testing for RSV was performed locally, at the discretion of the health care provider. Adverse events were not collected and analyzed separately for purposes of this registry. Palivizumab is contraindicated in children who have had a previous significant hypersensitivity reaction to palivizumab. Cases of anaphylaxis and anaphylactic shock, including fatal cases, were reported following initial exposure or re-exposure to palivizumab. Other acute hypersensitivity reactions, which might have been severe, were also reported on initial exposure or re-exposure to palivizumab. Adverse reactions occurring greater than or equal to 10 percent and at least 1 percent more frequently than placebo are fever and rash. In postmarketing reports, cases of severe thrombocytopenia (platelet count <50,000/microliter) and injection site reactions were reported.

Results

From September 2000 through May 2004, the registry collected data on 19,548 infants. The analysis presented injection rates and hospitalization rates for all infants by month of injection and by site of first dose (pediatrician's office or hospital). The observed number of injections per infant was compared with the expected number of doses based on the month the first injection was given. Over 4 years of data collection, less than 2 percent (1.3%) of enrolled infants were hospitalized for RSV. This analysis confirmed a low hospitalization rate for infants receiving palivizumab prophylaxis for RSV in a large nationwide cohort of infants from a geographically diverse group of practices and clinics. The registry data also showed that the use of palivizumab was mostly consistent with the 2003 guidelines of the American Academy of Pediatrics for use of palivizumab for prevention of RSV infections. As the registry was conducted prospectively, nearly complete demographic information and approximately 99 percent of followup information was captured on all enrolled infants, an improvement compared with previously completed retrospective studies.

Key Point

A simple stratified analysis was used to describe the characteristics of infants receiving injections to help prevent severe RSV disease. Infants in the registry had a low hospitalization rate, and these data support the effectiveness of this treatment outside of a controlled clinical study. Risk factors for RSV hospitalizations were described and quantified by presenting the number of infants with RSV hospitalization as a percentage of all enrolled infants who were hospitalized. These data supported an analysis of postlicensure effectiveness of RSV prophylaxis, in addition to describing the patient population and usage patterns.

For More Information

Leader S, Kohlhase K. Respiratory syncytial virus-coded pediatric hospitalizations, 1997-1999. Ped Infect Dis J. 2002;21(7):629–32.

Frogel M, Nerwen C, Cohen A. et al. Prevention of hospitalization due to respiratory syncytial virus: Results from the Palivizumab Outcomes Registry. J Perinatol. 2008;28:511–7.

American Academy of Pediatrics—Committee on Infectious Disease. Red Book 2003: Policy Statement: Revised indications for the use of palivizumab and respiratory syncytial virus immune globulin intravenous for the prevention of respiratory syncytial virus infections. Pediatrics. 2003;112:1442-6.

References for Chapter 13

1. Sedrakyan A, Marinac-Dabic D, Normand SL, et al. A framework for evidence evaluation and methodological issues in implantable device studies. Med Care. 2010 Jun;48(6 Suppl):S121-8. PMID: 20421824.
2. Cole P. The hypothesis generating machine. Epidemiology. 1993 May;4(3):271-3. PMID: 8512992.
3. Yusuf S, Wittes J, Probstfield J, et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA. 1991 Jul 3;266(1):93-8. PMID: 2046134.
4. National Cancer Institute. Surveillance Epidemiology and End Results. http://seer.cancer.gov. Accessed August 27, 2012.
5. Schneeweiss S, Gagne JJ, Glynn RJ, et al. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. Clin Pharmacol Ther. 2011 Dec;90(6):777-90. PMID: 22048230.
6. Rothman K, Greenland S, Lash TL, eds. Modern Epidemiology. 3rd ed. New York: Lippincott Williams & Wilkins; 2008.
7. Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley & Sons; 1987.
8. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. Am J Epidemiol. 2004 Jul 1;160(1):34-45. PMID: 15229115.
9. Rubin DB. Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. Proceedings of the Section on Survey Research Methods, American Statistical Association 1978. pp. 20-34.
10. Burton A, Altman DG. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. Br J Cancer. 2004 Jul 5;91(1):4-8. PMID: 15188004. PMCID: 2364743.
11. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol. 1995 Dec 15;142(12):1255-64. PMID: 7503045.
12. Hernan MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002 Jan 15;155(2):176-84. PMID: 11790682.
13. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. New York, NY: Springer; 2009.
14. U.S. Food and Drug Administration. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research. Standards for Data Management and Analytic Processes in the Office of Surveillance and Epidemiology (OSE). March 3, 2008. http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/StaffPoliciesandProcedures/ucm082060.pdf. Accessed August 15, 2012.
15. Swihart BJ, Caffo B, James BD, et al. Lasagna plots: a saucy alternative to spaghetti plots. Epidemiology. 2010 Sep;21(5):621-5. PMID: 20699681. PMCID: 2937254.
16. Hennekens CH, Buring JE, Mayrent SL. Epidemiology in medicine. 1st ed. Boston: Little, Brown and Company; 1987.
17. Kleinbaum DG, Kupper LL, Miller KE, et al. Applied regression analysis and other multivariable methods. Belmont, CA: Duxbury Press; 1998.
18. Aschengrau A, Seage G. Essentials of epidemiology in public health. Sudbury, MA: Jones & Bartlett; 2003.
19. Rosner B. Fundamentals of biostatistics. 5th ed. Boston: Duxbury Press; 2000.
20. Higgins J, Green S.The Cochrane Collaboration. The Cochrane Handbook for Systematic Reviews Of Interventions. 2006. http://www.cochrane.org/sites/default/files/uploads/Handbook4.2.6Sep2006.pdf. Accessed August 15, 2012.
21. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000 Sep;11(5):550-60. PMID: 10955408.
22. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006 Jan 26;354(4):353-65. PMID: 16436767.

23. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol. 2003 Aug 1;158(3):280-7. PMID: 12882951.

24. Sturmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006 May;59(5): 437-47. PMID: 16632131. PMCID: 1448214.

25. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):253-9. PMID: 16611199. PMCID: 1790968.

26. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology. 2009 Jul;20(4):512-22. PMID: 19487948. PMCID: 3077219.

27. Reeve BB, Potosky AL, Smith AW, et al. Impact of cancer on health-related quality of life of older Americans. J Natl Cancer Inst. 2009 Jun 16;101(12):860-8. PMID: 19509357. PMCID: 2720781.

28. Brodie BR, Stuckey T, Downey W, et al. Outcomes with drug-eluting stents versus bare metal stents in acute ST-elevation myocardial infarction: results from the Strategic Transcatheter Evaluation of New Therapies (STENT) Group. Catheter Cardiovasc Interv. 2008 Dec 1;72(7):893-900. PMID: 19016465.

29. Shuhaiber JH, Kim JB, Hur K, et al. Survival of primary and repeat lung transplantation in the United States. Ann Thorac Surg. 2009 Jan;87(1):261-6. PMID: 19101309.

30. Grabowski GA, Kacena K, Cole JA, et al. Dose-response relationships for enzyme replacement therapy with imiglucerase/alglucerase in patients with Gaucher disease type 1. Genet Med. 2009 Feb;11(2):92-100. PMID: 19265748.

31. Brookhart, M. Alan. Instrumental Variables for Comparative Effectiveness Research: A Review of Applications. Slide Presentation from the AHRQ 2008 Annual Conference (Text Version); Rockville, MD: Agency for Healthcare Research and Quality; Jan, 2009. http://www.ahrq.gov/about/annualmtg08/090908slides/Brookhart.htm. Accessed August 14, 2012.

32. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. Journal of the American Statistical Association. 1996;91(434):444-55.

33. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf. 2010 Jun;19(6):537-54. PMID: 20354968. PMCID: 2886161.

34. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biostat. 2007;3(1):Article 14. PMID: 19655038. PMCID: 2719903.

35. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006 Jul;17(4):360-72. PMID: 16755261.

36. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. J Epidemiol Community Health. 2005 Jun;59(6):443-9. PMID: 15911637. PMCID: 1757045.

37. Holden JE, Kelley K, Agarwal R. Analyzing change: a primer on multilevel models with applications to nephrology. Am J Nephrol. 2008;28(5):792-801. PMID: 18477842. PMCID: 2613435.

38. Diez-Roux AV. Multilevel analysis in public health research. Annu Rev Public Health. 2000;21: 171-92. PMID: 10884951.

39. Leyland AH, Goldstein H. Multilevel modeling of health statistics. Chichester, UK: John Wiley & Sons, LTD; 2001.

40. Varadhan R, Segal JB, Boyd CM, et al. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. Journal of clinical epidemiology. 2013;66(8):818-25.

41. Palmer AJ. Health economics--what the nephrologist should know. Nephrol Dial Transplant. 2005 Jun;20(6):1038-41. PMID: 15840678.

42. Neumann PJ. Opportunities and barriers. Using cost-effectiveness analysis to improve health care. New York, NY: Oxford University Press; 2004.

43. Edejer TT-T, Baltussen R, Adam T, et al. Making choices in health: WHO guide to cost-effectiveness analysis. Geneva: World Health Organization; 2004.

44. Drummond M, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. 3rd ed. New York: Oxford University Press; 2005.

45. Muennig P. Designing and conducting cost-effectiveness analyses in medicine and health care. New York: John Wiley & Sons, LTD; 2002.

46. Haddix AC, Teutsch SM, Corso PS. Prevention effectiveness: a guide to decision analysis and economic evaluation. New York: Oxford University Press; 2003.

47. Gold MR, Siegel JE, Russell LB, et al. Cost-effectiveness in health and medicine: the Report of the Panel on Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.

48. Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. Health Technol Assess. 2005 May;9(20):1-92, iii-iv. PMID: 15899148.

49. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol. 1999 Jun 1;149(11):981-3. PMID: 10355372.

50. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med. 1991 Apr;10(4):577-81. PMID: 2057656.

51. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003 Nov 1;158(9):915-20. PMID: 14585769.

52. Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 2007 Mar;16(3):241-9. PMID: 17252614.

53. Haynes RB, Sackett DL, Guyatt GH, et al. Clinical epidemiology. 3rd ed. New York: Lippincott Williams & Wilkens; 2005.

54. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol. 2012 Jun;41(3):861-70. PMID: 22253319. PMCID: 3396320.

55. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999 Jan;10(1):37-48. PMID: 9888278.

56. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 Sep;15(5):615-25. PMID: 15308962.

57. National Cancer Institute. "SEER: Surveillance Epidemiology and End Results." http://seer.cancer.gov/about/factsheets/SEER_brochure.pdf. Accessed August 15, 2012.

58. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998 Dec 5;317(7172):1572. PMID: 9836663. PMCID: 1114388.

59. Kleinbaum DG, Klein M. Survival analysis: a self-learning text. 2nd ed. New York: Springer; 2005.

60. Twisk JWR. Applied longitudinal data analysis for epidemiology – a practical guide. Cambridge, UK: Cambridge University Press; 2003.

