Final Report

The Feasibility of Using Electronic Health Records (EHRs) and Other Electronic Health Data for Research on Small Populations

September 2013

Prepared by:
Kelly Devers, PhD
Bradford Gray, PhD
Christal Ramos, MPH
Arnav Shah, BA
Fredric Blavin, PhD
Timothy Waidmann, PhD

URBAN INSTITUTE

Acknowledgement

Disclaimer

Table of Contents

Abstract

Background. Many small populations have distinctive health and health care needs but have been difficult to study in survey research.

Objective. This report is part of a project funded by the Assistant Secretary for Planning and Evaluation to explore the feasibility of using electronic health record (EHR) and other electronic health data for research on small populations. The first part of the report illustrates the challenges and limitations of using existing federal surveys and federal claims databases for studying small populations. The second part explores the potential of the increasingly available EHR and other existing electronic health data to complement federal data sources, as well as potential next steps to demonstrate and improve the feasibility of using EHRs for research on small populations.

Methods. We use four example small populations throughout the report to illustrate a range of health and health care needs and considerations for research: Asian subpopulations; lesbian, gay, bisexual, and transgender populations; rural populations; and adolescents with autism spectrum disorders. We conducted interviews with experts on the health, health care and research needs for these small populations, as well as with experts on current efforts to use EHR and other electronic health data for research. Findings are based on these interviews, literature, and feedback from a technical expert panel.

Results. Challenges to studying small populations using federal survey data include their small size, uneven distribution, and lack of standardized ways to identify population members. The growing availability of EHR and other existing health information has the potential to help overcome some of these challenges, given a number of conditions are met to be able to use these data for research. These include technical, legal, and organizational conditions that each come with their own challenges. However, these challenges are being addressed by researchers around the country who have begun to use EHR and other electronic health data for research on small populations, particularly from organized delivery systems and research networks. Potential next steps may include improving data quality through validation studies and clinician engagement, development of research methods using a combination of data sources, efforts to improve the legal framework under which this type of research is regulated, and pilot studies on specific small populations.

Conclusions. There is great potential for using EHR and other existing electronic health data to study small populations. As with federal survey data, EHR data may be better suited for some types of research than others, and the context within which the data was collected must be kept in mind. Secondary use of existing electronic health data is challenging traditional views of research methods, privacy, and research collaboration. To further tap the potential use of these data for research on small populations, the Department of Health and Human Services could work with stakeholders to identify and prioritize key next steps and the potential role that public and/or private funders can play.

Executive Summary

Why Study Small Populations?

A vast body of research shows important differences among segments of the population on virtually all aspects of health and health care. These segments may be defined by characteristics such as race, ethnicity, sexual orientation, geography, health conditions or other factors. It is important to understand the needs of these populations in order to better provide patient-centered, culturally appropriate care. Being able to customize care to best serve the needs of different segments of the population is a critical step between the management of population health and personalized medicine. Documenting differences among these segments is an essential starting point for a wide array of policies and interventions to improve peoples' health. Although much of what we know about the health of the U.S. population comes from national surveys conducted by the federal government, there are major limitations on the use of federal survey data, particularly for studying small populations.

The needs of four example populations and the limitations in studying them using federal survey-based research are explored in the first part of the report. These examples include Asian-American subpopulations; the lesbian, gay, bisexual, and transgender (LGBT) population; adolescents with autism spectrum disorders (ASDs); and residents of rural areas. These populations were selected based on conversations with a number of federal agencies to provide a broad range of pressing health and health care questions and challenges in studying small populations. An additional consideration was to explore populations that are not so small that obtaining sufficient information about them would be infeasible now or in the near future. Due to the specific health care needs as well as the limitations in studying these small populations using survey data, there has been much interest in exploring alternate data sources that can be used for research, such as electronic health record (EHR) data and other existing electronic health data, which are explored in the second part of this report. The report is based on published information, interviews with experienced experts and comments from a technical expert panel.

Limitations in Using Federal Survey Data for Research on Small Populations

There are a number of strengths to using federal survey data for research, such as the ability to generalize findings at a national level or across large populations. However, a number of limitations exist, such as the cross-sectional nature of the data, weaknesses with self-reported data, and selection bias. In general, problems stem from the size of these segments relative to the total population due to the small likelihood that an adequate number will be included in the sample to study. These segments may also be less likely to particulate in federal survey research or difficult to identify when they do.

To illustrate the challenges facing research on small populations, this report focuses on four case examples:

Asian-American subpopulations. Challenges exist in obtaining adequate sample sizes to conduct analysis on Asian Americans overall, and even more for subpopulations. However, instances where

subpopulation analysis has been possible reveal major differences in health. There is also a lack of consistent race/ethnicity categories used in data collection.

Lesbian, gay, bisexual, and transgender population. Many of the health issues and research challenges facing this population are related to stigma, which has caused hesitation in collecting data on LGBT status and has prevented this population from identifying themselves. In addition, there is a lack of standard definitions by which to identify this population through surveys, as questions regarding behavior, attraction, and identity all result in different responses and each have important implications for health.

Adolescents with autism spectrum disorders. While much research has concentrated on diagnosis of these disorders during childhood, little is known about health and health care during the transition to adulthood for individuals with ASDs, a time period that is critical to their future well-being. The cross-sectional nature of most surveys and inconsistency in how disability is measured among children and in adults makes it impossible to follow this population over time in most existing survey data.

Rural populations. Geographic isolation and low population density has limited both economic opportunities and access to health care services for rural populations, who face the health care needs of an aging population as well as unique environmental health issues. Variations in how to define the boundaries of rural areas (which may not always align with county -boundaries—the smallest geographic unit used in most surveys) also complicate studying this population.

Potential Uses of Existing Electronic Health Data

Electronic health records and other types of electronic health information have the potential to revolutionize the health and health care research enterprise. In addition to creating a source of rich information about large numbers of people (so-called "big data"), the electronic medium offers faster and cheaper means of accessing, extracting, linking, and using health data for a variety of purposes, such as quality and efficiency improvement and research. For example, EHRs and other information technology can be used to identify target patient subpopulations and provide information for research databases.

EHR-based data may be useful for research on small populations that may differ from the majority in ways that affect their health and that have been difficult to study with traditional methods and data sources such as federal surveys and claims data. General surveys often include too few people from particular demographic or clinical subpopulations for production of valid and reliable results, and they face limits in the amount and type of information they can collect. Claims data may not provide needed clinical detail and may be distorted by the purpose for which it was created (i.e., to obtain payment).

The second part of the report explores the potential use of EHRs and other electronic data sources to improve research on small populations that have been difficult to study. While "research" can take many forms, we define the term broadly in this report, as our primary purpose is to consider how EHR data can potentially be used to study the health and health care needs of small populations as illustrated by the four subgroups, including making comparisons to the larger population or other subgroups as needed. As described in Part I of this report, the priority research questions of interest about small n populations are varied, including topics traditionally addressed through clinical, pharmaceutical, health services, public health, public policy, and evaluation research. EHR data, alone or in combination with other forms of data, may be better suited for some purposes than others. Additionally, increasing interest in quality

improvement provides opportunities to harness EHR data for research on small n populations but may also present some challenges. We discuss the issue of the "fit" between the purpose and nature of the research on small n populations and the potential use of EHR data further throughout this report.[1]

We continue to use our four example small populations to illustrate both the potential and the challenges in using EHR and other electronic health data for research in Part I of this report. This part is organized around the conditions needed to conduct EHR-based research on small populations, describing both barriers and facilitators.

The Growing Availability of Electronic Health Data

The Institute of Medicine sees EHRs as an essential part of a "learning health care system," and many believe they are critical for the success of medical homes, accountable care organizations, and other provider payment and delivery system reforms resulting from the Affordable Care Act. The use of EHR data for research depends first of all on the adoption and use of EHRs by health care providers. Over the past decade or so, early adopters of EHRs have begun to tap their potential for clinical, epidemiological, and health services research. These early adopters have included HMOs, large multispecialty medical groups, and large hospital-owned and operated systems that employ physicians and operate other facilities along the care continuum. Some have now started or participate in EHR-based research networks, often with federal support. Federal stimulus funds under the Health Information Technology for Economic and Clinical Health Act has resulted in growing number of providers that use EHRs, and this increases the size and variety of the populations that can be studied. For example, more federally qualified health centers, small physician practices, and critical access and safety net hospitals are adopting and using EHR technology resulting in more information about traditionally vulnerable patient populations.

The current level and rate of increase in EHR adoption and use by providers suggests that the health care industry may be approaching a "tipping point," that is the moment of critical mass where ideas, products, and behaviors spread like viruses."[2] The use of EHRs to capture, organize, and use information for purposes of quality and efficiency improvement as well as research is not just the expectation or norm among the "innovators" but increasingly the expectation and norm for entire health care industry.

Information available in EHRs

Information in EHRs comes from both patients and care providers. Information such as demographic and other background information may be collected directly from the patient using a form or questionnaire they fill out at the registration desk, in the waiting room, or through a patient portal. Data entered during the office visit by the clinician may include reason for the visit, height, weight, vital signs, patient-reported symptoms and characteristics (such as behavior and lifestyle), diagnoses, treatments and tests ordered, and medications prescribed. In addition, data from the pharmacy, laboratory, and radiology are often incorporated into the EHR. Claims and billing information may also be integrated with an EHR. There is the potential to identify some small populations using information that is typically recorded in an EHR such as demographics and diagnosis.

Having this information directly entered into the computer can transform the research enterprise, making data available in close to real time, facilitating the identification of patients with characteristics of interest, eliminating the need for data entry, and reducing reliance on patient recall as is required in

survey research. EHRs also include a level of clinical detail on the process of care that is not available in federal survey or claims data. Having such detail about all patients in a health system also allows for identification of small populations, such as those with rare conditions.[3] EHRs also provide information on patients who may not otherwise be included in research because they would not meet the requirements to participate in a clinical trial.[4]

Unlike federal survey data, however, EHR data are not collected or structured for research. Repurposing information collected for other purposes always presents challenges. Even though EHRs do include information that can facilitate research on small populations, a number of technical, legal, and multi-institutional conditions must be in place in order for this research to reach its full potential.

Technical conditions required for research using EHR and other electronic health data

To use EHR data and other electronic health data for research, information it contains must be extracted and formatted for research. The information in an EHR is collected to assist clinicians and health care organizations in their day-to-day work, providing documentation required by law, for billing, and to inform provider decision-making for care of individual patients. For these purposes, there is often no need to ensure that information is entered in a uniform fashion, or to plan for the ability to pull selectively certain information from the system, to be able to aggregate data, or to identify certain groups of patients. The cost of converting this information into databases suitable for research purposes is substantial.

A major limiting step required for using data from EHRs for research is the ability to extract it from the EHR system. While an EHR system is where information is entered, it is not the place where the data can be cleaned, reformatted, and analyzed. Extraction can require a large staff of programmers, and ease of doing so depends on the system and vendor used.[5] Some organizations have created a central warehouse where EHR, billing system, registration system, labs, and radiology systems are extracted, pooled together, and linked. Others have developed software to automate extraction or to query their EHR systems for selected records based on patient characteristics needed for analysis.

The major difficulty for both data extraction and research is that much of the content of EHRs has not been entered in a standard format. Desired information may be in free text that was entered by the clinicians to record their observations and assist with their decision-making. Some estimates say only 20 percent of information in EHRs is coded and put into structured fields, meaning most of the information is in free text. However, there has been great progress in the development of techniques to classify unstructured data. Algorithms and software have been developed for natural language processing (NLP) to take a clinician's free text and create standard categories. However, some experts caution that NLP is at best a partial solution. In many cases, it may be more efficient and may produce more accurate data to ask the patient for the desired information or to use other data sources rather than trying to find it in the free text.[6]

In addition to lack of standardization, there are major concerns regarding the accuracy and completeness of data entered into EHRs. Research requires high quality and complete data for reaching valid conclusions. Compared to paper charts, electronic health records have been found to hold significant errors—in part, because many clinicians have not been accustomed to using a computer as part of their daily workflow during this transitional period from paper to electronic medical records. In addition to

typos and spelling errors, errors of omission and commission have been found in medication lists and in problem lists where chronic and acute conditions are documented.[7] In addition, cultural or financial barriers to access may prevent certain populations from receiving care, reducing the representativeness of EHR data available for research.[8] There is also the issue of patients moving in and out of health care and EHR systems—either because they have stopped receiving care or have gone to another system. Such movement makes it difficult to create cohorts and to make reliable inferences about them.[9] However, increasingly integrated models of health care delivery may present opportunities to study a more complete picture of a patient's care.

Finally, the skills required to conduct research using EHR data are highly technical and specialized. This includes information technology, clinical and research skills needed to prepare the data, conduct analysis, and interpret findings in light of the context in which the data was collected. Individuals with this combination of expertise are currently in short supply.

Legal conditions required for research using EHR and other electronic health data

In addition to requirements for data extraction and analysis, there are legal requirements that complicate the repurposing of EHR data for research. Traditional research regulated by Institutional Review Boards that comply with federal laws can complicate the reuse of data collected for another purpose, and measures taken to protect privacy and data security may need to be reconsidered when using EHR data for research. Such data may have the potential to address additional research questions as the information accumulates over time. There is ongoing debate about complications created by legal requirements governing privacy and human subjects research.

Governance processes specifying who owns, controls, and regulates the data must also be in place in order to use EHR data for research. While HIPAA, the Common Rule, and state laws currently provide the major guidance regarding how health data can be used for research, each organization must determine how it will remain in compliance and how patient data can be used. Data governance requires major resource investments and cooperation within and across organizations.

Organizational conditions required for research combining multiple data sources

Because of the limitations of data from any single organization, there is great interest in combining data from multiple organizations. Data that is in electronic form can facilitate this. However, there are complexities in using EHR data for multi-institutional research. A mechanism is needed for data sharing. There are two major ways that data can be shared across multiple institutions: through a consolidated warehouse where a copy of the data from each institution is stored, or through some form of "distributed" network in which each organization retains its own data but data from each cooperating organization can be queried and produce research results. Centralizing data in a warehouse may increase efficiency when standardizing and querying the EHR data, but it requires resources to build and maintain and presents a number of privacy and governance concerns.[10] The alternative—a virtual data warehouse in which data remain in separate locations—avoids the need for investment to build a separate infrastructure and simplifies the issues of data ownership and may better serve to protect privacy. However, it requires each

participating organization to have the infrastructure to store data. Both methods for sharing data require significant infrastructure development, both technically and organizationally.

Ongoing funding for research infrastructure is needed but most grants and contracts pay for specific, discrete studies. However, in recent years the availability of this funding has increased. For example, this year the Patient-Centered Outcomes Research Institute is investing $68 million to support the initial development of a National Patient-Centered Clinical Research Network to build the capacity needed support comparative effectiveness research.[11]

In addition, for studies that include data from multiple organizations, approval must be obtained from multiple Institutional Review Boards, adding to the time and resources needed to conduct the research. Also, a process is needed to ensure the quality of multisite data for research.[12] Research among multiple institutions is facilitated by the interoperability of their EHR systems, which remains underdeveloped. Without interoperability, a large amount of effort is needed to make data comparable and combinable. Major health systems, some EHR vendors, and federal incentives are promoting standardized data fields and formats across different EHR systems. Research agencies also have the opportunity to promote standardization through their funding decisions. Incentives for meeting "meaningful use" standards will also like have some effect, and in combination with other levers and incentives, the availability of standardized EHR data for research should continue to increase.[13]

As noted above, a number of research networks have also developed to facilitate research using data from multiple institutions (see Table II.2 in Part II). These include practice-based research networks of primary care practices, as well as other networks such as community health centers, HMOs, or cancer care providers who are collaborating to facilitate research. A major benefit of research networks includes the wealth of clinical information available through their EHRs. Often the organizations within a network are already either sharing a common EHR system or have worked to develop some form of centralized or distributed data warehouse for research purposes. Research on small populations is increasingly feasible as networks of EHRs with common structures and formats have developed, including a larger number of patients from multiple health care systems.

Other data sources may be linked with EHR data to provide additional information for research. Commonly linked administrative databases include disease and immunization registries, claims files, survey data, provider files, vital statistics (e.g., birth and death records), and area-level data.[14] Additional clinical information such as genetic, care management, and social network information also have the potential for linkage with EHR data for research. The use of multiple data sources may both serve to validate electronic health data as well as increase the amount of information available on target study populations.

Potential for Future Research on Small Populations

Despite existing challenges to meeting the conditions needed to use EHR and other electronic health data for research, our interviews and literature review illustrate that innovative solutions are being developed through a variety of publicly supported and private efforts. In particular, a number of large delivery systems and research networks have made substantial steps forward in developing the infrastructure and methods needed to conduct this type of research.

Experts in the field have suggested ways to move forward in the field of research using EHR or other electronic health data in general and/or ways to study specific small or minority populations. These suggestions can be categorized as potential studies aimed at data validation, new tools and methods for mining and extracting data, descriptive studies around specific populations, and outcomes research. There were also recommendations to explore the types of research for which EHR data are best suited, as well as ways that it can be used in combination with other data sources for research, including survey data. In addition to potential studies, there have been recommendations for efforts to engage clinicians in order to improve the quality of data available for EHR research. Providing education around the importance of the data may motivate physicians to enter data into structured fields rather than free text. Opportunities also exist to update the current legal framework that regulates use of electronic health data for research to both promote patient ability to make meaningful choices while minimizing the burden on both patients and researchers.

In order for research using EHR and other electronic health data to reach its full potential both in general and with small populations, engagement of key stakeholders must continue. Many of these stakeholders are working to identify critical next steps and promising pilots through an effort led by the Assistant Secretary for Planning and Evaluation (ASPE), including the development of this report with the input of technical experts. Other key stakeholders include government agencies, EHR vendors, health plans, providers, researchers, and consumer/patient groups, which all play an important role in achieving the conditions needed for research using EHR and other electronic health data.

Table ES.1. Major Conditions Required for Research Using EHR and Other Electronic Health Data on Small Populations

Condition	Challenges	Solutions Being Tested
Technical		
Data extraction	Requires IT skills, data storage, vendor cooperation, identification of desired records and variables	Central data warehouse within an organization, software to extract data from distributed data systems
Processing unstructured data	Highly heterogeneous, use of acronyms and appreciations, may include typing and spelling errors	Tools for natural language processing
High-quality, complete data	Errors of omission and commission, data limited to population receiving care from the organization, who may also receive care elsewhere that is not included; generalizability	Careful interpretation of results, linkage to other data sources, use of data from integrated delivery systems and research networks
Privacy and Security		
Protection of patient privacy	Informed consent required for traditional research too burdensome for EHR-based research and may result in biased samples when only consenters included, information needed to identify small populations may be a threat to privacy for individuals	Obtaining general consent from patients for research using EHR data, use of de-identified data, classifying analysis as quality improvement rather than research
Governance	Resource investment and cooperation needed for infrastructure specifying who owns, controls, and regulates the data for research use	HIPAA provides some guidance, some organizations have developed a separate institute or company to conduct research
Combining Multiple Data Sources		
Data sharing	Creating central warehouse for multiple organizations is resource intensive to build, maintain, and govern, privacy and data ownership concerns	Virtual/distributed data warehouses, practice-based research networks, regional health information exchange
EHR interoperability	Large variety of EHR systems and vendors, lack of standards	Federal incentives, voluntary consensus standards, efforts across organizations and vendors to standardize

Table ES.2. Ability of Federal Survey and EHR/Other Electronic Health Data to Address Challenges in Studying Small Populations

Challenge	Survey Data	EHR and Other Electronic Health Data
Sampling Challenges		
Small size of population	Difficult to obtain an adequate sample when sampled randomly	Larger sample (although not random) increases the potential to obtain enough records from a small population
Uneven distribution across the country of some small populations	Difficult to obtain an adequate sample when randomly sampled	Can use data from providers where the targeted subpopulation is concentrated
Information Challenges		
Ability to identify members of small populations	Lack of consistent categories used to classify members makes this challenging. Also, at times categories are not granular enough to identify specific small populations	Same, although natural language processing and use of multiple electronic data sources has shown some promise to help identify certain small populations. Challenges exist training providers and staff to collect needed information
Detail available to understand health and health care needs	Limits to survey length and self-reported information make level of detail low	Large volume of detailed information available, documented by providers, registration staff, and patient
Validity of data	Relatively strong, although there are weaknesses with self-reported information	Varies by type of electronic health data as providers document information for non-research purposes
Research Challenges		
Ability to study small populations over time	Cross sectional nature of most surveys does not allow this	Longitudinal nature of electronic health records well suited to follow populations over time
Need for different types of research	Data collection designed for generalization across the broader population and for hypothesis testing	Better suited to study unique populations than for generalization, as well as for descriptive or hypothesis generating research
Privacy	Access to information needed to identify small populations may risk ability to identify individuals	Secondary use of EHR and other electronic health data for research is challenging in the current legal framework

Part I: The Challenge of Small Populations for Research on Health and Health Care: Examples from Four Under-Studied Populations

Introduction to Part I

A vast body of research shows important differences among segments of the population on virtually all aspects of health and health care, including patterns of disease and disability, use of services, and quality and outcomes of care. Documenting such differences is an essential starting point for a wide array of policies and interventions to improve peoples' health. Biological, cultural, historical, and socioeconomic differences among different segments of the population may create distinctive patterns of health care needs and differences in the use of and responses to medical services. Understanding the patterns and differences is impossible unless researchers can separate and compare data from various segments of the population. That is difficult when those population segments are small or difficult to identify. This is a particular concern when the small population in question has special vulnerabilities or may be subject to inequitable treatment. To date, the federal government's very substantial data collection efforts have not generated adequate data about some subpopulations because of their small size or their distribution (either great concentration or lack of concentration) or because of insufficiently standardized ways of identifying the population in a survey context.

The small size of some populations means they may not be included in numbers sufficient for separate analyses in federal surveys. Also, information identifying some small populations may not be routinely included in the medical records and insurance claims that are another source of data. To illustrate the different research and methodological challenges facing research on small populations, this report focuses on four case examples—Asian-American subpopulations; lesbian, gay, bisexual, and transgender (LGBT) populations; adolescents with autism spectrum disorders (ASDs); and residents of rural areas. This report is about why research is needed about small populations such as those that we have chosen and about the challenges that small populations pose for research; we make no attempt here to report comprehensively on the health and health care needs of the four populations. We also recognize that many other relatively small populations may have special health care needs or pose particular challenges to the health care system. Our cases are illustrative of a more general set of issues.

Advocacy organizations, as well as some researchers and policymakers, have pushed for the collection of more data about various small populations, including the examples we focus on in this report. With the growing use of electronic health data in the provision of medical care, the possibility that such data might be used for research that complements or supplements existing federal data collection activities merits consideration. That is the topic of Part II of this report. For purposes of this report, we define "research" broadly as addressing issues traditionally addressed through clinical, pharmaceutical, health services, public health, public policy, and evaluation research.

Methodology for Identifying and Exploring Small Populations in This Report

In selecting our example small populations, we targeted those that would illustrate a broad range of health and health care questions, as well as challenges encountered in conducting research to answer them, with existing federal data sources and potential with electronic sources generated in medical care.

Small populations that need study share characteristics with what are typically considered underserved populations: "poor; uninsured; have limited English language proficiency and/or lack familiarity with the health care delivery system; or live in locations where providers are not readily available to meet their needs."[15] To focus our study, we consulted with government officials at the Agency for Healthcare Research & Quality (AHRQ), the Center for Disease Control's (CDC's) National Center for Health Statistics (NCHS), and the Health Resources and Services Administration (HRSA) about populations about which information requests have been received that could not be answered from existing federal data sources. We have also reviewed some related National Institutes of Health (NIH) projects, like the Health Care System Research Collaboratory program.

Once the four study populations were selected, we reviewed past federal surveys regarding the extent to which they could be identified in available data sources, and we examined existing literature for information about their characteristics, health and health care issues, as well as reasons why they have been difficult to study in existing federal surveys and with other sources of data.

In addition, we conducted tailored interviews with 16 expert informants whose work has focused on one of our small populations (see Table I.1). Topics in the interview guide were based on issues and concerns raised in available literature and by organizations that serve the populations in question. An initial purposive sample of experts was identified from published sources, advice from the governmental sources mentioned above, and the research team's knowledge of the field, followed by some snowballing based on suggestions by the experts we were interviewing. Each person gave permission to have the interview recorded, and the interviews were summarized thematically. Particular attention was paid to areas of convergence and divergence among interviews, as well as between interviews and the literature.

Limitations in Federal Survey Data

There are a number of strengths to primary survey data compared to other primary data sources (e.g., focus groups, case studies) and secondary data (e.g., administrative and claims data). Survey data allows the researcher more control over who is included (i.e., sample frame and sample), the kinds of information that is collected from them (e.g., data domains, elements or specific questions), and key aspects of data elements (e.g., standardization and quality) compared to administrative, claims, or other secondary data sources. Consequently, it is often easier to generalize to the nation or other large populations and to replicate survey research.

All research approaches and data sources have limitations, and that is true of survey research. Although many important research questions (e.g., about outcomes of treatment or the consequences of being uninsured) require longitudinal data, most surveys are designed to collect cross-sectional data at a point of time. The Medical Expenditure Panel Survey (MEPS) is a two-year panel and a rare example of a

study that attempts to follow cohorts (of households) over time. Such efforts are few and expensive. There are also limitations regarding the kinds of data that can be collected via survey research. For health matters, for example, surveys most often are limited to collecting self-reports about individual's overall health status, so the resulting data do not include the kinds of clinical information (e.g., about diagnoses, service and procedures, laboratory results, drugs, genetic information) needed for some kinds of studies. Selection bias, which results from survey respondents' decisions about whether to participate or not, can lead to misleading data.[16] Self-reported survey data have weaknesses resulting, for example, from limitations in knowledge or from recall bias. Finally, with the exception of highly specialized studies, surveys generally obtain data from too few people to break out separate results for small populations. As a result even valid inferences drawn about the population (or major segments thereof) based on well-designed survey samples may not apply to small populations such as we are considering in this report.

General problems with small populations do not necessarily stem from the absolute size of the population, but rather its size relative to the total population (or sampling frame) from which the survey sample is drawn. Sample sizes calculated to collect information on the general population of Americans often lack ability to accurately detect small populations. This problem only increases when wanting to study specific health conditions within these small populations. There are standard approaches to increasing the chances of including people from small populations, such as using a list of group members to specifically target or screening questions to increase representation of the groups. However, these strategies are not typically used in national surveys.

Standard "solutions" for getting adequate numbers for analysis from small populations include oversampling[17] and combining data from multiple years. But oversampling subgroups may require the researcher to screen out large numbers of people who do not fit the category in order to obtain the sought-after number of those who do. This becomes more costly as the target group's presence in the population being screened becomes smaller and as the number of needed subgroups (e.g., age, gender, or those using different languages) increases. The smaller a group's presence in the population being screened, the more calls are needed to obtain the desired number of respondents. Combining data from multiple years becomes problematic if year-to-year changes are taking place within that population or if survey questions change. A third alternative, sampling from an organization that specializes in service to the population in question, raises questions of representativeness.

In general, the limitations of national surveys for studying small populations can be summarized as issues related to coverage of the target population and issues related to data collection.[18] These issues as they relate to our four example populations are presented in Table I.2 and are discussed in greater detail later in this report.

Frame problems

Surveys typically use a list of landline telephone numbers and/or addresses as the frame from which the sample will be drawn. Certain population segments (e.g. migrant workers) may be underrepresented if their members disproportionately lack a landline phone or stable/documented address. (The increased use of cellular phones has presented general challenges and issues for survey research.)[19] Federal household surveys typically select their samples by first selecting a sample of geographic areas, then households within those areas, and finally individuals within those households. Target populations that are

geographically segregated, such as remote rural communities or neighborhoods where an Asian subpopulation may be concentrated,[20] they may be underrepresented in the sample if their geographic area is not selected.

Data collection problems

Even if members of small populations are included in the sample, challenges remain in collecting information through a survey questionnaire. These challenges include:

Unit Nonresponse
Certain populations may be less likely to participate in a survey even if invited. For instance, functional limitations may prevent individuals with autism from participating, and proxy respondents are typically used. Even greater challenges occur in getting individuals to repeatedly respond to a survey as is needed to study health issues over time, such as through transition into adulthood.[21] In addition, most surveys are conducted in English and perhaps Spanish, making it difficult for some non-English speakers in Asian subpopulations to participate.[22] Some federal surveys, such as National Health and Nutrition Examination Survey, National Health Information Survey, and Medical Expenditure Survey address this issue by having translation options available for Asian subpopulations, or allow family members to answer for respondents.

Item Nonresponse
Some members of small populations may be unwilling to answer certain questions around sensitive topics (e.g., citizenship or immigration status, risky behaviors, cultural norms and mores, where one works and lives) due to privacy and other concerns. There have been efforts to address this challenge; for example, the National Survey of Family Growth has adopted the use of audio computer-assisted self-interviewing technology, which allows for respondents to listen to a set of prerecorded questions through a computer and input their answers to collect sensitive information, such as drug use. In some cases, sensitive information may be needed to identify the subpopulation in the survey data or to answer the pressing health and health care questions about it. In terms of using survey data to study health issues, there may also be health conditions or behaviors that individuals are less willing or able to disclose in a survey. Which survey method is used may make a difference, with some people more willing to make sensitive disclosures online or in written surveys rather than in a telephone survey, particularly if interviewer hesitancy or other non-verbal communication creates discomfort.[23]

Instrumentation
Even when individuals are willing to answer each question on a survey, it is often difficult to design questions that collect the desired information. For instance, the variety of definitions used to understand each of the four small populations discussed in this report make it difficult to design questions that will identify them.[24] Rare characteristics or conditions may not be included as response options, or may be included in a larger category (such as "Asian" or "conditions on the autism spectrum"), making more granular analysis of sub-categories impossible. There is also lack of alignment in how key questions are asked in different national surveys or over time, affecting comparability and ability to combine these data sources. In addition, there are cognitive limitations in people's ability to understand, remember and self-report much of the information needed to study health issues, such as diagnoses[25] and other detailed clinical information, as well as what services were used and when. There are a number of federal efforts to address these limitations in national survey data. As discussed later, Section 4302 of the Affordable Care Act (ACA) required the adoption of data collection standards on race, ethnicity, sex, primary

language, and disability status in national population health surveys sponsored by HHS. Under the auspices of the Department of Health and Human Services Data Council, the data standards are being implemented in the major surveys.

To illustrate the need for research on small populations and the challenges that such populations pose for research, the following section summarizes the health care needs of these populations and discuss the limitations of the sources of data commonly used by researchers. We do so to illustrate the need for research; a comprehensive examination of the health and health care needs of these populations is beyond the scope of this report. It should also be noted that there is great heterogeneity—for example, by age, gender, or place of residence—within the small populations we have selected, as there will be in any population. Small numbers is a problem that confronts many research efforts that would explore variations *within* small populations, as well as in attempts to make comparisons with other, often larger populations.

In a Part II of this report, we consider the potential usefulness of electronic health information collected by health care providers as a source of data about these four groupings. The intent of this part of the report is to describe the challenges of doing research on small subpopulations and consider the extent to which past limitations might be overcome by the growing use of electronic technologies within the health care system, even if the organizations that have successfully implemented such technologies are not typical.

Population #1: Asian-American Subpopulations

"Asians" are one of the five race categories that must be used in the federal government's surveys and administrative forms under rules of the Office of Management and Budget, but the Asian-American population is quite internally diverse. The 15.5 million Asian Americans who compose about 4.4 percent of the American population include more than 50 different Asian ethnicities and 100 languages. Asian Americans are concentrated in urban areas, particularly in California, New York, and Texas. Which Asian-American subpopulations are found in particular areas varies. Urban areas in California like Los Angeles and San Francisco, as well as eastern areas like New York City have larger Chinese populations than any other Asian subpopulation, while urban areas in Texas have higher concentrations of Asian Indians and Vietnamese.[26] Other local concentrations of Asian subpopulations can increasingly be found throughout the country.[27] Between 2000 and 2010, there was a 46 percent increase in the Asian-American population, making them the fastest growing racial group.[28]

It has been well documented that racial and ethnic minorities receive lower quality health care than non-minorities even after accounting for access-related factors,[29] but little of the research on racial/ethnic disparities has focused on Asian Americans. Their health care needs remain poorly understood due to inconsistent definitions used in data collection, lack of disaggregated data about ethnic subgroups, and the uneven geographic distribution of the Asian-American population.[30]

The commonplace view of Asian Americans as self-sufficient, educated, and upwardly mobile fails to recognize the health needs of Asians overall, as well as their diversity in terms of ethnic background, country of origin, length of time in the United States, and other factors that may affect health and health care.[31]

Figure I.1, which comes from the Palo Alto Medical Foundation Research Institute's Pan Asian Cohort Study (National Institutes of Health, National Institute of Diabetes and Digestive Kidney Diseases grant 5R01DK81371), which primarily utilizes electronic health record (EHR) data, shows diabetes prevalence among men in the San Francisco Bay area and provides a vivid example of the differences in health problems among sub-groups of the Asian-American population.[32] The prevalence rate among Filipino men is more than three times that of Japanese men. It is apparent from these and other data, that health needs vary greatly *within* what is often treated in research as a single racial population.[33]

Figure I.1. Pan Asian Cohort Study—Preliminary Findings for Diabetes Prevalence
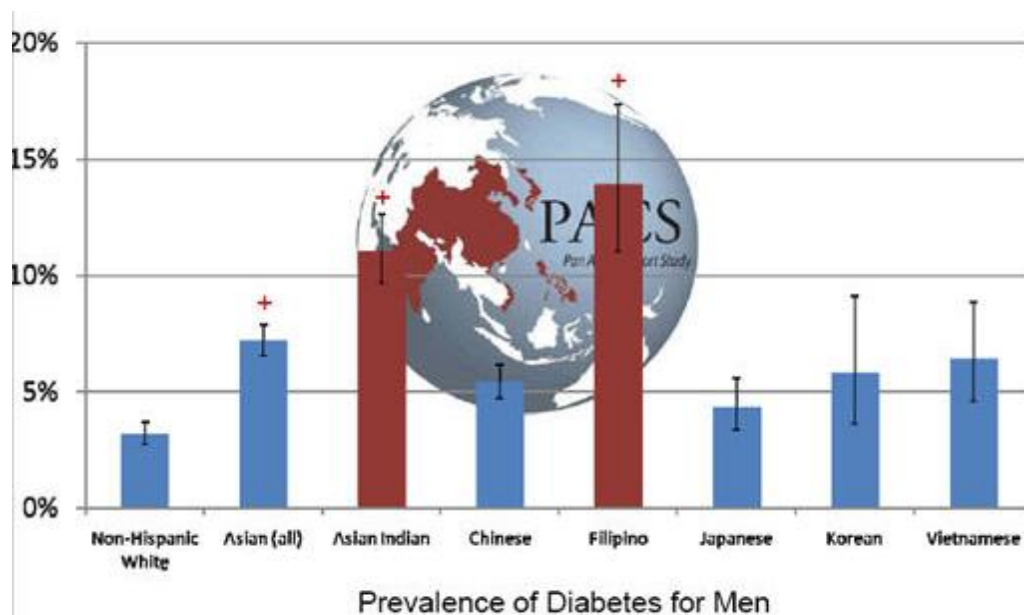
Source: Pan Asian Cohort Study. "Preliminary Findings for Diabetes Prevalence." Palo Alto Medical Foundation. Accessed March 1, 2013. http://www.pamf.org/pacs/men.jpg.

There is also evidence of health care–related differences within the Asian-American population. Asian immigrants to the United States are less likely than U.S.-born Asians to have health insurance and use health care services.[34] Linguistic isolation (living in a household in which no one above age 14 speaks English) may contribute to this. About one-quarter of Asian Americans live in linguistically isolated households, with rates ranging from 10 percent among Filipinos to 45 percent of the Vietnamese.[35] Not surprisingly, linguistically isolated households tend to be of low socioeconomic status and have poorer access to care and more depravation of various kinds than do households in which English is spoken. New immigrants from all countries tend to locate near earlier immigrants. This pattern may facilitate access to various kinds of culturally specific goods and services but may produce isolation from the larger society as well as shared exposure to any environmental risk factors that are proximate to their locale.[36]

The language barriers and cultural differences associated with immigrant status create various complexities, including communications difficulties with health care providers, advice that is inconsistent with cultural beliefs and practices, and dissatisfaction with or distrust of medical advice.[37] Imperfect language translation and nuance can create confusion. Language and cultural isolation of immigrant or

non-English speaking groups may present barriers to care-seeking and treatment.[38] Behavioral health issues—stress, smoking, domestic violence, alcohol abuse—may also be associated with these factors.

There is need for better information about subpopulations of Asian Americans, as can be can be illustrated by considering the examples of Vietnamese and Filipinos in the United States.

Vietnamese Americans

The majority of the 1.7 million ethnic Vietnamese Americans trace their origins to the mass exodus that followed the Vietnam War. Concentrations of Vietnamese Americans can be found in California, Texans, Washington, Florida, and Virginia.[39] Vietnamese Americans have a lower median income than do Asian Americans overall.[40] Moreover, the circumstances under which they entered this country left much of this population with a sense of cultural, economic, political, psychological, and social upheaval that continues to affect their health today.[41]

Information about the health problems of the Vietnamese-American population is limited. There is evidence that Vietnamese women have higher rates of ulcers, stroke and diabetes compared to women in other Asian subpopulations.[42] Vietnamese-American women also have cervical cancer rates that are three times that of Asian-American and Pacific Islander women overall.[43] Notably, low levels of knowledge of the Pap test have been found among Vietnamese-American women[44] who also have low cervical cancer screening rates.[45] Health beliefs and attitudes towards gynecological exams, as well as concerns over cost contribute to low screening rates among Vietnamese Americans.[46,47,48]

The 2007 California Health Interview Survey (CHIS), which oversampled Asian subpopulations and was administered in five languages (including Mandarin, Cantonese, and Vietnamese), provides evidence that language barriers and health illiteracy are particularly important problems in this population. Vietnamese were more likely than Chinese to have limited English proficiency (38.5 percent vs. 27.4 percent), and limited English proficiency was strongly related both to low health literacy and poor self-reported health status.[49] Almost two-thirds (64 percent) of the Vietnamese who had limited English proficiency reported themselves to be in poor health, by far the highest level among the five racial/ethnic groups for which separate data could be broken out in the survey. By comparison, 39 percent of Chinese with limited English proficiency reported "poor" health, while the rate among whites, of whom more than 99 percent were proficient in English, was 13 percent.

Filipino Americans

Filipino are the third-largest Asian subpopulation in the United States (after Americans of Chinese and Indian backgrounds), with 2.6 million people and concentrations in California, Hawaii, Illinois and New York.[50,51] Reflecting a history of Spanish and American rule, Filipinos have a unique blend of Eastern and Western culture, including Hispanic surnames and English and Spanish as official languages. However, more than 120 languages are spoken among ethnic subgroups of the Philippines, and a substantial minority of Filipino-American's speaks Tagalog, which is the 4th most frequently spoken language at home in the United States (2007), although most Tagalog speakers also speak English.[52] Filipinos have migrated to the United States throughout the 20th century and earlier, many for economic opportunities in an English-speaking environment. Thus, the transition for Filipino immigrants may in general have been less severe than for Vietnamese immigrants.

Despite largely successful assimilation in the United States and the highest high school graduation rate of any Asian sub-group, Filipino Americans face a number of health issues. They have higher rates of diabetes[53] and coronary heart disease [54] than whites. Filipino women also have greater risk of stroke.[55] In addition, Filipino women have the highest rates of cancer, epilepsy, and rank highest in drug use and smoking among Asian-American women subpopulations. However, they also have significantly better self-rated mental health.[56] Use of "traditional" medicine is particularly prevalent among first-generation Filipino Americans, particularly those who obtain care during visits to their home country. Examples of traditional medicine include touch/therapy massage, spiritual healing, and use of natural remedies such as herbs, oils and spices.[57]

Coverage of Asian-American subpopulations in federal data collection

The best information about Asian-American subpopulations comes from the U.S. Census, but little information is collected there about health and health care. The Current Population Survey and American Community Survey (ACS) do collect information on health insurance that can be broken down by subpopulation. The ACS also collects information on disability. The Census Bureau has recently released criteria around an option for federal agencies to use the ACS as a sampling frame for follow-on surveys for rare populations, potentially allowing for further data collection from Asian subpopulations or other small populations as identified through the ACS.[58] However, these follow-on surveys are expensive, and, as is further discussed below, there remain challenges in identifying some Asian subpopulations through the census.

Limited health information about Asian-American subpopulations is available in some federal surveys, including the National Health Interview Survey (NHIS), the National Health and Nutrition Examination Survey (NHANES), the MEPS, and the Early Childhood Longitudinal Survey (see Table I.3). However, within a racial group (Asians) that comprises only 4.4 percent of the populations, sample sizes of subpopulations are often too small to permit meaningful data analysis, particularly when co-variates such as age, sex, or region are factored in. Also, a sampling bias arises in surveys that collect data only in English and Spanish, as is the case with most national surveys.[59] For the first time, the most recent NHANES survey oversampled Asians (including Koreans) in larger cities and worked with the Asian community and advocacy groups for outreach.[60] However, a lack of interviewers able to conduct the survey in the appropriate languages and other factors like cultural attitudes and beliefs about participating in surveys may have limited participation from Asian subpopulations, thus lowering the response rate for Asian subpopulations.[61]

Data about Asian-American subpopulation groups are even more limited in other federal surveys. None were collected previously, for example, in the CDC's Behavioral Risk Factor Surveillance System (BRFSS), the National Household Education Survey, the Survey of Income and Program Participation, National Survey of Family Growth, National Immunization Survey, or Medicare Current Beneficiary Survey, although many federal surveys are being updated to include this information going forward. There is also variation by state in what they collect in their National Vital Statistics, which identify Chinese, Japanese, Hawaiian, and Filipino in 50 states, but identifies other Asian subpopulations such as Vietnamese and Korean only in nine states (in which two-thirds of the Vietnamese and Korean subpopulations reside).[62]
Some states may collect data on Vietnamese and Koreans, but the sample sizes are too small to produce valid or reliable estimates, so they do not report figures for them at all.

Some other surveys have collected data about at least some Asian-American subgroups. The federally funded National Latino and Asian American Study collected data in 2002–03 from a nationally representative sample about the mental health needs of two rapidly growing populations. The Asian-American sample was stratified into Chinese, Vietnamese, Filipino, and Other Asians, and data were collected in Chinese, Vietnamese, and Tagalog as well as English and Spanish.[63,64] The California Health Interview Survey (CHIS), modeled after the National Health Interview Survey (NHIS), sought to include hard-to-reach populations and collected data in several Asian languages.[65] Some other state or city-based surveys, such as New York City Community Health Survey, have included information on Asian-American subpopulations.

In addition to survey-based studies, studies are beginning to appear that have used EHR data to study Asian subpopulations.[66,67] This topic is the focus of the second part of this report.

Limitations of available data sources

Recognizing the health needs of and health-related differences among, Asian-American subpopulations, various researchers, policy makers, and advocates of Asian Americans have called for more consistent and standardized collection of data on Asian subpopulations. The challenges faced getting adequate data to study the health and health care of Asian-American subpopulations include language barriers, small numbers, and differences from project to project in how groupings are defined and combined. The first two of these problems interact with each other. Although costly, it is possible to collect data in multiple languages, and some surveys have done so. But the problem of small numbers adds complications. The Asian-American population is itself small, and its subpopulations and language groups are of course even smaller.

Under the Paperwork Reduction Act, the Office of Management and Budget uses race and ethnicity standards in its review of federal agency requests to collect data through surveys and forms. For the most part, surveys conform to the standard categories. Additional granularity is encouraged when feasible, but always must permit aggregation to the appropriate categories prescribed in the standard. Because administrative data are not always reported by individuals themselves, rather collected by providers or other parties, the level of consistency may not match surveys. The aim however is to strive to meet the standard when possible. Determinations about level of granularity are made in the context of an expectation about whether a particular data collection activity is likely to generate a sufficient response.

Standards continue to evolve. In 1997, OMB revised federal data collection standards to separate Asians and Native Hawaiians. More recently the ACA directed HHS to establish standards for the collection of race, ethnicity, sex, primary language, and disability status. An effort led by the HHS Data Council produced a set of guidelines for surveys that expands the standards.[68] As new and existing surveys are presented for review and approval, these standards are now being implemented. A similar effort is under way to recommend guidelines for administrative data.

In addition to efforts spurred by the ACA, other federal, state, and private initiatives could generate improved data. Federal Meaningful Use requirements do specify collection of race and ethnicity categories required in specific geographic areas based on the population make-up.[69] Thus, medical records-based information about Asian subpopulations is likely to be collected only in locales where concentrations of those populations exist.

By the mid-2000s nearly 80 percent of hospitals were collecting race/ethnicity data from their patients, with teaching, urban, and hospitals in states with mandates to collect racial/ethnic data more likely to collect and report the data (such as state requirements that patient demographic information be included in hospital discharge data).[70] There is less information about the collection of such information by other providers, and there has been doubt and confusion about how best to collect it. The Institute of Medicine has advised that such data should be collected from patients themselves, rather than by clerical observation, and most hospitals reported doing so. Most hospitals were using the OMB categories but up to 10 percent were using finer categories based in part on local circumstances. 78 percent of hospitals that collected race/ethnicity data used the category "Asian", 25 percent used "Pacific Islander" and fewer collected more granular Asian categories.[71] A 2009 IOM committee report highlighted several efforts to improve hospital collection of race and ethnicity data, including a Robert Wood Johnson Foundation initiative that required participating hospitals to systematically collect such data and use it to stratify quality measures. The IOM report notes that other hospitals have successfully collected race and ethnicity data for the purpose of linking them to quality measures. In 2007, Massachusetts required all hospitals in the state to collect race and ethnicity data on patients with an inpatient stay, an observation unit stay, or an emergency department visit.[72]

There have been many efforts to improve Medicare race and ethnicity data collection. CMS has supported various efforts, such as annual updates from Social Security data, quarterly updates on American Indians and Alaska Natives from the Indian Health Service, and requesting self-reporting of race through mailings.[73] Researchers have used Census surname lists that allow them to more correctly impute race/ethnicity codes.[74]

The categories used to characterize racial/ethnic groups present additional problems. Groups like the Association of Asian Pacific Community Health Organizations have worked to standardize definitions for collecting data on Asians across organizations to better understand their health service use.[75] The problem of categories has distinctive features among Asian-American subpopulations. The U.S. Census reports data for six Asian-American subcategories as well as "Other Asian" with a write-in box (see Figure I.2), but the use of so many categories may not be practical for many data collection purposes. In addition, Asians from the same subpopulation may describe themselves differently when given the opportunity to fill in the open ended box for "Other Asian." The federal Office of Management and Budget has adopted standard racial/ethnic categories for federal data collection, but they have not been uniformly adopted by the many different entities that collect survey or administrative data.[76] Moreover OMB's five racial and one ethnic (Hispanic/Latino or not) category are considered by some researchers and advocacy organizations to be insufficient for understanding disparities and targeting quality improvement (QI) efforts. In considering the collection of race, ethnicity, and language data, an 2009 Institute of Medicine committee recommended adding questions about (a) English language proficiency, (b) preferred spoken language for health care, and (c) "granular ethnicity," defined as "a person's ethnic origin or descent, 'roots' or heritage, or place of birth of the person or the person's parents or ancestors."[77]

Figure I.2. Reproduction of the Question on Race from the 2010 Census

Source: U.S. Census Bureau, 2010 Census questionnaire.

Changes in the categories used in data collection create difficulties in documenting trends. In 1997, the OMB revised federal data collection standards to make separate categories of (a) Asians and (b) Native Hawaiian and Other Pacific Islanders (NHPI). However, race and ethnicity data collection is not mandatory across government programs and often uses inconsistent categories where it has been implemented. A study in the early 2000s compared Medicare enrollee data with self-reported race and ethnicity in Medicare's Consumer Assessment of Health Plans (CAHPs) survey. The enrollment data matched only 55 percent of the people who self-reported as Asian, in part because many Asians were coded as "other" in the enrollment data.[78] Other studies have also found that Asians are commonly misclassified or classified as "unknown" race.[79] Some researchers have used preferred language selected for Medicare mailings and surname data from the Census Bureau to impute missing data for Asians,[80] although common Hispanic surnames for Filipinos make this problematic, as do some last names (e.g. Lee and Park among Koreans). Birthplace or parent's country of birth has also been used as a proxy for ethnicity, as in the national SEER cancer registry, but nativity and ethnic identification are not always synonymous.

In sum, various cultural, socioeconomic, and historical factors mean that there are variations in many aspects of the health of people from the various Asian subpopulations, but the research on their health needs and the care that they receive has been limited. Survey research has been limited by the small size of the subpopulations and by language barriers, as well as by other general limitations (e.g., self-reported, clinical detailed needed for certain studies). Research from administrative and medical records data has faced practical issues in the collection of recommended data on race/ethnicity and related issues (e.g., country of origin or month in country, language, etc.). The geographic concentration of some subpopulations may facilitate survey data collection at the state or local level and enhance the feasibility

of medical record based research from health plans and providers that serve that population, but only if data collection goes beyond the standard racial/ethnic categories and data are collected as recommended (e.g., self-reported versus what clerks or clinicians assume). Generalization from certain geographic locations is hazardous, since the Asian communities on the West Coast, East Coast, and elsewhere differ in terms of their immigration histories and various social, economic, political, and even health-related characteristics.[81]

Population #2: Lesbian, Gay, Bisexual, and Transgender People

The health and health needs of lesbian, gay, bisexual, and transgender people are not well documented. Even basic information is hard to come by. As a recent Institute of Medicine report puts it, "it has been an ongoing challenge for researchers to collect reliable data from sufficiently large samples to assess the demographic characteristics of LGBT populations."[82] This project mainly focuses on the health and health needs of lesbian, gay, and bisexual people. The transgender population has a host of separate issues around classification, health problems, and provider relations that are not well researched.[83]

To start with the basics, federal and non-federal survey-based estimates of numbers of lesbian, gay, bisexual, and transgender people have varied by gender, over time, and according to survey methods and question wording (see TableI.4 in the Appendix to Part I). Recent estimates puts the percentage of the adult population who identify as homosexual, gay, lesbian, or bisexual at about 3.5%).[84] No such information is available about transgender people. The percentage of adults who identify themselves as lesbian, gay, or bisexual to survey researchers is smaller than the percentage who report having same sex partners or who report some desire for or attraction to a person of the same sex. The small size of LGBT populations and the sensitivity of results to the wording of questions are among the challenges to studying health issues in these populations via survey research. However, there are many indications that such research is needed.

Health needs of the LGBT population—what's known

In its 2011 report *on The Health of Lesbian, Gay, Bisexual, and Transgender People*, the Institute of Medicine (IOM) summarized available evidence about health and health care issues faced by these populations in childhood/adolescence, early/middle adulthood, and later adulthood.[85] The experience of stigma, discrimination, and violence is reported across the life course, as are elevated rates of HIV/AIDS among men, particularly young black men, who have sex with men. Among LGBT youth (as compared to heterosexual youth), there are higher risks for or rates of (a) suicide ideation and attempts; (b) depression, (c) smoking, alcohol consumption, and substance use; (d) homelessness; and (e) victimization through violence and harassment.

Elevated rates of suicidal ideation and attempts and depression have also been reported among LGBT people in early/middle adulthood, along with more mood and anxiety disorders, higher rates of smoking, alcohol and substance use, and experience of stigma, discrimination, and violence. Lesbians and bisexual women appear to use fewer preventive health services than heterosexual women and to have higher rates of obesity and breast cancer. Gay men and lesbians are also less likely than their heterosexual peers to be parents.

Evidence is more limited about later adulthood, but the greater experience of stigma, discrimination, and violence continues, although a degree of "crisis competence" and resilience may also develop. Lesbian and gay people in later life are also less likely than heterosexuals to have, and to receive care from, adult children. The IOM found some evidence of negative health outcomes among transgender people as a result of long-term hormone use. There is also evidence that individuals from same-sex couples have worse health care experiences in terms of access and satisfaction than do different-sex married couples.[86]

Experts concerned about the health of the LGBT population are frustrated by the thin body of available research and data.[87] The IOM report emphasizes the limitations of available research about the health and health care of LGBT people, noting that most evidence pertains to lesbians and gays; that evidence about racial and ethnic minorities is particularly limited, and that most research is not based on probability samples, raising questions about generalizability. To improve understanding of LGBT health, the report pointed to the need for (a) more demographic data on these populations (and minority subpopulations) across the life course, (b) research on the influence of social influences (e.g., families, schools, workplaces, community organizations) on the lives and mental health of LGBT people, and (c) research on barriers to care that disproportionately affect LGBT people, and research on the effectiveness of interventions designed to address health inequities and negative health outcomes experienced by LGBT people.[88] The IOM also called for development of standardized measures of sexual orientation and gender identity, for data on the LGBT population to be collected in federally-funded surveys, and for information on sexual orientation and gender identity to be collected in electronic health records.[89]

Factors affecting the health care of and research on the LGBT population

Stigma—the "inferior status, negative regard, and relative powerless that society collectively assigns to individuals and groups that are associated with various conditions, statuses, and attributes" —was identified by the IOM as a major factor that affects access to or use of medical care by LGBT people.[90] Unfortunately, stigma and its effects also complicate research into the health and health care needs of this population.

Stigma may take the form of negative behavior—epithets, shunning, discrimination, and violence— toward the stigmatized group. Health providers themselves may hold negative beliefs and attitudes that create discomfort for LGBT people in health care situations.[91] Notably, some of the conditions for which this population may be at higher risk—for example, psychological and substance abuse problems—also involve an element of stigmatization.

Stigma and its effects are central to many questions regarding the health and health care of the LGBT population, but they also make good research into those very questions more difficult. Experience with and anticipation of stigma-related attitudes and behavior may affect the willingness of people from the LGBT population to self-identify in survey research.[92]

The effects of stigma may also make people reluctant to seek needed care and to withhold important information when they do so. The content of medical records may also be affected by health professionals' lack of knowledge about some health care needs of LGBT patients, or other subgroups that are unfamiliar to the provider.[93] As the IOM report noted, health professionals do not necessarily know

what questions to ask about a patient's sexual history or be comfortable in doing so. Provider biases or lack of education may affect the questions they ask or the information they document.

Data about LGBT status may be affected by the fact that sexual behavior and gender identity can change over time,[94] adding a dimension to longitudinal research. Repeating questions about sexual behavior and gender identity in longitudinal studies, as well as date stamping information in electronic health records, are examples of potential approaches to address this specific issue.[95]

Care-seeking behavior is also affected by not having health insurance, and the IOM report cites several studies showing that LGBT people and their children are more likely than heterosexuals to lack health insurance.[96] A majority of large employers now provide health benefits to the same-sex partners of employers, but this has been much less common among small employers. The implementation of the Affordable Care Act should substantially address the high uninsured rates among LGBT people.

Barriers to research on the LGBT population

Collecting valid and reliable survey data about LGBT populations has been complicated by several problems. There has been a historical reluctance to seek information about sexual orientation and gender identity in national health-related surveys. Being part of a same-sex couple has been used as a crude fallback measure in one study that used data from the Medical Expenditure Panel Survey to study LGBT peoples' health care experiences.[97] However, the reluctance to collect relevant information in national surveys appears to be changing. For additional federal surveys that may be used to identify members of the LGBT population, see Table I.3. Asking questions about sexual orientation, gender identity, and behavior is crucial to identifying this population.[98] After several years of research and literature review, the National Center for Health Statistics has adopted a basic question and two follow-up questions regarding "sexual identity" ("Do you think of yourself as….") for use in the 2013 Health Interview Survey.[99] In the clinical context, however, sexual behavior may be more important than orientation or identity. Specific risk factors are associated with same sex sexual behavior no matter whether individuals self-identify as gay, while people who can be identified as gay may be stigmatized no matter what their sexual behavior may be. These multiple dimensions may raise the need for multiple questions, depending upon the purpose of the research.[100]

The smaller surveys that have included questions about sexual orientation, same-sex sexual behavior, and gender identity have varied in their focus and measures used. The choice of language in survey questions matters—affecting, for example, the extent to which respondents will identify themselves as lesbian or gay.[101]

The reluctance of some LGBT people—particularly but not only adolescents—to identify themselves as such to researchers has also made survey research more difficult, and it could complicate the collection and research use of relevant information in electronic health records, as we will discuss in the Part II of this report. It is possible that such reluctance will decrease over time as societal acceptance of LGBT people increases. Challenges in identifying LGBT populations in both surveys and EHRs likely differ by age, gender and sexual orientation—for example, gender identity may be better measured as a scale rather than a categorical question for women—who tend to have greater fluidity in their gender identity.[102] Bisexuals may be the least likely to identify themselves as they are less likely to be "out" in their workplace or to health care providers than other people in this population. Sexual *behavior* is particularly

relevant to health concerns among men, but collecting information about sexual orientation and gender identity is also important for research on the impact on health and health care of discrimination, stigma, and stress.[103]

A third problem that is particularly important in survey research (though it could also arise in EHR-based research) is the difficulty of obtaining high quality samples of small populations. As previously discussed, small numbers and the need to break the resulting sample into smaller units by sex, age, or race/ethnicity (and perhaps other factors) create the need for oversampling or combining years of data. Getting sufficient numbers of small categories in survey research of the general population is both inefficient and expensive. Because of the lack of good alternatives, researchers may draw samples from people who have had contact with organizations whose missions focus on the LGBT population. The representativeness of such samples is not known.

The records of service providers are also a potential source of data related to health and health care, but to date information about patients' gender identity has not been a routine, structured field in medical records. Vanderbilt University Medical Center found that the time between when patients were first seen and when their LGBT status appeared in medical records averaged 30 months. This may be due to fluidity of sexual orientation, because patient were not comfortable disclosing the information, or because the provider didn't ask about or document the information. Little is known about the extent to which questions about sexual behavior are asked in clinical encounters or recorded in medical records, although some methodological research that attempts to identify sexual orientation, gender identity, and sexual behavior using the narrative notes or unstructured electronic health record data with natural language processing (NLP) software is under way.[104] Training medical and administrative staff about asking and recording information related to sexual orientation is a substantial task.[105] Medical records will become much more useful for research as the health care industry moves from paper to electronic form and develops other strategies and tools to structure data or mine unstructured data. We explore this potential in Part II of this report.

Population #3: Adolescents with Autism Spectrum Disorders

Autism spectrum disorders (ASDs) are a group of developmental disabilities that range from mild to severe and are characterized by social impairment, difficulty communicating, and repetitive motions or other unusual behaviors.[106] These characteristics are usually noticeable before the age of 3 and remain as a lifelong chronic condition with both medical and psychological implications.[107] ASDs include autistic disorder, Asperger's disorder, pervasive developmental disorder–not otherwise specified (PDD-NOS), Rett syndrome, and childhood disintegrative disorder.[108] Based on 2008 data from the 14 sites in its Autism and Developmental Disabilities Monitoring Network, the Centers for Disease Control estimates 1 in 88 8-year-old children have ASDs.[109] Prevalence in these sites had increased 23 percent from two years earlier and 78 percent since 2002. Although there is disagreement about whether the true prevalence has increased (since guidelines for diagnosis have changed, more services are available, and awareness of ASD has increased), the CDC numbers are based on evaluation records, not parental reports. Measuring ASD prevalence continues to be a challenge due to the complexity of the disorder, the lack of consistent and reliable diagnostic standards, and changes in the definition of such conditions.[110] ASD prevalence is about five times higher in boys than in girls (ratio of 4.5 boys to 1 girl). Prevalence is also significantly

higher among non-Hispanic white children than among black and Hispanic children. Intellectual ability is highly variable, with 38 percent reported as intellectually disabled, 24 percent as borderline, and 38 percent with average or above average intellectual ability.

There are controversies about what should be included in the category of autism spectrum disorders. The NIH classifies Rett syndrome as an ASD, but some argue that it is more similar to non-autistic spectrum disorders such as fragile X syndrome or Down syndrome. Unlike other ASDs, Rett syndrome is also almost always in girls.[111] There is also debate over whether Asperger's disorder is a separate disorder or simply a less severe form of autism.[112] The next revision of the American Psychiatric Association's *Diagnostic and Statistical Manual (DSM)* will drop individual classifications for autistic disorder, Asperger's disorder, childhood disintegrative disorder and PDD-NOS, grouping all of them under "autism spectrum disorder"—a term that is already widely used. APA has said this change will help "more accurately and consistently diagnose children with autism." Rett syndrome will be dropped from the DSM altogether. There is concern among the Asperger's and Rett communities that these changes will result in a loss of identity among individuals with these specific disorders and that it may affect health insurance coverage and school funding for special education.[113]

The exact causes of ASDs remain unknown, but research suggests genetics and environment both play important roles. Researchers are studying factors such as family medical conditions, parental age and other demographic factors, exposure to toxins, and complications during birth or pregnancy. CDC and IOM studies have found no link to childhood immunizations.[114,115,116,117]

Health and health care issues

Among children with various developmental disabilities, autism has been found associated with the highest levels of health and functional impairment indicators. Over 95 percent of children with autism also have co-occurring conditions such as attention deficit disorder, attention deficit-hyper activity disorder, learning disability, mental retardation, stuttering, and other developmental delays.[118] Children with autism are also at elevated risk for depression, anxiety, and behavioral problems,[119] often as a result of difficulty being understood or bullying.[120]

Children with ASDs are also more likely than other children to be obese and to have a variety of conditions—respiratory disorders, food and skin allergies,[121] epilepsy, schizophrenia, bowel disorders, cranial anomalies, type 1 diabetes, muscular dystrophy, and sleep disorders.[122] As a result, children with ASDs use more health care services, therapy, counseling, and medication than children without ASDs.[123,124] Prevalence of prescription medications for children with ASD is high: surveys indicate one-half to two-thirds are prescribed at least one medication of any type, and about 45 percent prescribed at least one psychotropic medication. The most commonly prescribed psychotropic medications are antidepressants, stimulants, and antipsychotics.[125]

The significant amount of care needed for many children with ASDs means many of their parents have needed to reduce or stop work to provide care, spending an average of 10 hours per week providing or coordinating care. As a result, families of children with ASDs are more likely to report financial problems and to need additional income to support their child's medical care compared to families with children with other special health care needs that do not involve emotional, developmental, or behavioral problems. Among children with special health care needs, children with ASD were much more likely to have unmet health care needs for specific health care services and family support services. Having a

medical home has been found to help reduce the financial burden on families of children with ASDs.[126] However, children with ASDs are less likely than children without ASDs to receive care within a medical home.[127]

Transition to adulthood

Most research on ASDs focuses on the identification, assessment, and treatment of children. Few studies examine their transition into the adult world.[128] The health care transition between adolescence and adulthood requires planning in order to maximize lifelong functioning and well-being. This process would ideally include ensuring uninterrupted, developmentally appropriate health care services as the person moves from adolescence to adulthood.[129] For those with ASDs, there are a number of special considerations for this transitional period. The transition period from pediatric to adult care and from child to adult special services will have lifelong implications for their education, employment, social activities, and health.[130] Because their conditions range in severity, a wide range of individualized adult services and supports is needed for this population.[131]

Two key aspects of transition planning for teens with ASDs are helping them take increased responsibility for their health care, and plan for the transfer of care from a pediatric to an adult provider. Unfortunately, providers who care for adults often lack training and experience in dealing with this transitioning population.[132] For those whose disability is impaired enough to interfere with the ability to make financial or medical decisions, parents can file for a petition to maintain guardianship.[133] Most individuals diagnosed with autism during childhood remain dependent into adulthood on their parents or caregivers for support in education, accommodation, and occupational situations.[134]

Teens with ASDs who are transitioning to adulthood need help in understanding their disability, opportunities to talk about topics such as safety, substance abuse and sexuality, education about how to take medications and make routine health care appointments, and continual insurance coverage. An adult provider also needs to be identified, and the adolescent's medical records transferred.[135] None of this is simple.

Unfortunately, health care transition planning is not common for youth with ASDs.[136] One national survey found only 14 percent had a discussion with their pediatrician about transitioning to an adult provider, and fewer than 25 percent had discussed retaining health insurance.[137] Being from a racial or ethnic minority, having low income, being from a non-English speaking family, and not having a medical home reduces the odds that youth with ASDs will receive comprehensive transition services.[138] Even within medical homes, both parents and pediatricians have reported dissatisfaction with the time and resources dedicated to this transition.[139]

Extent of coverage in current federal data collection activities

Existing data collection efforts have several foci and purposes, and their strategies reflect the challenges just mentioned. The Centers for Disease Control and Prevention collect data about the prevalence of ASDs through 14 sites in the Autism Developmental Disabilities Monitoring Network, which identify 8-year-old children with ASDs and other developmental disabilities through record review every other year. Most national health-related surveys do not have a longitudinal design, making it impossible to follow cohorts of youth with ASDs as they transition to adulthood. However, the Department of Education conducts several longitudinal studies, including as the National Longitudinal Transition Study (NLTS),

the Early Childhood Longitudinal Survey (ECLS), and the National Household Education Surveys (NHES). These surveys focus on services that children and youth receive in school and the effects of childhood disability on adult outcomes.[140] The National Longitudinal Study follows a national sample of students who were 13 to 16 years old in 2000. The Early Childhood Longitudinal Study includes 3 cohorts of children who were either followed from birth through kindergarten or from kindergarten through grade school. The survey asks ASD-related questions of the early childhood (9 months-kindergarten) cohort. The National Household Education Survey collects information from adults on learning at all ages among members of their household, from early childhood through school-age and adulthood, capturing a sample of adequate size for national and regional estimates.

The major sources of health-related data have been the National Longitudinal Survey of Adolescent Health (a longitudinal, cohort study that began in the 1994–1995 school year to follow a nationally representative sample of adolescents in grades 7–12 in the United States) and the National Survey of Children's Health (NSCH, collected from a random sample of households) which focuses on physical limitations, symptoms, and diagnoses. The NSCH also collects information on medications prescribed, services used, and more general questions on health and health care.[141] The State and Local Integrated Telephone Survey (SLAITS) is also an important source of information on children with special health needs, such as those with ASDs. It is collected to supplement national data by providing more detailed information from states (see Table I.3).

There is a survey question that asks parents if their child has ever been diagnosed with autism, Asperger's disorder, pervasive developmental disorder, or other autism spectrum disorder, but only one yes or no response for the list of conditions is recorded. The validity of parental reports of such diagnostic information is also open to question and affects the official counts of ASD prevalence.[142]

Survey data collection about children and adolescents on the autism spectrum faces a distinctive set of challenges. First is the problem of small numbers, with ASDs, according to CDC estimates, occurring in only about one of every hundred households that might be contacted for a survey. Second, the condition can be difficult to diagnose (the state-of-the-art diagnostic regime takes several hours to administer) and diagnostic criteria are evolving. There are thus concerns about the validity and reliability of reported cases. The issue of missed diagnoses is a large problem, but may be decreasing as available services, support, and ASD awareness grow.[143] Third, as is the case with many child health issues, data cannot be collected directly from the affected individuals because of their age or the nature of their disorders. Data must be collected from a proxy, generally a parent, school, or service provider. Clinicians typically identify children with ASDs when they fail to meet specific developmental progress milestones, or when certain behavioral characteristics are observed. Federal surveys identify children with ASDs by asking the parent, "Has a doctor or health professional ever told you that your child has autism?" The alternative sources of data may not apply the same diagnostic criteria. There may be tendencies to over-report because of eligibility for sought-after special programs and resources. Conversely, there may be under-reporting in demographic groups not aware of these services, as well as in the past when awareness around ASDs was less prevalent.[144] Shifting definitions and lack of biologic markers for ASDs has made identification difficult,[145] and getting a consistent definition applied across respondents is no small challenge.