61. Newman SC. Biostatistical methods in epidemiology. New York: John Wiley & Sons, LTD; 2001.

Chapter 14. Modifying and Stopping Registries

1. Introduction

Most, if not all registries, should undergo periodic critical evaluation by key stakeholders to ensure that the objectives are being met. When registry objectives are no longer being met or when clinical or other changes affect the registry (e.g., changes in treatment practices, introduction of a new therapy), the registry may need to be adapted or to stop collecting new data. This chapter is divided into two sections. The first section describes possible reasons for a registry transition and issues that should be considered in planning and implementing a transition. The second section discusses factors that may lead to the determination to stop a patient registry. Case Examples 28, 29, 30, and 31 describe a variety of registry transitions.

2. Registry Transitions

A wide variety of factors may drive the decision to proceed with a registry transition. For example, a registry may need to transition to a new technology platform to remain functional for its participants, or a registry that was designed to study the natural history of a disease for which there was no effective treatment may change its purpose when a new product or therapy becomes available in the market. Other scenarios in which a transition may be necessary include changes in funding sources and stakeholders (e.g., funding for a government-sponsored registry may end, resulting in transition to private ownership, such as to a professional association) or the introduction of new regulatory requirements (e.g., a registry may need to be adapted to fulfill a postmarketing commitment). Because many different factors may contribute to a registry's transition, transitions are highly variable in scope and resource requirements.

This section focuses on issues that are of particular significance in a major registry transition, defined as a change in the purpose, sponsor, and/or technology platform, all of which will have a substantive impact on the ongoing conduct of the registry. Less ambitious transitions (e.g., changes in data elements on preexisting case report forms) are not specifically covered here; however, parts of this section (e.g., data analysis) are pertinent to such transitions.

While the considerations for a major registry transition are similar to those for the launch of a new registry, there are several distinguishing features. First, a registry transition is facilitated by an existing registry and the collective experience of conducting that registry. The existing registry can essentially serve as the starting point for creating a prototype of the revision. The planning and design of the registry transition should also benefit from lessons learned in operating the existing version of the registry. What has worked well, and what has been problematic? What challenges have been encountered at every level, from staff entering data at the participating sites to the analyst creating reports? Indeed, one or more of these issues may be contributing factors in the decision to proceed with the registry transition. Even if this is not the case, the transition provides an opportunity to address these issues. Registry transitions also present unique challenges distinct from the development of a new registry. In particular, transferring data collected in an existing registry to the revised registry (i.e., data migration) can be a complex and resource-intensive process. (See Case Example 29.)

Despite these differences, the steps in the execution of a major registry transition are analogous to those involved in the launch of a new registry. Therefore, the section is organized in accordance with the general framework for developing a new registry, with a planning and design phase, an implementation phase to carry out the project plan, and an assessment of the potential impact on data management and analysis.

2.1 Planning and Design

The planning and design of a registry transition begins with an assessment phase, in which the

need for a transition is considered. Articulating the purpose(s), determining if a major registry transition is an appropriate means of achieving the purpose(s), and assessing the feasibility of a registry transition are important considerations, as such projects require a significant commitment of resources and have associated risks. Chapter 2, Section 2.2 describes the assessment phase for a new registry, much of which is directly relevant to the consideration of a major registry transition. If the assessment leads to a decision to move forward, then the planning and design of the transition can proceed with the creation of a project charter, formation of a transition team, and development of a comprehensive project plan that encompasses governance, ethical and legal issues, and technology considerations.

2.1.1 Creating a Project Charter

The creation of a project charter is often a useful starting point in planning a transition. A project charter typically includes the following information:

- Overview of the transition
- Purpose/justification for the transition
- Goals and objectives of the transition
- Business case for the transition (if applicable)
- Identification of major stakeholders
- Assumptions and constraints (organizational, environmental, and external)
- Potential risks
- Milestones/deliverables or high-level timeline
- Budget
- Transition team members
- References to source documents, if applicable (e.g., new clinical practice guidelines)

2.1.2 Forming a Transition Team

The next step is to assemble the transition team, which will be responsible for planning and implementing the registry transition. It is important to include key stakeholders and to think broadly about the talent and expertise needed to accomplish a successful transition. In general, the transition team should include the following members:

- *Sponsor/funding organization representative*: Ensures that the team has the resources necessary to complete the project and keeps the sponsor apprised of any issues that may affect the timeline or budget for the transition.
- *Project manager*: Accountable for all aspects of the transition, including timely escalation of issues for resolution.
- *Clinical expert*: Provides guidance on changes that affect the clinical content of the registry (e.g., changes in purpose and data collection), and provides input on data migration, as needed.
- *Epidemiologist/biostatistician*: Provides guidance on changes that affect the study design and analysis plans (e.g., changes in purpose, data collection, data management, and data migration).
- *Data management expert*: Provides guidance on changes that affect data collection, data storage, or data quality assurance.
- *Legal/ethical expert*: Provides guidance on how changes affect the legal and ethical construct of the registry (e.g., contract with funding source(s), contracts with participating sites, contracts with vendors, and data sharing agreements), and identifies any ethical issues (e.g., need for institutional review board review, changes to informed consent documents, need to re-consent participants).
- *Other representatives*: Depending on the nature of the transition, other representatives may be included on the transition team, such as (1) a principal investigator or study coordinator from a participating site with experience entering data into the registry to provide guidance on feasibility and burden of data entry, (2) a technical expert to help guide a transition to a new technology platform, (3) a patient advocate to provide the patient's perspective, and/or (4) representatives of other databases that are linked to the registry (e.g., a related registry or substudy).

Once the transition team has been assembled, it is critical to achieve consensus on the rationale and the overarching goal(s) for the registry transition. Open discussion at this stage may identify

unanticipated barriers, which can be addressed proactively in the transition planning. Gaining the full support of the transition team will increase the likelihood of a successful registry transition.

2.1.3 Developing a Project Plan

The next step for the transition team is to develop a detailed project plan encompassing timeline and budget. The transition project plan should be thoughtful, complete, and realistic. As with all projects of this magnitude and complexity, disagreement among stakeholders over scope, cost overruns, and time delays may occur. These predictable issues should be anticipated as much as possible, and risk mitigation strategies considered. The project plan should also consider other sources of risk specific to the transition (e.g., unexpected issues with technology compatibility, delays in obtaining institutional review board approval, and disputes related to ownership issues). Chapter 2 provides more information on project planning considerations.

The project plan should also address staffing issues. The transition may require new expertise and skills that alter staffing requirements. Training existing employees or hiring appropriately skilled personnel may be necessary. Planning for additional workload on the registry staff during the actual transition is also an important consideration, as they may be operating and supporting the existing registry while working on the transition to the modified registry.

Other issues that should be considered in transition planning relate to governance, ethical concerns, legal matters, data collection, and technology. These issues are discussed in more detail in the following sections.

2.1.4 Governance Issues

Nearly all registry transitions will require an internal and external governance structure to manage and approve changes, whether the transition relates to the scientific objectives of the registry, technology changes, or data access. The transition team is one important component of the governance structure. Chapter 2, Section 2.6 reviews the governance considerations for the planning of a new registry, many of which are relevant to a registry transition. Some additional considerations are addressed below.

- *Scientific advisory board governance during the transition*: Many registries have scientific advisory boards that oversee the conduct of the registry. These boards may also play a role in governance during a registry transition and provide external perspective for the considerations and future objectives for a registry transition. Membership of the scientific advisory board should be reviewed to ensure the key stakeholders that are involved in the transition are represented. During the registry transition, the scientific advisory board can also act as an advocate of change by publicly supporting the transition and helping to engage and motivate clinicians at the participating centers. External stakeholders, such as patient advocacy groups and regulatory agencies/ health authorities, may also be informed of the transition and, depending on the goals of the transition, potentially enlisted as additional public advocates for the registry transition.
- *Governance of data access*: Registry transitions will also require revisiting the registry's data access policies and procedures. If a data access committee is already in place, the committee should be charged with (1) determining how changes in the registry will affect the policies and procedures for accessing data, and (2) reviewing the operational plan for executing analysis plans with respect to the registry transition. Furthermore, if the transition involves a change in registry stakeholders, the procedures for conducting analyses and developing publications should be re-examined. New stakeholders may need to be involved in the prioritization of analysis plans, conduct of analyses, and/or the review of scientific abstracts and manuscripts.

2.1.5 Ethical and Legal Issues

The major ethical and legal issues for registries focus on data privacy, patient confidentiality, and ownership of and access to the data. These issues, covered comprehensively in Chapter 7, should also be carefully considered during a registry transition. It is important to note that interpretations of the

pertinent laws and regulations are numerous and varied, leading to inconsistent application among institutions, which may affect multicenter registries.[1] Hence, input from legal counsel and regulatory authorities should be sought when planning a registry transition. Some common legal or ethical concerns that may arise during registry transitions are reviewed here.

An early step in the registry transition planning process is consideration of the need for institutional review board (IRB)/ethics committee (EC) review. If the purpose of the registry is unchanged and no new data are being collected, IRB/EC review may not be necessary—subject to ethical guidelines and the requirements of the individual IRBs/ECs. However, IRB/EC review would likely be required in certain transitions; for example, if new data will be collected through contact with patients, if the new data that will be collected includes identifiable personal information, or if the data will be used in a different manner than previously communicated to patients (45 CFR §46.102(f)).

Especially when the purpose of the registry has changed, a registry transition may involve extending the followup period of the initial cohort. In these circumstances, recontacting patients or using their identifiers may be necessary to collect the longer-term data. For example, a cardiac assist device registry may have been established initially to determine perioperative safety. However, new safety concerns associated with longer term implantation may prompt a change in the purpose of the registry. Medical records, death indices, and patient interviews may be required in order to collect the longer-term followup data. This new data collection effort would likely require IRB/EC review.

Consideration should also be given to whether any changes will be required in the informed consent process (e.g., obtaining revised consents from existing subjects, obtaining new consents for registries that do not currently have such consents). If consent was obtained for registry participation initially, re-consenting may be needed, especially when the registry transition will result in (1) longer or otherwise different followup than what was originally agreed to by patients, (2) direct contact with patients to obtain new data, (3) collection of biological samples or linkage of existing specimens to registry data, (4) the use of data from deceased participants, or (5) linkage of the participant's data to other databases. If the planned registry modifications involve patients for whom the feasibility of obtaining consent would require unreasonable burden or situations where the consenting process would potentially introduce an unacceptable level of bias,[2-4] discussions with local IRBs/ECs should be undertaken to see if the consent can be waived. Chapter 8 discusses these issues in more detail.

2.1.6 Data Collection

A major component of the registry transition project plan should be a thorough evaluation of current and future data collection needs. The project plan should allocate time for epidemiologists and clinical experts to jointly review the current registry case report form (CRF). It is of paramount importance that the relevance of the current set of data elements is reviewed, in light of what is known about new hypotheses to be tested. During this review, some data elements may deemed irrelevant and may not be required moving forward. When considering the collection of additional covariates and outcomes, particular attention must be given to balancing the scientific relevance of the new data elements with the logistical burden on participating centers.