There are additional challenges in identifying ASDs in racial and ethnic minorities because some observed characteristics may be attributed to cultural norms and communication barriers related to immigration status or ethnicity.[146] These challenges may result in under diagnosis of ASDs in minority groups.[147] One study found pediatricians were more likely to suspect autism when judging clinical vignettes of ethnic majority cases than among minorities.[148]

Obtaining data on youth with ASDs (or other disabilities) as they transition to adulthood faces additional complexities. The activities of daily living that may be used as measures of disability differ for youth and adults, and youth and adults are eligible for different programs and program participation may be used as an indicator of disability.[149]

Population #4: Residents of Rural Communities

Depending on the definition used—particularly degree of proximity to urban areas—the proportion of the U.S. population described as rural ranges from 17 to 49 percent.[150,151,152] Rural communities are far from uniform, but they are generally less densely populated and more geographically isolated than urban areas. These characteristics result in limited access to services and economic opportunities.[153] Compared to the rest of the population, people in rural areas are more likely to live in poverty as a result of low wage jobs and less likely to be highly educated.[154] Many rural areas face declining numbers due to the out-migration of younger residents.

Health issues

Rural communities are generally older populations and have higher rates of chronic conditions.[155] People in rural counties are more likely than their urban counterparts to face food insecurity (i.e., reports of problems regarding quality, variety, or desirability of diet or eating patterns[156]) which is associated with risks of diabetes and obesity.[157] Rural populations are more likely to report fair to poor health status[158] and to have higher rates of mortality, disability, and smoking and lower rates of physical activity.[159]

A culture of independence and self-reliance in many rural areas presents challenges to the implementation of public health programs,[160] as well as to treatment for mental illness and substance abuse.[161] While the prevalence of mental illness does not seem to differ between rural and urban areas (although documentation is poorer in rural areas), suicide rates are higher in rural communities.[162] In addition, rural youth and young adults have higher use of alcohol and methamphetamines than their urban counterparts, with the degree of use increasing with degree of rurality.[163] Although heightened awareness has increased enforcement, the production of methamphetamines has flourished in isolated rural settings due to the availability of abandoned buildings and anhydrous ammonia, a common fertilizer used by farmers as well as a key ingredient in methamphetamine production.[164]

Rural residents in some parts of the country face environmental health risks associated with agriculture, mining, and industrial pollution. Contaminated water is a risk in communities that rely on well-water, which is not subject to the Safe Water Drinking Act and therefore lacks monitoring and regulation.[165] Rural counties with known sources of water pollution and air pollution have higher rates of cancer mortality, and rural coal-mining areas have higher overall mortality rates.[166]

Despite the known environmental health risks, rural health departments are less likely than urban departments to provide environmental surveillance, inspections, regulation, and licensing services or to employ environmental specialists or epidemiologists. Lack of resources has prevented many rural communities from developing the environmental workforce needed to address many of their environmental health risks.[167]

Rural health issues differ based on a community's dependence on farming, as well as based on other characteristics. While farming once characterized most rural counties, by 2000 the portion of rural counties dependent on farming had declined to 20 percent.[168] The shift from family farms to large corporate farms added environmental health risks.[169] The potential for farm injuries, antibiotic-resistant infections from livestock production, exposure to pesticides, diesel, and solvents also accompany agricultural production, and are associated with cancer, respiratory health issues, reproductive outcomes and neurological disorders.[170]

Gaps in health insurance coverage are also an issue, particularly in rural counties that are not adjacent to an urban county, where nearly a quarter of residents were uninsured in 1998 and where employment in small businesses that do not offer health benefits is particularly common.[171] One challenge in examining rural health issues is determining which urban-rural differences are due to distinct rural factors and which are due to the demographics of the people living there, such as employment characteristics and age. One analysis of BRFSS data found that once the analysis controlled for these factors, some urban-rural differences were reduced or even disappeared.[172]

There is also an important racial/ethnic component of rural health. Rural communities along the U.S./Mexico border, where nearly 67 percent of U.S. Hispanic residents live, are affected by social factors related to border crossing. The border population has been growing faster than the overall U.S. population growth, and many border areas lack much of the economic and health care infrastructure needed to support this growth, making access to health care a particular challenge.[173] The growth of racial and ethnic minorities in non-border rural communities due to immigration and migrant or seasonal farm workers has also been accompanied by growing health disparities as these communities have yet to develop the capacity to overcome cultural and language barriers.[174] In addition, nearly half of the U.S. Native American population, in which rates of alcoholism and substance abuse are particularly high, lives in rural areas (compared to 23 percent of the U.S. white population).[175]

Access issues

A fundamental challenge facing small communities is the high cost per capita of providing health services.[176] Investments to make services available in sparsely populated areas produce services for fewer people than do similar investments made in more populated areas.

Because of the economic base needed to support expensive medical services, access to services becomes increasingly difficult as communities become smaller and more isolated. Rural residents may lack access not only to specialty services and tertiary care but also to such basic services as emergency care, primary care, mental health and substance abuse treatment, and dental care. Realities regarding economies of scale mean that some services needed for people in rural areas are only available in urban areas.[177]

Attracting and retaining clinicians in rural areas remains a challenge due to isolation, limited health facilities, and lack of educational opportunities for their families.[178] The primary workforce shortage in rural areas has continued to worsen even as researchers and policy-makers have sought for solutions.[179] Various federal programs target the need for improved access in rural areas, such as additional support and incentives for clinicians, clinics and hospitals in rural areas.[180] New care models expand use of physician assistants and nurse practitioners (a requirement to be federally qualified as a Rural Health Clinic). Telemedicine has been used to increase access to various services, including mental health, emergency care, and pharmacy, in remote areas.[181] Health IT has also been used to provide access to specialty care, facilitate communication between rural primary care teams and specialists, monitor patients remotely,[182] and provide linguistic congruity in care to Hispanic patients.[183] However, adoption of technology such as telemedicine and EHRs has been slowed by lack of broadband Internet connectivity in many rural areas.[184,185] To help these communities progress, the Federal Communications Commission FCC is investing up to $400 million to expand broadband access to rural health care providers.[186]

The disadvantage faced by rural health care providers in terms of resources as well as the digital divide has created concern among advocacy organizations for rural providers' ability to participate in a number of federal opportunities, particularly requirements for EHR adoption to meet meaningful use standards. Providers without sufficient Internet access may receive a hardship exception from meeting these standards, but concerns remain over the widening adoption gap between rural and urban health care providers.[187,188] The National Rural Health Association recommends further timeline extensions and resources to help rural facilities adopt EHR technology.[189]

Various health professionals, rural health advocates, and states have been heavily involved in discussions about these definitions, along addressing with agricultural and environmental concerns. Part of the difficulty is lack of adequate data about where clinicians are practicing, and the physicians assistants and nurse practitioners who play an important role in providing care in many underserved areas are not always identifiable in claims data[190] There is also lack of agreement over how to define and count the various types of primary care providers[191] and whether and how mid-level providers should count relative to physicians.[192] HRSA is working with many of these stakeholders to create a minimum data set that would allow for better workforce tracking and planning.[193] There has also been joint effort by HRSA's Office of Rural Health Policy and the Department of Agriculture's Economic Research Service (ERS) to define frontier geographic areas in order to identify and target policies at the most remote areas.

Data issues regarding rural health

Rural areas are covered in federal surveys by the U.S. Census Bureau, the Appalachian Regional Commission, the Environmental Protection Agency, the Department of Agriculture, the Centers for Disease Control and Prevention, the Agency for Healthcare Research and Quality, the Health Resources and Services Administration, and the Substance Abuse and Mental Health Services Administration.[194] Larger surveys such as the Current Population Survey, American Community Survey, and National Health Interview Survey have census tract or county identifiers that allow for identification of rural populations. However, smaller surveys such as the National Health and Nutrition Examination Survey or the National Survey of Family Growth allow for estimates only at the national level (see Table I.3).

But the relatively small size of the rural population when combined with its diversity creates a distinctive problem. Rural areas differ from each other in many ways, including in their racial/ethnic and

socioeconomic composition and their proximity to urban areas. It may be important to know whether the rural respondents to a survey are from a Texas border county or Litchfield County, Connecticut. But adding a geographic identifier such as county or zip code to a data set raises concerns that this information could be combined with other collected data to make possible the identification of individuals from whom data were collected. The data security practices that agencies have developed to forestall this possibility make rural research much more difficult and expensive. Some federal agencies restrict researcher access to data needed to study some populations.

For example, the public use files from some federal surveys do not include certain information such as zip codes or date or birth that might be used to identify individual data subjects, because of statutory requirements to protect personally identifiable information. The excluded information may be needed to study certain populations. Researchers can gain access to the excluded information only by going to the designated data use center for the agency that collected the data (e.g., the Agency for Healthcare Research and Quality or the National Center for Health Statistics), paying a user fee, and analyzing the data on the agency's own computers. There are also restrictions, designed to protect individual privacy, on what researchers can take with them when they leave the agency's offices. Such restrictions constitute a significant logistical and financial barrier to research on small subpopulations using data from large federal surveys. In an effort to address this issue, NCHS has tried approaches such as providing remote access options for researchers to analyze restricted data.[195]

A second problem for rural research is that minority populations, particularly those facing language barriers, are under-represented in many surveys of rural areas. For example, there has been an influx of Southeast Asian refugees in meatpacking communities in Iowa and some other states, filling jobs once occupied by Mexican and Central American workers who departed after federal immigration law enforcement increased. These towns are struggling to provide language services for these new refugees and often have difficulty identifying the languages that are being spoken.[196] These challenges also exist for data collection, creating gaps in information on the health and health care needs of rural racial and ethnic minorities.

Lack of consensus and consistency on how to define "rural" makes identification of rural populations within federal data difficult, even where geographic identifiers are available. There does seem to be agreement that no one definition can suffice for all instances and that the definition used should align with the goals and needs at hand.[197,198] More than two dozen rural definitions are currently used by federal agencies, each identifying different populations as rural (see Table I.5 for a list of the most commonly used taxonomies).[199] While geographic isolation and population size/density are common elements among these definitions, there is variation in whether administrative (such as municipalities), land-use (such as population size), or economic concepts (such as commuting areas) are used to define the boundaries or rural areas. County-level, economic definitions (such as nonmetropolitan areas) are most commonly used in rural research because of the availability of county-level data.[200]

Discussion/Conclusion

This report has focused on need for health information data about small populations and the challenges that meeting that need has posed for researchers. To explore these challenges we considered populations

defined by four types of characteristics—sexual orientation and behavior, geography, race and ethnicity, and a health-related condition—that were selected to illustrate the range of problems that face researchers when using existing federal surveys (see Table I.6). In a Part II of this report, we examine the potential of data based on electronic health records and related electronic data sources to complement these surveys and overcome some of the problems researchers have historically faced.

In each of our four illustrative populations, we have presented evidence of distinctive health and health care issues that could usefully be better understood by research. Some of these issues pertain to problems and concerns that may characterize the population itself—as with the high rates of diabetes among Filipino Americans, the distance from specialty care that some rural populations face, or the problems posed by the transition to adulthood for adolescents with autism spectrum disorders. Some issues pertain to possible differences and possibility disparities from other populations or the population at large regarding health conditions, services, or outcomes of care.

Research to address questions about small populations depends on several things. The most fundamental is the ability to identify the population of interest in the data. The second is having data on the independent and dependent variables of interest, as well as relevant co-variates (e.g., education, income) that need to be controlled for. Third, the value of many data sources can be enhanced if researchers are able to link to other data sources. Such linkage requires availability of a unique identifier or a matching algorithm that uses multiple variables. Fourth, some research questions require longitudinal data in which data about the same people can be linked over time. Finally, given resource realities and constraints, ways are needed to conduct research as efficiently and effectively as possible. Primary data collection strategies for getting sufficient numbers of people from small populations can be very expensive.

Some national health survey data sets (including the National Survey of Family Growth, National Health and Nutrition Examination Survey, National Health Information Survey, and Behavioral Risk Factor Surveillance System) contain information about the LGBT population or Asian subpopulations. Although such data may be collected, issues exist that make it difficult to use for research on small populations. Information (e.g., zip codes) that is needed to characterize an individual's degree of rural-ness is not available in federal public use data sets because of concerns that deductive identification of individual people might be possible. Additionally, validity concerns can be raised about information reported by a parent in household surveys about a condition such as a child's autism. Survey data may also not include the dependent variables and co-variates needed to answer questions about the health and health care of small populations. Data analysis also requires sufficient numbers, and this can be a problem in survey research and secondary data analysis for people in categories that appear only in small numbers in a large population. This is particularly true when co-variates are considered. The common solutions for this problem all have important drawbacks.

Combining data from surveys conducted in multiple years may yield a sufficiently large analytic sample, but it can produce misleading results if changes are occurring within the population over time. Oversampling a small population in survey research is often feasible, but it can be expensive. Two-stage sampling, starting with a targeted survey, and then a follow-up survey of the target population, can be expensive, and can only be used when the target population is stable and easily identified.[201] Web-based surveys are another potential approach, but these are also limited by self-selection bias (due to high nonresponse rates), representativeness issues, and concerns about the reliability and validity of the data

collected.[202,203] Finally, focusing the study on a region or setting in which there is a concentration of people who fit the category is an oft-used option for obtaining sufficiently large numbers, but the resulting data may not be representative of the larger population.

Available data sources also have other important limitations. Federal survey research is typically cross-sectional, lending itself poorly to research questions that have a longitudinal dimension. Additionally, survey domains, questions, and response categories may change over time, limiting the ability to use the data longitudinally. Data based on insurance claims may permit data analysis that has a longitudinal dimension, but insurance claims do not typically include information that would permit identifying someone as from a LGBT or an Asian-American subpopulation and the data are limited to billed services from particular payers.

In sum, policymakers, advocates, or researchers interested in the health and health needs of small populations encounter various barriers to research using existing federal surveys.

A great deal of hope has been placed in the possibility that electronic information generated in the patient care process in organizations that have electronic health records will provide data that can be used for research on small populations, even though the organizations that collect such information at this time are hardly representative. Electronic health records and associated electronic data (e.g., patient reported health behavior or laboratory or prescription information) have a number potential benefits, such as the possible inclusion of large numbers of individuals from small populations, the collection of rich information about key process of care and outcome variables of interest, the potential for longitudinal study of cohorts of people (e.g. regarding outcomes of care), and the ability to do these relatively inexpensively.

In Part II of this report, we explore these possibilities on how electronic health records and other electronic data can be used to strengthen research on these patient populations.

Appendix to Part I

Table I.1. Key Informant Interviews

Pre-Interviews (to identify target populations)
Agency for Healthcare Research & Quality
- Steve Cohen, PhD, Harvey Schwartz, PhD, Cecilia Casale, PhD, Ed Lomotan, MD, Gurvaneet Randhawa MD, Jim Branscome, Joel Cohen, PhD

National Center for Health Statistics
- Virginia Cain, PhD, Vicki Burt, Don Malec, PhD

Maternal and Child Health Bureau, Health Resources and Services Administration
- Bonnie Strickland, PhD, Michael Kogan, PhD, Mary Kay Kenney, PhD, Marie Mann, MD

Office of Rural Health Policy, Health Resources and Services Administration
- Aaron Fischbach, Curt Mueller, PhD, Michelle Goodman, Tom Morris, Michael McNeely, Sarah Bryce

Target Population Interviews
LGBT
- Judith Bradford, PhD, The Fenway Institute
- Gary Gates, PhD, UCLA School of Law's Williams Institute
- Stewart Landers, JD, John Snow, Inc.
- Harvey Makadon, MD, National LGBT Health Education Center, The Fenway Institute
- Shane Snowdon, Human Rights Campaign

Asian Americans
- Priscilla Huang, JD, Asian & Pacific Islander American Health Forum
- Latha Palaniappan, MD, Palo Alto Medical Foundation
- Marguerite Ro, DrPH, Public Health Dept., Seattle and King County, WA
- Chau Trinh-Shevrin, DrPH, Center for the Study of Asian American Health, Department of Medicine, NYU

Adolescents with Autism Spectrum Disorders
- Debra Lotstein, MD, UCLA School of Medicine
- Margaret (Peggy) McManus, National Alliance to Advance Adolescent Health
- Megumi Okumura, MD, UCSF School of Medicine
- Julie Lounds Taylor, PhD, Vanderbilt University School of Medicine

Individuals Living in Rural Areas
- Amy Brock-Martin, DrPH, South Carolina Rural Health Research Center
- David Hartley, PhD, University of Southern Maine
- Erika Ziller, PhD, University of Southern Maine
- Ira Moscovice, PhD, University of Minnesota
- Keith Mueller, PhD, University of Iowa

Table I.2. Limitations of National Surveys for Small Populations

Population	General Problem: Small n relative to frame	General Problem: Lack of approaches to increase sample	Frame Problem:* Telephone number frame	Frame Problem:* Area frame samples	Data Collection Problem: Unit nonresponse	Data Collection Problem: Item nonresponse	Data Collection Problem: Instrumen-tation
Asian Americans	X	X		X	X		X
LGBT	X	X				X	X
Adolescents on the autism spectrum	X	X			X	X	X
Rural populations	X	X	X	X	X		X

* These frame problems refer to specific challenges to constructing sampling frames based on telephone numbers or geographic areas. See the "Limitations in Survey Data" section for more information on general problems obtaining an adequate frame for small sample size groups relative to the rest of the population.

Table I.3. The Ability of Key National Surveys to Study Four Target Populations

Data Set	Avail-ability	Sample Size	Population #1 Race	Population #1 Ethnicity/Nativity	Population #2 Sexual Orientation/ Behavior	Population #3 Health/ Disability Status	Population #4 Geographic Identifier
Current Population Survey (CPS)	19xx-2011	**2011, 19-64**: 121,520	White, Black, American Indian /Aleut /Eskimo, Asian, Hawaiian /Pacific Islander, and two or more races. Asian can be further classified into subgroups.	Hispanic origin (detailed), birthplace (state or country), mother's birthplace, father's birthplace, year of immigration, citizenship status	N/A	Self-reported health status, work disability, activity/functional limitations	State identifier; metro status; metro area identifier; some counties identified
American Community Survey (ACS)	Years with health insurance question: 2008-2011	**2010, 19-64**: 1,806,189	White, Black, American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Korean, Vietnamese, Japanese, Other Asian or Pacific Islander, Other Race, two major races, three or more major races	Hispanic origin (detailed), birthplace (state or country), parent's birthplaces, ancestry, year of immigration, year naturalized, citizenship status, language spoken at home, English fluency	N/A	Activity/functional limitations, work disability	State, super-PUMA, PUMA, metro status, metro area, Appalachian region, county sample drawn from

Data Set	Avail-ability	Sample Size	Population #1 Race	Population #1 Ethnicity/Nativity	Population #2 Sexual Orientation/ Behavior	Population #3 Health/ Disability Status	Population #4 Geographic Identifier
National Health Interview Survey (NHIS)	1997-2011	**2010, 19-64:** 54,177 full file; 21,396 sample adults	White, Black, American Indian, Alaska Native, Asian (subgroups: Chinese, Japanese, Vietnamese, Filipino, Asian Indian, Korean, other), Native Hawaiian or other Pacific Islander (Guamanian, Samoan, other). Asians were oversampled in the 2006-2009 surveys.	Hispanic ethnicity (detailed), number of years in U.S., citizenship status, global region of birth	**Starting in 2013:** http://www.hhs.gov /news/press/2011pr es/06/20110629a.ht ml	**See NHIS documentation:** Various health status, health condition, activity limitation, and health behavior variables	Region identifiers on public use; access to Census tract/block level and state identifiers at RDC
Medical Expenditure Panel Survey (MEPS)	19xx-2010	**2010, 19-64:** 21,596	Race/ethnicity data collected during the NHIS interview are available (MEPS draws sample from persons interviewed in prior NHIS survey).	Hispanic ethnicity (detailed), born in U.S., number of years in U.S., citizenship status	N/A	**See MEPS documentation:** Self-reported health status, health condition, activity limitation, and health behavior variables	Region only on public use; access to more detailed level at RDC

Data Set	Avail-ability	Sample Size	Population #1 Race	Population #1 Ethnicity/Nativity	Population #2 Sexual Orientation/ Behavior	Population #3 Health/ Disability Status	Population #4 Geographic Identifier
SLAITS-National Survey of Children with Special Health Care Needs	July 2009 - March 2011;	**2009-11, 0-17**: 40,242 detailed CSHCN interviews	White, Black, other, multiple (In some states, Hawaiian/PI, Asian, American/Alaskan Native can be identified)	Hispanic ethnicity, citizenship, child born in U.S. and number of years, parents born in U.S. and number of years	N/A	**See documentation:** health condition/limitation/disability; behavioral, developmental, and emotional health variables; special health care needs	State, MSA status
National Health and Nutrition Examination Survey (NHANES)	1999-2012	**2009-10, 19-64**: 4,861	White, Black, American Indian/Alaska Native, Asian, Native Hawaiian/Pacific Islander, other. Respondents asked to classify themselves as Asian Indian, Chinese, Filipino, Korean, Vietnamese, Japanese, Other Asian or Pacific Islander	Hispanic ethnicity, country of birth, citizenship status, length of time in U.S.	**Yes:** http://www.cdc.gov /NCHS/nhanes/variable_tables/sexual_ behavior.htm **Cognitive testing report:** http://wwwn.cdc.gov/qbank/report/Miller_NCHS_2001NHANESSexualityReport.pdf	**See documentation:** Medical examination data, health status, health conditions, behavioral health, etc…	National

Data Set	Avail-ability	Sample Size	Population #1 Race	Population #1 Ethnicity/Nativity	Population #2 Sexual Orientation/ Behavior	Population #3 Health/ Disability Status	Population #4 Geographic Identifier
National Survey of Family Growth	2006-2010	**2006-2010:** ~10,000 men and 12,000 women, 15-44 years old	White, Black, Hispanic, Asian, Pacific Islander	Hispanic ethnicity (Mexican vs. all other)	**Sexual identity and attraction:** http://www.cdc.gov /nchs/nsfg/abc_list_ s.htm#sexualorienta tionandattraction	Men's and women's health as related to family life, marriage and divorce, pregnancy, infertility, use of contraception.	The geographic scope of the study is national. Detailed geographic identifiers are available on the restricted access contextual data file.
Behavioral Risk Factor Surveillance System (BRFSS)	1995-2011	**2010, 19-64:** 292,502	White, Black, Hispanic, American Indian or Alaska Native, and Asian or Pacific Islander	Hispanic ethnicity	About 19 states have had a question one time or other, but not necessarily every year. In 2014 there is an approved optional module on sexual orientation and gender identity.	Self-reported health status, condition specific measures, diet, physician activity, functional limitations	State (typically), MSA
National Survey on Drug Use and Health (NSDUH)	1994-2011	~60,000	White, Black, Hispanic, American Indian or Alaska Native, Native Hawaiian, other Pacific Islander, Chinese, Filipino, Japanese, Korean, Indian, Vietnamese, other Asian	Hispanic ethnicity	1996: "During the past 12 months, have you had sex with only males, only females, or with both males and females?" Currently testing 2 questions on sexual orientation to be added in 2015[204]	Drug and alcohol use, health care use, health conditions, mental health, health insurance	State (typically), urban/rural

Data Set	Avail-ability	Sample Size	Population #1 Race	Population #1 Ethnicity/Nativity	Population #2 Sexual Orientation/ Behavior	Population #3 Health/ Disability Status	Population #4 Geographic Identifier
National Immunization Survey	1994-2012	2010: 17,004	White, Black/African American, American Indian, Alaska Native, Asian, Native Hawaiian, Pacific Islander, Other	Hispanic, Mexican, Mexican-American, Central American, South American, Puerto Rican, Cuban/Cuban American, Spanish-Caribbean, Other Spanish/Hispanic	N/A	N/A	National, State, and selected large urban areas
SLAITS - Survey of Adult Transition and Health	2001, 2007	1,865	N/A ("derived"?)	Hispanic	N/A	Self-reported health status, disability, special health care needs, activity limitations,	State, region, MSA
SLAITS - National Survey of Children's Health	2003, 2007-2008, 2011-2012	2011-2012: 91800	White/Caucasian, Black/African-American, American Indian/Native American, Alaska Native, Asian, Native Hawaiian, Pacific Islander, Other	Hispanic	N/A	Various disabilities and conditions, including autism, Asperger's disorder, pervasive developmental disorder, or autism spectrum disorder	State, MSA

Data Set	Avail-ability	Sample Size	Population #1 Race	Population #1 Ethnicity/Nativity	Population #2 Sexual Orientation/ Behavior	Population #3 Health/ Disability Status	Population #4 Geographic Identifier
Medicare Current Beneficiary Survey	1991-	16,000 per year	American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Some Other Race. More granular racial/ethnic categories will be added in 2014.	Hispanic	N/A	Self-reported general health, functional limitations	National
National Latino and Asian American Study	2002-2003	2,554 Latinos and 2,095 Asian Americans	Chinese, Vietnamese, Filipino, Other Asians (others subpopulations collected but too small for subgroup analysis)	Puerto Rican, Cuban, Mexican, Other Latinos	N/A	Various psychiatric disorders	National
National Longitudinal Study of Adolescent Health (Add Health)	1994-95, 1996, 2001-02, 2007-08	2008: 15,701			Same-sex relationships, sexual behavior	Self-reported health status and physical exam	

Data Set	Avail-ability	Sample Size	Population #1 Race	Population #1 Ethnicity/Nativity	Population #2 Sexual Orientation/ Behavior	Population #3 Health/ Disability Status	Population #4 Geographic Identifier
National Adult Tobacco Survey	2009-2010	118,581	Non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, non-Hispanic other (including American Indian or Alaska Native, Native Hawaiian or Pacific Islander, multiracial, or some other race)	Hispanic	Heterosexual-straight; lesbian, gay, bisexual, or transgender (LGBT); or not specified. A new version of this survey is in the field that no longer captured transgender after 2010.	General health, cigarette smoking, other tobacco use, cessation, secondhand smoke, chronic diseases	National, State

Table I.4. Estimated Percentage of People by Sexual Orientation and Behavior from Selected Federal and Non-Federal Sample Surveys

This table does not display the most recent estimates, but rather is presented to illustrate how federal and non-federal survey-based estimates of numbers of lesbian, gay, bisexual, and transgender people have varied by gender, over time, and according to survey methods and question wording. For more discussion, see the "Population #2: Lesbian, Gay, Bisexual, and Transgender People" section in Part I.

Survey	Ages	Percent of Men Identifying as Homosexual, Gay, Lesbian, or Bisexual	Percent of Women Identifying as Homosexual, Gay, Lesbian, or Bisexual	Percent of Men Reporting Same-Sex Partners	Percent of Women Reporting Same-Sex Partners	Percent of Men Reporting Some Same-Sex Desire or Attraction	Percent of Women Reporting Some Same-Sex Desire or Attraction
National Survey of Sexual Health and Behavior, 2010	18+	6.8	4.5	—	—	—	—
General Social Survey, 2008	18+	2.9	4.6	—	—	—	—
General Social Survey, 2008	18 - 44	4.1	4.1	10.0	10.0	—	—
National Survey of Family Growth, 2002	18 - 44	4.1	4.1	6.2	11.5	7.1	13.4
National Health and Social Life Survey, 1992	18 - 59	2.8	1.4	7.1	3.8	7.7	7.5

Notes: Estimates are based on small sample sizes, resulting in large confidence intervals around the estimates; see the text for details. Also, differences in estimates can occur because of sampling error (that is, the estimates in the table are based on probability samples) and nonsampling error, errors due to differential nonresponse and coverage, differences in the target population (the cohorts surveyed), differences in the survey questionnaires used, year of implementation, mode of administration, and the survey respondent.
ORIGINAL SOURCE: Institute of Medicine. "The Health of Lesbian, Gay, Bisexual, and Transgender People." March 31, 2011. http://www.iom.edu/Reports/2011/The-Health-of-Lesbian-Gay-Bisexual-and-Transgender-People.aspx
Table Sources: Herbenick et al. (2010), Table 1, for results from the NSSHB; Gates (2010), Figures 1 and 7, for results from the GSS; Mosher et al. (2005), Tables 12 and 13, for results from the NSFG; Laumann et al. (1994a), Table 8.2, for results from the 1992 NHSLS.

Table I.5. Common Rural Taxonomies Used by the Federal Government

Taxonomy	Unit	Urban Definition (rural is what's left)	Limitation
OMB Metropolitan and Nonmetropolitan Taxonomy	Counties	Defines metropolitan areas as counties with 1 or more urbanized area (based on population size) and counties economically tied to that core, measured by commuting to work.	County boundaries may over- or under-bound urban core
USDA Economic Research Service Urban Influence Codes (UIC)	Counties	Builds on OMB metro and nonmetro dichotomy to create continuum based on population size and adjacency/nonadjacency to metro counties	Frequently used for research but not for federal or state policy
Census Bureau Rural and Urban Taxonomy	Census-tract	Urban clusters based on population size	Limited health-related data available at the census tract level, which is not stable over census years
Rural/Urban Commuting Area Taxonomy (RUCA)	Census-tract	Based on work commuting flows	Difficult to link to health data, often collected at the county or zip code level. A zip-code based version has been developed for this purpose, but is complex to use.

Source: Summarized from Hart 2005.[205]

Table I.6. Potential Areas for Further Research

Population	Subpopulation	Health Issue	Challenges in Studying with Existing Federal Survey Data
Asian subpopulation	Vietnamese women	Cervical cancer	Difficulty disaggregating Vietnamese women and self-report of cervical cancer diagnosis
	Filipino	Diabetes	Difficulty disaggregating Filipino and self-report of diabetes diagnosis
Lesbian, Gay, Bisexual, Transgender	Lesbian women	Obesity	Limited data collected on sexual identity and self-reported weight
	LGBT Youth	Mental health	Limited data collected on sexual identity or potential unwillingness to respond to survey questions around mental health
Rural	Minorities	Access to care	Language barriers prevent adequate representation
Autism spectrum disorders	Adolescents in transition to adulthood	Transition to adulthood	Lack of longitudinal data and inconsistent definitions of disability between children and adulthood

Part II: The Potential Use of Electronic Health Records and Other Electronic Health Data to Improve Research on the Health and Health Care of Small Populations

Introduction to Part II

Patients' health records and other electronic health information are an essential part of care, documenting critical issues such as their history, preventive care, diagnostic tests, and diagnoses and treatments over time. Health records also facilitate information sharing among physicians, other health professionals, and provider organizations that may be involved in a patient's care. Containing key information regardless of where and from whom the patient receives care, health records can also be fairly comprehensive as well as longitudinal. Comprehensive integrated health records support the continuity and timeliness of care, which can in turn represent higher quality and less costly care.

Given the rich information contained in health records, much medical and health services research has been based on them, solely or in combination with other types of data (e.g., survey, claims). However, the traditional medium (i.e., paper and pen) in which health records have been created as well as organized and managed (i.e., paper file folders in a filing cabinet) has limited their usefulness for research. The manual process of identifying and obtaining the relevant records from one or more providers, abstracting the information contained in them, and creating a database for analysis is time-consuming, expensive, and fraught with potential errors and problems.[206]

The increased adoption and use of electronic health records (EHRs) and other forms of electronic health information have the potential to revolutionize research, overcoming many historical constraints. The new medium (electronic) in which health records are created, organized, and managed (computer hardware and software) result in "big data" (a lot of detailed data on a large number of people) and potentially faster and cheaper means of using medical records for research. For example, EHRs and other information technology can facilitate the identifying patients with a particular diagnosis or receiving certain services, obtaining their records, extracting information, and creating a database needed for analysis. Additionally, recent developments like EHR certification standards, 'Meaningful Use" (MU) criteria, tools like natural language processing (NLP) software, and electronic health information exchange (HIE) infrastructure (e.g., email, Internet, cloud) and standards (e.g., HL7) have the potential to improve the reliability and validity of EHR data as well as their comprehensiveness and longitudinality. As the Institute of Medicine (IOM) notes, EHRs and other electronic health data provide the information infrastructure to support a "learning health care system" that continuously and relatively quickly turns data into information to guide ongoing improvement efforts and research.[207]

Research on "small n" populations is an important area where EHR and other electronic data have the potential to complement existing data sources and methods, perhaps revolutionizing the research process. By "small n" populations, we mean subpopulations that are much less common than the "average," "typical" or "majority" population and may differ from them in important ways (e.g., disease prevalence, treatment). For a variety of reasons, small n populations have been difficult to study with traditional methods and data sources, such federal surveys and claims data sets.

As described in Part I of this report, there are important limitations to the use of federal surveys for the health and health care needs of small n populations. These surveys may include too few people in important demographic or clinical subpopulations (e.g., race/ethnicity, sexual orientation/gender identity, location, or clinical condition) to produce valid and reliable findings. Additionally, the surveys may not contain items or questions specific to the population of interest or on co-variates needed as controls (e.g., education, income, years in country, primary language). Finally, surveys may have a lot of missing or inaccurate data about sensitive topics that raise privacy concerns (e.g., sexual behavior).

Claims data from public or private health insurers or research agencies (e.g., AHRQ HCUP data) provide sources of data for research on some small n populations. However, these data have a number of limitations as well, primarily because they have been generated to obtain payment. Depending on the payment method, providers may be more or less motivated to submit comprehensive and accurate claims. Additionally, many important clinical details, as well as patient-reported information, do not appear in claims, although efforts are currently under way to try to enhance claims data with EHR and other types of data (e.g., laboratory and pharmacy data, death certificates or other vital records) for research purposes.[208] Finally, claims data from particular health plans and providers may not provide comprehensive or longitudinal information because patients may change health plans and providers or see providers that are not part of the same organized delivery system.

The purpose of this report is to explore the potential use of EHRs and other electronic information to improve research about small populations, alone or in combination with other data sources. While "research" can take many forms , we define the term broadly in this report, as our primary purpose is to consider how EHR data can potentially be used to study the health and health care needs of small populations as illustrated by the four examples or sub- groups, including making comparisons to the larger population or other sub-groups as needed. As described in Part I, the priority research questions of interest about small n populations are highly varied, including topics traditionally addressed through clinical, pharmaceutical, health services, public health, public policy and evaluation research. In some cases, even basic descriptive information about certain small populations remains unavailable due to current limitations with data and research methods. The Institute of Medicine has described different approaches to collecting evidence that may be more or less appropriate to address different types of research questions.[209] In a similar way, EHR data, alone or in combination with other forms of data, may be better suited for some purposes or types of research than others. Additionally, increasing interest in quality improvement provides opportunities to harness EHR data for research on small n populations but may also present some challenges. We discuss the issue of the "fit" between the purpose and nature of the research on small n populations and the potential use of EHR data further throughout this report.

To explore this potential, we focus on four small n populations that have been difficult to study using conventional methods and source of data—the LGBT population, Asian-American subpopulations, adolescents with autism spectrum disorders, and residents of rural areas. Each of these groupings has distinctive health or health care needs that have been difficult to study for reasons that include small numbers, sensitivity or validity of some reported information (problems in both survey data and data based on medical records or claims), and concerns about confidentiality when separate data elements could be combined to identify particular individuals in a data set.

Using EHR-based information for research on small n populations shares many challenges with all research that would use such information, but, as we will discuss, some special issues arise with small n populations. The four on which we focus illustrate a range of challenges in using EHR and other electronic health information for research. For example, information about the race/ethnicity information that is increasingly being collected in structured data fields in EHRs may not necessarily include smaller ethnic categories and categories may different across health systems. Information about sexual orientation, gender identity, and sexual behavior, if collected at all, is frequently located in the clinician's notes or other unstructured data fields because of the potential discomfort and stigma historically associated with LGBT status or certain types of sexual behavior. But, natural language processing (NLP) of that unstructured data could be used to identify lesbian, gay, and bisexual individuals, or patient surveys could be administered through a patient portal or on an iPad in the waiting room and input or streamed into the EHR. A combination of structured (age, diagnoses, medications) and unstructured EHR information could be used to identify adolescents with autism spectrum disorder (ASD) and/or also be combined with claims and/or educational records. Finally, providers located in rural areas could be identified and recruited for research on the health and health care needs of rural residents and other issues, but rural providers are less likely to have an EHR and the ability to exchange health information, and privacy concerns arise because of the possibility that individuals in a sparsely populated areas could be identified if rural zip codes are included in the data.

To explore the potential strengths and limits of using EHR data for research on small n populations, alone or in combination with other data, this report covers four general topics. First, we provide a brief description of the methods and data used for the report and briefly discuss the need for research on small n populations. Second, we describe the increasing adoption and use of EHRs among physicians and hospitals, the kinds of data available in them, and the major issues encountered in using them for research within a single health care organization, such as federally qualified health center, physician group, or large organized delivery system. Third, we describe some additional challenges to conducting research with EHR data from multiple health care organizations and/or in combining EHR and other data sources. Finally, we conclude with a discussion of the implications for HHS, including some potential next steps for exploring and improving the use of EHR and other data for research on these and other small n populations.

Methodology

We conducted semi-structured telephone interviews with 22 expert informants experienced with use of electronic health data for research—in some cases specifically with our four target populations. Initial interviewees were identified through research team knowledge and literature, followed by a snowball sampling technique where interviewees suggested of other relevant experts. Interviewees came from organized delivery systems, universities, private research institutions, and a supplier of health information technology (HIT) (see Table II.1) and were leaders or participants of a number of well-established research networks that use EHRs for research (see the Appendix to Part II). Topics in the interview guide were based on literature as well as on the specific experience represented by each interviewee. They included the advantages and challenges of using EHR data for research, the types of research for which EHR data has the most potential, issues related to sharing data between organizations, and consent, privacy, data security, and confidentiality.

We also conducted a targeted review of literature review that explored technical, legal, and organizational issues related to EHR-based research. Our informants identified additional published and unpublished materials for us to read and review, including websites, materials from major projects using EHRs, and presentations at conferences or other meetings. Using these materials as a starting point, we identified search terms and utilized PubMed and other databases to find other relevant literature. This search resulted in 118 articles in the peer reviewed and gray literatures. See the References in Part II section for a list of citations.

The Need for Research on Small Populations

Research has found differences among segments of the population on nearly all aspects of health and health care. The ability to identify and document such differences is an essential starting point for improving people's health. The four small populations that we selected illustrate a range of unanswered health and health care questions as well as the challenges in conducting research to answer these questions, both with existing federal data sources and potentially with EHR data. While small relative to the U.S. population, these populations have each reached a size where research on their health and health care needs has become both increasingly important and increasingly possible, particularly as new data sources are becoming available. Members of these groups are eager to be recognized and to better understand the particular characteristics and needs of their populations.

These populations were identified based on discussions with government officials at the Assistant Secretary for Planning and Evaluation (ASPE), Agency for Healthcare Research & Quality (AHRQ), and the Center for Disease Control and Prevention's National Center for Health Statistics (NCHS), and the Health Resources and Services Administration (HRSA), who have all received requests for better information about populations that have been difficult to study in existing federal surveys. Here we provide a brief overview of the distinct characteristics, health and health care needs of our four example populations. More detail can be found in Part I of this report.

Asian subpopulations such as Filipinos and Vietnamese

Asian Americans are the fastest growing racial group,[210] making up about 4.4 percent of the American population but including more than 50 different ethnicities and 100 languages.[211] Language and cultural barriers to accessing health care are important concerns generally among immigrant populations, but their health and health care needs are poorly understood due to lack of disaggregated data about ethnic subgroups.[212] But there is evidence that various ethnic subpopulations have distinct patterns of disease and health care use. For example, one study found the prevalence of diabetes was three times higher among Filipino men than among Japanese men.[213] Other research has shown how Vietnamese women to have both higher cervical cancer rates—the highest among Asian-American women—but also low screening rates. [214]

Small numbers relative to the total population, uneven geographic distribution, and language barriers combine to make it difficult to obtain adequate samples of Asian-American subgroups in national surveys. In claims data or health records, subpopulations may remain difficult to identify because ethnicity and language are not routinely or accurately collected. These factors, along with the time and cost of manual data abstraction, have been barriers for records-based research.

Lesbian, gay, bisexual, and transgender people

The health and health care needs of lesbian, gay, bisexual, and transgender (LGBT) people are not well documented, and even basic survey-based estimates of the size of these populations are inconsistent. However, there is evidence that experiencing stigma, discrimination, and violence are common among LGBT populations, and this has significant implications for this population's health and access to care. For example, elevated rates of suicidal attempts, depression, and substance use have been reported among LGBT youth as well as for those in early/middle adulthood compared to their heterosexual counterparts. Elevated rates of HIV/AIDS among men, particularly young black men who have sex with men, has been a concern for many years. There is also evidence that lesbian and bisexual women use fewer preventive services than heterosexual women and have higher rates of obesity and breast cancer. The associated stigma may make LGBT individuals hesitant to seek care, or to withhold information from their provider when they do.[215] Therefore, information needed to identify this population in medical records is seldom there. Some experts believe that LGBT people may be more willing to identify themselves in a written or online survey compared to a face-to-face encounter. At present, however, there is no well-validated way to reliably collect data on LGBT populations, and numbers vary depending on whether information is collected on behavior, identity, or relationships. In addition, small numbers relative to the whole population make it difficult to obtain adequate samples for basic analyses, much less if split by age or gender, although there is evidence the subgroups of LGBT populations have distinct health care needs.

While transgender people have much in common with LGB populations, they also experience a number of distinct challenges with their health and health care. Although we have included them with LGB populations for illustrative purposes, there are additional issues regarding research for transgender populations that we were unable to fully cover in this report.

Adolescents with autism spectrum disorders

Autism spectrum disorders (ASDs) are a group of developmental disabilities characterized by difficulty communicating and repetitive motions or other unusual behaviors, and range from mild to severe.[216] ASDs are lifelong chronic conditions that often require significant medical and psychological care. Over 95 percent of children with autism also have co-occurring conditions such as attention deficit disorder, learning disability, or mental retardation.[217] Children with autism are also more likely to experience depression, anxiety, and behavioral problems,[218] often as a result of difficulty being understood or bullying.[219] As a result, children with ASDs use much more health care services, therapy, counseling, and medication than children without ASDs.[220,221] The prevalence of prescription medications for children with ASD is high—with the most commonly prescribed drugs being psychotropic medications, antidepressants, stimulants, and antipsychotics.[222]

Most research on ASDs focuses on children, but the health care transition between adolescence and adulthood is a particularly vulnerable period for this population as they move from pediatric to adult care and from child to adult special services.[223] However, transition planning for this population is not common. [224] This transition has been difficult to study because most national health-related surveys do not have a longitudinal design, making it impossible to follow youth with ASDs over time. In addition, because the condition is difficult to diagnose and diagnostic criteria have evolved over time, there are concerns about the validity and reliability of case reported in parental surveys. There may be opportunities to use health records alone or in combination with other records (e.g., education, social

service) to study people with ASDs over time, although the lack of biologic markers and shifting definitions of ASDs may continue to pose challenges in identification, even using clinical data.

Residents of rural areas

Rural communities are generally less densely populated and more geographically isolated than urban areas, often limiting economic opportunities. The out-migration of younger residents has left many of these communities with declining and generally older populations. In addition to the higher rates of chronic conditions associated with age, rural populations are more likely than urban residents to report fair to poor health status[225] and to have higher rates of mortality, disability, and smoking and lower rates of physical activity.[226] The rural residents of some parts of the country also face environmental health risks associated with agriculture, mining, and industrial pollution. Access to health care services is a serious concern as many rural communities lack the economic resources needed to support expensive medical services. Difficulty attracting and retaining clinicians further limits access to care. Telemedicine has the potential to help with some access problems, but Internet connectivity and adoption of HIT lag behind in many rural areas.

Research on rural populations has been by small numbers in some research activities and by a lack of consistency in defining rural populations. More than two dozen definitions are used for different purposes by federal agencies, with criteria ranging from population size/density to land-use to commuting distance. In addition, although granular geographic identifiers (such as county and zip code) are needed to examine rural communities, such variables about individuals are not included in public-use data sets because of concerns that those living in sparsely populated areas could be identified.

The Growing Availability of Electronic Health Data

For electronic health records to help solve the challenges of conducting research on small n populations, several conditions need to be present. The first is a critical level of adoption of relatively advanced EHRs by a range of providers (e.g., primary care physicians, specialists, hospitals, laboratory, and pharmacy) so that information about sufficient numbers of "small n" populations will be included. The second is having EHRs that not only support day-to-day patient care work, but that contain information that is sufficiently valid and reliable to support research. The transformation of information in EHR systems into databases that are of research quality requires extensive validation work. Experience in carrying out the needed quality control work is accumulating, as we will discuss below. Also critical is the ability to exchange the data within and across organizations, which requires both interoperability and the infrastructure for exchanging data. There are other conditions that must be met—such as systems to ensure the consent, privacy, and security that facilitates the sharing and use of the data while maintaining consumers' and patients' participation and trust—which we discuss later in the report. Here, we focus on aspects of these first three conditions and how recent legislation and health reform is facilitating more widespread adoption and use of EHRs and information exchange. While all of these conditions may not yet be fully in place among providers that treat small populations, it is important to begin thinking about research capabilities and infrastructure needs as the availability of these data are growing. In this report, we have reviewed the work of those who are on the cutting edge of using EHR data for research as a guide to understanding what may be more widely feasible in the future, and to provide lessons on how current challenges can be overcome in using this type of data for research on small populations.

The Health Information Technology for Economic and Clinical Health Act (HITECH) became law in 2009 as a part of the American Recovery and Reinvestment Act. HITECH made an estimated $27 billion available to enable eligible health professionals and hospitals to adopt, implement, or upgrade EHRs to achieve the "meaningful use" of HIT, as defined by the Office of the National Coordinator (ONC). The intent of meaningful use standards is to improve quality and efficiency of care through widespread implementation and use of EHRs among providers participating in the Medicare or Medicaid EHR payment incentive programs administered by the Center for Medicare and Medicaid Services (CMS). Meaningful use is defined through the regulatory rule-making process in three stages, ultimately resulting in a set of criteria for how EHRs must be used. As of August 2013, 56 percent of registered eligible professionals and 77 percent of registered eligible hospitals had received payment for meeting the meaningful use criteria.[227]

The HITECH legislation also established the Regional Extension Center (REC) and state health information exchange (HIE) programs.[228] A total of 62 RECs provide technical assistance to "high priority" providers (e.g., physicians in small practices) to help them implement EHRs and achieve meaningful use. The HIEs work to facilitate data exchange among care providers within a region through a number of mechanisms.

The CDC's National Ambulatory Medicare Care Survey (NAMCS) provides the best information about the extent of physician adoption of EHRs. Based on an expert consensus, NAMCS defines a "basic" EHR system for physicians as having the electronic capability for managing patient demographic information, patient problem lists, patient medication lists, clinical notes, and orders for prescriptions, and for viewing laboratory and imaging results.[229] In 2012, NAMCS estimates show that 40 percent of office-based physicians used an electronic medical or health record (EMR/EHR) that met the criteria of a basic system, up from 22 percent in 2009 (a 48 percent increase).[230] Earlier multivariate analysis results indicate that primary care physicians are more likely than other physicians to adopt and use EHRs, and that those practicing in large groups, in hospitals or medical centers, and in the Western region of the United States were more likely to adopt and use EHRs relative to their respective counterparts.[231]

Regarding EHR adoption in hospitals, in 2008, the ONC started funding an annual IT survey by the American Hospital Association. In 2012, approximately 44 percent of non-federal acute care hospitals reported having EHRs that meet the criteria of a basic system, defined as having a set of eight clinical functions (patient demographic information, patient problem lists, patient medication lists, discharge summaries, lab and radiologic reports, diagnostic test results, and orders for medications) deployed in at least one hospital unit.[232,233] This was an increase from 16 percent in 2009.[234] Small, public, and rural hospitals were less likely than larger, private, and urban hospitals to have a basic EHR system. Similar—or slightly better—adoption patterns were found on a recent survey of children's hospitals.[235]

Data related to health information exchange among hospitals and physicians is limited. Estimates from the AHA indicate that few hospitals are using EHRs to exchange health information: only 11 percent of hospitals reported in 2010 that they exchange key clinical information with other providers.[236] However, a recent study found that hospitals' exchange of health information with other providers and hospitals outsider their organization has increased by 41 percent since 2008.[237] A recent survey estimates that approximately 15 percent of children's hospitals exchanged health information electronically.[238] Data are not available about the extent of health information exchange among office-based providers.

Despite the significant progress toward adoption of EHRs by physicians and hospitals, a significant number of obstacles have presented themselves. Barriers identified in recent review of some 60 publications included design and technical concerns, ease of use, interoperability, privacy and security, costs, productivity, familiarity and ability with EHR, motivation to use EHR, patient and health professional interaction, and lack of time and workload.[239] Implementation challenges were reported among all types of users (e.g., public, patients, providers, and managers), but particularly among small, public, and rural providers.[240]

In sum, HITECH has provided focus and a major "spark" for the adoption and use of EHRs and the exchange of health care information, and considerable progress has been made. Additional incentives for the adoption and use of EHRS came from provisions of the Affordable Care Act (ACA) and include value-based purchasing, patient centered medical homes (PCMHs), and accountable care organizations (ACOs). Some geographic areas and types of provider or organized delivery systems that serve small n populations have reached a tipping point of having sufficient EHR adoption and exchange capacity to support research on some small population. Below, we discuss in further detail what kinds of information is or is not readily available in current EHRs and the implications for research on small populations.

Information Available in an Electronic Health Record

To be useful for research on small populations, EHRs much include information identifying individuals as fitting into those populations, as well as information about their health and health care. For example, even if members of an Asian subpopulation were identifiable using EHRs, if they rarely seek health care or tend to seek care from places where there is less EHR penetration, or if language is a barrier to communication when they do seek care, limited information may have been recorded on their actual health and health care.

Much relevant information is routinely collected in EHRs in the process of patient care. In 2003, the Institute of Medicine identified eight core functions that EHR systems should be capable of performing in order to promote safety, quality and efficiency in health care. These functions include:[241]

- health information and data

- result management

- order management

- decision support

- electronic communication and connectivity

- patient support

- administrative processes and reporting

- reporting and population health

Additional functions common to EHRs include alerts for clinical preventive services, drug-drug interactions and drug allergies. Organizations have taken several approaches to obtaining a system with the needed functionalities. Purchasing a comprehensive system (often referred to as the "single-vendor

strategy") has been the most common approach among U.S. hospitals,[242] but some piece together elements from different systems (e.g., scheduling, billing, and EHRs) and there is variation in what information is included in EHRs in different organizations.

EHRs typically include a patient's demographic information, personal and family medical history, allergies, immunizations, medications, health conditions, contact and insurance information, as well as a record of what has occurred during visits with the provider.[243] Information may be collected both at sign in at the registration desk and during the visit with the provider.

Patient-reported data

Basic contact, insurance, and demographic information about patients is collected at the registration desk or in the waiting room. Patients may also be asked for pertinent information about their health. Some providers use iPads or computer kiosks that allow patients to enter information directly into their EHR. Some also have patient portals that allow patients to view their information and to communicate with their health care providers. These can be set up to directly interface with the EHR,[244] creating source of information within the EHR. At this stage of EHR use, all patients are not equally likely to use patient portals; minority patients may be less likely to use them and younger patients more likely.[245]

One benefit of collecting some information directly from patients through a written or computerized telephone questionnaire or patient portal is that it gets around the difficulty of getting staff to ask patients for information about such topics as race/ethnicity or sexual orientation.[246] While challenges remain with how to word questions in order to identify LGBT populations, the bigger challenge remains training providers and other staff to ask the questions when there are common biases that may prevent them from wanting to ask or document this information.[247] Both UC Davis and Vanderbilt health systems are beginning to collect information about patient's sexual orientation and have opted to use patient portals for doing so.[248] Given the opportunity to answer questions from home, patients may be more comfortable reporting certain information. Added benefit of reporting from home is that family members may help if there are language barriers. Geisinger Health System has started using patient portals to collect information about existing medications, and this information gets put into the EHR. Patient reporting may both save clinician time and include information that would not otherwise get entered. Vendors have developed tools such as clinical prediction rules and analytics engines to prompt clinicians based on information a patient enters.[249]

In recent years, there has been increasing effort to promote standardized collection of race, ethnicity and language data by registration staff in response to policy initiatives as well as accreditation requirements. Efforts often include staff training and patient education. For example, the Hospital Association of Rhode Island received funding for a five-hospital pilot to improve collection of race and ethnicity data. Its pilot included input from stakeholders on which granular ethnicity categories should be collected, standard interview scripts for staff to collect patient information, and materials to educate patients on why they were collecting the data.[250]

Clinical encounter data

Data collected during office visits and entered by the clinician into patient records during a visit may include reason for the visit, height, weight, vital signs, patient reported symptoms and characteristics (such as behavior and lifestyle), diagnoses, treatments and tests ordered, and medications prescribed.

Information the pharmacy, laboratory and radiology are often incorporated into the EHR. This should include test results and imaging from other systems.

Clinical information may be entered in a structured format where the clinician can select from standard, predetermined categories such as diagnosis or procedure codes or medication list. Clinicians may also enter information in free-text notes in their own words or the patient's words. For a condition such as autism spectrum disorder, relevant information may be entered as a diagnostic code or in free text about symptoms suggest the diagnosis or about patient or parental reports of such a diagnosis in the past. Diagnostic information may also be implied by the clinician's prescription choices.

Although the use of electronic health records creates opportunities for standardizing much patient care information by setting requirements for data fields, many clinicians prefer to record information in the unstructured manner that was used when entering information into paper charts. Many clinicians have traditionally audio-recorded their notes from the visit, and voice recognition software can now transcribe audio-recording into free-text fields in the EHR.[251] This preference may disappear over time as younger medical students who grew up using computers enter clinical practice. Whether information in an EHR is structured or unstructured has important implications for research, which will be described later in this report, but today most information contained in EHRs is unstructured.

Claims/billing information

Many providers have electronic practice management systems that handle functions like scheduling, billing, and collections. Such systems are increasingly being integrated with electronic health records. Although this is being done for practice management purposes, it can make the overall data system more useful for research. Billing systems can have more complete diagnostic and procedure information than do EHRs.

Figure II.1. Example: Potential Structure and Information in an EHR
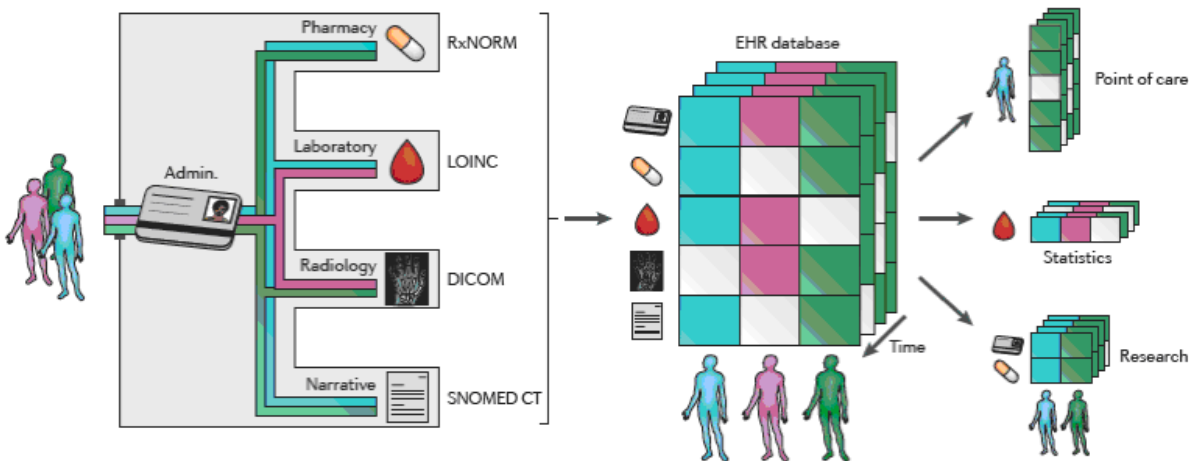
Source: Jensen PB, Jensen LJ, and Brunak S. Mining electronic health records: towards better research applications and clinical care. Nature Reviews, June 2012 (13): 395-403.

Availability of Information to Identify Small Populations

Some small populations may be identifiable using information that is now typically recorded in EHRs. Residents of rural areas may be identifiable by the address and zip code information that is collected for billing purposes, although not all providers collect updated address information at each visit, so some of this information may not be up to date. [252] In addition, lack of EHRs in rural practices and hospitals limits the availability of electronic health data on rural populations.[253] While rural providers are increasingly adoption EHR systems, there will remain the problems of interconnectivity and interoperability. There is also evidence that critical access and small hospitals are at risk of failing to meet Meaningful Use criteria, which suggests there may continue to be limited data available on rural populations,[254] even where EHRs are adopted. Therefore, conducting rural health research using EHR data may remain for the time being in the hands of a few integrated health care delivery systems with EHRs and data warehouses that serve large rural populations, which may not be representative of rural populations in general. Some of these organizations have been able to drill down within their rural populations for research or quality improvement purposes. For example Intermountain Healthcare has looked at rural patients with 3 or more chronic conditions,[255] and Kaiser Permanente Northwest (KP-NW) has looked at rural Hispanic patients with Spanish as their primary language, among whom drug seeking behavior has been a particular problem. This population mostly receives its care through the Oregon Community Health Information Network (OCHIN) of federally qualified health centers (FQHCs), to which the KP Foundation Health Plan gave $1 million to purchase the Epic electronic health record software, so this network and KP are now collaborating on research. Since OCHIN hosts the EHR for nearly all the FHQCs in Oregon and the FHQCs are attempting to create a single medical record for each unique individual (rather than a separate record for each clinic visited by a patient), it is possible to identify drug-seeking behavior by patients who attempt to obtain opiate-containing drug products from multiple FQHCs at the same time.[256]

Adolescents with autism spectrum disorders may also be identified using date of birth and diagnostic information in the EHR. However, the autism diagnosis may appear in free text rather than in structured

fields in the EHRs.[257,258] Even within structured fields, a number of diagnostic codes can indicate someone has an ASD. Kaiser Permanente in Northern California has developed a list of valid autism diagnoses based ICD codes and who made the diagnosis.[259] There is also variability within or across provider organizations regarding who can authoritatively diagnose ASDs, as well as on the tests and benchmarks that are used. Diagnoses of ASD are often made at psychological testing sites that are separate the patient's health care organization, particularly for those with higher incomes, and this may affect whether ASD appears in the organization's EHR. Regardless of a family's ability to pay, diagnosis of ASDs is also often made by school psychologists, especially at kindergarten intake. Providers of ASD patients' medical care are not necessarily skilled at diagnosing conditions such as ASDs.[260]

An additional challenge when studying any adolescent population is that EHRs have generally been designed for adult populations, and pediatric EHRs thus far are not yet as robust. AHRQ and CMS are currently working to strengthen pediatric EHRs with key data elements. However, this work is still in the early stages. EHR and other electronic health data may be particularly important in moving forward research on pediatric medicine, a field where clinicians and families have typically depended on findings from adult clinical trials. A number of pediatric primary care practice-based research networks have developed that are beginning to explore the use of electronic health data for research.[261] For example, Pediatric Research in Office Settings (PROS) is the American Academy of Pediatrics' practice-based research network and has begun an EHR-based sub-network called ePROS. This sub-network was funded through the American Recovery and Reinvestment Act of 2009 and is being built to develop and test the infrastructure needed to conduct pediatric research using EHR systems. It includes providers from diverse practice settings across different states and using a variety of vendors, with plans to expand the sub-network substantially within the next one to two years.[262]

Using EHR information to identify patients who are members of specific Asian subpopulations or the LGBT population remains challenging at present. The broad OMB race/ethnicity categories are increasingly collected in health care settings, but recording information in medical records about patients' membership in subpopulations such as Filipino or Vietnamese rarely happens. There are also variations in how "Asians" get recorded, sometimes along with Pacific Islanders (as per the OMB categories) and sometimes under "Other." Indeed and more generally, the race/ethnicity information in medical records is of variable quality because standardization requires a degree of staff training that does not always occur.[263]

Because the Americans with Disability Act requires health care providers make interpreters available where needed, language information that may identify some Asian subpopulations may be in some organizations' EHRs. KP-NW collects information about primary language spoken at home as well as need for translation services, and has standardized this variable across health plans so someone could easily look up language sub-groups, such as patients who speak Tagalog.[264] At University of Vermont, refugee and immigrant patients have been identified through billing data where interpreters were used.[265] Another approach to identifying racial and ethnic minorities may be use of last names as proxies.

Sexual orientation is almost never collected or entered into patient records, although a few organizations have begun to do so. Therefore, it is important for this and other characteristics not to impute null values where the fields are blank. UC Davis Medical Center has started using a form to collect information for entry into EHRs about patients' sexual orientation as well as gender now and as

assigned at birth.[266] Some such information may already be available in provider notes based what patients may have said about behavior, attraction, or sexual identity. But there has been no standard way to collect this information, so it is difficult to create structured fields for this information. Some EHR vendors such as Epic do have fields to capture information about sexual partners and this can be used to run reports based on the sex of partners. Epic has expressed interest in receiving input from users on how to collect sexual and gender identity in its EHRs.[267] The HMO Research Network's virtual data warehouse has also incorporated sexual orientation as a variable, although they believe there is significant under-reporting of these data across participating health plans. An additional challenge even if this information is being collected is that sexual orientation may change over time, so the information in an EHR may or may not be up to date. This challenge also makes it difficult to identify transgender populations because gender is typically collected only once.

The availability of different types of information in an EHR provides multiple possible approaches that can be used to identify a population, and the potential to improve accuracy when these approaches are used in combination. For example, while there are limitations to using diagnosis to identify patients with ASDs, looking also at the ICD-9 codes and medications may provide information to supplement or validate the diagnostic information. However, some of these types of information may be more accessible and more highly valid in an EHR than others.[268]

For example, while ICD-9 codes tend to be readily available, it is variable how reflective they may be of the patient's actual diagnosis. Information on family and social history are generally incomplete and of low quality. However, information such as vital signs (blood pressure, weight, etc.) tend be collected relatively frequently and recorded accurately. Lab results are not always available in an EHR, but when they are they provide highly reliable information and may also be a better indication of what the clinician was thinking than the diagnostic code. EHRs also keep fairly accurate record of what was prescribed, which may also serve to validate the diagnosis (for example, if prescribed insulin, the patient likely has diabetes). However, prescriptions may be less useful to study utilization considering up to 40 percent of prescriptions are never filled.[269]

Characteristics of EHR and Other Electronic Health Data That Make Them Useful for Research

EHR and other electronic health data are increasingly utilized for quality measurement and improvement, but until recently, the potential benefit of EHRs for research has not received much attention outside a few innovative, early adopting health care organizations. However, the use of EHRs for quality improvement has provided a foundation for extracting and formatting EHR data so it can be usable for other purposes, including research. In an EHR-based system, all quality improvement activities are implemented using the EHR. The wealth of information being collected has the potential to facilitate great leaps forward in both the scope and efficiency of clinical, health services and policy research.[270] But, the answer to the fundamental question of whether EHR data are currently good enough for research on small n populations may depend on the definition of research and/or the specific kinds of research of interest. While EHRs may be well-suited for some types of research, it may be poorly suited for other kinds of research, and while the field has recognized this concept of "fit" between purposes and data it is still

working through for which kinds of research EHRs and other electronic health data are currently well-suited and where further work is needed.

Health services research has been defined as "the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately our health and well-being."[271] For example, EHR data has great potential value for comparative effectiveness research (CER) about drugs, medical devices, tests, surgeries, or ways to deliver health care.[272] However, CER may require more precise and complete information than is necessarily found in EHRs and so may require additional investment to insure that the data quality in a given system is adequate to the specific type or aims of the research. However, even less precise and complete information may be useful to identify patient populations or potential areas for further study.

Today's medical and pharmaceutical research largely consists of relatively small clinical studies using highly selected patients with only one health condition. Findings based on such study participants may have limited generalizability to patients in the real world who often have multiple conditions. The large volume of information going into an EHR creates the possibility of examining rich clinical information about large numbers of patients over time. While EHR-based research may not replace traditional methods of advancing medical knowledge and faces a number of challenges, there are examples in which innovative health systems and researchers have begun to demonstrate its potential for research. Data analytics engines have been developed to mine warehouses of EHR data, to provide the information about how patients with certain characteristics respond to a given medication or treatment.[273]

Analyses of data that have been collected in routine patient care have the potential to greatly increase the speed at which research can move forward. For example, researchers at MetroHealth Medical Center in Cleveland, Ohio were able in 11 weeks to study patient characteristics associated with venous thromboembolic events over 13 years among almost one million patients.[274] Without EHR data, the resources required to recruit and follow so many patients over time would have been incomparably greater. Research to identify risks missed in clinical trials may be conducted through analysis of EHR data—such as Kaiser Permanente's review of internal medical records that revealed the connection between Vioxx and cardiac complications.[275] A benefit of EHR data is that once you identify a population, there may potentially be years of data already available rather than having to wait many years to collect the information, particularly in organized delivery systems.[276]

The fact that EHR data are already computerized and is available in real time substantially increases the efficiency of research, eliminating the need for extraction from paper records and data entry. Rather than being spent for data collection, resources can go towards programming and database work to prepare EHR data for analysis.[277] The data are also timelier than claims or survey data, where there is often a significant lag involved in collecting and processing the data. Data collection in real time also eliminates the need for patients to recall something that happened in the past such as is often required in survey research.[278] EHRs also include much detail about processes of care that isn't available in claims data, as well as information on the uninsured. HRSA has made a substantial effort to invest in data capabilities of safety net providers for this reason—and research networks such as CHARN provide an opportunity to better understand populations where there might otherwise be very limited information. Use of clinical data from EHRs can also help reduce or mitigate traditional coding problems with claims and other administrative data.[279]

The availability of medical record data about all patients in a health system also allows for identification of small subpopulations where identifying information is available in the EHR, such as those in uncommon demographics or with rare conditions.[280] Information may be present about patients who might not otherwise be included in research because they would not meet the narrow requirements for participation in a clinical trial.[281] For example, EHR data has been used for observational comparative effectiveness research among patients with hard to detect co-morbidities, to identify patients for recruitment for interventions, and for population management research.[282] The population covered by an EHR system may provide more representative information than comes from traditional research samples.[283] As use of EHRs increase and efforts continue to improve interoperability of EHR systems and to create networks for pooling data, future research may be based or on actual populations rather than small samples.[284]

Another important aspect of EHRs is their longitudinal nature, which allows populations of patients to be followed efficiently over time so that, for example, outcomes of treatment can be studied. In contrast, surveys collect information at one point in time, typically asking if someone was ever diagnosed or currently has a condition. However, diagnoses change over time. For example, at KP-NW every diagnosis has a date stamp that begins an episode of care, and an end date is also recorded when the episode is resolved. In the EHR, a health problem list is available in a centralized place that displays a patient's entire history of diagnoses received, as well as whether each is ongoing or has been resolved (as opposed to needing to review thousands of pages in a thick chart to get this information). In addition, the recent change that allows children to remain on their parent's insurance coverage through age 26 increases the likelihood that they will remain in a given record system through their transition to adulthood, making it possible to follow those with a condition such as ASD through this transition.[285] As the number of years covered by an organization's EHR system increase, opportunities will grow for research that covers multiple generations of family members.[286] With longitudinal data, there is the potential to make causal inferences, while this is not possible with cross sectional data. However, other factors must be carefully considered in interpreting longitudinal EHR data, such as organizational or national changes that may account for the observed change. For example, an increase in smokers among EHR data may result from increased documentation due to incentives for meaningful use rather than an actual increase in smokers.[287]

A limitation of EHR data, in comparison to survey data, is that the information is not collected or structured for research, which presents a number of challenges for research. While EHRs do include information of great potential value for research on small populations, a number of conditions at the technical, legal, and organizational level must be in place for such research to reach its full potential. These conditions and related challenges in meeting them are described in the following sections of this report, which are organized by these three categories. Technical conditions such as the need to convert EHR data into an analyzable format, legal conditions such as agreement over standards of privacy, and organizational conditions such as the infrastructure needed to share data across multiple institutions will be reviewed. Examples from our interviews and the literature of organizations that have begun to use EHR data for research demonstrate how conditions are coming together to allow the research opportunities to move forward. However, as we discuss in the conclusion, hurdles remain and additional steps are needed in order to take advantage of the opportunities at hand.