Additional considerations may arise if a registry transition involves one of the following specific circumstances:

- *Collection of Biological Samples*: Biobanks, defined as facilities that store biological material (e.g., serum, genomic material, pathology specimens) from humans, are increasingly popular additions to registries.[5] The addition of a biobank raises many logistical issues, which are outside the scope of this chapter. However, it should be noted that the addition of a biobank will likely require changes in the informed consent. Some biobanks have used general consents to cover future analyses of the biological material and integration into the registry, but there is significant concern about these broad consent

documents. Some commentaries on this issue have suggested that such broad consents are more appropriate when limited to a specific disease entity, thereby allowing for studies examining diagnosis, mechanisms of disease, risk factors, and treatment outcomes.[6-8] Chapter 8 discusses these issues in more detail.

- *Pediatric Registries*: If a registry enrolls pediatric participants and the registry transition involves extending the followup period, consideration should be given to whether participants need to give consent when they reach an eligible age. This is particularly important for those registries that plan to add a biobank or link to other databases as part of the transition process. There is considerable debate regarding the ethics of parents enrolling their children in research studies. More discussion on this topic can be found in Chapter 7. It is also important to note that for all registries, the right to withdraw is inherent;[8, 9] see Chapter 8.
- *National to International Registry*: Some registry transitions may extend the geographic scope of a registry. For example, a U.S.-based registry may add participating sites in Europe. When the registry scope extends beyond national borders, additional ethical and legal concerns must be addressed. Each country may have different legislation and restrictions for the collection and processing of subject information and its use for research. Adequate time and additional resources to investigate these requirements should be factored into the project plan. Moreover, if Federal funds are used in the registry transition, additional steps may be involved in the expansion of the registry. In particular, some registries may be collecting data on vulnerable international populations for which additional privacy protection safeguards may be necessary. Federal guidelines for performing international research should be consulted as part of the planning process.

2.1.7 Data Ownership and Licensing

A number of scenarios exist in which ownership of registry materials must be delineated, including the ownership of the interface, platform, infrastructure, and data. During a registry transition, particularly one involving a change in stakeholders, a careful review of agreements or contracts should be performed to determine if modifications are needed. In some cases, the registry transition may involve moving data from one platform to another. Hence, data ownership may need to be clarified. For example, a professional organization may determine that the vendor maintaining its registry is performing below expectations and may select a new vendor to house and run the registry. Depending on the terms of the prior agreements, it may or may not be possible to import the historical data into the new vendor's system.

Registry data are often collected using electronic or paper CRFs that may have intellectual property protections, including copyright, trademark, and patent. If continued use of these forms is planned, measures should be taken to ensure that the appropriate permissions for use are still applicable when the registry transitions.

2.1.8 Data Access

In addition to data ownership, ongoing data access is an important consideration. The new and ongoing registry stakeholders should consider whether the previous stakeholders should have access to the previously collected data as well as to the data collected in the future. Federal and academic stakeholders may need to execute technology transfer agreements (e.g., material transfer agreements) or other contractual agreements in order to access the data.

2.1.9 Changes in Funding

Registry transitions may also include changes in funding. For example, a registry that was initially funded through a government grant may be transitioned to a professional association or industry partner. When funding sources change, the role of the funding entities should be clearly delineated to ensure that there is no real or perceived threat to privacy or data confidentiality.

In some cases, a change in funding may require contract modifications in anticipation of potential conflicts between the new stakeholders and the remaining stakeholders. For example, industry may elect to partially fund a registry that is also

receiving Federal funding from a regulatory agency. Contracts may need to be modified to clearly delineate how each set of funds will be spent. The new Chapter 24 on public-private partnerships provides more information on these issues. As with all contracts involving Federal funds, attention should be given to regulations governing their appropriate use. Additionally, changes in funding may alter the locus of legal rights and obligations. It is important to have unambiguous conversations with stakeholders and associated contractual agreements that clearly delineate the rights of the funding entities.

When data are transferred from one owner/sponsor to another, the liability associated with the protection of subjects' information should be clarified. Consideration should be given to indemnification clauses in data transfer agreements. Oftentimes, the data transfer agreements detail that the new sponsor of the registry will accept all liability for use of the data previously collected by the transferring sponsor. The data transfer agreement should also contain a clause that the new sponsor agrees to use the data properly. In these circumstances, the liability would be assumed by the new sponsor if a breach of information occurred whereby subject-level information was relayed to an outside party. If the new sponsor is a Federal entity, however, there are regulations that prohibit the Federal Government from indemnifying others (e.g., Anti-Deficiency Act).

2.1.10 Contracts With Vendors

Issues may arise with vendors (including inadequate performance of duties, loss of financial solvency, or escalating cost of renewing the contract), necessitating a transition to a new vendor. In light of these potential outcomes, it is necessary to draft contracts that consider these scenarios and contain provisions to address them. For example, if a registry is being transitioned to a new, fledgling company, consideration should be given to establishing an escrow account for the registry. This account would cover the cost of ensuring that the data remain accessible to the sponsoring body. Moreover, it would prevent the registry from being part of the estate if the company is unable to meet its contractual obligations. Establishing the escrow account would increase the cost of the initiative for the sponsor, the vendor, or both, and should be considered when planning the transition. In addition, contracts should contain explicit clauses that guarantee the transmission of data to a new vendor when the contract expires or if the vendor defaults on the contract.

2.1.11 Technology Considerations

A registry designed to collect long-term followup data will inevitably undergo technology changes. Platforms for electronic data capture (EDC) may be upgraded (for example, through version updates within a system), or the registry sponsor may select a different third-party vendor to host the EDC system. Upgrading the EDC system and technology platform may enable more frequent data entry from participating centers, rather than annual or semi-annual data reporting under previous technology environments. Such changes have implications for training plans for participating centers (see Section 2.2.2, Training). Technology considerations relevant to linkage of a registry to an electronic health (medical) records (EHR/EMR) database or other database and for collection of patient-reported outcomes are covered in Chapters 15 through 18. In transitioning to a new registry technology platform, it is important to clearly define software requirements in order to avoid design flaws, which are costly to correct after project completion. Soliciting input from various stakeholders (e.g., data entry personnel, clinical experts, data analysts) may be helpful to validate the proposed design of the new registry. The proposed design should be presented to them in an easy-to-understand format (e.g., a prototype) rather than a detailed requirements document, which may be more difficult to comprehend. Setting aside time for user-acceptance testing or pilot testing may also be useful to identify issues before the transition is complete.

One of the earliest and most important decisions in transitioning to a new technology platform is whether to develop the platform in-house or use an external vendor. Each approach has advantages and disadvantages. The in-house approach requires personnel with the appropriate expertise and the

infrastructure to support such a project. Development tools widely used by software companies should be employed, if possible, to mitigate the risk of experiencing shortages of qualified personnel for ongoing support and maintenance of the application. Organizations that do not have the internal resources and expertise to develop a registry application in-house usually turn to external vendors. Selecting a registry vendor is an important strategic decision for an organization, particularly for sponsors who anticipate operating the registry for many years. Some factors that should be considered in selecting a registry vendor are outlined in Table 14–1.

Table 14–1. Considerations in selecting a registry vendor

• Develop detailed requirements for the new registry before issuing a request for proposals. The requirements may be modified later to align with the vendor's framework for development, but having complete requirements early in the process will allow for a more accurate timeline and cost estimate.
• Gather as much information as possible about the potential vendor by contacting existing clients and asking detailed questions about communication, timelines, budget, and post-release support.
• Ask an independent expert to evaluate and analyze the technology platforms and technology expertise of the potential vendor.
• Ask the potential vendor to be specific with their cost estimates. Avoid vendors that cannot provide concrete estimates.
• Know the hosting and maintenance fees of the existing registry and compare them to the hosting and maintenance estimates from the potential vendor.
• Assess the security policies and procedures established by the vendor and ensure that they comply with the industry standards and best-of-breed practices.
• Assess the ability and willingness of the potential vendor to transfer registry data (both to transfer historical data into their registry application and transfer the data out from their registry application if the registry changes vendors in the future).
• Learn about the vendor's experience in importing data from other sources of medical information using standard interfaces (e.g., Health Level Seven [HL7], Clinical Data Interchange Standards Consortium [CDISC]), and also about their ability to build custom interfaces. A list of existing and emerging standards in the field is maintained by U.S. Food and Drug Administration.
• Consider the vendor's international experience, including translation and help-desk support, if pertinent to the planned transition.
• Discuss policies related to data access, including how participating sites can access their own data and how the registry team can obtain data sets for analysis.

Once a vendor has been selected and the features of their technology platform are known, it is important to assess the hardware, software, and browser configurations at the participating sites, as these may affect performance of the registry application. It is also important to ensure that the participating sites have access to the optimal configurations on which the application has been tested and validated. Requesting a technology contact person at each of the participating sites may be helpful to facilitate working through these issues during the transition.

Another technology consideration is transitioning personnel involved in data entry at participating sites from an existing registry to the new registry. This requires an analysis of security levels in order to transfer users to the appropriate permission level in the revised registry. In some cases, users can be transferred electronically from the existing to the new registry application, but in other cases, they must be added manually. The transition team must develop a plan for accomplishing the transfer that minimizes the effort at the participating sites, but ensures that only valid users can access the

registry at the appropriate permission level. Of note, a registry transition provides an opportunity to assess the activity level of users at the participating sites and their ongoing need to access the registry.

A final technology consideration pertinent to a transition relates to the closeout of an existing registry. Generally, the closeout should be scheduled well after the anticipated launch of the new registry, as timelines on such complex projects are often delayed. The existing registry may also be useful in validating successful data migration into the new registry.

2.2 Implementation

Once a transition plan has been developed and the decision has been made to move ahead, it is important to communicate with stakeholders about the plans, train registry participants on the changes and support them through the launch, and assess the impact of the transition on data management and analysis activities.

2.2.1 Communication

Communicating with all stakeholders is critical during a registry transition. The transition team should develop a communication plan that defines who is responsible for communicating what and to whom. The frequency and mode of communication should be established with a particular sensitivity to key stakeholders. Since the registry transition will likely disrupt workflow at the participating sites, communicating the rationale for the change, the timeline, and the impact on users is important. Any change in expectations or incentives for participation should be fully explained. It is important to anticipate and respond to questions and concerns from participating sites, knowing that change can lead to stress and anxiety. In most circumstances, the communication plan will focus on retaining participating sites through the transition. However, a registry transition provides an opportunity to evaluate participating centers to decide whether all of them should be retained. A transition may also be an ideal time to recruit additional sites.

2.2.2 Training

The development and implementation of a robust training program prior to the registry transition will facilitate the rollout of the revised registry and improve the quality of data collected. Training needs will vary according to the scope of the registry transition. For example, a technological change that affects the user interface, functionality, and/or organization of the data elements will likely require extensive training, whereas a transition related to a change in purpose with minimal impact on data entry should require less training. When developing a training program, the key elements of adult learning theory[10] should be kept in mind, and several questions should be addressed:

- Who is the intended audience? Determining the audience will have a significant impact on the design and implementation of the training program. For example, internal staff training will differ from that of external registry participants, and the training program for clinicians will likely differ from that designed for data entry personnel.
- What are the learning objectives? The learning objectives should drive the development of the curriculum. What do the people involved in the registry need to know to be successful during and after the transition? The focus should be on what will change and why, and what the impact of the changes on registry participants will be.
- What information is needed to meet the learning objectives? High-level overviews and detailed documents are useful to help participants with varying levels of interaction with the registry understand the changes. The creation of a reference guide that clearly describes what changes were made and why each change was made will be extremely helpful to some registry participants.
- What are the best mechanisms for disseminating the information? People respond differently to various learning environments and techniques. Depending on the size of the registry, training may be offered in various ways, some of which are described below:

  - Conference calls can be effective for smaller groups and allow for open discussion.
  - Webinars can be useful when larger groups are involved and the training activity includes visual presentation.
  - Face-to-face meetings are frequently effective since the learner is less likely to be distracted.
  - One-on-one training sessions are usually well received, since the training can be customized to the individual learner. However, this approach is costly.
  - User's guides, manuals, FAQs, and other documents can be posted on a Web site, or hard-copy materials can be distributed to participants.
- What is the best approach to ensure that learning has occurred? It is important to confirm that the training program is successful, in order to avoid issues with retention and data quality after the transition launch. Learning assessment approaches include tests (e.g., the completion of a sample data collection form or other task), surveys, and direct feedback. Feedback from the learning assessments should be incorporated into the training program, as needed. Pilot testing may also be useful for refining and strengthening the training program before launch.

2.2.3 Supporting Participants Through the Registry Launch

In addition to a robust training program, sufficient personnel and resources should be assigned to respond to input and inquiries from registry participants following the launch of the revised registry. Accessibility of the support team is very important during this critical period of the transition. Planning for the registry launch should delineate how users can submit questions or concerns (e.g., by email or by calling a support desk), who will be the first responders, and how complex issues will be escalated for further evaluation. Many straightforward questions (e.g., problems logging on) can be resolved quickly and efficiently. However, it is important to carefully assess all input from participants since they may uncover problems with the revised registry that have been missed during testing. Such problems may require immediate attention not only from support personnel, but also from the developers of the registry application. At some defined point in time (e.g., 1 to 3 months after launch), a broader analysis of all of the questions and comments from participants may be helpful in prioritizing any further changes to the registry.

2.2.4 Data Management

Technological changes may require changes to the database/data warehouse used to store the registry data. Database or data warehouse transfers are complex processes that involve a number of steps, including creating a new database layout, mapping the legacy data to the new database layout, and transferring the data with rigorous quality controls to ensure that the transfer is successful. Database transfers also need to be conducted in accordance with any regional IRB or EC approvals to ensure that the privacy of any patient-level data is maintained. The size and complexity of the registry as well as the extent of the changes in the CRFs will determine the complexity of the data mapping process. The data fields known to users of the registry might be collected in different contexts (e.g., with added specificity or new dependencies between data elements on CRFs), and these differences must be considered in the data mapping process. Relatively small changes in the wording of a question on the CRF, or creating an additional category on an existing item (e.g., expanding categories of ethnicities) may introduce ambiguities in the mapping of the existing data set to the new environment. In other instances, significant changes to the definition of an outcome variable will typically require review and adjudication of prior cases to establish longitudinal consistency across the data set (for further detail, see Section 2.3.3. below). For these reasons, input and evaluation of the impact of the migration on future registry outcome analyses from subject matter experts, including epidemiologists and clinical experts, along with documentation of decision rules that were established during the epidemiological and clinical review, will be needed in the data mapping process. The effort and expense involved in the

data migration is often underestimated and adequate time must be allocated during the project planning and in establishing timelines. Despite careful attention to detail, this activity often becomes an iterative process, with data mapping, data importation, and quality control checks that lead to corrections in the data mapping, re-importation of the data, et cetera.

Many practical issues should be considered when transferring a database. First, it is important to document the rationale for adding, modifying, or deleting data fields, so that this information can be communicated to stakeholders and registry participants. Second, careful consideration should be given to the future impact of changes. Certain changes may make it difficult to link prior data sets with the new data sets. For example, adopting a new, broader definition may mean that data can only be linked in one direction, as shown in Table 14–2.

Table 14–2. Impact of definition changes on data linkage

Old Definition	New Definition	Linkage Direction
Death: A mortality that occurred in the hospital within 30 days of the procedure.	Death: a mortality that occurred within 30 days of the procedure, whether in the hospital or not.	Deaths in the old data set fall within the parameters of the new definition. However, deaths according to the new definition would not necessarily apply to the old definition, since they include mortalities post-hospitalization.

When making changes to the data structure, the following questions should be considered:

- Will existing queries (i.e., questions raised by a data manager and issued back to the participating centers regarding a data entry issue) need to be rewritten for the new data set?
- Will existing reports (e.g., percent of patients with a laboratory value above a certain number) need to be revised for the new dataset?
- Will more server space be needed to house the data?
- How can the impact of the changes on the processes affected by the new data structures be minimized?

It is also important to determine what metadata (e.g., long name, short name, data type/data format, and permissible values) are important to capture for each field and how the transition will affect the metadata.

2.3 Data Analysis

A registry transition may introduce many data analysis considerations that require the input of epidemiologists and/or biostatisticians. Transitions that involve new hypotheses or technological changes can present enormous challenges to the continuity and validity of the analyses. The issues range from the handling of new data elements to the introduction of selection bias or recall bias if the cohort definition evolves during the transition.

2.3.1 Changes in Cohort Definition

A registry transition may involve a change in the inclusion or exclusion criteria for patient participation, thus shifting the definition of the study cohort. These changes can occur under a number of scenarios, such as if the registry moves from a disease-based cohort (i.e., no inclusion criteria for receiving a particular treatment) to focusing on a cohort of patients with the disease who receive a specific therapy or class of therapies (i.e., inclusion criteria now requires patients to be receiving a treatment). Cohort definitions may also change based on geography (e.g., if a registry transitions from a national to a global catchment area). This introduces the possibility of geographic differences in disease severity or treatment patterns, which may require thorough documentation of baseline clinical status in order to stratify or perform covariate adjustment, if necessary.

Other changes in the cohort definition may occur if the registry transitions from having broad participation by centers to a limited set of centers (e.g., physicians who are associated with large specialty care clinics). A registry transition that results in such a change in the cohort definition has the potential to introduce selection bias into the registry by focusing the enrollment and ongoing followup of subjects on a potentially more severely affected group of patients. As enrollment and followup occur, epidemiologists should be actively involved to assess whether selection bias has been introduced. Comparisons of demographic and baseline clinical features of subjects before and after the transition may be sufficient to assess the degree of bias introduced and to understand which factors or variables can be considered for stratification or covariate adjustment. Advanced methodologies such as comorbidity indices or propensity score analyses may be necessary to adequately adjust for the changes in the cohort over time.

2.3.2 Introducing New Data Elements

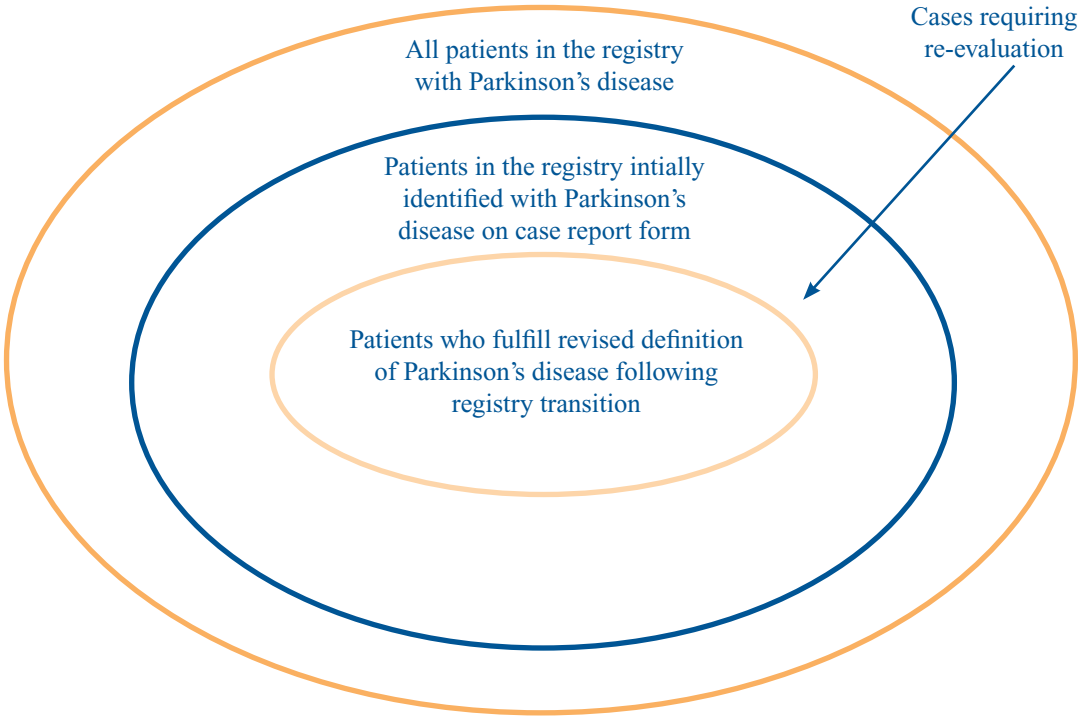
As scientific advances improve the understanding of a particular disease or new treatments become available, new hypotheses will likely be formed. In order to test new hypotheses, it may be necessary to add data elements and/or refine the definition of existing data elements. When adding new data elements, common data elements and validated instruments should be used when possible. (See Chapter 4.) it may be necessary to validate the new data elements, through source document verification of the original medical records, laboratory tests, or diagnostic reports. Results of source document verification may show there are discrepancies in the accuracy of new data elements being captured. For example, investigators interested in collecting data on heart failure as an outcome may find variation in how the definition of heart failure is applied across contributing centers. While the refinement of definitions for data elements occurs, analyses on the outcome variables may still take place. However, methods of quantitative sensitivity analysis may be necessary to understand the degree to which misclassification of variables may introduce bias into the analytic results. Results of source document verification efforts can be used as inputs into quantitative sensitivity analysis to directly estimate the sensitivity and specificity of the outcome variable.