Technical Conditions Required for Research Using EHR and Other Electronic Health Data

In order to use information in EHRs for research, it is first necessary for a number of technical conditions to be in place, such as the ability to extract and format data for research, as well as to address issues with missing data and data quality. As with claims data, the information in EHRs was not collected for research purposes. Whereas claims data are collected and entered in ways that help to maximize revenues, information is entered in EHRs to support provide patient care and to fit into clinical routines and workflows.[288] In addition to assisting clinicians and health care organizations in their day-to-day work, the information that goes into EHRs provides documentation that is required by law, that is used for billing, and that informs, patient care decisions. For these purposes, there is not necessarily a need to ensure data are entered in a uniform fashion or to create the capacity for selectively pulling certain information from the system, aggregating data, or identifying certain groups of patients. The cost of converting the information contained in EHRs into databases suitable for research purposes is substantial and requires specific expertise.

Data extraction

Using data from EHRs for research requires extraction from an organization's EHR system so that the data can be cleaned, reformatted, and analyzed. These steps require a substantial staff of programmers; their numbers depend on the system and vendor used.[289] Some organizations create a data warehouse to store extracted data for secondary use—records in such a warehouse have a different architecture than an EHR, which is designed for clinical transactions.[290] An organization may even have multiple data warehouses with the same data but in different forms to support various strategic functions, including resource strategic planning, resource scheduling and inventory control. Part of the problem is that various user groups often do not agree on the definition of variables, acceptable reliability rates and the list of variables to be extracted. However, these functions require data in a different format than exists in an EHR.[291] For example, to facilitate access to information about any given patient, the design of an EHR may include many tables with a lot of linking, allowing clinicians to retrieve only certain information on a patient quickly, such as problem list or prescriptions. However, for research it is more useful to have all of this information in one large flat file.

This can be handled in various ways. Intermountain Healthcare has developed a central data warehouse where all information from its EHR, billing system, insurance product, registration system, and laboratory and radiology systems are pooled and linked. Data sets for research are then extracted from this warehouse rather than the EHR so that research does not interrupt the clinical care process or slow down the EHR.[292] Rather than pooling to and extracting from a central location, Geisinger extracts data from 13 databases (including one EHR database and 12 databases from other clinical and administrative systems) and puts those into a separate database designed for research and quality improvement.[293] New York City's Health and Hospital Corporation (HHC) has data warehouses for each of its component hospital and community health systems from which aggregate data can be pulled. HHC has compiled several registries, such as a registry of some 60,000 diabetics that contains information that is used to track patients and improve outcomes.[294]

Intellectual property issues may be involved. Epic sells a data management product that extracts data from organizations' internal files. However, because Epic considers these files to be intellectual property, client organizations are not allowed to share the internal variable names without permission from Epic. This restriction has been such an impediment that Kaiser Permanente is changing variable names used for many years that have Epic names.[295] There are concerns that as large vendors such as Epic have gained market power, they are able to charge high prices while providing inflexible products and requiring additional costs for each functionality added to the EHR system.

Some research using EHR data has occurred by extracting a subset of data needed for the specific study either by manually identifying the desired records and/or variables, or by querying the system so it automatically retrieves the desired information. For example, a researcher may want to extract the records of adolescent patients with autism spectrum disorders. However, the information needed to select desired records may not be easily available for the computer to identify. While age is likely available to identify adolescents, diagnostic information is often not readily available on ASDs. In addition, not all systems were built to be queried. For example, Montefiore Medical Center in Bronx, New York, found that its system was not structured to be queried, and they needed to develop software to enable them to pull data for analysis from the system.[296]

Studies comparing the accuracy of automated versus manual extraction of EHR data on quality measures has found that the electronic method resulted in and underestimate of the rate of recommended care. For instance, the number of patients that received a clinical preventive service or who met a recommended treatment goal was undercounted when the automated method was used.[297,298] These findings suggest there are risks along with efficiencies in using automated extraction of EHR data for research purposes.

Part of the challenge is that the information needed to identify selected patient characteristics (e.g., autism spectrum disorder) may be spread across multiple fields but not expressed directly. For example, Kaiser Permanente developed and validated a software algorithm to detect episodes of pregnancy in patients EHRs. This algorithm searched for indicators of pregnancy in diagnosis and procedure codes, laboratory tests, pharmacy dispensing, and imaging procedures that are typical of pregnancy. Although using medical records to identifying which patients are pregnant seems straightforward, they found that it is not so easy to automate this synthesis of multiple data points from different sections of a patient chart, which is also difficult to do manually.[299]

Processing free-text data

Data extracted from EHRs must be converted to an analyzable format. The major difficulty for both data extraction and research is that a large portion of the data in EHRs has not been entered in a coded format. Desired information may be in free text that was entered by the clinicians to record their observations and assist with their decision-making. Even diagnoses may be put into free text by physicians because coding it is not needed for their day-to-day work. Some diagnoses (including perhaps ASD) may not be entered because of stigma concerns. Thus, relying on coded fields alone to identify patients with certain diagnoses may result in incomplete and perhaps biased representation.[300] As part of an evaluation of its mental health integration program, Intermountain Healthcare looked for use of a depression metric among patients who received care at its organization. Intermountain found that even when mental health services

were described in physicians' notes, the corresponding data elements were often missing from the structured fields in the EHR.[301]

Free-text data are difficult to use in research they are highly heterogeneous, describing patients with similar characteristics or conditions in different ways. This variation makes it difficult to identify for data analysis patients with shared characteristics. The text may also not conform to standard grammar, may use acronyms and abbreviations, and may include typing and spelling errors. A clinician's assessments may also be recorded as tentative, and the information may be context specific from subject to subject. A disease may be mentioned when it has been "ruled out." Recording the nuances in each case both makes the information valuable for clinicians' work and difficult to use for analysis.[302]

Active efforts are under way to find methods to overcome the limitations of unstructured data, and there has been great progress in developing algorithms and software for natural language processing with which to create standard categories from free text inserted into EHRs by clinicians. Researchers have been able to identify some populations searching for certain words or phrases in the free text of EHRs. For example, Dr. Jesse Ehrenfeld from Vanderbilt University developed and validated tools for natural language processing to identify LGBT individuals from their EHR data in order to determine whether such patient characteristics might be affecting diagnosis, treatment, and health outcomes. This process involves searching records for key terms such as "lesbian" or "bisexual," but also looking for other indicators such as patients listing a same-gender emergency contact with a different last name. He reports that the initial search algorithm resulted in a false positive rate or 22 percent, but that after refining the algorithm to identify negation words for exclusion, only 3 percent of those identified as LGBT using the algorithm had been incorrectly classified as such.[303]

One systematic literature review of clinical coding and classification processes to transform natural language into standardized data found these processes had varying degrees of success.[304] In general, the reliability of natural language processing programs appears to be better where variables are narrowly and consistently defined.[305] Types of coding were found to fall into two primary groups: those that map text to existing classification systems such as international classification of disease (ICD) or current procedural terminology (CPT) codes, and those such as Dr. Ehrenfeld's that used a coding scheme developed for a specific study to look for the presence or absence of certain terms or phrases.[306]

Despite the success of some efforts to covert free text into coded data, some experts caution that natural language processing should not be considered a magic bullet. Natural language processing requires computers that are very large and fast in order to process free text in a reasonable amount of time. In many cases, it may be more efficient and accurate to ask patients for the desired information rather than searching for it in the free text.[307] Also, billing, lab, pharmacy or radiology databases may be better sources of diagnostic information than free text and may worth exploring before turning to natural language processing of the free text in EHRs. These utilization databases tend to be more structured than the problem notes recorded in the EHR.[308]

Other unstructured data includes scanned images, including radiology images but also PDFs of letters or records from other providers that have been scanned or faxed and then uploaded to the EHR. While useful for a clinician to open and view, converting them into something codable takes great effort and computing power. This issue is a whole sub-field of informatics by itself.[309]

Missing data and data quality

In addition to lack of standardization, the accuracy and completeness of data entered into EHRs are major concerns for research, since high quality and complete data are needed for drawing valid conclusions. Data quality has often been called into question when EHR data have been used for quality assessments. Compared to paper charts, electronic health records have been found to hold significant errors—in part because during this transitional period, many clinicians have not been accustomed to using a computer as part of their daily workflow. In addition to typos and spelling errors, errors of omission and commission have been found in medication lists and in problem lists where chronic and acute conditions are documented.[310] Information entered in an EHR may also be affected by billing considerations. For example, some clinicians may not see the need to add secondary diagnoses for complex patients, if doing so would not affect the DRG payments. Such omissions may result in researchers' underreporting certain diagnostic complexities.[311]

Because EHRs today may not reliably provide a complete picture of a patient's health, researchers should guard against drawing conclusions as though they were complete, such as assuming that the absence of mention means that a particular characteristics, condition or treatment are not present. For clinical purposes, a physician may be more likely to record problems than improvement, particularly if there is no need for follow-up, but a researcher would need that information.[312] In addition, some research that relies on EHR data may be skewed because the data do not include people who are unable to obtain care because of access barriers resulting from lack of insurance or differences in language or culture.[313] This is a particular issue for the transgender population, which is often uninsured or seeks services that insurance does not cover, such as hormonal therapies, that have often been obtained outside the health care system.[314] There is also the issue of patients moving in and out of EHR systems—either because they have stopped receiving care or have gone to another health care provider. For Asian subpopulations, they may even be going between countries and receiving care and taking medications they have obtained abroad. The mobility of populations can make it difficult to create cohorts and to make reliable inferences about them.[315]

The need for certain types of patient such as those with ASDs to see multiple providers (including mental health and medical providers) also makes it challenging to get a complete picture of someone's health care through an EHR. Children may also receive testing for ASDs through the educational system that may not be shared with the child's pediatrician. Although this challenge is related to the bigger issue of how the health system is organized, further development of the ability to share information among providers will be important in studying small populations. However, there remains the challenge of a patient may go to that do not have electronic data (such as some long term care facilities), making it more difficult to integrate the information into the patient's electronic record with his or her primary care provider.[316]

However, increasingly integrated models of health care delivery should present opportunities to gain more complete pictures of patients' care for study. In an integrated delivery system, a single organization provides most or all of a patient's care across multiple settings. Integrated systems tend to be particularly advanced in the functionality and use of the EHR systems as a mechanism by which they can coordinate care across multiple settings. Therefore, a number of those interviewed for this report work in such organizations, and many examples we mention in this report come from integrated delivery systems. Shared EHR systems have permitted an increasing number of health care organizations to operate as

virtual systems even though they are not a single organizational entity. This creates new opportunities to study patient care across multiple settings.

With the recent growth of accountable care organizations (ACOs) and the accompanying needed data sharing, researchers may increasingly be able to capture information about patients regardless of where they receive care. For example, because Essentia Health in the upper Midwest is an ACO, it has electronic access to patient information no matter where among the collaborating organizations they receive care, and Essentia can successfully request this information from other providers as a condition of getting paid for services for patients covered by the ACO contract.[317]

The growth of ambulatory networks connected with hospitals also facilitates this type of data sharing. For example, the Pediatric Research Consortium (PeRC) at Children's Hospital of Pennsylvania (CHOP) is able to match outpatient data from CHOP's primary care network with hospital data for patients who have received care in both. However, information is not available about care received in other settings, so the EHR system is most useful for the subset of patients who receive sub-specialty care within CHOP as opposed to the whole network.[318]

Restricted data

At times a portion of the medical record is restricted or separated from the rest of the patient's information if it is viewed as sensitive in order to protect the patient's privacy. This may be of particular concern for small populations where there may be an associated stigma, such as ASDs or LGBT populations. Patients with ASDs often receive care from mental health providers, and it is common for some or all of this information to be restricted. Even if it is included in the medical record, researchers may need special permission to be able to use it for a study—particularly as mentally disabled or cognitively impaired persons are considered vulnerable populations and therefore are a protected class of human subjects when research is considered by institutional review boards. This is an issue not only for EHR data, but for claims data as well—where any substance abuse claims must be removed when the data are used for research.[319]

Legacy systems

Because most EHR systems are relatively new, the number of years of available patient data varies by organization; information needed to look at a patient over time may be in paper charts or legacy electronic systems and not available for EHR-based research. Physicians in organizations that have upgraded their EHR systems may be able to login to the old system to access critical patient information stored there, but the information might not be readily available for research. The alternative ways to link legacy data into new systems all require time and resources.[320]

Needed expertise

The skills required to conduct research using EHR data are highly technical and specialized. A team of information systems staff is needed to support an EHR data warehouse to support care delivery, and translation to a research database requires another set of technical experts. This research informatics team must include programmers and analysts who build and maintain a research-focused warehouse.[321] Higher education has yet to catch up with programs designed to provide training around these skills, which would require links between business and medical schools.[322] The leader of this team must possess both IT skills

and clinical expertise, and these individuals are in short supply as well, particularly as both the fields of medicine and technology have been quickly evolving.

It is also crucial that individuals conducting EHR research have knowledge of research methods specific for EHR data because a unique longitudinal data set is being repurposed. Expertise needed include statistical expertise to format and analyze the data, and the ability to interpret findings while considering how the data were collected and formatted, as well as any limitations connected to the patient population and the context. These considerations require individuals with expertise around organizational and policy history that may affect how data was recorded. For example, an organization's decision to train staff on the collection of race/ethnicity data, whether for internal purposes or to comply with policy or accreditation requirements, may explain a perceived growth in the number of patients they serve from a certain Asian subpopulation over time. Changes in the system, personnel, and social history need to be documented and considered when interpreting data. Therefore, it is important that data warehouses and networks collaborate with their participating organizations and providers.[323]

Privacy and Security Conditions Required for Research Using EHR and Other Electronic Health Data

In addition to technical requirements for data extraction and analysis, there are legal requirements that complicate the repurposing of EHR data for research. Privacy and security may be of particular concern for small populations, where individuals may be easily identified with just a few variables. In addition, particularly where there may be issues with stigma, individuals from small populations may not want to be identifiable by their employer, school, or others who may access the data. Institutional review boards are used to requiring that data are used for only one project for which patients consent, and that identifiable data are destroyed at the end of the study. Such requirements create barriers for the use of EHR-based data for research. Usual practices for protecting privacy and security may need to be reconsidered when EHR-based data are to be used for research. This data source will have increasing potential to answer additional research questions as more information is collected over time. Alternatives to study-by-study review and consent requirements will need to be found if the potential of EHR-based data is to be realized.

Legal landscape

Presently, the two federal laws most relevant to the use of electronic health data for research are the Health Insurance Portability and Accountability Act (HIPAA) and the Common Rule.[324] In addition, there are state laws that govern the use of health data tend to go beyond the protections provided by HIPAA. While HIPAA allows covered entities (including most health care providers) to access, use and disclose identifiable personal health information for treatment, payment, and health operations (including quality improvement), the HIPAA Privacy Rule requires informed consent be obtained from individuals to use this information used for research. The Common Rule covers research conducted using federal funding from certain agencies, and defines research as "systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to general knowledge." Application of these two laws broadly defines what is legally considered research today.

The original HIPAA legislation was passed in 1996—before the use of EHR based data for research was foreseen. Concern is growing about how much the HIPAA rules and their local application may deter important research based on secondary use of patient records.[325] The HIPAA omnibus rule was changed earlier this year with the intention of increasing protection and control of personal health information, particularly in light of the growth of electronic data. Individual rights are expanded so patients can ask for a copy of their electronic medical record, as well as instruct their provider not to share their information with their insurance company if they pay in cash. In addition, the new rule aims to reduce individual burden by allowing the use of their health information for future research purposes.[326] This however does not address the need for consent for secondary uses of already collected data for research.

There is ongoing legal/ethical debate about the role of restrictions based on HIPAA and human subjects' protection in governing the use of EHRs for research, as well as on the blurring line between the use of the information for quality improvement and for research. The IOM has suggested that in a learning health care system, the distinction between research and quality improvement or other internal uses is artificial, and the laws remain unclear on this difference as well. Out of caution, IRBs tend to treat all secondary uses of data as research—a practice supported by publication policies of many academic journals that require IRB approval for results to be published.[327] Other countries such as the UK and Canada are in the midst of similar debates around balancing the need to protect privacy with secondary uses of data for research. Some countries such as Denmark have concluded that database-driven research should be allowed without the consent typically needed to protect research subjects because of its contribution to the common good without disrupting people's everyday lives. Because studies entirely based on national registries or clinical databases can be done without patient consent, a growing number of population-based studies using EHR data are being done in Denmark.[328,329]

In addition, where there is lack of clarity or knowledge on the details of the laws, researchers tend to air on the more conservative side where they perceive there may be a potential issue for their IRB. At times, it is even unnecessary to go through the IRB but it is done with intentions of being cautious—but also creating unnecessary expense and patient and provider burden that at times are not legally necessary.

Opportunities for patients to make meaningful choices

While the intent of informed consent is to respect patient autonomy, it has been argued that the public benefit of health research is greater, particularly if adequate provisions for protecting data confidentiality are present.[330,331,332] The burden would be intolerable if patients had to be re-contacted for consent for each new research use of a database that contained their records. The ability for patients to now give consent for future research given the update to HIPAA may help relieve this burden. However, patients may want their information to be used for certain purposes and not others, or change their mind over time. Interestingly, there is some evidence that patients view the use of medical records to be part of the health care routine and a necessary part of receiving good treatment rather than considering it in terms of the costs and benefits of participation in research.[333] There are a full range of practices that can help patients make a meaningful choice, such as transparency around how their information will be used, who will use it, and allowing patients access to their own data.[334]

The benefits of seeking individual informed consent before using their EHR-based data for research are increasingly seen as coming at too high an administrative burden on research.[335,336] Of even greater concern is the potential for bias when records of patients who have not consented are excluded. One

national survey found strong support and willingness to share one's electronic health information for research,[337] and evidence is accumulating that patients who refuse to agree to the use of their records in research differ in various ways from those who agree. A recent review of 17 such studies from around the world (including 5 from the United States) found differences by age, sex, race, education, income, and health status between patients who did and did not consent to the use of their medical records for research.[338] Such differences could bias research results or limit generalizability of findings. This could be particularly problematic in research on small populations. In addition, there are specific issues with including child populations (such as adolescents with ASDs) in research because they are not legally able to provide informed consent, which implies understanding of the potential risks of participating in research. Parents must provide consent on their behalf, but may uncomfortable with their children being included in research studies. Until recently, children were rarely included in medical studies. Agencies such as the FDA are making an effort to educate parents on the importance of including children research.[339]

Both HIPAA and the Common Rule have been criticized for over-emphasizing patient consent rather than providing more comprehensive opportunities for patients to make meaningful choices.[340] Organizations that conduct a lot of research using EHR data have taken a number of approaches to issues of meaningful choice and protecting patient privacy. These approaches include obtaining general consent from patients at the time care is being provided for the use of their records for research, standardizing IRB documents, classifying studies as quality improvement rather than research, and using de-identified data. For example, Essentia Health asks patients to sign a general consent form each year to use their data for research purposes. Only 1–2 percent of Essentia's patients have been opting out, and those who opt out don't appear to be different from those who do not demographically. This general consent applies only to research conducted within the health system and its research institute, and IRB approval is needed for use of the data for research.[341] Geisinger Health System requires IRB approval for each research project, but has standardized the needed documentation to streamline the process. They also take additional steps to protect patient information, such as altering dates in the copy of the data used for research to protect confidentiality.[342]

For Kaiser Permanente, when someone signs up to be a member, they are informed that their data will be used for "approved research purposes." Members may request to be excluded from all future research projects or from all genetic research. IRB approval is not needed when identifying information in EHR-based studies is used only to make linkages and then removed.[343] Vanderbilt has also granted a waiver of consent under the IRB Common Rule to allow research on LGBT patients without consent since the data are de-identified after extraction. However, patients do have the opportunity to opt out of studies.[344] New York's Health and Hospital Corporation makes only de-identified data available to researchers.[345]

Other health systems such as Intermountain Healthcare and UC Davis conduct some studies that are classified as quality improvement rather than research, and these do not require IRB approval or informed consent.[346] Such classifying of studies as serving operational purposes may avoid the privacy protections needed for research (defined as intended to generate generalizable knowledge new findings for publication), there are tradeoffs. If the activity is conducted for quality improvement or other in-house purposes, the investigator may lose ability to set priorities, be unable to invest the time needed for a rigorous study, or to candidly share findings externally. This disincentive to share knowledge externally prevents much of this type of work from contributing to a learning health care system.[347] On the other

hand, analytics performed for internal uses such as quality improvement may have the benefit of leveraging available data facilitate studies that are quicker and less costly than traditional research. [348,349]

De-identified data

HIPAA's Privacy Rule does not regulate de-identified data, and it specifies that data can be de-identified using safe harbor criteria (the removal of 18 specified data fields that could be used to identify an individual) or statistical methods (demonstrating extremely small statistical risk that an individual could be identified). Statistical methods are less commonly used because the description is vague and there remains lack of a standard approach.[350] In addition, individuals with the knowledge needed to make an expert determination that the statistical risk is sufficiently small are in short supply. However, some organizations such as Vanderbilt's Multicenter Perioperative Outcomes Group, a consortium of 30 medical centers aggregating EHR data, patient reported outcomes and administrative outcomes,[351] have opted to seek this expert determination instead after finding use of the safe harbor criteria to be more challenging, particularly when pooling data from multiple centers. The Privacy Rule does allow the alternative of using a limited data set that includes certain geographic and date information considered important for patient-centered outcomes research, but then requires a data use agreement between the data holder and the recipient. Researchers at Kaiser Permanente have found limited data sets to be useful for research when the length of time between events can be included where full dates are not allowed.

While eliminating the need for informed consent, de-identifying data may remove the information needed to identify small populations. For instance, removal of geographic identifiers makes it impossible to identify residents of rural communities. In addition, de-identified data complicates linkage of patient records from multiple sources, such as with lab or pharmacy data if not integrated into the EHR or across multiple institutions where the patient may receive care.

Governance

Governance processes specifying who owns, controls, and regulates the data must also be in place in order to use EHR data for research. Data governance is generally understood to include legal and regulatory concerns, the structure and role of governance bodies, IRB issues, properties of data, data sharing considerations, business issues, stakeholder engagement and participation, and sustainability.[352] Institutions may designate committees or have designated employees responsible for these issues. Data governance has also been described as the process designated for the data steward (such as a health care organization) to carry out its responsibilities. A data steward has fiduciary responsibilities toward the data, or has been trusted with information that patients consider private. The role of a data steward continues to evolve both conceptually and legally, particularly as health care data have potential not only for research, but are already used for many purposes in the public interest such as for quality monitoring and improvement.[353] There remains a lack of coherent policies and standards to help govern the secondary use of health data.[354]

In the absence of specific governance structures for research processes, some organizations such as New York's Health and Hospital Corporation have developed a data warehouse and use the data for quality improvement; their data are used less frequently for research.[355] However, building this infrastructure is resource intensive and obtaining funding for this type of development may be difficult for health systems. One of the reasons Essentia developed a separate research institute was because grants are often unwilling to pay for programming at the site of day to day operations.[356] Geisinger has also

developed a separate Research Center which is based on an honest broker system where researchers can request to look at a topic (such as diabetes and a specific genome), and then the broker runs the database and shares the results.[357] Some health systems are creating new companies that house and mine their electronic health record data and to combine them with other sources such as EHRs from other health care organizations. Two examples of health systems with such companies are Montefiore (Emerging Health Information Technology) and MetroHealth (Explorys).

Organizational Conditions Required for Research Combining Multiple Data Sources

Because of the previously mentioned limitations with using data from a single organization's EHR for research, the ability to combine EHR data with other electronic data sources is often needed to strengthen study results, particularly for small populations. Combining EHR data across institutions can allow for a larger sample size to increase the likelihood of being able to study small populations, as well as offer a more complete picture of patients that receive care in more than one place. While providing additional information, using data from multiple data sources for research does come with an additional set of challenges and requires a number of organizational conditions be in place, as described in this section. Examples of multi-organizational efforts such as research networks are described below where organizations are already working together to overcome these challenges. In addition, a number of other data sources that may be combined with EHR data to further facilitate research on small populations are described at the end of this section.

Using EHR and other electronic health data from multiple organizations

In order to conduct research with data from multiple organizations, a rationale and a mechanism are needed for organizations to share the data. The technical and legal issues associated with data sharing have received considerable attention throughout the implementation of provisions in the HITECH Act to promote health information exchange to improve the quality of care. There are two major ways that data can be share across multiple institutions: through a consolidated warehouse where a copy of the data from each institution is stored, or through some form of distributed network where the data remains stored with each organization but can be queried to retrieve standardized results from multiple databases. An additional criticism of the current legal framework surrounding human subjects research is the lack of guidance around the technical architecture of databases, although they may involve creating multiple copies of a patient's data.[358]

While centralizing data in a warehouse may increase efficiency when standardizing and querying the EHR data, it requires resources to build and maintain. In addition, there are privacy and governance issues associated with creating a copy of patient information and storing it outside the organization when these data were collected for the organization's use in caring for the patient.[359] Also, as the data are centrally combined from multiple organizations, it becomes further removed from the different organizational contexts where the data were collected that must be considered when interpreting the data, such as changes in how the data was collected and documented over time. In addition, centralized data warehouses may be less flexible as all required data elements must be contributed by the organization in advance and then remain in the warehouse, giving organizations less control over which data they want to contribute for what purposes.[360]

As an alternative to creating a central warehouse or database, a virtual data warehouse may be created where data remains in separate home locations. This alternative may be more viable as it bypasses the need for investment outside the organization in building a separate infrastructure, and also simplifies the issues of data ownership. Virtual warehouses are easier to implement and more private because data remain at the collaborating organizations (referred to as a distributed network). Secure, remote analysis of these separate databases occurs through a central portal that queries and distributes results. Organizations may decide which data they are interested in contributing and what studies they want to participate in. One common type of distributed network is a federated research network, where separate, heterogeneous databases from multiple organizations make up the distributed network and each organization retains control of its own data. [361,362] For example, ePROS is creating a federated database that links data from multiple organizations in order to allow for queries of de-identified patient data.[363] Often the databases include standardized content areas, data dictionaries, and methods to define individuals. [364] While more efficient than a centralized model, investment is still needed in the administrative and governance infrastructure to maintain security and ensure appropriate use of the query function.[365] A number of distributed research networks are being piloted to support clinical effectiveness research (CER).[366,367]

Figure II.2. Example: The Cancer Research Network (CRN) Virtual Data Warehouse
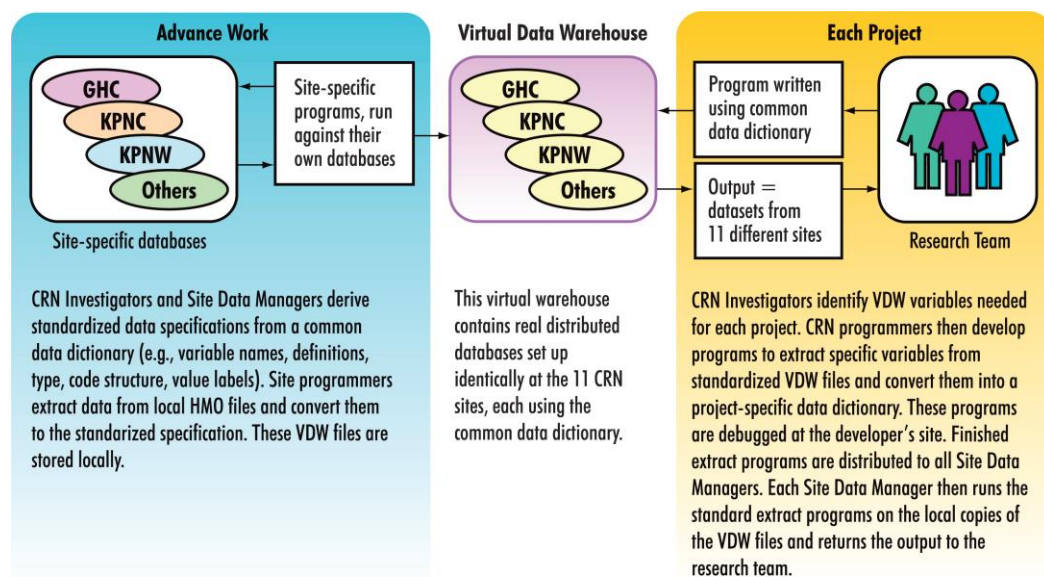
Source: Hornbrook et al. Building a Virtual Cancer Research Organization. *Journal of the National Cancer Institute Monographs*. 2005 (35), 12-25.

However, there are some reasons an organization may select a centralized warehouse instead of a virtual one. For example, the Community Health Applied Research Network (CHARN) chose a centralized data network because to house the data where it originates as in a virtual network, each participating organization needs to have its own infrastructure. However, because CHARN's participants are community health centers that have limited resources, they lacked the capacity to make a virtual network an option. Cost would also be a significant barrier for each community health center to maintain its data locally. Finally, data quality was a consideration when CHARN selected a centralized database. Because of the variability among community health centers, were they to request data from each center it would be difficult to know what types of problems there may be in terms of outliers, omissions and

commissions in the data. Therefore, they decided it would be simpler to look at the data all together. The issues faced by community health centers may be common among other under-resourced organizations that provide care for certain small populations, such as health care organizations in rural areas.[368]

An additional alternative to a distributed warehouse where data are still contributed for central analysis is to have distributed analytics. This approach is being used by the Massachusetts eHealth Institute, where participating organizations to contribute just the minimum information that is needed. While this approach addresses a lot of privacy related concerns, it does require participating organizations to conduct some of their own analytics before contributing their results.[369]

No matter which method is chosen for sharing data, each strategy requires significant infrastructure development, both technically and organizationally. One study of research teams that have developed such infrastructure to support CER identified a number of challenges, including the substantial effort required to establish and sustain partnerships for data sharing, understanding the strengths and limitations of their clinical information platforms, and the need for rigorous methods to ensure data quality across multiple sites.[370] Another study involving interviews with multi-site research initiatives around data governance found a number of challenges related to data governance, but also found these initiatives are using strategies to address these barriers such as capitalizing on pre-existing relationships, beginning with smaller studies and then expanding, developing legal and policy documents with broad input, exchanging de-identified data only, and structuring governance bodies with broad representation.[371] It is important that each organization contributing data is represented in the analysis as well in order to provide context on how the organization has changed, which affect how the data are interpreted. Particularly for those who care for certain small populations, the organizations that care for them are likely unique as well and need to be able to provide that context. The uniqueness of each organization may result in quality issues once their data are combined, even if data from the individual organizations are of high quality on their own.[372]

Funding for research infrastructure development is rare, as currently most grants and contracts pay for specific, discrete studies. However, in recent years the availability of this funding has increased. For example, the American Recover and Reinvestment Act of 2009 allocated $100 million to building infrastructure to use electronic clinical data for CER, patient-centered outcomes research, and quality improvement.[373] In addition, in 2013 the Patient-Centered Outcomes Research Institute is investing $68 million to support the initial development of a National Patient-Centered Clinical Research Network to build the capacity needed support CER. There are currently three funding opportunities related to building this national network.[374]

In addition, for studies that include data from multiple organizations, approval may have to be obtained from multiple Institutional Review Boards, adding to the time and resources needed to conduct the research. Where organizations are from different states, there may also be different state laws governing health information to which each organization must comply. Some approaches to minimizing this burden have included careful distinctions between quality improvement and research-driven interventions, particularly where projects are low-risk. Negotiation of an arrangement where a central or lead IRB with particular expertise in the area first reviews the study and then other IRBs can accept their review may also be another solution.[375] In addition, where research is conducted across distributed databases using methods such as distributed regression, the only information exchanged is statistical

results rather than the underlying data. This technical strategy is one solution to protecting patient privacy. However, an issue with small populations is that unique individuals relative to their surrounding population can potentially be identified. In fact, some researchers are finding that people may re-identify themselves, even when given privacy protection.[376]

Finally, a process is needed to ensure the quality of multisite data for research, including prioritization of variables and dimensions of quality for assessment, development and use of standardized approaches to assessment, iterative cycles of assessment within and between sites, targeted assessment of data known to be vulnerable to quality problems, and detailed documentation of quality to inform data users—particularly in determining whether the data are fit for use in CER studies.[377] Ideally, these efforts should be shared among the collaborating organizations on a continuous basis to keep pace with new versions of existing software and the introduction of new software to manage health care processes.

Interoperability of EHR systems

Research among multiple institutions is facilitated by interoperability of their EHR systems. In its absence, a large amount of effort is needed to integrate data. One of the reasons that building the infrastructure to share data is so challenging from a technical standpoint is the lack of interoperability among different EHR systems. Just among providers who have been able to demonstrate they are meaningfully using their EHRs based on the criteria specified under the Medicare EHR incentive payment program, 333 different EHR vendors have been used, although consolidation is occurring in the EHR industry with the top 5 vendors increasing being used by a larger share of providers.[378] While the industry continues to consolidate, the wide variety of systems currently in use has led to two major challenges: 1) Syntactic interoperability, or the ability for systems to communicate with one another to exchange data; and 2) Semantic interoperability, or the ability for systems to understand the data exchanged. The ability to exchange data is more easily solved. However, differences in vocabulary and classifications are a more difficult problem, particularly when trying to identify members of small populations across multiple institutions.[379] Even within a single organization's EHR, standardizing the data is a challenge. This challenge is amplified across multiple organizations. Even for seemingly well-defined concepts there is variation. For example, what one system may call "high blood pressure" another system may call "elevated blood pressure."[380] Or, systems may use different race/ethnicity categories.

There are a number of efforts to create standards for EHR data, including the Health Level Seven International's (HL7) Continuity of Care Document. HL7 is the global authority on standards for interoperability of health information technology. In partnership with ASTM International—another developer of voluntary consensus standards, the Continuity of Care Document was developed to foster interoperability by promoting standardization across systems through the use of templates representing typical sections of a patient's EHR.[381] While progress is being made in moving toward interoperability standards, the current set of standards are not at a level that solves many of the problems of researchers we talked to. Many of those we interviewed have been working with their vendors and other health care organizations as well to develop strategies for sharing data despite the lack of a single standard, universal approach to interoperability.