2.3.3 Impact on Existing Cases of an Outcome

A registry transition may lead to redefining an outcome in order to increase sensitivity and specificity. For example, a registry that has been collecting data about the onset of Parkinson's disease as an outcome measure may transition to more stringent inclusion and exclusion criteria. Although this may result in increased validity of the outcome, the statistical power of the analyses from the registry may be compromised, as there will likely be fewer patients meeting the case definition going forward. Patients who have already been identified in the registry as cases may require re-evaluation (and possibly re-adjudication) to determine if their clinical scenario fulfills the revised selection criteria.

Figure 14–1 shows the potential impact of a change in an outcome (e.g., case definition of Parkinson's disease) following a registry transition. Note that the smaller cohort size following the registry transition may reduce statistical power and cases that met original case definition may require re-evaluation.

Figure 14–1. Potential impact of a change in outcome



2.3.4 Impact of Patient-Reported Outcomes

Registries frequently include patient-reported outcomes, such as those reported on the SF-36® health survey, or activities of daily living. It is important to note and characterize whether the type of patients who are reporting self-assessments to the registry and their disease severity or outcome status are changing over time because of the transition. If such instruments are introduced during a registry transition, patients may begin to preferentially recall events, which can lead to a bias in the outcomes. In addition, if the registry transitions from a purely disease-based registry to a therapy- or product-based registry, patients who become aware of this change may begin to report their health status more or less favorably.

Table 14–3 illustrates the possible consequences of transitioning from a disease-based registry to a focus on patients with the disease exposed to a particular therapy. Prior to the transition, the risk of the outcome among exposed and unexposed patients was similar. Following the transition, there are more exposed patients, and, for the purposes of illustrating the impact of bias, assume awareness of the registry transition results in exposed patients preferentially reporting onset of a particular outcome. Because of this preferential report, the risk is approximately 25 percent greater among the exposed as compared with before the transition. The apparent risk ratio is now 1.46 comparing exposed to unexposed.

Table 14–3. Possible consequences of a change in registry focus

	Exposed	Unexposed
Before Transition		
Cases with specific patient-reported outcome	70	50
Total patients	450	375
Cumulative incidence per 100	15.6	13.3
Risk ratio	1.17	
Following the Transition*		
Cases with specific patient-reported outcome	175	50
Total patients	900	375
Cumulative incidence per 100	19.4	13.3
Apparent risk ratio	1.46	

* The emphasis on enrolling patients who have been exposed to the therapy leads to an apparent 25 percent increase in the incidence of cases among the exposed.

2.3.5 Comparative Effectiveness Analysis

A registry may transition from a disease-based cohort to one that is focused on specific treatment(s) in order to establish comparative effectiveness studies between multiple treatments. A greater emphasis on baseline covariate data may be required in this situation, and epidemiologists should be involved to identify the key variables that would account for differences in disease severity among the treatment groups in order to mitigate biases such as confounding by indication. Epidemiologists must also be involved in planning the statistical analysis, which may require matching or other multivariate statistical techniques.

2.3.6 Biostatistics and Statistical Power

Statistical power must be considered in registry transitions that lead to changes in the size of the cohort and/or the extent of followup. For example, a transition that focuses the registry on a smaller number of participating centers may diminish the number of new enrollees, but provide the benefit of an extended length of followup. The transition may eventually provide a greater number of exposed patients who develop the outcome(s) of interest. Biostatisticians should be involved in assessing the impact of changes in cohort accrual on statistical precision of the analyses. Previously specified hypotheses of interest may no longer be testable from the standpoint of statistical power. Alternatively, consideration of statistical power for newly specified hypotheses following the transition may provide an assessment of the extent of enrollment and followup required for robust future analyses.

2.4 Summary of Registry Transition Considerations

Many registries will undergo a major transition at some point in their life cycle, most often related to a change in purpose, sponsor, and/or technology platform. A major registry transition is a complex and resource-intensive process with associated risks. Careful and comprehensive planning will maximize the probability of success. However, unexpected challenges may still occur during the implementation phase. The transition team should be prepared to react to circumstances as they arise and modify the project plan accordingly. This chapter has reviewed the steps involved in the execution of a registry transition, including the planning and design, implementation, subsequent impact on data management and analysis issues. Table 14–4 presents a checklist of key issues that may be helpful to readers who are considering a major registry transition.

Table 14–4. Checklist of key considerations for a registry transition

Planning and Design Phase

1. Determine if a registry transition is appropriate and feasible.
   a. Has the purpose of the transition been clearly articulated?
   b. Is a transition an appropriate means of achieving the purpose?
   c. Is the transition feasible from a resource perspective?
2. Organize a transition team.
   a. Has a transition team been assembled with all necessary areas of expertise?
   b. Is the team in agreement on the rationale for and goals of the transition?
   c. If applicable, how will partner organizations be involved in the transition planning and design?
3. Develop a transition project plan.
   a. Does the project plan cover timeline, budget, and staffing?
   b. Have "lessons learned" from operating the current registry been considered and addressed in the transition plan, as necessary?
   c. Have major risks been identified and risk mitigation strategies considered?
4. Engage advisory boards and other stakeholders.
   a. Is the scientific advisory board in agreement with the rationale for and goals of the transition?
   b. Are any changes needed to the scientific advisory board to ensure that appropriate areas of expertise for the transition are represented?
   c. Will changes to the data access policies and procedures be necessary?
5. Consider legal and ethical issues.
   a. Will the changes require review/approval by an IRB or EC?
   b. Do the changes require informed consent, or does the existing informed consent form need to be updated?
   c. Is the registry expanding to collect data in new countries? If so, what additional ethical and legal considerations must be addressed?
   d. Are any changes needed to existing contracts or agreements?
6. Assess the potential impact of technology changes.
   a. Does the transition involve changing to a new technology? If so, have the hardware, software, and browser configurations been assessed at participating sites to ensure that the new technology will perform well?
   b. Is there a plan for transferring personnel (usernames/passwords) from the previous system to the new system?
   c. Will a new registry vendor be selected? If so, have potential vendors been thoroughly assessed (see Figure 14–1)?

Implementation

1. Share information on the transition with registry participants and stakeholders.
   a. Is there a communication plan that clearly defines who should communicate what information to whom and at what time?
   b. Who will answer questions about the transition?

Table 14–4. Checklist of key considerations for a registry transition (continued)

2. Train registry participants and support them through the launch.
   a. Have training plans been developed for registry staff and participants?
   b. Is there sufficient registry staff to carry out training for participants?
   c. Have registry materials (e.g., user guides, data definitions) been updated?
   d. Has a plan been developed to support participants after launch of the revised registry?

Data Management and Data Analysis

1. Develop a plan for data migration.
   a. Is data mapping or migration necessary?
   b. Are the timeline and budget sufficient for data migration, which is often an iterative, complex process?
   c. Is there a clear rationale for adding, modifying, or deleting each data field?
   d. Have the implications of changes to the data structure been carefully considered?
2. Determine how the transition may affect data analyses.
   a. Did the transition change the definition of the study cohort? If so, has the potential for selection bias or recall bias been assessed?
   b. Have new or modified data elements been reviewed to determine if participants are reporting this information correctly?
   c. Have outcome measures been redefined? Will existing cases of the outcome require re-adjudication?
   d. If comparative effectiveness research is planned, will additional baseline covariates be needed for the analyses?
   e. Will the transition affect the statistical precision of the analyses?

3. Planning for the End of a Patient Registry

Once a registry is in place, how long should it continue? What are reasonable decision criteria for stopping data collection? This section considers the issues related to stopping a patient registry study and suggests some guidelines. Although the specific answers to these questions will vary from study to study, general types of considerations can be identified. The discussion here focuses on registries intended to assess specific safety or effectiveness outcomes rather than those intended to assess health care operations such as continuous quality improvement. See also Case Example 28.

3.1 When Should a Patient Registry End?

3.1.1 Stopping an Experiment

The principles regarding rules for stopping a study mostly stem from the need to consider stopping an experiment. Because experiments differ from registries in crucial ways, it is important to distinguish between the issues involved in stopping an experimental study and in stopping a nonexperimental study. In an experiment, the patient's treatment is determined by the study protocol, which typically involves random assignment to a treatment regimen. In a nonexperimental study, patients are treated according to the treatment protocol devised by their own clinicians, typically uninfluenced by the study. In a randomized trial of a new therapeutic

agent or a field trial for a vaccine, the size of the study population is ordinarily set in the study protocol, based on assumptions about the expected or hypothesized results and the study size needed to reach a reasonable scientific conclusion. Ordinarily this planned study size is based on power calculations, which require as input the criteria for statistical significance, the effect size anticipated, the baseline occurrence rate of the study outcome, and the relative size of the study arms. Because of inherent problems in relying on statistical significance for inference,[11, 12] the study size preferably will be planned around estimation of effect and the desired level of precision. In a study intended to provide some reassurance about the safety of an agent, the study size may be planned to provide a specific probability that the upper confidence bound of a conventional confidence interval measuring an adverse effect would be less than some specified value, given a postulated value for the effect itself (such as no effect). In the latter situation, if no effect is anticipated, a power calculation is not only unreasonable but is not even possible, whereas planning a study on the basis of precision of estimation is always possible and always reasonable.

Stopping an experiment earlier than planned is an important decision typically made by an advisory group, such as a data safety and monitoring board, which is constituted to monitor study results and make decisions about early stopping. In a biomedical experiment, the investigator has a greater ethical obligation than in a nonexperimental study to safeguard the well-being of study participants. This is because the investigator is administering an intervention to study participants that is expected to affect the probability that study participants will experience one or more specific health outcomes.

Equipoise is a widely accepted (but, unfortunately, not universally accepted) ethical precept regarding human biomedical experimentation.[13] Equipoise requires that at the outset of the study, the investigator has a neutral outlook regarding which of the study groups would fare better. A strict interpretation of equipoise requires each of the study investigators to be in a state of equipoise. An alternative view, referred to as "clinical equipoise," is that equipoise can be achieved at the group level, with the enthusiasm of some investigators for the prospects of the study intervention being balanced by the skepticism of others.[14] Whichever interpretation of equipoise is adopted, most investigators agree that if equipoise becomes untenable as study results accumulate, the study should be stopped to avoid depriving some study participants of a potential benefit relative to what other participants receive.

For an advisory board to decide to stop a study early, there must be solid evidence of a difference between the groups before the planned study endpoint is reached. Such stopping decisions are usually based on ethical concerns, as scientific considerations would seldom dictate an early stop to a study that had been planned to reach a specific size. Advisory boards must base stopping decisions on analyses of accumulating study data, which are usually formally presented at regular meetings of the review board. Statistical concerns have been raised about biases that can arise from repeated analyses of accumulating data.[15] To offset these concerns, many experiments are planned with only a limited number of interim analyses, and the interpretation of study results takes into account the number of interim analyses.

3.1.2 Stopping a Fixed-Length Nonexperimental Study

Like experiments, most nonexperimental studies also have a fixed time for their conduct and a planned size that reflects goals analogous to those in experimental studies. Nevertheless, the ethical concerns that motivate stopping an experiment before its planned completion do not have a direct counterpart in nonexperimental studies. Nonexperimental studies have ethical concerns, but they relate to issues such as data privacy, intrusive questioning, or excessive inducements for participation rather than to concerns about intervention in the lives of the participants. Although it is theoretically reasonable that an investigator could choose to stop a nonexperimental study for ethical reasons, those reasons would presumably relate to ethical problems that were discovered in the course of the study but were unrecognized at the outset rather

than to an early conclusion regarding the study goal. For example, the investigator in a nonexperimental study could learn from an interim analysis that the association between the exposure and the outcome under study was much stronger than anticipated. Unlike the experimental setting, however, the investigator in a nonexperimental study is not administering the exposure to any of the study subjects and thus has no responsibility to the study subjects regarding their exposure.

The discovery of an ethical problem during the conduct of a nonexperimental study is therefore possible but extremely rare. Because the findings from an interim analysis should not lead to discontinuation of a nonexperimental study, there is little motivation to conduct interim analyses for nonexperimental studies that have been planned with a fixed size and period of execution. If there is some considerable time value to the findings, such as to inform regulatory action, it might be worthwhile to conduct an interim analysis to get an early appraisal of study findings. Unless there is an appropriate outlet for releasing interim findings, however, it is possible that early findings will not circulate beyond the circle of investigators. In most circumstances, such analyses are hard to justify in light of the fact that they are based on a smaller amount of data than was judged appropriate when the study was planned; thus the originally planned analysis based on all the collected data will still need to be conducted. Unless there is a clear public health case to publicize interim results, journal policies that require that data to be published have not been previously published may inhibit any release of preliminary findings to news media or to journals.

3.1.3 Stopping an Open-Ended Study

Although patient registries may be undertaken with a fixed length or size, or both, based on study goals relating to specific safety or efficacy hypotheses, many such studies are begun as open-ended enterprises without a planned stopping point. For example, patient registries without specific hypotheses may be undertaken to monitor the safety of patients receiving a novel therapy. The Antiepileptic Drug Pregnancy Registry, established in 1997, is an example of an open-ended registry that focuses on a set of specific endpoints (congenital malformations) among a subset of patients (pregnant women) taking a class of medications (antiepileptic drugs).[16] It has no fixed stopping point.

Measuring the frequency of rare endpoints demands large study sizes. Therefore, a monitoring system that includes rare endpoints may have to run for a long while before the accumulated data will be informative for low-frequency events. On the other hand, the lower the frequency of an adverse event, even one with serious consequences, the smaller is the public health problem that a relative excess of such events would represent.

Traditional surveillance systems are intended to continue indefinitely because they are intended to monitor changes in event frequency over time. For example, surveillance systems for epidemic infectious diseases provide early warning about outbreaks and help direct efforts to contain such outbreaks. In contrast, a patient registry is not a true surveillance system, since most are not intended to provide an early warning of a change in outcome frequency. Rather, most patient registries are intended to compile data on outcomes associated with novel treatments, to supplement the sparse data usually available at the time that these treatments are considered for approval by regulatory agencies. For example, a regulatory agency might mandate a patient registry as a condition of approval to supplement safety information that was submitted during the application process.

How long should such a registry continue? Although it is not possible to supply a general answer to this question, there is little reason to support a registry continuing indefinitely unless there is a suspicion that the treatments or treatment effects will change over time. Otherwise, the time should come when the number of patients studied suffices to answer the questions that motivated the registry. The Acyclovir Pregnancy Registry, which began in 1984, was stopped in 1999. Its advisory committee concluded: “The [Acyclovir Pregnancy] Registry findings to date do not show an increase in the number of birth defects identified among the prospective reports [of exposures to acyclovir]

when compared with those expected in the general population. In addition, there is no pattern of defects among prospective or retrospective acyclovir reports. These findings should provide some assurance in counseling women following prenatal exposure [to acyclovir]."[17] The consensus was that additional information would not add materially to the information that already had been collected, and thus the registry was closed down.

To avoid uncertainty about the fate of an open-ended study, it would be sensible to formulate a specific goal that permits a satisfactory conclusion to data collection. Such a goal might be, for example, the observation of a minimum number of specific adverse events of some type. Even better would be to plan to continue data collection until the upper bound of a confidence interval for the rate or risk of the key outcome falls below some threshold or until the lower bound falls above a threshold. Analogous stopping guidelines could be formulated for registry studies that are designed with a built-in comparison group.

3.2 Decisions on Stopping and Registry Goals

Ideally, stopping decisions ought to evaluate data from a registry against its stated goals. Thus, the registry protocol or charter should include one or more specific and measurable endpoints against which to judge whether the project should continue or stop. Without that guidance, any decision to discontinue a registry may appear arbitrary and will be more readily subject to political considerations. In cases where there are no measurable endpoints to use in making the decision, it is important that any final reports or publications linked to the registry include a clear discussion of the reasons for stopping it.

Registry goals will vary according to the motivation for undertaking the project and the source of funding. Product-specific registries may be created as postapproval regulatory commitments. For products about which there are limited preapproval safety data, the wish for additional comfort about the product's safety profile can be translated into a measurable goal. Such a goal might be to exclude the occurrence of life-threatening or fatal drug-related events at a certain frequency. For example, the goal could be to establish a specified level of confidence that unexplained hepatic necrosis in the 3 months following drug exposure occurs in less than 1 patient in 1,000. Alternatively, the goal might be to provide a more precise estimate of the frequency of a previously identified risk, such as anaphylaxis. Ideally, this goal should be formulated in specific numeric terms. With specific goals, the registry can have a planned target and will not be open ended.

If a registry study does not have a single or very limited set of primary objectives, a stopping point will be more challenging to plan and to justify. Even so, with measurable goals for some endpoints, it will be possible to determine whether the registry has achieved a core purpose, indicating a reasonable stopping point. Conversely, a registry that fails to meet measurable goals and appears to be unable to meet them in a reasonable time is also a candidate to be stopped. For example, if the registry faces unexpectedly low patient accrual, it should be stopped, as was done with the Observational Familial Adenomatous Polyposis Registry Study in Patients Receiving Celecoxib.[18] This study enrolled only 72 patients in 4 years, out of a planned 200 during 5 years. Another reason to consider stopping is incomplete or poor-quality information. Poor-quality data are of particular concern when the data regard sensitive or illegal behavior, such as self-reported information on sexual practices.[19] Decisions about stopping a registry because of low enrollment or inadequate information are made simpler with clearly stated goals regarding both features of the study. The criteria for useful quantity and quality of information should be specified at the outset. How well the study meets the criteria can be assessed periodically during data collection.

A registry may outlive the question it was created to answer. For example, if use of the product is superseded by another treatment, the questions that drove the creation of the registry may no longer be relevant, in which case it may best be retired (see Case Example 45). For medical devices, for example, newer technology is continuously replacing the old, although safety issues for older technology may motivate continuing a registry of

an outmoded technology. A related issue arises when the question of interest evolves as data collection proceeds. Stopping or continuing the registry depends on whether it can address the changing goal or goals. That, in turn, depends on whether the governance of the registry provides adequate flexibility to refocus the registry in a new direction.

The decision to stop a registry may also depend on mundane considerations such as cost or staffing. For long-running registries, eventually the value of new information may face diminishing returns. Some registries have central core staff, deeply committed to the registry, who serve as its historical memory. Departure of such individuals can cripple the registry's function, and a decision to stop may be appropriate. Similarly, a cohort of engaged investigators may disperse over time or lose interest in the registry. Funding sources may dry up, making it impossible for the registry to function at a level that justifies its continued existence.

A thorny question concerns how a registry can continue with altered ownership or governance. Suppose a registry is formed with multiple stakeholders, and one or more withdraws for the reasons described above. For example, when the implantable cardioverter defibrillator (ICD) registry was formed, it came about in response to a CMS Coverage with Evidence Development decision. The Heart Rhythm Society and the American College of Cardiology developed the registry with funding from industry to help institutions meet the need for registry participation for payment purposes, and they layered quality improvement and research goals onto that mandate.[20] The resulting registry was rapidly integrated into more than 2,000 institutions in the United States. If CMS determines that the ICD registry is no longer needed for its purposes, the registry must determine if it will continue as a quality improvement program and whether to add other stakeholders and funding sources or participation drivers (such as manufacturers, insurers, or other government agencies such as FDA).

3.3 What Happens When a Registry Ends?

Stopping a registry might mean ceasing all information collection and issuing a final report. An intermediate decision that falls short of a full stop might involve ceasing to accrue new patients while continuing to collect information on existing participants. This step may be useful if the registry goals are in the process of changing. If a registry is to be stopped, the archiving rules should be checked and followed, so that those who need to consult the data for questions not fully addressed in reports or publications can get their answers later, provided that the charter of the registry allows it. Following German reunification in 1990, it was determined that the East German National Cancer Registry, which had received detailed reports on 2 million cancer cases from 1954 to 1990, was in violation of West German privacy laws, and the data were quarantined. In the more usual case, orderly archiving of the data in anticipation of later access should be part of the close-down procedure, in a manner consistent with the charter under which the data were collected.[21]

A slightly different scenario occurs when the registry has a single sponsor whose purposes have been achieved or determined to be unachievable, and the sponsor decides to end the registry. Is there an obligation to patients or participating providers to continue the registry because some value (e.g., quality improvement, data for other comparisons) can still be derived? It is difficult to argue that the sponsor has an ongoing financial responsibility once the registry has achieved or failed to achieve its primary purpose, especially if this has been spelled out in the protocol and informed consent. Yet one can argue that, to the extent that it is feasible and affordable to engage other stakeholders in discussions of potential transitioning of the registry to other owners, this approach should be encouraged. Nontrivial issues of data ownership, property, confidentiality, and patient privacy would need to be satisfactorily addressed to make such transitions possible, and therefore it is always best to consider this possibility early on in registry planning. Both the National Registry of Myocardial Infarction, sponsored by Genentech, Inc., and the OPTIMIZE-

HF registry, sponsored by GlaxoSmithKline, successfully completed transitions to other organizations (American College of Cardiology and American Heart Association, respectively) when those registries were concluded, providing their participating hospitals with the ability to continue the quality improvement efforts begun under those registries.[22, 23]

There is no clear ethical obligation to participants to continue a registry that has outlived its scientific usefulness. In fact, altering the purpose of a registry would be complicated unless the original registry operators were interested in doing so. For instance, if a registry is to be transferred, then it should be a restricted transfer (presumably a gift) to ensure that the permissions, terms, and conditions under which it was compiled continue to be satisfied. The participants should be notified and should determine if they will continue participation and allow their data to be used for this new purpose.

There are a few potential reasons to consider preserving registry data once the registry developers have determined that it should end. One reason is that the data may be capable of producing a recognized public health benefit that will continue if the registry does. Another situation may be that the registry has historical importance, such as a registry that tracks the outbreak of a novel infectious disease that may provide insight into the transmission of the disease, if not now, then sometime in the future. Longitudinal collections of data may also be useful for hypothesis generation.