In addition, five major health systems, including Intermountain Healthcare, Geisinger Health System, Group Health Cooperative, Kaiser Permanente and Mayo Clinic have created the Care Connectivity Consortium as a pioneer effort and have achieved interoperability across multiple vendors to enable the

sharing of patient information.[382] While primarily motivated by wanting to provide a model by which EHR data can be shared across institutions to improve patient care, the ability of health systems to overcome interoperability challenges will also have significant benefits for research.

Those we interviewed felt that major vendors and federal incentives can both play important roles in promoting standardized data fields and formats across different EHR systems. For example, if Epic includes sexual orientation and gender identity in its system, that could lead to it becoming an industry standard. However, some smaller vendors may not invest in including these fields in their products unless it is added to Meaningful Use criteria.[383] Meaningful Use requirements as well as quality reporting requirements for accreditation and recognition programs do all have the potential to help lead to greater standardization and interoperability across systems.[384] While Meaningful Use presents only minimum requirements for standardization, physicians have the added incentive to do more because it enhances the value of their practices to potential purchasers.[385]

Research agencies also have the opportunity to promote standardization through what they fund. Although Meaningful Use itself may only do so much, in combination with other levers and incentives, the availability of standardized EHR data for research will likely continue to increase.[386] In addition to interoperability across EHRs, there is the need to integrate supply chain, financial, and clinical data to provide a fuller picture. For an organization like the Health and Hospitals Corporation, which includes hundreds of systems, many decisions and definitions used by each individual component of the system do not align once information is brought together. For example, in terms of defining a visit or encounter, a clinician may only consider a patient to be discharged if they are alive, but from a financial standpoint, a discharge is some who is alive or dead. Or, the name of the same doctor may be entered differently in different systems (for example, whether the last name is listed first or second, whether the title Dr. is included, etc.). Going back and standardizing the data across systems is a lot of additional work. In the long run, it will be important to align these different types of systems as well.[387]

Practice-based research networks

Practice-based research networks (PBRNs) have facilitated much of the research using EHR data from multiple institutions. PBRNs are groups of primary care clinicians and practices that work together to answer community-based health care questions as well as to translate research findings into practice. AHRQ has devoted funding to support PBRNs through targeted grant programs as well as by supporting a resource center, learning groups and conferences. The DARTNet Institute is a growing collaboration of PBRNs (currently including nine of them) that is building a national collection of data from electronic health records, claims, and patient-reported outcomes for the use of quality improvement and research.

Research networks can make a wealth of clinical information available for research through their EHRs. The organizations within a network are often already either sharing a common EHR system or have worked to develop some form of centralized or distributed data warehouse for research purposes. In addition to PBRNs, there are other research networks that expand beyond primary care practices. The Cancer Research Network, a collaboration of integrated delivery settings funded by the National Cancer Institute of the National Institutes of Health, is another example of a network created to facilitate research. Still another example is the Community Health Applied Research Network (CHARN), a network of community health centers and universities established to conduct patient-centered outcome research among underserved populations. Members of CHARN include Kaiser Permanente Center for Health

Research (which serves as the coordinating center), the Association of Asian Pacific Community Health Organizations (AAPCHO), Fenway Health in Boston, OCHIN in Oregon, and the Alliance of Chicago Community Health Services.

Research on small populations is increasingly feasible as networks of EHRs with common structures and formats have developed. There is also the potential to link data across systems to identify a cohort of interest.[388] For example, within the Cancer Research Network, any of the individual health plans will likely include the numbers of patients needed for research on any of the five to seven most common cancers. However, for pediatric cancers or rarer cancers, data must be pooled from multiple medium sized sites or perhaps the two KP California regions to obtain sufficient number of cases for research. Most rare cancers require use of data from California, where KP has 4 million members in its EHR system.[389]

One challenge for PBRNs is that securing permission from individual practices and their vendors to access their server can take some time to make sure everyone is comfortable with the arrangement.[390] Even after practices agree to participate, data use agreements must be established that are specific enough to provide protection, but flexible enough to accommodate research. Often additional, unanticipated data elements are required for research, requiring the revision of data use agreements, as well as working with IRBs at multiple institutions.[391]

EHR vendors have not yet played a big role in networks, which have mostly been built either by health systems or grant funded. However, it appears vendors are currently trying to better understand this space since there is a potential business model. While the involvement of vendors may provide additional resources and help move forward network technology, there is the danger that as the data becomes perceived as more valuable, it may make data sharing more difficult. This may also pose a threat to the current public/private partnership where the data collection occurs in the private sector without public and private sector researchers paying them to do so.[392]

Regional health information exchanges

While initially envisioned as another major source of patient data, it is unclear what role regional health information exchanges will play in the future of EHR-based research. One of the original purposes of the Office of the National Coordinator of Health IT was to facilitate the development of regional health information organizations (RHIOs) that would facilitate health information exchange among stakeholders in their region's health care system. These RHIOs were intended to provide the infrastructure for a national health information exchange. However, their development has faced a number of barriers, including many of challenges mentioned in this report in EHR-based research, particularly lack of resources for infrastructure.[393] Further removal from the day-to-day patient care would make data quality and interpretation an additional challenge when using data from these regional exchanges for research. There have been examples, however, where regional health information exchanges have provided data for regional quality improvement efforts.[394]

Linking EHR and other electronic health data with other data sources

A number of other data sources may be linked with EHR data to provide additional information for research, as well as to validate information in the EHR available to identify and study small populations. Data linkage requires that at least one common identifier be available in both sources that can be used to link records. Unique identifiers that are commonly used to link data at the patient level include social security numbers, health insurance claim numbers, and medical record numbers. Hospital or area level identifiers may also be used for linkage to organizational or geographic level data. Commonly linked administrative databases include disease registries, claims files, survey data, provider files, and area-level data.[395] Additional clinical information—such as genetic, care management, and social network information—also has the potential for linkage with EHR data for research. Several examples of additional data sources for EHR-based research are described below.

Patient Registries
An electronic data source that may be useful for research in combination with EHRs are patient registries, where uniform data are collected from multiple institutions in a central database for a population defined by a particular disease, condition, or exposure. This data may be directly pulled from EHRs or require manual entry based on information from the patient's record. Registries are a simpler form of consolidated data. They include only a core set of relevant data elements for a specific purpose. Registries may be local, such as immunization registries or vital statistics departments that collect birth and death data. Death records may be particularly important because death is often difficult to determine from an EHR. There are also national registries, such as the CDC's National Program of Cancer Registries, and the National Cancer Institute collects information on diagnosed cancer cases and cancer deaths simply to measure incidence and mortality.[396] The Institute's tumor registry adheres to national and accreditation standards and has specialized staff that pour through records in local registries looking for evidence of cancer, including blood cancers. Although labor intensive, it is currently more accurate to use a manual process to determine which records should be included in the registry. In contrast, an automated process to query the registry for records of interest may be used if the records included are already well validated. Local registries are often able to accept EHR data and accept edits from providers. One complication is that at times, data can be corrected in the registry but not in the EHR source data. Registries may collect some patient demographic data in order to determine whether certain populations bear a disproportionate burden of the disease.

Information from registries has been linked to EHR data in order to identify patients with specific conditions. For example, in one study a tumor registry was linked to the Cancer Research Network's distributed data warehouse to identify cancer cases. Race and ethnicity in this study were extracted from cancer registries as well. This study was able to look across eight years of data to examine whether someone's health care utilization increases directly prior to diagnosis of a new primary cancer.[397] The ability to look back to before patients were diagnosed with a certain condition is another unique benefit of research using EHR data and has the potential to improve our ability to identify patients who are at greatest risk of disease to improve targeting for preventive interventions.

Registries can also be linked to EHR data for data validation, such as was done in one study that linked clinical databases with a cancer registry to confirm cases of cancer. In this particular study, they found that 98.9 percent of cases overlapped. The use of multiple data sources presents opportunities to

improve data quality for research. For example, addition of death data from a cancer registry to the clinical database allowed for more accurate stage-specific and overall survival figures.[398]

While registries and EHRs can combine to provide a fuller picture, like EHRs, patient registry data may be incomplete as well. It remains a challenge both to motivate clinicians to participate in registries and to facilitate easy transfer of information from patient records into the registry.[399] Some studies have suggested there may be systematic bias when using only records that can be matched between multiple data sources, such as EHRs and registries. A review of the literature around this topic found a number of patient or population factors such as age, sex, race, geography, socio-economic status and health status that may be associated with incomplete data linkage. This association may result in a systematic bias among clinical outcomes reported from such studies.[400]

An additional limitation of some registries such as the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) registries is that they do not identify the recurrence of cancer. Researchers at Kaiser Permanente are trying to address this gap by looking for utilization clusters in claims as well as digital images to identify recurrence. The potential to use pattern recognition to analyze digital images may increase the accuracy of automated approaches to identify cancer incidence for registries and other purposes, potentially finding more than the human eye could have recognized.

In addition to registries, other systems that exist for surveillance purposes may provide useful electronic information. For example, the FDA's Mini-Sentinel Network is a large multi-system collaboration to track exposure to specific drug products and to conduct case-control studies to identify unexpected adverse events. Participating sites agreed to make their patient medical records available to verify any statistically-identified associations. Because this effort is classified as public health surveillance, no IRB compliance is required.

Genetic Data
As the field of genomics has rapidly evolved in recent years, the routine generation of genetic data for individual patients has received much attention from the general public. The clinical utility is now limited by current inability to effectively process, store, update and interpret genetic data while protecting patient privacy.[401] However, efforts have begun to integrate genetic data into EHRs,[402,403] opening many additional possibilities for research. For example, the mining of EHRs with genetic data may reveal previously unknown disease correlations based on patient genetic make-up.[404]

The National Health and Nutrition Examination Survey (NHANES) has collected DNA specimens from participants from 1999 to 2002, which may be used for secondary analysis and can be linked with the survey data. For permission to use the data, researchers may submit proposals to the Centers for Disease Control's Research Data Center (RDC) for approval, and analysis must occur at a RDC location.[405] In a study funded by the NIH, Kaiser Permanente in California has been able to link genetic information with its EHRs. By collecting saliva from 100,000 members, Kaiser has examined the associations between genetics and smoking and drinking habits as well as body mass index.[406] While these saliva samples were expressly collected for research purposes, there have been other instances where blood or other bio specimens collected for medical purposes were reused for research.[407] Instances such as these bring to light the need for clearer consensus and guidelines about the appropriate secondary use of information collected for clinical purposes. One example that may serve as a potential model is the open-consent framework used for the Personal Genome Project, where consent implies research

participants accept that their data could be included in a public, open-access database with no guarantee of anonymity and confidentiality.[408]

Other Data Sources
A number of other data sources provide opportunities for linkages with EHRs. For example, claims data in the Healthcare Cost and Utilization Project (HCUP) databases now feature new linkage capabilities, including ability for linkage to clinical data from labs, trauma registries, EMS data and nurse staffing data.[409] AHRQ has sponsored a number of clinical data pilots to demonstrate the feasibility of linking hospital lab data with HCUP data.[410] Claims data may be an important supplemental source when studying insured populations because it can provide information on care provided across health systems. It may also currently be more useful to identify utilization such as visits or procedures better than EHRs. Although many health care organizations are now using EHRs to bill, EHRs likely only include their own claims, requiring claims for care received elsewhere to be obtained from another source such as the payer.[411] The increase of digital data in all health care settings presents numerous opportunities for research.

In addition, the emergence of care management software programs that track weight, exercise, and medication adherence provide additional information that some providers are entering into EHRs. These programs may download data from pedometers to measure aerobic activity,[412] and have been used for employee incentive programs run by employers or insurance companies. There remains much potential to develop interfaces whereby these types of programs can directly link to EHR systems. There has also been interest in incorporating personal health data from social networking websites and applications on mobile devices into health records for medical care as well as research and public health surveillance. For example, entries on Twitter about disease outbreaks have been correlated with official public surveillance data (although both reflect public concern rather than actual documentation of disease). Or, tracking consumers' online behavior could be linked with bioinformatics. However, use of this data for such purposes presents complications in terms of privacy and consent as online, the lines between public and private are increasingly blurred.[413]

Linking to state and county data sources has allowed some of the organizations we interviewed to better understand their patient population.[414] KP often links its data to the California Department of Developmental Services' database for its ASD patients. However, they are unable to link to the patient's educational records due to state laws.[415] The ability to link EHR data to public school records would be ideal for research on autism spectrum disorders because individuals are often identified in both places and in theory should be managed jointly between the pediatrician and the school.[416] Linking to outside data sets also allows research on the population level, for which Essentia has linked its EHR to publicly available state and county data.[417] State employee health plans such as the California Public Employees' Retirement System (CalPERS), which covers active and retired state and local government employees and their family members, may also be a potential data source of demographic and administrative information, diagnosis as well as information on spending.[418]

There have been a number of recent federal efforts to increase the availability of social, demographic, and behavioral data using a variety of data sources. AHRQ has recently awarded grants from the American Recovery and Reinvestment Act to enhance race/ethnicity information in statewide hospital encounter databases, another source of patient information. State grantees are taking a number of approaches to enhancing data, from standardizing, educating and auditing hospitals as they report R/E/L

data to revising administrative codes to include a mandate.[419] Also, CMS has recently commissioned a study to examine the barriers to collecting social and behavioral data from EHRs for Stage 3 of the meaningful use program, and how to overcome these obstacles. This study will identify the core social and behavioral domains that should be included in an EHR, possibilities for linking EHRs to public health departments, social service agencies, and other non-health care organizations, as well as case studies where such links have been established and how privacy issues were addressed.[420]

In addition, as EHR adoption increases, EHR data plays an increasingly important role in national health surveys such as the National Ambulatory Medical Care Survey (NAMCS), which collects information on practice characteristics and patient visits by abstracting data from a sample of patient medical records from each participating practice. While previously limited to national and regional estimates, the Affordable Care Act has funded a sample increase that will allow for state-based estimates of clinical preventive services.[421] This survey also collects information on EHR adoption, as previously described.

Potential for Future Research on Small Populations

Despite existing challenges to meeting the conditions needed to use EHR data for research, the experts we interviewed provided examples of innovative ways barriers were being overcome. Additionally, they were cautiously optimistic that some other barriers could overcome in relatively short time frames, potentially resulting in a "tipping point" or "major paradigm shift" in how clinical and health services and policy research is conducted in the not so distant future. Specifically, the experts we interviewed had a number of suggestions for ways to move forward in the field of EHR-based research in general and/or ways to study specific small or minority populations. These suggestions can be categorized as potential studies aimed at data validation, new tools and methods for mining and extracting data, descriptive studies around specific populations, and outcomes research. There were also a number of recommendations around engaging and encouraging collaboration among key stakeholders (clinicians, small populations, and vendors) to improve the quality of data collected, as well as on improving the legal framework and other policy issues around secondary uses of electronic health data.

Data validation

The most commonly suggested types of studies were those aimed at further examining the strengths and limits of EHR data, as well as identifying potential methods to strengthen the data for research use. Research networks such as the HMO Research Network,[422] Community Health Applied Research Network,[423] and Practice-Based Research Networks or DARTNet may be good places to conduct this kind of research because of the volume and variety of data they have available and the expertise they have already been developing through other projects and studies. The Health Care Systems Collaboratory was also identified as a good place to start for these types of projects because participants are advanced and can demonstrate the potential of EHR-based research.[424]

A potential related area for research included the development and testing of various patient surveys and/or completed instruments, including perhaps a catalog of items patients could self-report that would be integrated into the EHR and combined with other data. For example, it has been shown that patients will accurately report their height so it does not need to be measured by the nurse, but patients are less

likely to accurately report their weight.[425] A study to examine whether meaningful use has increased documentation of targeted variables was also suggested.[426] In one such study, Kaiser is conducting targeted patient interviews as patients left a doctor's office to see if they are smokers (a meaningful use measure) and whether the doctor talked to them about it, giving them a better sense of how to interpret their EHR data. The use of interviews and other methods of directly hearing from the patient are an important form of validation because although electronic health data can provide a lot of information, the only way to in know how a patient feels is to talk to him or her, or the caregiver. The collection of health-related quality of life data and/or patient experience data provides additional information from the patient's perspective.

One suggestion for research funders from the technical expert panel was to take some studies that have been conducted on small populations using survey methods and to release requests for proposals to see if there is anyone who could look at the same issue and population using EHRs or other electronic health data, allowing for a comparison of results between methods. Similar rapid response requests for proposals could be used when there is a pressing issue for a particular small population that EHR networks could potentially examine. There were also potential studies suggested among those interviewed to examine the validity of data used to identify specific small populations for research, such as:

- A large, prospective study to understand how sexual orientation and gender identity data captured in EHRs differs from patient views[427]

- Research to identify how patients are identified as having an ASD and the data elements needed to study ASD patients, both to assess what data are available and how complete these data are[428]

- Examination of the potential of natural language processing to identify ASD patients[429] and sexual orientation[430]

- Studies on whether and how physicians are collecting information around sexual practices and sexual orientation[431]

New tools and/or methods

As several examples briefly described in the report illustrate, the field is developing a variety of new methods and/or tools to identify priority small n populations in EHR databases and transform key EHR data into analytic files for research. For example, researchers described algorithms or natural language processing software that more reliably and validly identified small n populations of interest and ways to use well-validated surveys to collect key information and integrate it into the EHRs. They also described a variety of different kinds of databases and some of their relative strengths and weaknesses. These and other kinds of tools could be further developed and the significant experience gained from current projects be capitalized on to develop a clearer picture of the strengths and weaknesses of different approaches for extracting and using the data from a variety of perspective and the conditions under which one may be relatively advantageous or likely to succeed.

There is also work being done to explore new methods that can incorporate the use of EHRs and other electronic health data into more traditional methods of research, as well as to better understand what types of studies EHR data may or may not be best suited for. There is a need to further develop research study

designs in order to study small populations. While randomized controlled trials have traditionally been the "gold standard," there is growing agreement that this discipline must evolve, particularly to be able to focus trials on specific subgroups to look for differences. For example, the HSC Collaboratory has been exploring the use of EHRs for more pragmatic, real world approaches to clinical trials. While these approaches may not produce results that are generalizable, for research on small populations in particular there is a lot to be learned if they can be studied as the unique group that they are when the opportunity is available to use quasi-experimental models. Ease of access to the population may also provide opportunities to study of small, unique populations that may be concentrated in certain areas or in a health system or plan where there is good data. For example, Kaiser Hawaii may provide opportunities for research on Asian subpopulations as it serves a large concentration of Asians and has had good ethnicity data for years.

In addition, there should be considerations over what would be a useful control group for studies on small populations. Using controls from within the same electronic health data set may be advantageous because any bias in the data is likely not systematically skewed to the control. Although these biases may not be quantifiable, they can at least be described qualitatively in light of knowledge of the limitations of the data.[432]

It would also be helpful to identify ideal study components where EHRs and other electronic health data can help supplement other information that is collected, such as to provide utilization information for clinical trials, or to help develop high risk cohorts. EHRs may offer a viable first stage screening for proxies, such as use of a treatment as a proxy for having a rare condition. EHRs may be helpful in identifying these research questions, potentially by examining the distribution of comorbidities, or how delivery of care differs across subpopulations. There may also be ways to combine EHR and other types of data such as survey data. Some examples may include using EHRs to identify a population for a more targeted survey, or conducting a survey and then supplementing that information with what is available in medical records. Using a combination of data sources may also facilitate more effective identification of small populations. In addition, while geospatial approaches have typically been used to study rural populations, they may also be useful to study other small populations because they are often not evenly distributed throughout the country.[433]

Descriptive studies

There were also a number of suggested studies using EHR data to better understand the health and health care of specific small populations. For example, Kaiser has used sophisticated sampling with its EHR data to stratify patients into various subgroups according to how likely they are to have COPD—presumably, this could be done with other health outcomes. These studies could serve to examine how various subpopulations fare relative to the majority population and to identify disparities in order to address them. Some examples include:

- **Health:** studies to examine comorbidities of adults with ASDs,[434] or common diagnoses among different Asian subpopulations[435]

- **Social determinants of health:** studies to better understand the patient complexity and risk associated with social determinants of health barriers (e.g., limited English proficiency, poverty level, insurance status) among different Asian subpopulations, many of whom are immigrants[436]

- **Health care utilization:** studies to examine use of pediatric services by adolescents with ASDs during the transition to adulthood,[437] use of psychotropic and ADHD medication among young children with ASDs,[438] as well as referrals to mental health services and outside behavioral diagnostic testing[439]

- **Enabling services:** studies to examine the impact of supportive health services (e.g. insurance eligibility, interpretation, case management) on health for Asian subpopulations[440]

- **Quality:** research around the receipt of recommended care by Asian subpopulations, LGBT, and other minority or disadvantaged groups[441]

- **Patient experience:** use of satisfaction surveys linked to encounter data to examine the experience of LGBT patients[442]

Outcomes research

Finally, a number of interviewees pointed to the potential of EHR data to be used for research examining outcomes, and how these outcomes may differ for different sub-groups of the population. This would include examining the outcomes of medications, types of treatments or care processes,[443] interventions such as smoking cessation or medications,[444] and new models of care such as telemedicine for rural patients.[445]

The information in EHRs is well suited for research around clinical topics, health services, delivery system issues, and quality of care. The volume of information makes it useful for high-level, broad utilization benchmarking as well as for more detailed information on small populations.[446] The ability to identify small populations also presents an opportunity for comparison studies to identify disparities in health and/or health care that may be experienced by certain groups, such as differences in access or quality of care. These data are also useful for descriptive epidemiology that looks at the prevalence and trends of certain conditions over time by certain demographic or other characteristics,[447] as well as quality improvement research to improve care for certain populations.[448]

EHRs also provide a unique opportunity to look for undiagnosed conditions. For example, CHARN is looking for people with possible undiagnosed hypertension by identifying people in EHRs who have high blood pressure but have not gotten tested for hypertension. They are then targeted for testing and therapeutic intervention.[449]

Stakeholder engagement and collaboration

In addition to potential studies, those interviewed recommended efforts to further engage key stakeholders to improve the quality of data collected, as well as to direct the research agenda for using electronic health data to study small populations. In particular, clinician engagement was recommended in order to improve the quality of data available for EHR research. Providing education about the importance of the data may motivate physicians to enter data into structured fields rather than free text. An additional incentive may be to provide feedback on their data quality along with reports around the quality of care.[450] Encouraging clinicians to use their data will lead to improvement as they identify and address errors. Obtaining trust from participants is a big issue—for example, a representative from CHARN

interviewed is aware the participating community health centers (CHCs) are still watching to make sure the coordinating center is not just writing reports using their data rather than engaging the CHCs in research.[451] Information could also be provided to help them manage their patient populations more effectively so they can see the usefulness of high quality data. For example, reports could identify complex chronically ill patients for follow-up.[452] Engaging clinicians in the development of research may help identify research questions that help address the challenges they face in clinical practice. Also, practices that participate in research networks should be supported monetarily and in terms of infrastructure to make sure they are collecting the data that researchers want. Relationship building is required, as well as some benefit to the providers from the data in order to obtain their buy-in and support. Some interviewees also suggested being purposive regarding what types of practices contribute data for research—partnering with those who are interested in using their EHRs to generate evidence, and practices with patient populations who might otherwise be underrepresented in research, such as those serving children or ethnic minorities.[453]

In addition to engaging providers who treat small populations, engaging the small populations themselves is important to improve the quality of data collected. One recommendation from the technical expert panel was to work with the LGBT community to develop ways to respectfully identify them, as well as to gain consensus around what information to collect and what categories to use. With HHS piloting questions to identify the LGBT population on national surveys, there may be an opportunity to compare these findings with EHR-based methods of identifying LGBT patients. Another suggestion was to convene a task force to identify the data needed to study small populations. Establishing common data elements for each population, such as specific demographic variables, may also be a task for such a task force. Vendors must also be engaged around the need for common data elements, as well as to promote the development of EHRs that support a learning health care system.[454]

The legal framework and other policy issues

Although the technical expert panel identified a potential role for the federal government in disseminating best practices on how research has been successfully conducted thus far within the legal framework, there was agreement that in the long run, these "work-arounds" would not be sufficient. Elements of the law that have been suggested as ripe for revision include the over-emphasis on informed consent over other fair information practices, preferential treatment of quality improvement and other internal uses over research, and lack of guidance around network architecture, governance and IRB structure.[455] There is also opportunity for the government to educate the public around the benefits of using their health data for research and the barriers that over-protection of privacy pose to progress in the fields of medical and public health research. Privacy concerns that prevent patients from allowing their data to be shared also leads to a number of health risks, such as errors that occur when a patient's multiple providers do not know what each other are doing. While the younger generation has grown up in the age of social media and may have fewer concerns around privacy, recent events such as the publicity around PRISM (the National Security Agency's electronic surveillance program mining telecommunications data) have brought to light existing public concerns around privacy.

Implementation of policies aimed at closing the digital divide experienced by rural and safety net providers such as the HITECH Act will also improve the availability of electronic health data to study small populations. The need for a business model for EHRs in rural practice remains. The development of subscription-based EHRs operated over secure web portals and requiring only web appliances in the

physician's office may be one solution. Further development of networks like CHARN and support for such networks to learn from the experiences of more well-resourced research enterprises such as Kaiser or the HMO research network is also important for studying these populations. The government may also consider supporting the development of decentralized data warehouses and other IT infrastructure to link health systems in specific geographic areas, such as underserved urban areas or sparsely populated rural areas. Funding the development of "Centers of Research Excellence" to support the development of EHR-based research on small populations may also help build infrastructure.

Finally, closing gaps that occur when children age out of their parent's insurance will improve the continuity of electronic information available to study small populations over time. While additional opportunities and subsidies to purchase insurance through the Affordable Care Act may help address gaps in coverage, there must also be efforts by delivery systems to close gaps in information. Development of personal health records and more robust information exchanges as incentivized in the HITECH Act will help. Simpler solutions exist as well, such as providing patients with a copy of their information that they can share with new providers. This has been done in cancer care and may be helpful to adolescents with ASDs as they transition to adulthood as well.

Summary and Conclusions

Relative to other federal data sources like surveys and claims databases, as well as paper charts, electronic health records have some major strengths. These include: the potential to reach larger samples of individuals, perhaps in some cases approaching the majority of the population or subpopulations of interest; the inclusion of many types of clinically rich, detailed information; the potential inclusiveness and longitudinality of some data sets; and, the ability to link EHR data to other data sources, including patient self-reported information on a variety of issues such as behavior, functioning, or health status and other outcomes. Additionally, the change in medium from paper and pen to computer hard ware and software facilitates the identification, extraction, and sharing of data on a scope, scale, and speed heretofore not possible. Finally, ARRA HITECH funding has stimulated more providers to adopt and use EHRs and ongoing efforts in this area and implementation of health reform is likely to give providers additional incentives to invest in and use EHRs.

While some significant barriers remain, many of the conditions required for harnessing the power of EHRs for a research on the health and health care needs of the American people and key small n populations are present or closer to being realized. Our interviews and literature review illustrate that innovate solutions are being developed through a variety of publicly supported and private efforts. Moreover, these innovative solutions provide concrete examples of how thorny governance, privacy and security, technical, and other barriers might be overcome. They also allow for a "cataloging" of lessons learned from various approaches and potential next steps.

Toward that end, interviewees and our own thinking result in a number of possible suggestions for moving the field forward. They can broadly be described as additional "environmental scanning" to identify promising approaches, convening of HHS agencies and possibly other groups via a public-private partnership framework to identify possible next steps and their prioritization, support for targeted EHR method and data project or specific research projects using EHR data alone or in combination with other

data, and strategic planning and coordination within HHS on ways to proceed in the shorter and longer term.

For example, the research for this report has identified some of the major recent efforts in various HHS departments that have touched on the potential use of EHR data for research, implicitly or explicitly. However, we have not had the opportunity to fully catalogue or mine these programs for "lessons learned." A more comprehensive and detailed identification and mining of innovative examples would be potentially very valuable to the field.

Similarly, we have identified and spoken with the leaders of some of the major federal and/or private research efforts to date and gotten some opportunity to get their thoughts on key areas for further work. Additional input will be gathered from a sub-set of them serving as TEP members. However, a broader group of researchers with complementary and diverse areas of expertise could be convened to weigh in on priorities and next steps. In addition, other major stakeholders such as provider and professional associations could be convened to discuss the issues that the use of EHRs for research as well as operations and related issues (i.e., quality and efficiency improvement) raise. EHRs are currently used for ongoing care and operations, and it is not clear whether and to what extent providers and professionals understand how they can help ensure that such data are useful for research and what might motivate them to become more engaged in and invested in improving the data for ongoing research. In other words, what is the business case for providers and professionals to engage in and/or participate in research that uses EHRs and/or what conditions would make them more interested and able to do so.

As noted above, interviewees identified specific projects that could be pursued. While some of these projects could be described more as EHR data and methods projects, such as EHR data validation studies or studies related to the strengths and weaknesses of different database approaches, others are more focused on particular priority target population or small n population and their health and health care needs. However, right now, many federal funding solicitations do not explicitly call for projects that innovate with respect to EHR data and methods and/or attempt to use it for research for research on specific priority populations.

Finally, drawing on the first two general steps, HHS could develop a broad plan for moving the field forward and/or specific mechanisms and projects that could be pursued to leverage the investments already made in EHR infrastructure, methods, and research. Given the potential scope a scale of the efforts needed, as well as the need to involve a variety of private organizations (e.g., health plans, organized delivery systems) in these efforts, it can be very difficult to determine where to begin and some pathways and mechanisms to facilitate progress. However, it seems clear that a locus of leadership and coordination of effort would be helpful in and of itself. There are pockets of substantial activity but currently no clear organization, department, or mechanism for pulling these pieces together within HHS or between HHS and other potential private partners, particularly with respect to the use of EHR data for research. This is clearly loci of leadership for other areas related to EHRs, such as CMS and ONC for the adoption and use of EHRs to improve quality and efficiency, and private organizations (e.g., health plans, organized delivery systems, vendors, professional associations) are highly engaged and involved in that process. Perhaps there could be an equivalent effort around the use of EHR data for research, which pulls together clinical and health services and policy researchers, key federal agencies, and other private organizations.

In sum, EHRs hold great promise to advance research on a number of topics and populations, particularly small n populations. Although there are numerous barriers, the adoption and use of EHRs is increasing fairly rapidly for many reasons, including ARRA HITECH and health reform and there is tremendous energy and enthusiasm in pockets of the research community about ways to further harness EHRs for research. This report has identified and described some prior federal efforts and related projects, ways they are working to overcome these barriers, and general next steps. Further work will be done by the TEP to identify more specific areas and possible priority areas and ways these general approaches could be more concrete and actionable by HHS alone or in some cases in conjunction with private partners such as foundations and/or associations or networks of major health plans, organized delivery systems, and professional associations.

Appendix to Part II

Table II.1. Key Informant Interviews

Using EHR Data - Target Populations
Asian Americans
- Rosy Chang Weir, PhD, Association of Asian Pacific Community Health Organizations (AAPCHO)

Adolescents with Autism Spectrum Disorders
- Lisa Croen, PhD, Division of Research, Kaiser Permanente Northern California, Kaiser Permanente Autism Research Program

Lesbian, Gay, Bisexual, and Transgender People
- Edward Callahan, PhD, UC Davis, School of Medicine
- Jesse Ehrenfeld, MD, Vanderbilt Program for LGBTI Health, Vanderbilt University School of Medicine

Individuals Living in Rural Areas
- Tom Elliott, MD, Essentia Institute of Rural Health (EIRH)

Using EHR Data—Small Populations in General
- Philip Alberti, PhD, Association of American Medical Colleges
- Robert Califf, MD, Duke University (NIH Health Care Systems Research Collaboratory Coordinating Center)
- Louis Capponi, MD, New York City Health and Hospitals Corporation
- Kaytura Felix, MD, HRSA (co-program director for CHARN)
- Russ Glasgow, PhD, National Cancer Institute (NIH Health Care Systems Research Collaboratory)
- Patricia Franklin, MD) University of Massachusetts Medical School (FORCE-TJR)
- Erin Holve, PhD, AcademyHealth (Electronic Data Methods Forum)
- Mark Hornbrook, PhD, Kaiser Permanente Northwest's Center for Health Research
- Harold Luft, PhD, Palo Alto Medical Foundation Research Institute
- Mary Ann McBurnie, PhD, Kaiser Permanente Center for Health Research (leads CHARN Central Data Management Coordination Center)
- Wilson Pace, MD, University of Colorado, Denver (DARTNet)
- Lucy Savitz, PhD, Intermountain Healthcare
- James Walker, MD, Siemens Medical Solutions, Inc.
- Richard Wasserman, MD, Pediatric Research in Office Settings (PROS), American Academy of Pediatrics and University of Vermont , and Alex Fiks, MD, Pediatric Research Consortium, Children's Hospital of Philadelphia
- David West and Lisa Schilling, University of Colorado (DARTNet and SAFTINet)
- James Younkin, Keystone Health Information Exchange

Table II.2. Select Networks/Organizations Discussed in Part II

Research Network/Collaboratory	Participating Organizations Interviewed for Report	Funding sources	Description
Community Health Applied Research Network [CHARN][1]	AAPCHO, Kaiser Permanente Center for Health Research (Coordination Center)	HRSA	Network of federally qualified health centers and universities created to conduct patient-centered outcome research among underserved populations. Made up of four research node centers and one data coordinating center. Was originally funded in 2010.
HMO Research Network[2]	Kaiser Permanente—Northern California, Kaiser Permanente - Northwest, Essentia Institute of Rural Health, Palo Alto Medical Foundation Research Institute	Membership fees for network infrastructure. Participating systems apply for federal grants/contracts for specific research projects.	Consortium of 18 participating health care delivery systems focused on comparative effectiveness studies and translational health services research. Uses a Virtual Data Warehouse. Has been in operation since 1994.