In creating a registry, the investigators should plan what will happen to data when the registry ends. If a public health benefit might be realized from registry data, then archiving of registry data is a potential answer. Decisions must be made by the registry owners in careful consideration of other stakeholders, potential costs, and privacy and security concerns.

3.4 Summary of Considerations for Planning for the End of a Registry

Experimental studies, such as clinical trials or field trials, come with a high ethical burden of responsibility, which includes periodically reevaluating the ethical basis for continuing the trial in the light of interim results. Consequently, trials require interim analyses and data safety monitoring boards, which decide whether the study should be stopped for ethical reasons. In nonexperimental studies, there is much less motivation to conduct interim analyses because there is no ethical motivation to do so. There is also no reason to appoint a data safety monitoring board, although any study could appoint an external advisory board. If nonexperimental studies are planned to be of fixed length or fixed study size, they can be conducted as planned without interim analyses, unless the time value of an early, interim analysis is important enough to compensate for the added cost of conducting it and the tentativeness of the findings, which are based on only a subset of the planned study data.

Even if a patient registry is undertaken as an open-ended project without a fixed endpoint, it need not continue forever. Unlike true surveillance efforts, patient registries of novel therapies are not intended to monitor changes in occurrence rates over time. Rather, they are conducted to assemble enough data to evaluate associations that could not be evaluated with the limited data available at the time of new product approval. Therefore, reasonable goals should be set for the amount of information to be collected in such registries, based on specific endpoints of interest. These goals can and should be cast in specific terms regarding data quality, study enrollment, and precision of the estimates of specific measures that the registry is intended to describe.

Case Examples for Chapter 14

Case Example 28. Determining when to stop an open-ended registry

Description	The Bupropion Pregnancy Registry was an observational exposure-registration and followup study to monitor prenatal exposure to bupropion and detect any major teratogenic effect.
Sponsor	GlaxoSmithKline
Year Started	1997
Year Ended	The registry closed to new enrollments on November 1, 2007, and continued to follow existing cases through March 31, 2008.
No. of Sites	Not applicable
No. of Patients	1,597

Challenge

Bupropion, an antidepressant with the potential for prenatal exposure, was labeled with a pregnancy category C by the U.S. Food and Drug Administration (FDA) due to prior animal data. The manufacturer established a prospective pregnancy registry to monitor pregnancy exposures to bupropion for any potential increased risk of congenital anomalies. Because the purpose of the registry was postmarketing safety surveillance, the duration of the registry was open ended. The registry had collected data on more than 1,500 exposed pregnant women over 10 years when a potential signal suggestive of a bupropion-related increase in cardiovascular birth defects emerged.

Proposed Solution

The advisory committee reviewed the registry data to assess the potential signal. However, due to the potential bias from the large percentage of cases lost to followup (35.8%), retrospective reports, and incomplete descriptions of the reported cardiovascular defects, it was not possible to determine the credibility of the potential signal using registry data alone. Further, the sample size was not adequate to reach definitive conclusions regarding the absolute or relative risk of any specific birth defects in women using bupropion during pregnancy (as the registry was powered only to examine the rate of birth defects overall) and was unlikely to achieve its goal as structured.

The advisory committee recommended a study to expedite the accumulation of pregnancy outcome data among women exposed to bupropion during pregnancy. In response, a large, claims-based, retrospective cohort study was conducted. This study enrolled 1,213 women exposed in the first trimester and did not confirm a consistent pattern of defects (Cole et al., 2007). The prevalence of cardiovascular defects associated with first-trimester exposure to bupropion was 10.7 per 1,000 infants.

Results

The advisory committee reviewed the evidence and concluded that the signal did not represent an increased risk. The committee recommended discontinuation of the registry based on findings from the retrospective cohort and 10 years of surveillance through the registry. The committee took the position that sufficient information had accumulated to meet the scientific objective of the registry. The high lost-to-followup rate was also taken into consideration. The registry closed to new enrollments on November 1, 2007, and continued to follow existing cases through March 31, 2008.

Key Point

In a registry without a specified end date or target size, it is important to periodically review the registry data to determine if the registry has met its scientific objectives and to ensure that the registry purpose is still relevant.

Case Example 28. Determining when to stop an open-ended registry (continued)

For More Information

Cole JA, Modell JG, Haight BR. et al. Bupropion in pregnancy and the prevalence of congenital malformations. Pharmacoepidemiol Drug Safety. 2007;16:474–84.

The Bupropion Pregnancy Registry. Final Report: September 1, 1997–March 31, 2008. Issued August 2008. http://pregnancyregistry.gsk.com/bupropion.html. Accessed June 6, 2012.

Alwan S, Reefhuis J, Botto LD, et al. Maternal use of bupropion and risk for congenital heart defects. Am J Obstet Gynecol. 2010 Jul;203(1):52.e1-6. Epub 2010 Apr 24.

Case Example 29. Challenges in transitions and changes in data collection

Description	The Cystic Fibrosis Foundation (CFF) Patient Registry is a rare-disease registry that collects data from clinical visits, hospitalizations, and care episodes to track national trends in morbidity and mortality, assess the effectiveness of treatments, and drive quality improvement in care for patients with cystic fibrosis (CF).
Sponsor	Cystic Fibrosis Foundation
Year Started	1986
Year Ended	Ongoing
No. of Sites	110 CFF-accredited care centers in the United States
No. of Patients	More than 26,000

Challenge

The CFF Patient Registry collects information on more than 26,000 patients with cystic fibrosis who receive care at one of over 110 CFF-accredited care centers across the United States. The registry collects demographic and diagnostic information, lung function, nutritional status, respiratory microbiology, and other indicators of disease progression, as well as prescribed medications and hospitalizations. Registry data dating back to 1986 are used to track national trends in morbidity and mortality, to assess the effectiveness of treatments, and to drive quality improvement. The CFF Patient Registry has evolved through multiple iterations, including the most recent transition in 2010. This registry transition, prompted by the implementation of a new technology platform and the need to update and expand the data collection fields, imposed operational challenges on the CFF working group tasked with the registry transition. Challenges included modifying data collection fields, mapping historical data, and ensuring usability for the sites.

Proposed Solution

Planning for the latest transition began in 2007, when CFF began searching for a new vendor with an FDA-compliant (21 CFR, Part 11) Web platform. After selecting a vendor, CFF began the process of evaluating all data fields in the registry. Six working groups of subject matter experts were convened to review data fields from the existing registry and make suggestions for improvements, additions, and deletions (in the areas of genetics, pulmonology, microbiology, cystic fibrosis–related diabetes, transplantation, and infant care/ambiguous diagnosis). CFF staff created an online prototype of the registry that allowed reviewers, including registry physicians and subject matter experts, to see and comment on all data fields and functions planned for the new registry platform. The recommendations of these groups were vetted by CFF staff in order to balance data entry burden for the care center staff and usefulness for future research.

After selecting data fields for the new version of the registry, focus shifted to designing online case report forms and mapping historical data into the new registry. For example, the old registry

Case Example 29. Challenges in transitions and changes in data collection (continued)

Proposed Solution (continued)

platform collected information on 81 distinct genotype mutations in addition to an "other, specify:" option. The genetics working group expanded the list of available mutation variables to 269, and data previously entered into the "other, specify:" field were mapped forward into these new mutation variables. Once mapping was complete, there were several iterative migrations of historical data into the new registry platform. CFF staff carefully audited the import process and identified errors prior to the launch of the live registry.

During a preliminary test period, users were able to log in to the new registry, populated with "dummy" data, and become acquainted with the new functionality, format, and questions in the registry. During a second test period, users viewed their actual patient data in the new registry, and helped to identify and resolve several minor errors in data migration. When the registry went live in April 2010, users were supported with online training manuals and videos, FAQ documents, Webinars, and live support.

Results

Three months after the successful addition of new fields and data migration to the new registry, a survey was distributed to all sites to solicit user feedback and identify areas for further change. The response rate was high, with 70 percent of users providing feedback on their experiences with the new platform. Responders ranged from new users to experienced sites, and results indicated that sites were still gaining familiarity with the new registry. Based on the survey responses, several modifications to the platform were made to improve the functionality of the system.

Key Point

The transition of the CFF Patient Registry to a new technology platform illustrates the importance of a well-planned registry transition. Early allocation of human and financial resources, collaboration with sites, and a realistic timeline allowed for stepwise development. This approach minimized disruption to users and ensured the integrity of the data migration process.

For More Information

www.cff.org

Case Example 30. Transitioning from startup to ongoing registry funding with public and private partners

Description	Rheumatoid Arthritis Comparative Effectiveness Research (RACER) is a disease registry designed to assess comparative and cost effectiveness of existing and new biologic drug therapies for rheumatoid arthritis.
Sponsor	University of Pittsburgh; National Institutes of Health; Genentech, Inc.
Year Started	2010
Year Ended	Ongoing
No. of Sites	4 clinics in the Pittsburgh area
No. of Patients	More than 1,000

Challenge

RACER was established by researchers at the University of Pittsburgh to determine relative effectiveness of existing versus new and expensive biologic therapies for rheumatoid arthritis; compare biologic therapies in terms of mechanistic and biomarker measurements to predict optimal treatment; and test whether practical treat-to-target strategies can improve treatment management of the disease. Clinical outcomes assessed include disease activity, quality of life, function and work productivity, and biomarker and mechanistic outcomes including C-reactive protein, rheumatoid factor, and cyclic citrullinated peptide auto-antibody levels.

There is no protocol-mandated visit schedule, and providers determine the visit schedule appropriate for each patient. Registry data are extracted from the site's electronic health record (EHR) system and supplemented with patient-reported outcome (PRO) measures and some physician-reported measures at each visit. The laboratory values are measured using a blood sample collected at each visit.

Initial funding for the registry was provided in 2009 through a grant from the National Institutes of Health (NIH), which allowed the registry to be launched, patients to be enrolled, and data collection to begin. However, the grant mechanism was for a finite period of 2 years, and university researchers saw value in continuing data collection. Researchers sought to continue registry operations and disseminate registry data and results to interested parties.

Proposed Solution

Prior to the end of NIH funding, the registry, which now had data on more than 1,000 patients, began discussions with Genentech, Inc, a biotechnology company with two marketed biologics for rheumatoid arthritis. Initially, the discussions focused on data sharing and acquiring access to the registry data set with an intent to collaborate and generate real-world effectiveness of rheumatoid arthritis treatments. Although Genentech was not mandated by a regulatory agency to collect postmarketing data on their product, they expressed interest in being able to collect this data for scientific purposes with minimal investment in redundant or duplicative registry infrastructure. Genentech valued the work that RACER had done already, especially EHR integration, the collection of PROs and quality-of-life measures, and the potential for nested substudies within the larger registry database.

Discussions between the registry and Genentech turned to possible models for sustaining registry operations while granting Genentech access to the valuable comparative effectiveness data being collected by the registry.

Results

Plans are now in place to fully transition the funding of the registry to Genentech after the original NIH contract has ended. Genentech will continue to have full access to registry data on an ongoing basis, but will have an arms-length relationship to the operations of the registry. The researchers and registry staff at the University of Pittsburgh will retain full control of registry operations and ownership of the registry data set.

Case Example 30. Transitioning from startup to ongoing registry funding with public and private partners (continued)

Results (continued)

The University of Pittsburgh researchers and Genentech will meet regularly to collaborate on analyses, manuscripts, and abstracts. Work is currently being planned on a protocol for a nested substudy to evaluate the relative efficacy of the four clinically relevant treatment options in patients failing their first tumor necrosis factor (TNF) antagonist treatment.

Key Point

Registries can benefit from public and private partners to secure funding during both the startup and maintenance phases of the registry. Members of a public-private partnership may join and leave as funding becomes available and shared interests intersect. This partnership enables effective public-private collaboration to advance science using observational research.

For More Information

Patel AM, Amity CL, Frydrych LM, et al. The 2010 rheumatoid arthritis criteria versus the 1987 rheumatoid arthritis criteria: Will the real criteria please stand up! Arthritis Rheum. 2011 Oct;63(10):S48-9.

Soejima M, Patel AM, Goudeau D, et al. Effect of disease modifying anti-rheumatic Drugs (DMARDs) on anti-CCP2 and anti-citrullinated protein antibody (ACPA) levels during longitudinal assessments in rheumatoid arthritis patients. Arthritis Rheum. 2011 Oct;63(10):S146-7.

Soejima M, Zhijie Z, Jones DM, et al. Prognostic and diagnostic significance of autoantibodies to citrullinated proteins (ACPA) in patients with a scleroderma-rheumatoid arthritis (SSc-RA) overlap syndrome. Arthritis Rheum. 2011 Oct;63(10):S332-3.

Lupash D, Patel AM, Amity CL, et al. Comparison of the Patient-Based Routine Assessment of Patient Index 3 in usual care of rheumatoid arthritis to the Physician-Based Disease Activity Score-28 Joint Count and Clinical Disease Activity Index. Arthritis Rheum. 2011 Oct;63(10):S834.

Patel AM, Amity CL, Frydrych LM, et al. Comparison of rheumatoid arthritis patient characteristics from randomized controlled trials to a registry designed for rheumatoid arthritis comparative effectiveness research. Arthritis Rheum. 2011 Oct;63(10):S837-8.

Case Example 31. Modifying a registry due to changes in standards of care

Description	The GOLD reGISTry was a prospective, multicenter, 5-year global disease registry designed to collect information on patients with advanced and localized gastrointestinal stromal tumors. The registry collected diagnostic, treatment, and outcomes information in order to identify and compare practice patterns worldwide and assist practitioners in making treatment decisions as standards of care evolved.
Sponsor	Novartis Oncology
Year Started	2007
Year Ended	2012
No. of Sites	More than 200
No. of Patients	1653

Challenge

When it was launched in 2007, the 5-year GOLD reGISTry enrolled only patients with advanced gastrointestinal stromal tumors (GIST). This population was of interest to researchers because standards of care for advanced GIST were not as clearly defined and widely used as the standard of care for localized GIST, which was complete surgical excision. The sponsor expected that the outcomes data collected from advanced GIST patients would be valuable in helping to refine standards of care for these patients.

In 2008 and 2009, Gleevec®/Glivec® (imatinib mesylate) received FDA and European Medicines Agency (EMA) approval for adjuvant use in localized GIST after tumor resection. This approval, combined with emerging clinical trial data, prompted new interest in collecting diagnostic, treatment, and outcomes information from patients with localized GIST.

Proposed Solution

The sponsor had selected a steering committee with engaged key opinion leaders who provided guidance for the study and encouraged flexibility in study design to allow for potential changes. In 2009, the steering committee convened and determined that the registry would begin collecting data on patients with localized GIST, in addition to those with advanced disease who were already enrolled in the registry. The study team drafted a protocol amendment to include the localized GIST population and allowing assessment of physician adherence to new clinical guidelines published by the European Society of Medical Oncology the same year. The data management and statistical analysis plans were also revised to allow for the incorporation of the new data.

Significant efforts were then directed at site engagement, including abstract submissions and publicity through the key opinion leaders. The registry also maintained site interest through interim study summaries presented at professional congresses. The sponsor had limited monitoring resources available to accommodate the new patient population, so study designers developed a plan that used remote monitoring and training, reserving onsite visits for research-naïve sites or for-cause audits. This allowed monitors to focus on those sites that required more training and allowed these sites to gain clinical research experience in an observational study.

Results

The registry enrolled 1,653 patients in the two populations within four years: more than 1,000 with advanced GIST, and more than 500 with localized GIST. The steering committee played an important role in the recruitment and retention of sites, highlighting the importance of the study through publications and interim summaries presented at scientific and professional congresses throughout the enrollment period. As the planned 5-year study period has been completed, the sponsor is now in the process of locking the registry database in preparation for final analyses.

Case Example 31. Modifying a registry due to changes in standards of care (continued)

Key Point

Changes in standard of care can significantly impact the design of a study as new treatments are approved or new patient populations become of interest. Registry developers should anticipate that such changes might occur, and should consider what aspects of the registry could be most impacted. A steering committee well regarded in the field and knowledgeable about the disease and treatment can provide significant guidance during registry transitions and keep sites engaged as the changes are implemented.

For More Information

DeMatteo RP, Ballman KV, Antonescu CR, et al. Adjuvant imatinib mesylate after resection of localised, primary gastrointestinal stromal tumour: a randomised, double-blind, placebo-controlled trial. Lancet. 2009 Mar 28;373(9669):1097-104.

Casali PG, Jost L, Reichardt P, et al. Gastrointestinal stromal tumours : ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol. 2009 May;20 Suppl 4: 64-7.

Chacon M, Reichardt P, Gu J, et al. The GOLD reGISTry: A global observational registry collecting longitudinal data on patients with advanced GIST—Second annual summary. J Clin Oncol (Meeting Abstracts). May 2010; 28 (15 Suppl.): 10092.

References for Chapter 14

1. Nosowsky R, Giordano TJ. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule: implications for clinical research. Annu Rev Med. 2006;57:575-90. PMID: 16409167.
2. Littenberg B, MacLean CD. Passive consent for clinical research in the age of HIPAA. J Gen Intern Med. 2006 Mar;21(3):207-11. PMID: 16637821. PMCID: 1828090.
3. Al-Shahi R, Warlow C. Using patient-identifiable data for observational research and audit. BMJ. 2000 Oct 28;321(7268):1031-2. PMID: 11053151. PMCID: 1118832.
4. Tu JV, Willison DJ, Silver FL, et al. Impracticability of informed consent in the Registry of the Canadian Stroke Network. N Engl J Med. 2004 Apr 1;350(14):1414-21. PMID: 15070791.
5. Truyers C, Kellen E, Arbyn M, et al. The use of human tissue in epidemiological research; ethical and legal considerations in two biobanks in Belgium. Med Health Care Philos. 2010 May;13(2):169-75. PMID: 19936964.
6. Hauser RM, Weinstein M, Pool R, et al., eds. Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata. Washington, DC: National Academies Press; 2010.
7. Hofmann B. Broadening consent--and diluting ethics? J Med Ethics. 2009 Feb;35(2):125-9. PMID: 19181887.
8. Ries NM. Growing up as a research subject: ethical and legal issues in birth cohort studies involving genetic research. Health Law J. 2007;15:1-41. PMID: 19702179.
9. Ries NM, LeGrandeur J, Caulfield T. Handling ethical, legal and social issues in birth cohort studies involving genetic research: responses from studies in six countries. BMC Med Ethics. 2010;11:4. PMID: 20331891. PMCID: 2859353.
10. Knowles M. Adult Learning. In: Craing R, ed. The ASTD Training and Development Handbook. New York: McGraw-Hill; 1996. p. 253-64.
11. Rothman K, Greenland S, Lash TL, eds. Modern Epidemiology. 3rd ed. New York: Lippincott Williams & Wilkins; 2008.

12. Rothman KJ, Johnson ES, Sugano DS. Is flutamide effective in patients with bilateral orchiectomy? Lancet. 1999 Apr 3;353(9159):1184. PMID: 10210003.

13. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 1987 Jul 16;317(3): 141-5. PMID: 3600702.

14. Weijer C, Shapiro SH, Cranley Glass K. For and against: clinical equipoise and not the uncertainty principle is the moral underpinning of the randomised controlled trial. BMJ. 2000 Sep 23;321(7263):756-8. PMID: 10999914. PMCID: 1127868.

15. McPherson K. Statistics: the problem of examining accumulating data more than once. N Engl J Med. 1974 Feb 28;290(9):501-2. PMID: 4589874.

16. The antiepileptic drug registry. http://www.aedpregnancyregistry.org/. Accessed August 15, 2012.

17. Acyclovir Pregnancy Registry and Valacyclovir Pregnancy Registry Interim Report, December 1997. Glaxo Wellcome, RTP, NC 27709; as referenced on Web page titled, "GlaxoSmithKline Pregnancy Registries.". http://pregnancyregistry.gsk.com/acyclovir.html. Accessed August 15, 2012.

18. ClinicalTrials.gov. Observational Familial Adenomatous Polyposis Registry Study In Patients Receiving Celecoxib Compared to Control Patients. http://clinicaltrials.gov/ct2/show/NCT00151476. Accessed August 15, 2012.

19. Eng TR, Butler WT. The hidden epidemic: confronting sexually transmitted diseases. Institute of Medicine (U.S.) Committee on Prevention and Control of Sexually Transmitted Diseases. Washington, DC: National Academies Press; 1997.

20. Hammill S, Phurrough S, Brindis R. The National ICD Registry: now and into the future. Heart Rhythm. 2006 Apr;3(4):470-3. PMID: 16567298.

21. House of Commons Select Committee on Science and Technology. Science and Technology Sixth Report. July 27, 2000. Section on Cancer Research: Epidemiology and Publich Health Research. http://www.publications.parliament.uk/pa/cm199900/cmselect/cmsctech/332/33208.htm#a39. Accessed September 10, 2013.

22. New Program to Help Hospitals Improve Care for Health Failure Patients Saves Lives [press release]. http://www.prnewswire.com/news-releases/new-program-to-help-hospitals-improve-care-for-heart-failure-patients-saves-lives-54170737.html. Accessed August 28, 2012.

23. "American College Of Cardiology Foundation's NCDR Creates Network To Measure Patient Care." Medical News Today. www.medicalnewstoday.com/articles/55798.php. Accessed July 2, 2012.

U.S. Department of
Health and Human Services

Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850

AHRQ Pub. No. 13(14)-EHC111
April 2014
ISBN: 978-1-58763-432-1