---

[1] http://www.kpchr.org/CHARN/public/index.aspx?pageid=1
[2] http://www.hmoresearchnetwork.org/

Research Network/Collaboratory	Participating Organizations Interviewed for Report	Funding sources	Description
Cancer Research Network[3]	Kaiser Permanente Northern California; Kaiser Permanente Northwest	NIH	An NCI-funded initiative made up of 9 health care systems [serving close to 9 million members] and 6 affiliate sites to support cancer research based in non-profit integrated health care delivery settings. All participating sites are also members of the HMO Research Network. First funded in 1999.
Health Care Systems Research Collaboratory[4]	Duke University (Coordinating Center)	NIH	Collaboratory aiming to provide a framework of implementation methods and best practices for clinical research done by health care systems. Collaboratory aims to support high impact demonstration projects and provide leadership and technical research expertise.
Electronic Data Methods Forum[5]	N/A	AHRQ	Project that fosters exchange and collaboration between different AHRQ-funded projects aiming to build infrastructure and methods for collecting and analyzing prospective electronic clinical data.

---

[3] http://crn.cancer.gov/
[4] https://commonfund.nih.gov/hcscollaboratory/
[5] http://www.edm-forum.org/publicgrant/Home/

Research Network/Collaboratory	Participating Organizations Interviewed for Report	Funding sources	Description
Registry of Patient Registries[6]	N/A	AHRQ	This project aims to engage stakeholders in the design of a database system that can search existing patient registries in the U.S.; facilitate the use of common data fields; provide searchable summary results; be able to search existing data for research purposes; serve as a recruitment mechanism for new registries. The project was launched in 2012.
Practice-Based Research Networks[7]	DARTNet, SAFTINet, Pediatric Research in Office Settings, Pediatric Research Consortium	AHRQ	Networks of primary care providers and practices joining together to answer community-based health care questions and transform research findings into practice. Consists of 116 primary care PBRNs and 20 affiliate PBRNs (non-primary care and international networks).

---

[6] https://patientregistry.ahrq.gov/
[7] http://pbrn.ahrq.gov/

Research Network/Collaboratory	Participating Organizations Interviewed for Report	Funding sources	Description
ACTION II network[8]	Association of Asian Pacific Community Health Organizations Health and Hospitals Corporation of New York City University of Massachusetts Medical School Kaiser Permanente Northern California Kaiser Permanente Northwest Palo Alto Medical Foundation Research Institute American Academy of Pediatrics Vanderbilt University Medical Center	AHRQ	A network intended to promote innovation through field-based research in health care delivery by accelerating the diffusion of research into practice. Includes 17 partnerships and more than 350 participating organizations that provide health care to an estimated 50 percent of the U.S. population. ACTION II was initially funded in 2011. Its predecessor, ACTION,[9] was funded from in 2006-2010. Prior to ACTION, the Integrated Delivery System Network (IDSRN),[10] was funded from 2000-2005, and awarded nearly $26 million for 93 projects.

[8] http://www.ahrq.gov/research/findings/factsheets/translating/action2/index.html
[9] http://www.ahrq.gov/research/findings/factsheets/translating/action/index.html
[10] http://archive.ahrq.gov/research/idsrn.htm

Table II.3. Technical Expert Panel

Technical Expert Panel

- Jody Blatt, CMS Center for Medicare and Medicaid Innovation

- Jesse Ehrenfeld, MD, Vanderbilt University School of Medicine

- Thomas Elliott, MD, Essentia Institute of Rural Health

- Kaytura Felix, MD, Health Resources and Services Administration.

- David Hickam, MD, Patient-Centered Outcomes Research Institute

- Mark Hornbrook, PhD, Kaiser Permanente's Center for Health Research

- David Kaelber, MD, PhD, MetroHealth System

- Mary Kay Kenney, Health Resources and Services Administration

- Alice Leiter, JD, Center for Democracy & Technology

- Curt Mueller, PhD, Health Resources and Services Administration

- Mary Ann McBurnie, PhD, Kaiser Permanente Center for Health Research

- Wilson Pace, MD, Professor of Family Medicine, University of Colorado, Denver

- Shobha Srinivasan, PhD, National Cancer Institute.

- Michael Stoto, PhD, Georgetown University

- Phillip Wang, MD, PhD, National Institute of Mental Health

- Jonathan Weiner, DrPH, Johns Hopkins University's Bloomberg School of Public Health

References in Part II

1. Adler-Milstein J, Bates DW, and Jha AK. "U.S. Regional Health Information Organizations: Progress and Challenges." *Health Affairs*, 2009; 28(2):483–492.

2. Adler-Milstein J, Bates DW, and Jha AK. "Operational Health Information Exchanges Show Substantial Growth, but Long-Term Funding Remains a Concern." *Health Affairs*, 2013; 32(8):1–7.

3. Allen T. "Better Care through Sharing Electronic Medical Records." *Health Affairs* blog, September 4, 2012, http://healthaffairs.org/blog/2012/09/04/better-care-through-sharing-electronic-medical-records/.

4. Aligning Forces for Quality. "Reform in Action: Can Publicly Reporting the Performance of Health Care Providers Spur Quality Improvement?" April 2012. http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2012/rwjf400299.

5. American Cancer Society. "Cancer Facts & Figures 2013." Accessed February 28, 2013. http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-036845.pdf.

6. American Medical Informatics Association. Letter to ONC Re: 45 CFR Part 171, Nationwide Health Information Network: Conditions for Trusted Exchange Request for Information (RFI), 2012.

7. Andrews R. "Clinically-Enhanced Statewide Hospital Discharge Data: Practical Experience and Potential Value." Presented at AcademyHealth Annual Research Meeting, Baltimore, MD, June 23, 2013.

8. Agency for Healthcare Research and Quality. "What Is Comparative Effectiveness Research?" AHRQ website, accessed July 10, 2013. http://effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness- research1/

9. Arispe IE. "The National Center for Health Statistics: Adapting to Meet New Data Needs." Presented at AcademyHealth Annual Research Meeting; Baltimore, MD, June 2013.

10. Bahensky JA, Jaana M, and Ward MM. "Health Care Information Technology in Rural America: Electronic Medical Record Adoption Status in Meeting the National Agenda." *Journal of Rural Health*, 2008; 24(2): 101–5.

11. Bahensky JA, Ward MM, Nyarko K, and Li P. "HIT Implementation in Critical Access Hospitals: Extent ofImplementation and Business Strategies Supporting IT Use." *Journal of Medical Systems*, 2011; 35(4): 599–607.

12. Bellin E, Fletcher DD, Geberer N, Islam S, and Srivastava N. "Democratizing Information Creation from Health Care Data for Quality Improvement, Research, and Education–the Montefiore Medical Center Experience." *Academic Medicine: Journal of the Association of American Medical Colleges*, 2010; 85(8): 1362–68.

13. Belmont J, and McGuire AL. "The Futility of Genomic Counseling: Essential Role of Electronic Health Records." *Genome Medicine*, 2009; 1(5): 48.

14. Bennett KJ, Olatsi B, and Probst J. "Health Disparities: A Rural-Urban Chartbook." South Carolina Rural Health Research Center, June 2008, http://rhr.sph.sc.edu/report/%287-3%29%20Health%20Disparities%20A%20Rural%20Urban%20Chartbook%20-%20Distribution%20Copy.pdf.

15. Benson B. "Legacy EHR System and Data Lookup a Thing of the Past." HITECH Answers, accessed May 25, 2013, http://www.hitechanswers.net/legacy-ehr-system-data-lookup/.

16. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, and Brand CA. "Data Linkage: A Powerful Research Tool with Potential Problems." *BMC Health Services Research*, 2010; 10(1): 346.

17. Boyle CA, and Boulet SL. "Health Care Use and Health and Functional Impact of Developmental Disabilities among US Children, 1997–2005." *Archives of Pediatrics & Adolescent Medicine*, 2009; 163(1): 19–26.

18. Bradley CJ, Penberthy L, Devers KJ, and Holden DJ. "Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future." *Health Services Research*, 2010; 45(5p2): 1468–88.

19. Brown J, Syat B, Lane K, and Platt R. "Blueprint for a Distributed Research Network to Conduct Population Studies and Safety Surveillance." Effective Health Care Program Research Reports 27. Agency for Healthcare Research and Quality, June 2010. http://effectivehealthcare.ahrq.gov/reports/final.cfm.

20. Centers for Disease Control. "CDC Features - Providing Quality Cancer Data." Accessed May 25, 2013. http://www.cdc.gov/Features/CancerRegistries/.

21. Chan KS, Fowles JB, and Weiner JP. "Review: Electronic Health Records and the Reliability and Validity of Quality Measures: A Review of the Literature." *Medical Care Research and Review*, 2010; 67(5): 503–27.

22. Charles D, Furukawa M, and Hufstader M. "Electronic Health Record Systems and Intent to Attest to Meaningful Use among Non-Federal Acute Care Hospitals in the United States: 2008–2011." ONC Data Brief 1. Office of the National Coordinator for Health IT, 2012.

23. Clark S, and Weale A. "Information Governance in Health: An Analysis of the Social Values Involved in Data Linkage Studies." Economic and Social Research Council, 2011.

24. Cohn SP. "Update to Privacy Laws and Regulations Required to Accommodate NHIN Data Sharing Practices." Accessed June 21, 2007. http://www.ncvhs.hhs.gov/071221lt.pdf.

25. Conn J. "More Than 300 Vendors Share Ambulatory Care EHR Market." *ModernHealthcare*, October 24, 2012, http://www.modernhealthcare.com/article/20121024/NEWS/310249954.

26. Croen LA, Najjar DV, Ray GT, Lotspeich L, and Bernal P. "A Comparison of Health Care Utilization and Costs of Children with and without Autism Spectrum Disorders in a Large Group-model Health Plan." *Pediatrics*, 2006; 118(4): e1203–11.

27. Decker SL, Jamoom EW, and Sisk JE. Physicians in Non-Primary Care and Small Practices and Those age 55 and Older Lag in Adopting Electronic Health Record Systems. *Health Affairs*, April 2012. 10.1377/hlthaff.2011.1121.

28. Department of Health and Human Services. "New Rule Protects Patient Privacy, Secures Health Information." News release, January 17, 2013. http://www.hhs.gov/news/press/2013pres/01/20130117b.html

29. DesRoches CM, Campbell EG, Rao SR, Donelan K, Ferris TG, Jha A, Kaushal R, Levy DE, Rosenbaum S, Shields AE, and Blumenthal D. "Electronic Health Records in Ambulatory Care—A National Survey of Physicians." *New England Journal of Medicine*, 2008; 359(1): 50–60.

30. DesRoches CM, Charles D, Furukawa MF, Joshi MS, Kralovec P, Mostashari F, Worzala C, and Jha AK. "Adoption of Electronic Health Records Grows Rapidly, but Fewer Than Half of US Hospitals Had at Least a Basic System in 2012." *Health Affairs*, 2013; 32(8): 1–8.

31. DesRoches CM, et al. "Small, Nonteaching, and Rural Hospitals Continue to Be Slow in Adopting Electronic Health Record Systems." *Health Affairs*, 2012; 31. 10.1377/hlthaff.2012.0153.

32. DesRoches CM, Worzala C, and Bates S. "Some Hospitals Are Falling Behind in Meeting 'Meaningful Use' Criteria and Could Be Vulnerable to Penalties in 2015." *Health Affairs* 2013; 32(8): 1355–60.

33. Dreyer N. "Interfacing Registries with EHRs." Presented at AHRQ annual conference, September 14, 2009. http://www.ahrq.gov/news/events/conference/2009/dreyer/index.html.

34. Eastwood B. "6 Big Data Analytics Use Cases for Healthcare IT." CIO.com, April 23, 2013. http://www.cio.com/article/732160/6_Big_Data_Analytics_Use_Cases_for_Healthcare_IT.

35. Ehrenfeld J. "Identification of LGBT Patients and Health Disparities: Using Electronic Health Records." Presented at the Sexual Orientation and Gender Identity Data Collection in Electronic Health Records: A Workshop, Institute of Medicine, October 12, 2012.

36. Federal Trade Commission. "Fair Information Practices Principles." Accessed August 25, 2013. http://www.ftc.gov/reports/privacy3/fairinfo.shtm.

37. Felt U, Bister MD, Strassnig M, and Wagner U. "Refusing the Information Paradigm: Informed Consent, Medical Research, and Patient Participation." *Health* (London, England: 1997), 2009; 13(1): 87–106.

38. Field K, Kosmider S, Johns J, Farrugia H, Hastie I, Croxford M, Chapman M, Harold M, Murigu N, and Gibbs P. "Linking Data from Hospital and Cancer Registry Databases: Should This Be Standard Practice?" *Internal Medicine Journal*, 2010; 40(8): 566–73.

39. Fiks AG, Grundmeier RW, Margolis B, et al. "Comparative Effectiveness Research Using the Electronic Medical Record: An Emerging Area of Investigation in Pediatric Primary Care." *Journal of Pediatrics*, 2012; 160(5): 719–24.

40. Ford EW, Menachemi N, Huerta TR, and Yu F. "Hospital IT Adoption Strategies Associated with Implementation Success: Implications for Achieving Meaningful Use." *Journal of Healthcare Management / American College of Healthcare Executives*, 2010; 55(3): 175–88; discussion 188–89.

41. Furukawa MF, Patel V, Charles D, et al. "Hospital Electronic Health Information Exchange Grew Substantially in 2008–2012." *Health Affairs*, 2013; 32(8): 1346–54.

42. Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, and Steiner J. "Chart Reviews in Emergency Medicine Research: Where Are the Methods?" *Annals of Emergency Medicine*, 1996; 27(3): 305–308.

43. Gladwell M. *The Tipping Point: How Little Things Can Make a Big Difference*. New York: Little Brown, 2000.

44. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, and Baker DW. "Disparities in Enrollment and Use of an Electronic Patient Portal." *Journal of General Internal Medicine*, 2011; 26(10): 1112–16.

45. Gold M, McLaughlin C, Devers K, Berenson B, and Bovbjerg RR. "Obtaining Providers' 'Buy-In' and Establishing Effective Means of Health Information Exchange Will Be Critical to HITECH's Success." *Health Affairs*, 2012; 31(3): 514–26.

46. Goldberg SI, Niemierko A, and Turchin A. "Analysis of Data Errors in Clinical Research Databases." *AMIA Annual Symposium Proceedings*, 2008: 242–46.

47. Grande D, Mitra N, Shah A, Wan F, and Asch D. "A National Survey of Patient Preferences about Secondary Uses of Electronic Health Information." Presented at AcademyHealth annual research meeting, Baltimore, MD, June 25, 2013.

48. Gurney JG, McPheeters ML, and Davis MM. "Parental Report of Health Conditions and Health Care Use among Children with and without Autism: National Survey of Children's Health." *Archives of Pediatrics & Adolescent Medicine*, 2006; 160(8): 825–30.

49. Hamilton J. "Matching DNA with Medical Records to Crack Disease and Aging." NPR, All Things Considered, November 19, 2012. http://www.npr.org/blogs/health/2012/11/19/165498842/matching-dna-with-medical- records-to-crack-disease-and-aging.

50. HealthIT.gov FAQs. Accessed May 24, 2013. http://www.healthit.gov/providers-professionals/faqs/what-information-does-electronic-health-record-ehr-contain.

51. Hendricks DR and Wehman P. "Transition from School to Adulthood for Youth with Autism Spectrum Disorders: Review and Recommendations." *Focus on Autism and Other Developmental Disabilities*, 2009; 24(2): 77–88.

52. Hoeffel EM, Rastogi S, Kim MO, and Shahid H. "The Asian Population: 2010." 2010 Census Brief, 2012. http://www.census.gov/prod/cen2010/briefs/c2010br-11.pdf.

53. Hoffman S, and Podgurski A. "Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research." Social Science Research Network scholarly paper, September 7, 2011. http://papers.ssrn.com/abstract=1923187.

54. Holve E, Segal C, and Lopez MH. "Opportunities and Challenges for Comparative Effectiveness Research (CER) with Electronic Clinical Data." *Medical Care*, 2012; 50 (Suppl): S11–S18.

55. Honeycutt T, and Wittenburg T. "Identifying Transition-Age Youth with Disabilities Using Existing Surveys." Mathematica Policy Research, 2012. http://www.mathematica-mpr.com/publications/PDFs/disability/transition_age_youth_disabilities.pdf.

56. Hornbrook MC, Whitlock EP, Berg CJ, Callaghan WM, Bachman DJ, Gold R, Bruce FC, Dietz PM, and Williams SB. "Development of an Algorithm to Identify Pregnancy Episodes in an Integrated Health Care Delivery System." *Health Services Research*, 2007; 42(2): 908–27.

57. Hornbrook MC, Fishman PA, Ritzwoller DP, Elston-Lafata J, O'Keeffe-Rosetti MC, and Salloum RG. "When Does an Episode of Care for Cancer Begin?" *Medical Care*, 2013; 51(4): 324–29.

58. Hsiao CJ, and Hing E. "Use and Characteristics of Electronic Health Record Systems Among Office-Based Physician Practices: United States, 2011–2012." NCHS Data Brief 111. Hyattsville, MD: National Center for Health Statistics, 2012.

59. Hsiao CJ, Jha AK, King J, Patel V, Furukawa MF, and Mostashari F. "Office-Based Physicians Are Responding to Incentives and Assistance by Adopting and Using Electronic Health Records." *Health Affairs*, July 2013. Epub ahead of print.

60. Ingelfinger JR, and Drazen JM. "Registry Research and Medical Privacy." *New England Journal of Medicine*, 2004; 350(14): 1452–53.

61. Institute of Medicine. "Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records - Workshop Summary - Institute of Medicine." December 20, 2012. http://iom.edu/Reports/2012/Collecting-Sexual-Orientation-and-Gender-Identity-Data-in-Electronic-Health-Records.aspx.

62. Institute of Medicine. "The Health of Lesbian, Gay, Bisexual, and Transgender People." March 31, 2011. http://www.iom.edu/Reports/2011/The-Health-of-Lesbian-Gay-Bisexual-and-Transgender-People.aspx

63. Institute of Medicine. "Key Capabilities of an Electronic Health Record System: Letter Report." July 31, 2003. http://www.iom.edu/Reports/2003/Key-Capabilities-of-an-Electronic-Health-Record-System.aspx.

64. Institute of Medicine. "Knowing What Works in Health Care: A Roadmap for the Nation." Consensus Report, January 24, 2008. http://www.iom.edu/Reports/2008/Knowing-What-Works-in-Health-Care-A-Roadmap-for- the-Nation.aspx.

65. Islam NS, Khan S, Kwon S, Jang D, Ro M, and Trinh-Shevrin C. "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations: Historical Challenges and Potential Solutions." *Journal of Health Care for the Poor and Underserved*, 2010; 21(4): 1354–81.

66. Jain SH, Conway PH, and Berwick DM. "A Public-Private Strategy to Advance the Use of Clinical Registries." *Anesthesiology*, 2012; 117(2): 227–29.

67. Jensen PB, Jensen LJ, and Brunak S. "Mining Electronic Health Records: Towards Better Research Applications and Clinical Care." *Nature Reviews, Genetics*, 2012; 13:295-405.

68. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, Shields A, Rosenbaum S, and Blumenthal D. "Use of Electronic Health Records in U.S. Hospitals." *New England Journal of Medicine*, 2009; 360(16): 1628–38.

69. Jha AK, DesRoches CM, Kralovec PD, and Joshi MS. "A Progress Report on Electronic Health Records in U.S. hospitals." *Health Affairs*, 2010; 29(10): 1951–57.

70. Jones C, Parker T, Ahearn M, Mishra AK, and Variyam J. "Health Status and Health Care Access of Farm and Rural Populations." Economic Information Bulletin EIB-57. USDA Economic Research Service, 2009. http://www.ers.usda.gov/publications/eib-economic-information-bulletin/eib57.aspx#.UeRQ5o2cfts.

71. Kaelber D. "Clinical Research Informatics." Presentation to Case Western Reserve University, 2013.

72. Kaelber DC, Foster W, Gilder J, Lover TE, and Jain AK. "Patient Characteristics Associated with Venous Thromboembolic Events: A Cohort Study Using Pooled Electronic Health Record Data." *Journal of the American Medical Informatics Association,* 2012; 19(6): 965–72.

73. Kahn MG. "Data Model Considerations for Clinical Effectiveness Researchers." *Medical Care*, 2012; 50(7 Supplement 1): S60–S67.

74. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, and Stein JF. "A Pragmatic Framework for Single-Site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research." *Medical Care*, 2012; (50 Suppl): S21–29.

75. Katz N, Andrews R, Zingmond D, and Weiser T. "Statewide Initiatives to Improve Race Ethnicity and Language Data: Three Unique Approaches." Presented at Council of State and Territorial Epidemiologists annual conference, Pasadena, CA, June 10, 2013. https://cste.confex.com/cste/2013/webprogram/Paper1519.html.

76. Kern LM, Malhotra S, Barron Y, Quaresimo J, Dhopeshwarkar R, Pichardo M, Edwards AM, and Kaushal R. "Accuracy of Electronically Reported 'Meaningful Use' Clinical Quality Measures: A Cross-Sectional Study." *Annals of Internal Medicine*, 2013; 158(2): 77–83.

77. Kho ME, Duffett M, Willison DJ, Cook DJ, and Brouwers MC. "Written Informed Consent and Selection Bias in Observational Studies Using Medical Records: Systematic Review." *BMJ*, 2009; 338: b866.

78. Langworthy-Lam KS, Aman MG, and Van Bourgondien ME. "Prevalence and Patterns of Use of Psychoactive Medicines in Individuals with Autism in the Autism Society of North Carolina." *Journal of Child and Adolescent Psychopharmacology*, 2002; 12(4): 311–21.

79. Lohr K, and Steinwachs D. "Health Services Research: An Evolving Definition of the Field." *Health Services Research*, 2002; 37(1):15-17.

80. Luft HS. "Commentary: Protecting Human Subjects and Their Data in Multi-Site Research." *Medical Care*, 2012; 50 Suppl: S74–76.

81. Luft H. "Embedded Research: Doing Research on the Organization within Which You Work." Presented at AcademyHealth annual research meeting, Baltimore, MD, June 2013.

82. Lunshof JD, Chadwick R, Vorhaus DB, and Church GM. "From Genetic Privacy to Open Consent." *Nature Review Genetics*, 2008; 9(5): 406–11.

83. Massachusetts eHealth Institute. "PopMedNet: Distributed Data Network." Accessed August 25, 2013. http://mehi.masstech.org/what-we-do/hie/mdphnet/popmednet.

84. Merrill M. "Pilot Project for Distributed Research Network Will Use EHRs." August 12, 2008. http://www.healthcareitnews.com/news/pilot-project-distributed-research-network-will-use-ehrs.

85. McGinn CA, Grenier S, Duplantie J, Shaw N, Sicotte C, Mathieu L, Yvan L, Légaré F, and Gagnon M. "Comparison of User Groups' Perspectives of Barriers and Facilitators to Implementing Electronic Health Records: A Systematic Review." *BMC Medicine*, 2011; 9(46): 1–10.

86. McGraw D. "Data Governance Challenges and Opportunities in Health Services Research." Presented at AcademyHealth annual research meeting, Baltimore, MD, June 24, 2013.

87. McGraw D, and Leiter A. "A Policy and Technology Framework for Using Clinical Data to Improve Quality." *Houston Journal of Law & Policy,* 2012; 137–67.

88. McGraw D, and Leiter A. "Legal and Policy Challenges to Secondary Uses of Information from Electronic Clinical Health Records." AcademyHealth, 2012.

89. Mearian L. "How Big Data Will Save Your Life." Computer World, April 25, 2013. http://www.computerworld.com/s/article/9238593/How_big_data_will_save_your_life.

90. Miller E. "The National Center for Health Statistics' Linked Data Files: Resources for Research and Policy." Presented at AcademyHealth annual research meeting, Baltimore, MD, June 25, 2013.

91. Moiduddin A, and Stromberg S. "Health Information Technology in California's Rural Practices: Assessing the Benefits and Barriers." Oakland, CA: California Healthcare Foundation, 2009.

92. Multicenter Perioperative Outcomes Group website, accessed August 26, 2013. http://mpog.med.umich.edu/.

93. Murphy J. "ONC Program Update." Presented at NCVHS meeting, June 19, 2013. http://www.ncvhs.hhs.gov/130619p1.pdf.

94. Nakamura MM, Ferris TG, DesRoches CM, and Jha AK. "Electronic Health Record Adoption by Children's Hospitals in the United States." *Archives of Pediatrics and Adolescent Medicine*, 2010; 164(12): 1145–51.

95. Nass SJ, Levit LA, and Gostin LO. *Beyond the HIPAA Privacy Rule.* Washington, DC: National Academies Press, 2009.

96. National Institute of Mental Health. "A Parent's Guide to Autism Spectrum Disorder." 2011. http://www.nimh.nih.gov/health/publications/a-parents-guide-to-autism-spectrum-disorder/complete-index.shtml.

97. Noble S, Donovan J, Turner E, Metcalfe C, Lane A, Rowlands MA, Neal D, Hamdy F, Ben-Shlomo Y, and Martin R. "Feasibility and Cost of Obtaining Informed Consent for Essential Review of Medical Records in Large-Scale Health Services Research." *Journal of Health Services Research & Policy*, 2009; 14(2): 77–81.

98. NORC at the University of Chicago. "Howard University Hospital Diabetes Treatment Center—Using Multi-modal Health IT Tools to Improve Quality and Delivery of Care in an Urban Setting." June 2012,.http://www.healthit.gov/sites/default/files/pdf/HowardCaseStudyReport.pdf.

99. NORC at the University of Chicago. "Patient Care Management and Rewards Program—Promoting and Tracking Wellness Behaviors within the Context of an Existing Case-management Program." June 2012. http://www.healthit.gov/sites/default/files/pdf/AEH_CaseStudyReport.pdf.

100. Obel N, Omland LH, Kronborg G, Larsen CS, Pedersen C, Pedersen G, Sørensen HT, and Gerstoft J. "Impact of Non-HIV and HIV Risk Factors on Survival in HIV-Infected Patients on HAART: A Population-Based Nationwide Cohort Study." *PloS One*, 2011; 6(7).

101. Olsen L, Aisner D, and McGinnis JM, ed. *Roundtable on Evidence-Based Medicine, The Learning Healthcare System: Workshop Summary.* IOM Roundtable on Evidence-Based Medicine. National Academies Press, 2007. http://www.nap.edu/catalog.php?record_id=11903.

102. Pace WD, Cifuentes M, Valuck RJ, Staton EW, Brandt EC, and West DR. "An Electronic Practice-Based Network for Observational Comparative Effectiveness Research." *Annals of Internal Medicine*, 2009; 151(5): 338–40.

103. Palo Alto Medical Foundation. "The Pan Asian Cohort Study." PAMF website, accessed July 15, 2013. http://www.pamf.org/pacs/.

104. Parsons A, McCullough C, Wang J, and Shih S. "Validity of Electronic Health Record–Derived Quality Measurement for Performance Monitoring." *Journal of the American Medical Informatics Association*, 2012; 19(4): 604–609.

105. Patient-Centered Outcomes Research Institute. "Improving Our National Infrastructure to Conduct Comparative Effectiveness Research." PCORI website, accessed July 10, 2013. http://www.pcori.org/funding-opportunities/improving-our-national-infrastructure-to-conduct-comparative-effectiveness-research/.

106. Powell J, and Buchan I. "Electronic Health Records Should Support Clinical Research," Journal of Medical *Internet Research,* 2005; 7(1). doi:10.2196/jmir.7.1.e4.

107. Randhawa GS, and Slutsky JR. "Building Sustainable Multi-functional Prospective Electronic Clinical Data Systems." *Medical Care,* 2012; 50 (Suppl): S3–6.

108. Regenstrief Institute. "Regenstrief Institute Data Core." Regenstrief Institute website, accessed July 15, 2013. http://www.regenstrief.org/centers/research-resources/data-core/

109. Reisch, LM, Fosse JS, Beverly K, Yu O, Barlow WE, Harris EL, Rolnick S, Barton MB, Geiger AM, Herrington LJ, Greene SM, Gletcher SW, and Elmore JG. "Training, Quality Assurance, and Assessment of Medical Record Abstraction in a Multisite Study." *American Journal of Epidemiology*, 2003; 157(6): 546–51.

110. Rosenbaum S. "Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access." *Health Services Research*, 2010; 45(5 Pt 2): 1442–55.

111. Sabharwal R, Holve E, Rein A, and Segal C. "Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy." Issue brief. 2012. http://repository.academyhealth.org/edm_briefs/1.

112. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, and Detmer DE. "Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper." *Journal of the American Medical Informatics Association: JAMIA*, 2007; 14(1): 1–9.

113. Segal C, and Holve E. "Emerging Data Resources, Tools, and Publications from the ARRA-CER Infrastructure Awards." Presented at AcademyHealth annual research meeting, Baltimore, MD, June 2013.

114. Selker H, Grossman C, Adams A, et al. "The Common Rule and Continuous Improvement in Health Care: A Learning Health System Perspective." Discussion Paper. Institute of Medicine, 2011. http://www.iom.edu/Global/Perspectives/2012/CommonRule.aspx.

115. Shortliffe EH, and Barnett GO. "Biomedical Data: Their Acquisition, Storage, and Use." In *Biomedical Informatics*, edited by Shortliffe EH and Cimino JJ. Health Informatics. Springer New York, 2006. http://link.springer.com/chapter/10.1007/0-387-36278-9_2.

116. Snyder C. "Considerations for Using Patient-Reported Outcomes in Clinical Practice: A Case Study." Presented at AcademyHealth Annual Research Meeting, Baltimore, MD, June 2013.

117. Stanfill MH, Williams M, Fenton SH, Jenders RA, and Hersh WR. "A Systematic Literature Review of Automated Clinical Coding and Classification Systems." *Journal of the American Medical Informatics Association: JAMIA*, 2010; 17(6): 646–51.

118. Steiner C. "The Healthcare Cost and Utilization Project (HCUP): Linked Data Enhancements and Improved Analytic Capacity." April 10, 2013.

119. Trinidad SB, Fullerton SM, Ludman EJ, Jarvik GP, Larson EB, and Burke W. "Research Ethics. Research Practice and Participant Preferences: The Growing Gulf." *Science*, 2011; 331(6015): 287–88.

120. University of California, Davis, Health System. No date. "Patient Questions for Demographics."

121. U.S. Department of Agriculture, Stronger Economies Together (SET) program. USDA website, 2012.

122. U.S. Food and Drug Administration. "Should Your Child be in a Clinical Trial?" Accessed August 25, 2013. http://www.fda.gov/forconsumers/consumerupdates/ucm048699.htm.

123. Vayena E, Mastroianni A, and Kahn J. "Caught in the Web: Informed Consent for Online Health Research." *Science Translational Medicine*, 2013; 5(173): 173fs6.

124. Waidmann TA, Ormond BA, and Spillman BC. Potential Savings through Prevention of Avoidable Chronic Illness among CalPERS State Active Members. Urban Institute, 2012. http://www.urban.org/publications/412550.html.

125. Webster PS, and Sampangi S. "Report on Data Improvement Pilot on Patient Ethnicity and Race (DIPPER): Pilot Design and Proposed Voluntary Standard." Rhode Island Medical Journal, 2013; January.

126. Weissberg J. "Use of Large System Databases: Cox-2 Inhibitors." Presentation at The Learning Healthcare System, Institute of Medicine—Roundtable on EBM, May, 2006. http://www.iom.edu/~/media/Files/Activity%20Files/Quality/VSRT/S1bWeissbergReadOnly.pdf.

Citations

[1] See for example Kahn, MG, et al on data quality using a "fitness for use" concept and definition. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, and Stein JF. "A Pragmatic Framework for Single-Site and Multisite Data Quality Assessment in Electronic Health Record-Based Clinical Research." *Medical Care*, 2012; 50 Suppl: S21–29.

[2] Gladwell M. *The Tipping Point: How Little Things Can Make a Big Difference*. New York: Little Brown, 2000.

[3] Interview with Wasserman and Fiks.

[4] Interviews with West, Schilling, and Glasgow.

[5] Interview with Walker.

[6] Interview with Walker.

[7] Chan KS, Fowles JB, and Weiner JP. "Review: Electronic Health Records and the Reliability and Validity of Quality Measures: a Review of the Literature." Medical Care Research and Review, 2010: 67(5): 503–527.

[8] Interview with Hornbrook.

[9] Interviews with Hornbrook and Califf.

[10] Sabharwal R, Holve E, Rein A, and Segal C. "Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy." Issue briefs and reports, March 1, 2012, http://repository.academyhealth.org/edm_briefs/1.

[11] "Improving Our National Infrastructure to Conduct Comparative Effectiveness Research." PCORI website, accessed July 10, 2013. http://www.pcori.org/funding-opportunities/improving-our-national-infrastructure-to-conduct-comparative-effectiveness-research/

[12] Kahn MG, Raebel MA, Glanz JM, Riedlinger K, and Stein JF. "A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research." Medical Care, 2012; 50 Suppl: S21–29.

[13] Interview with Califf.

[14] Bradley CJ, Penberthy L, Devers KJ, and Holden DJ. "Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future." Health Services Research, 2010; 45(5p2): 1468–1488.

[15] Moiduddin A and Moore J. "The Underserved and Health Information Technology: Issues and Opportunities." Paper prepared for the Office of the Assistant Secretary for Planning and Evaluation and U.S. Department of Health and Human Services, November 2008. Available at http://aspe.hhs.gov/sp/reports/2009/underserved/report.pdf.

[16] Lavrakas PJ, and Sage Publications, *Encyclopedia of Survey Research Methods* (Thousand Oaks, Calif.: SAGE Publications, 2008), http://www.credoreference.com/book/sagesurveyr.

[17] Islam et al., "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations."

[18] Groves RM, Fowler Jr FJ, Couper MP, Lepkowski JM, Singer E, and Tourangeau R. Survey methodology. Vol. 561. Wiley, 2009.

[19] Kempf AM and Remington PL, "New Challenges for Telephone Survey Research in the Twenty-first Century," *Annual Review of Public Health* 28 (2007): 113–126. doi:10.1146/annurev.publhealth.28.021406.144059.

[20] Interviews with Huang and Mueller.

[21] Interview with Lotstein.

[22] Interviews with Trinh-Shevrin and Palaniappan.

[23] Interviews with Landers and Gates.

[24] Interviews with Hartley and Ziller, Trinh-Shevrin, Huang, Snowdon, Landers, and Gates.

[25] Interviews with Lounds, Taylor, and Okumura.

[26] Hoeffel EH et al., *The Asian Population: 2010*, 2010 Census Briefs, March 2012, http://www.census.gov/prod/cen2010/briefs/c2010br-11.pdf.

[27] Islam NS et al., "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations: Historical Challenges and Potential Solutions," *Journal of Health Care for the Poor and Underserved* 21, no. 4 (2010): 1354–81. doi:10.1353/hpu.2010.0939.

[28] Hoeffel et al., *The Asian Population: 2010.*

[29] "Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care - Institute of Medicine," accessed February 15, 2013, http://www.iom.edu/Reports/2002/Unequal-Treatment-Confronting-Racial-and-Ethnic-Disparities-in-Health-Care.aspx.

[30] Islam et al., "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations."

[31] Chen Jr MS and Hawks BL, "A Debunking of the Myth of Healthy Asian Americans and Pacific Islanders," *American Journal of Health Promotion: AJHP* 9, no. 4 (April 1995): 261–268.

[32] http://www.pamf.org/pacs/. A similar pattern can be seen among women.

[33] Wang EJ, Wong EC, Dixit AA, Fortmann SP, Linde RB, and Palaniappan LP. 2011. "Type 2 Diabetes: Identifying High Risk Asian American Subgroups in a Clinical Population." *Diabetes Research and Clinical Practice* 93(2): 248–54. doi:10.1016/j.diabres.2011.05.025.

[34] Clough J, Lee S, and Chae DH, "Barriers to Health Care among Asian Immigrants in the United States: A Traditional Review." *Journal of Health Care for the Poor and Underserved* 24, no. 1 (2013): 384–403. doi:10.1353/hpu.2013.0019.

[35] U.S. Census Bureau, American Community Survey (U.S. Census Bureau, 2006–2007).

[36] Interview with Huang.

[37] Interviews with Ro, Palaniappan, and Huang.

[38] Chen W, "Chinese Female Immigrants English-speaking Ability and Breast and Cervical Cancer Early Detection Practices in the New York Metropolitan Area," *Asian Pacific Journal of Cancer Prevention: APJC*P 14, no. 2 (2013): 733–38.

[39] U.S. Census Bureau, *The Vietnamese Population in the United States*: 2010, 2011, http://www.bpsos.org/mainsite/images/DelawareValley/community_profile/us.census.2010.the%20vietnamese%20population_july%202.2011.pdf.

[40] McCracken M et al., "Cancer Incidence, Mortality, and Associated Risk Factors among Asian Americans of Chinese, Filipino, Vietnamese, Korean, and Japanese Ethnicities." *CA: A Cancer Journal for Clinicians* 57, no. 4 (August 2007): 190–205.

[41] D Nguyen, "Culture Shock—A Review of Vietnamese Culture and Its Concepts of Health and Disease." *Western Journal of Medicine* 142, no. 3 (1985): 409–12.

[42] Appel HB et al. "Physical, Behavioral, and Mental Health Issues in Asian American Women: Results from the National Latino Asian American Study." *Journal of Women's Health* 20, no. 11 (November 2011): 1703–11. doi:10.1089/jwh.2010.2726.

[43] American Cancer Society, "Cancer Facts & Figures 2013." Accessed February 28, 2013. http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-036845.pdf.

[44] Ma, GX et al. "Correlates of Cervical Cancer Screening among Vietnamese American Women." *Infectious Diseases in Obstetrics and Gynecology,* 2012: 617234. doi:10.1155/2012/617234.

[45] Ibid.

[46] Ho, IK and Dinh KT, "Cervical Cancer Screening Among Southeast Asian American Women." *Journal of Immigrant and Minority Health/Center for Minority Public Health* 13, no. 1 (2011): 49–60. doi:10.1007/s10903-010-9358-0.

[47] Ma et al., "Correlates of Cervical Cancer Screening Among Vietnamese American Women."

[48] Gregg J, et al. "Prioritizing Prevention: Culture, Context, and Cervical Cancer Screening Among Vietnamese American Women." *Journal of Immigrant and Minority Health/Center for Minority Public Health* 13, no. 6 (2011): 1084–89. doi:10.1007/s10903-011-9493-2.

[49] Sentell T and Braun KL. "Low Health Literacy, Limited English Proficiency, and Health Status in Asians, Latinos, and Other Racial/Ethnic Groups in California," *Journal of Health Communication* 17 Suppl 3 (2012): 82–99. doi:10.1080/10810730.2012.712621.

[50] Sterngass J, *Filipino Americans* (Infobase Publishing, 2009).

[51] U.S. Census Bureau, "2011 American Community Survey." Accessed February 28, 2013. http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_11_1YR_B02015&prodType=table.

[52] Shin HB and Kominski RA, Language Use in the United States: 2007, American Community Survey Reports, April 2010, http://www.census.gov/prod/2010pubs/acs-12.pdf.

[53] Wang EJ et al., "Type 2 Diabetes: Identifying High Risk Asian American Subgroups in a Clinical Population." *Diabetes Research and Clinical Practice* 93, no. 2 (2011): 248–54. doi:10.1016/j.diabres.2011.05.025.

[54] Holland AT et al., "Spectrum of Cardiovascular Diseases in Asian-American Racial/ethnic Subgroups," *Annals of Epidemiology* 21, no. 8 (2011): 608–14. doi:10.1016/j.annepidem.2011.04.004.

[55] Ibid.

[56] Appel et al., "Physical, Behavioral, and Mental Health Issues in Asian American Women."

[57] Semics, LLC. "Culture and health among Filipinos and Filipino-Americans in Central Los Angeles." 2007. http://www.calendow.org/uploadedFiles/Publications/By_Topic/Disparities/General/Culture%20Health%20Among%20Filipinos.pdf.

[58] "Charter of the Interagency Council on Statistical Policy Subcommittee on the American Community Survey," August 10, 2012.

[59] Islam et al., "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations."

[60] Interviews with Huang and Trinh-Shevrin.

[61] Interview with Trinh-Shevrin.

[62] Islam et al., "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data."

[63] Appel et al., "Physical, Behavioral, and Mental Health Issues in Asian American Women."

[64] Huang B et al., "Chronic Conditions, Behavioral Health, and Use of Health Services among Asian American Men: The First Nationally Representative Sample," *American Journal of Men's Health 7*, no. 1 (January 2013): 66–76. doi:10.1177/1557988312460885.

[65] Islam et al., "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data."

[66] Holland et al., "Spectrum of Cardiovascular Diseases in Asian-American Racial/ethnic Subgroups."

[67] Wang et al., "Type 2 Diabetes."

[68] See the HHS Data Council website for details: http://aspe.hhs.gov/datacncl/

[69] Centers for Medicare and Medicaid. "Medicare and Medicaid EHR Incentive Program: Meaningful Use, Stage 1 Requirements Overview." 2010. http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/MU_Stage1_ReqOverview.pdf

[70] Hasnain-Wynia R, Pierce D, and Pittman MA, *Who, When, and How: The Current State of Race, Ethnicity, and Primary Language Data Collection in Hospitals* - The Commonwealth Fund, May 18, 2004, http://www.commonwealthfund.org/Publications/Fund-Reports/2004/May/Who--When--and-How--The-Current-State-of-Race--Ethnicity--and-Primary-Language-Data-Collection-in-Ho.aspx#citation.

[71] Hasnain-Wynia R and Baker DW, "Obtaining Data on Patient Race, Ethnicity, and Primary Language in Health Care Organizations: Current Challenges and Proposed Solutions," *Health Services Research* 41, no. 4 Pt 1 (2006): 1501–18. doi:10.1111/j.1475-6773.2006.00552.x.

[72] IOM Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement Board on Health Care Services, Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement, accessed February 28, 2013, http://www.ahrq.gov/research/iomracereport/.

[73] McBean MA. "Improving Medicare's Data on Race and Ethnicity." *Medicare Brief / National Academy of Social Insurance* no. 15 (October 2006): 1–7.

[74] Eicheldinger C and Bonito AJ. "More Accurate Racial and Ethnic Codes for Medicare Administrative Data." *Health Care Financing Review* 29, no. 3 (2008): 27–42.

[75] NORC, *Understanding the Impact of Health IT in Underserved Communities and Those with Health Disparities*, Briefing Paper, October 29, 2010, http://www.healthit.gov/sites/default/files/pdf/hit-underserved-communities-health-disparities.pdf.

[76] Interview with Huang.

[77] IOM Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement Board on Health Care Services.

[78] Bonito AJ et al., *Creation of New Race-Ethnicity Codes and Socioeconomic Status (SES) Indicators for Medicare Beneficiaries* (Agency for Healthcare Research and Quality, January 2008), http://www.ahrq.gov/qual/medicareindicators/.

[79] Kressin NR et al., "Agreement between Administrative Data and Patients' Self-reports of Race/ethnicity," *American Journal of Public Health* 93, no. 10 (October 2003): 1734–1739.

[80] Bonito et al., *Creation of New Race-Ethnicity Codes and Socioeconomic Status (SES) Indicators for Medicare Beneficiaries.*

[81] Interviews with Trinh-Shevrin and Palaniappan.

[82] Institute of Medicine, *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding - Institute of Medicine* (National Academies Press, 2011), http://www.iom.edu/Reports/2011/The-Health-of-Lesbian-Gay-Bisexual-and-Transgender-People.aspx.

[83] Mayer KH et al. "Sexual and Gender Minority Health: What We Know and What Needs to Be Done." *American Journal of Public Health* 98, no. 6 (2008): 989–95. doi:10.2105/AJPH.2007.127811.

[84] Gates, Gary J., *How Many People Are Lesbian, Gay, Bisexual, and Transgender?* The Williams Institute, UCLA School of Law, April 2011.

[85] Institute of Medicine, *The Health of Lesbian, Gay, Bisexual, and Transgender People.*

[86] Clift JB and Kirby J, "Health Care Access and Perceptions of Provider Care among Individuals in Same-Sex Couples: Findings from the Medical Expenditure Panel Survey (MEPS)." *Journal of Homosexuality* 59, no. 6 (2012): 839–50. doi:10.1080/00918369.2012.694766.

[87] Interviews with Gates and Snowdon.

[88] Institute of Medicine, *The Health of Lesbian, Gay, Bisexual, and Transgender People,* pp. 6–8

[89] Ibid., pp. 9–10.

[90] Ibid., p. 61067.

[91] Bird ST and Bogart LM, "Perceived Race-Based and Socioeconomic Status (SES)-Based Discrimination in Interactions with Health Care Providers," *Ethnicity & Disease* 11, no. 3 (2001): 554–63.

[92] Technical Expert Meeting discussion, July 24, 2013.

[93] Technical Expert Meeting discussion, July 24, 2013.

[94] Diamond, LM. "What we got wrong about sexual identity development: Unexpected findings from a longitudinal study of young women." *Sexual orientation and mental health: Examining identity and development in lesbian, gay, and bisexual people* (2005): 73–94.

[95] Technical Expert Meeting discussion, July 24, 2013.

[96] For example, see Ponce NA et al., "The Effects of Unequal Access to Health Insurance for Same-Sex Couples in California," *Health Affairs* (Project Hope) 29, no. 8 (August 2010): 1539–48. doi:10.1377/hlthaff.2009.0583.

[97] Clift and Kirby, "Health Care Access and Perceptions of Provider Care among Individuals in Same-sex Couples."

[98] Technical Expert Meeting discussion (July 24, 2013)

[99] The basic question is "Do you think of yourself as lesbian or gay; straight, that is, not gay; bi sexual; something else; don't know. Follow up questions probe the meaning of the last two responses. "Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records - Workshop Summary - Institute of Medicine," accessed February 28, 2013, http://iom.edu/Reports/2012/Collecting-Sexual-Orientation-and-Gender-Identity-Data-in-Electronic-Health-Records.aspx. Testimony of Kristen Miller, p. 32

[100] Technical Expert Meeting discussion, July 24, 2013.

[101] "Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records - Workshop Summary - Institute of Medicine," accessed February 28, 2013, http://iom.edu/Reports/2012/Collecting-Sexual-Orientation-and-Gender-Identity-Data-in-Electronic-Health-Records.aspx. Testimony of Kristen Miller, p. 32.

[102] Interview with Snowdon.

[103] Interview with Landers.

[104] Ehrenfeld JM, "Identification of LGBT Patients & Health Disparities Using Electronic Health Records," October 12, 2012, http://www.iom.edu/~/media/Files/Activity%20Files/SelectPops/LGBTdata/Ehrenfeld%20presentation.pdf.

[105] Technical Expert Meeting discussion, July 24, 2013.

[106] National Institute of Mental Health, A Parent's Guide to Autism Spectrum Disorder, 2011, http://www.nimh.nih.gov/health/publications/a-parents-guide-to-autism-spectrum-disorder/complete-index.shtml.

[107] Hendricks DR and Wehman P, "Transition From School to Adulthood for Youth With Autism Spectrum Disorders Review and Recommendations," *Focus on Autism and Other Developmental Disabilities* 24, no. 2 (2009): 77–88. doi:10.1177/1088357608329827.

[108] National Institute of Mental Health, *A Parent's Guide to Autism Spectrum Disorder.*

[109] Centers for Disease Control and Prevention. Prevalence of Autism Spectrum Disorders—Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. MMWR. March 30, 2012/61(SS03);1-19, CDC.

[110] Technical Expert Meeting discussion, July 24, 2013.

[111] "Understanding Rett Syndrome - About Rett Syndrome - International Rett Syndrome Foundation," International Rett Syndrome Foundation, accessed February 28, 2013, http://www.rettsyndrome.org/understanding-rett-syndrome/about-rett-syndrome.

[112] "Autism Society - Asperger's Syndrome," *Autism Society*, accessed February 28, 2013, http://www.autism-society.org/about-autism/aspergers-syndrome/.

[113] Kaufmann WE, "DSM-5: The New Diagnostic Criteria for Autism Spectrum Disorders" (presented at the 2012 Research Symposium - Autism Consortium, Boston, MA, October 24, 2012), http://www.autismconsortium.org/symposium-files/WalterKaufmannAC2012Symposium.pdf.

[114] Mrozek-Budzyn D, Kiełtyka A, and Majewska R, "Lack of Association between Measles-mumps-rubella Vaccination and Autism in Children: a Case-control Study," *The Pediatric Infectious Disease Journal* 29, no. 5 (2010): 397–400. doi:10.1097/INF.0b013e3181c40a8a.

[115] DeStefano F, et al., "Age at First Measles-Mumps-Rubella Vaccination in Children with Autism and School-Matched Control Subjects: A Population-Based Study in Metropolitan Atlanta," *Pediatrics* 113, no. 2 (February 1, 2004): 259–66. doi:10.1542/peds.113.2.259.

[116] Price CS et al., "Prenatal and Infant Exposure to Thimerosal from Vaccines and Immunoglobulins and Risk of Autism," *Pediatrics* 126, no. 4 (October 2010): 656–64. doi:10.1542/peds.2010-0309.

[117] Institute of Medicine, *Immunization Safety Review: Vaccines and Autism* (Washington, D.C: National Academies Press, 2004).

[118] Boyle CA and Boulet SL, "Health Care Use and Health and Functional Impact of Developmental Disabilities Among Us Children, 1997-2005," *Archives of Pediatrics & Adolescent Medicine* 163, no. 1 (2009): 19–26. doi:10.1001/archpediatrics.2008.506.

[119] Gurney JG, McPheeters ML, and Davis MM, "Parental Report of Health Conditions and Health Care Use among Children with and Without Autism: National Survey of Children's Health," *Archives of Pediatrics & Adolescent Medicine* 160, no. 8 (2006): 825–30. doi:10.1001/archpedi.160.8.825.

[120] National Institute of Mental Health, *A Parent's Guide to Autism Spectrum Disorder.*

[121] Gurney, McPheeters, and Davis, "Parental Report of Health Conditions and Health Care Use Among Children with and Without Autism."

[122] Kohane IS et al., "The Co-morbidity Burden of Children and Young Adults with Autism Spectrum Disorders," *PloS One* 7, no. 4 (2012): e33224. doi:10.1371/journal.pone.0033224.

[123] Gurney, McPheeters, and Davis, "Parental Report of Health Conditions and Health Care Use Among Children with and Without Autism."

[124] Croen LA et al., "A Comparison of Health Care Utilization and Costs of Children with and Without Autism Spectrum Disorders in a Large Group-model Health Plan," *Pediatrics* 118, no. 4 (October 2006): e1203–11. doi:10.1542/peds.2006-0127.

[125] Langworthy-Lam KS, Aman MG, and Van Bourgondien ME, "Prevalence and Patterns of Use of Psychoactive Medicines in Individuals with Autism in the Autism Society of North Carolina," *Journal of Child and Adolescent Psychopharmacology* 12, no. 4 (2002): 311–21. doi:10.1089/104454602762599853.

[126] Kogan MD et al., "A National Profile of the Health Care Experiences and Family Impact of Autism Spectrum Disorder Among Children in the United States, 2005-2006," *Pediatrics* 122, no. 6 (December 2008): e1149–58. doi:10.1542/peds.2008-1057.

[127] Kogan MD et al., "Prevalence of Parent-reported Diagnosis of Autism Spectrum Disorder Among Children in the US, 2007," *Pediatrics* 124, no. 5 (2009): 1395–1403. doi:10.1542/peds.2009-1522.

[128] Eaves LC and Ho HH, "Young Adult Outcome of Autism Spectrum Disorders," *Journal of Autism and Developmental Disorders* 38, no. 4 (2008): 739–47. doi:10.1007/s10803-007-0441-x.

[129] Cooley WC and Sagerman PJ, "Supporting the Health Care Transition from Adolescence to Adulthood in the Medical Home," *Pediatrics* 128, no. 1 (2011): 182–200. doi:10.1542/peds.2011-0969.

[130] Honeycutt T and Wittenburg T, *Identifying Transition-Age Youth with Disabilities Using Existing Surveys* (Mathematica Policy Research, July 10, 2012), http://www.mathematica-mpr.com/publications/PDFs/disability/transition_age_youth_disabilities.pdf.

[131] Hendricks and Wehman, "Transition From School to Adulthood for Youth With Autism Spectrum Disorders."

[132] Interviews with Lotstein and Lounds Taylor.

[133] Prichard L et al., Transitioning Teens with Autism Spectrum Disorders, Guide (Autism Consortium), accessed February 28, 2013, http://www.autismconsortium.org/attachments/Autism_Consortium_Reference_Guide_FINAL.pdf.

[134] Billstedt E, Gillberg IC, and Gillberg C, "Aspects of Quality of Life in Adults Diagnosed with Autism in Childhood: a Population-based Study," *Autism: The International Journal of Research and Practice* 15, no. 1 (2011): 7–20. doi:10.1177/1362361309346066.

[135] Prichard et al., *Transitioning Teens with Autism Spectrum Disorders*.

[136] More attention has been paid to the obstacles faced by youth with ASDs as they transition to college, work, and independent living. There is a body of literature around how schools should help students with ASD make this transition and federal measures in place to make sure schools plan for this transition because many students with ASDs receive special education services. (Hendricks and Wehman, "Transition From School to Adulthood for Youth With Autism Spectrum Disorders.")

[137] Cheak-Zamora NC, et al., "Disparities in Transition Planning for Youth with Autism Spectrum Disorder." *Pediatrics* (February 11, 2013). doi:10.1542/peds.2012-1572.

[138] Lotstein DS et al., "Transition Planning for Youth With Special Health Care Needs: Results From the National Survey of Children With Special Health Care Needs," *Pediatrics* 115, no. 6 (June 1, 2005): 1562–68. doi:10.1542/peds.2004-1262.

[139] Carbone PS et al., "The Medical Home for Children with Autism Spectrum Disorders: Parent and Pediatrician Perspectives," *Journal of Autism and Developmental Disorders* 40, no. 3 (2010): 317–24. doi:10.1007/s10803-009-0874-5.

[140] Livermore G et al., Disability Data in National Surveys, Office of the Assistant Secretary for Planning and Evaluation (Mathematica Policy Research: Office of the Assistant Secretary for Planning and Evaluation, August 25, 2011), http://www.aspe.hhs.gov/daltcp/reports/2011/DDNatlSur.shtml.

[141] Ibid.

[142] Technical Expert Meeting discussion, July 24, 2013.

[143] Technical Expert Meeting discussion, July 24, 2013.

[144] Technical Expert Meeting discussion, July 24, 2013.

[145] Centers for Disease Control, Prevalence of Autism Spectrum Disorders—Autism and Developmental Disabilities Monitoring Network, United States, 2006, Surveillance Summaries, December 18, 2009. http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5810a1.htm.

[146] Reijneveld SA et al., "Psychosocial Problems among Immigrant and Non-immigrant Children--Ethnicity Plays a Role in Their Occurrence and Identification," *European Child & Adolescent Psychiatry* 14, no. 3 (2005): 145–52. doi:10.1007/s00787-005-0454-y.

[147] Mandell DS and Novak M, "The Role of Culture in Families' Treatment Decisions for Children with Autism Spectrum Disorders," Mental Retardation and Developmental Disabilities Research Reviews 11, no. 2 (2005): 110–15. doi:10.1002/mrdd.20061.

[148] Begeer S et al., "Underdiagnosis and Referral Bias of Autism in Ethnic Minorities," *Journal of Autism and Developmental Disorders* 39, no. 1 (2009): 142–48. doi:10.1007/s10803-008-0611-5.

[149] Honeycutt and Wittenburg, *Identifying Transition-Age Youth with Disabilities Using Existing Surveys.*

[150] Cromartie J and Bucholtz S, "Defining the 'Rural' in Rural America," Amber Waves, June 2008, http://webarchives.cdlib.org/wayback.public/UERS_ag_1/20111129061030/http://ers.usda.gov/AmberWaves/June08/Features/RuralAmerica.htm.

[151] Crosby RA et al., *Rural Populations and Health: Determinants, Disparities, and Solutions* (John Wiley & Sons, 2012).

[152] Hart LG, Larson EH, and Lishner DM, "Rural Definitions for Health Policy and Research," *American Journal of Public Health* 95, no. 7 (2005): 1149–55. doi:10.2105/AJPH.2004.042432.

[153] Jones C et al., "Health Status and Health Care Access of Farm and Rural Populations," Economic Information Bulletin (USDA Economic Research Service, August 2009). http://www.ers.usda.gov/publications/eib-economic-information-bulletin/eib57.aspx.

[154] "USDA Economic Research Service - State Data," February 26, 2013, http://www.ers.usda.gov/data-products/state-fact-sheets/state-data.aspx?StateFIPS=00.

[155] Committee on The Future of Rural Health Care, *Quality through Collaboration: The Future of Rural Health Care* (Washington, DC: The National Academies Press, 2005).

[156] "USDA Economic Research Service - Definitions of Food Security," accessed May 6, 2013, http://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/definitions-of-food-security.aspx#.UYf_obWcfts.

[157] Halverson J et al., *Patterns of Food Insecurity, Food Availability, and Health Outcomes among Rural and Urban Counties* (West Virginia Rural Health Research Center, 2011). http://ask.hrsa.gov/detail_materials.cfm?ProdID=4700&ReferringID=4628.

[158] Bennett, KJ, Olatsi B, and Probst J, *Health Disparities: A Rural-Urban Chartbook* (South Carolina Rural Health Research Center, June 2008). http://rhr.sph.sc.edu/report/%287-3%29%20Health%20Disparities%20A%20Rural%20Urban%20Chartbook%20-%20Distribution%20Copy.pdf.

[159] Jones et al., *Health Status and Health Care Access of Farm and Rural Populations.*

[160] Crosby et al., *Rural Populations and Health.*

[161] Health Resources and Services Administration, *Mental Health and Rural America: 1994-2005* (Office of Rural Health Policy, 2005), ftp://ftp.hrsa.gov/ruralhealth/RuralMentalHealth.pdf.

[162] Ibid.

[163] Lambert D, Gale JA, and Hartley D, "Substance Abuse by Youth and Young Adults in Rural America," *Journal of Rural Health: Official Journal of the American Rural Health Association and the National Rural Health Care Association* 24, no. 3 (2008): 221–28. doi:10.1111/j.1748-0361.2008.00162.x.

[164] Grant KM et al., "Methamphetamine Use in Rural Midwesterners," *American Journal on Addictions/ American Academy of Psychiatrists in Alcoholism and Addictions* 16, no. 2 (2007): 79–84. doi:10.1080/10550490601184159.

[165] Roundtable on Environmental Health Sciences, Research, and Medicine, Institute of Medicine, *Rebuilding the Unity of Health and the Environment in Rural America: Workshop Summary* (Washington, DC: National Academies Press, 2006).

[166] Hendryx M, Fedorko E, and Halverson J, "Pollution Sources and Mortality Rates Across Rural-urban Areas in the United States," *The Journal of Rural Health: Official Journal of the American Rural Health Association and the National Rural Health Care Association* 26, no. 4 (2010): 383–91. doi:10.1111/j.1748-0361.2010.00305.x.

[167] Persily CA,. Beane JS, and Rice MG, "Environmental Workforce Characteristics in the Rural Public Health Sector." Policy Brief (West Virginia Rural Health Research Center, December 2011). http://publichealth.hsc.wvu.edu/wvrhrc/docs/2010_persily_policy_brief.pdf.

[168] Jones et al., *Health Status and Health Care Access of Farm and Rural Populations.*

[169] Roundtable on Environmental Health Sciences, Research, and Medicine, and Institute of Medicine, *Rebuilding the Unity of Health and the Environment in Rural America.*

[170] Ibid.

[171] Ziller EC et al., *Health Insurance Coverage in Rural America*, Chartbook (Kaiser Commission on Medicaid and the Uninsured, September 2003). http://www.kff.org/uninsured/upload/Health-Insurance-Coverage-in-Rural-America-PDF.pdf.

[172] Bensley L, "Results of BRFSS Analysis," October 2012.

[173] Brock Martin A et al., *Rural Border Health Chartbook.* (South Carolina Rural Health Research Center, January 2013). http://rhr.sph.sc.edu/report/SCRHRC%20Rural%20Border%20Health.pdf.

[174] Roundtable on Environmental Health Sciences, Research, and Medicine, and Institute of Medicine, *Rebuilding the Unity of Health and the Environment in Rural America.*

[175] Health Resources and Services Administration, "Mental Health and Rural America: 1994–2005."

[176] Hart, Larson, and Lishner, "Rural Definitions for Health Policy and Research."

[177] Technical Expert Meeting discussion, July 24, 2013.

[178] Committee on The Future of Rural Health Care, *Quality through Collaboration*.

[179] McEllistrem-Evenson A, *Informing Rural Primary Care Workforce Policy: What Does the Evidence Tell Us? A Review of Rural Health Research Center Literature, 2000–2010* (Rural Health Research Gateway, April 2011). http://www.ruralcenter.org/minnesota-web-recruitment/resources/informing-rural-primary-care-workforce-policy-what-does-evidence.

[180] Mikacevich S and Stensland J, Serving Rural Medicare Beneficiaries (MedPac, June 2012). http://medpac.gov/chapters/Jun12_Ch05.pdf.

[181] Ibid.

[182] NORC, Understanding the Impact of Health IT in Underserved Communities and Those with Health Disparities. CCHC Case Study.

[183] Ibid. Tele-psychiatry case study.

[184] DesRoches CM et al., "Small, Nonteaching, and Rural Hospitals Continue to Be Slow in Adopting Electronic Health Record Systems," *Health Affairs* 31, no. 5 (2012): 1092–99. doi:10.1377/hlthaff.2012.0153.

[185] Colias M, "Rural Areas Still Not Wired for Digital Health Care," H&HN, September 2012, http://www.hhnmag.com/hhnmag/jsp/articledisplay.jsp?dcrpath=HHNMAG/Article/data/09SEP2012/0912HHN_Inbox_Telemedicine&domain=HHNMAG.

[186] "FCC Chairman Announces Up to $400 Million Healthcare Connect Fund to Create & Expand Telemedicine Networks, Increase Access to Medical Specialists, FCC Will Begin Accepting Applications for the Healthcare Connect Fund Beginning Late Summer of 2013," press release, accessed February 28, 2013, http://www.fcc.gov/document/fcc-chairman-announces-400-million-healthcare-connect-fund.

[187] CMS, *Payment Adjustment and Hardship Exceptions Tipsheet for Eligible Hospitals and CAHs*. August 2012. http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/PaymentAdj_HardshipExcepTipsheetforHospitals.pdf.

[188] DesRoches et al., "Small, Nonteaching, and Rural Hospitals Continue to Be Slow."

[189] National Rural Health Association, *Electronic Health Record Implementation and Meaningful Use Adoption in Rural Hospitals and Physician Clinics*. Policy Brief. 2012. http://www.ruralhealthweb.org/go/left/policy-and-advocacy/policy-documents-and-statements/issue-papers-and-policy-briefs/.

[190] Morgan PA et al., "Missing in Action: Care by Physician Assistants and Nurse Practitioners in National Health Surveys," *Health Services Research* 42, no. 5 (2007): 2022–37. doi:10.1111/j.1475-6773.2007.00700.x.

[191] Grumbach K et al., "The Challenge of Defining and Counting Generalist Physicians: An Analysis of Physician Masterfile Data." *American Journal of Public Health* 85, no. 10 (1995): 1402–07.

[192] Randolph R et al., "Designating Places and Populations as Medically Underserved: A Proposal for a New Approach." *Journal of Health Care for the Poor and Underserved* 18, no. 3 (2007): 575. doi:10.1353/hpu.2007.0065.

[193] Trude L, "Standardizing the Patchwork of Data on the U.S. Health Workforce - Health Workforce News," Health Workforce News, April 12, 2011. http://www.hwic.org/newsletter/2011/04/standardizing-health-workforce-data/.

[194] For example, see Hendryx, Fedorko, and Halverson, "Pollution Sources and Mortality Rates Across Rural-urban Areas in the United States"; Lambert, Gale, and Hartley, "Substance Abuse by Youth and Young Adults in Rural America."

[195] See NCHS Research Data Center site for more details: http://www.cdc.gov/rdc/B2AccessMod/Acs230.htm.

[196] "Refugees in Iowa Introduce Unprecedented Language Barriers in Rural Communities," *Friends of Refugees*, accessed March 26, 2013, http://forefugees.com/2012/08/05/refugees-in-iowa-introduce-unprecedented-language-barriers-in-rural-communities/.

[197] Hart, Larson, and Lishner, "Rural Definitions for Health Policy and Research."

[198] Cromartie and Bucholtz, "Defining the 'Rural' in Rural America."

[199] Ibid.

[200] Ibid.

[201] Marpsat M and Razafindratsima N, "Survey Methods for Hard-to-Reach Populations: Introduction to the Special Issue" *Methodological Innovations Online* 5, no. 2 (2010): 3–16.

[202] Gunther Eysenbach and Jeremy Wyatt. "Using the Internet for Surveys and Health Research." *Journal of Medical Internet Research* 4, no. 2 (2002): e13. doi:10.2196/jmir.4.2.e13.

[203] Van Gelder MMHJ, Bretveld RW, and Roeleveld N. "Web-Based Questionnaires: The Future in Epidemiology?" *American Journal of Epidemiology* 172, no. 11 (2010): 1292–98. doi:10.1093/aje/kwq291.

[204] *Federal Register* Notices, "Proposed Project: 2013 National Survey on Drug Use and Health (NSDUH) Dress Rehearsal (OMB No. 0930–0334) — Revision." 78, No. 41, Friday, March 1, 2013. http://www.gpo.gov/fdsys/pkg/FR-2013-03-01/pdf/2013-04756.pdf .

[205] Hart et al., "Rural Definitions for Health Policy and Research."

[206] Shortliffe EH and Barnett GO. "Biomedical Data: Their Acquisition, Storage, and Use." In *Biomedical Informatics*, edited by Shortliffe EH and Cimino JJ. Health Informatics. Springer New York, 2006. http://link.springer.com/chapter/10.1007/0-387-36278-9_2; Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, and Steiner J. "Chart Reviews in Emergency Medicine Research: Where Are the Methods?" *Annals of Emergency Medicine*, 1996; 27(3): 305–08; Goldberg SI, Niemierko A, and Turchin A. "Analysis of Data Errors in Clinical Research Databases." *AMIA Annual Symposium Proceedings 2008*: 242–46; Reisch, LM, Fosse JF, Beverly K, Yu O, Barlow WE, Harris EL, Rolnick S, Barton MB, Geiger AM, Herrington LJ, Greene SM, Gletcher SW, and Elmore JG. "Training, Quality Assurance, and Assessment of Medical Record Abstraction in a Multisite Study." *American Journal of Epidemiology,* 2003; 157(6): 546–51.

[207] Olsen L, Aisner D, and McGinnis JM, eds. *Roundtable on Evidence-Based Medicine, The Learning Healthcare System: Workshop Summary (*IOM Roundtable on Evidence-Based Medicine), The National Academies Press (2007) http://www.nap.edu/catalog.php?record_id=11903

[208] Miller E. "The National Center for Health Statistics' Linked Data Files: Resources for Research and Policy." Presented June 25, 2013 at AcademyHealth Annual Research Meeting, Baltimore, Maryland; Andrews R. "Clinically-Enhanced Statewide Hospital Discharge Data: Practical Experience and Potential Value." Presented June 23, 2013 at AcademyHealth Annual Research Meeting; Baltimore, Maryland; Steiner C, "The Healthcare Cost and Utilization Project (HCUP): Linked Data Enhancements and Improved Analytic Capacity," April 10, 2013.

[209] Institute of Medicine. "Knowing What Works in Health Care: A Roadmap for the Nation." Consensus Report, January 24, 2008. http://www.iom.edu/Reports/2008/Knowing-What-Works-in-Health-Care-A-Roadmap-for-the-Nation.aspx.

[210] Hoeffel EM, Rastogi S, Kim MO and Shahid H. "The Asian Population: 2010." 2010 Census Briefs, March 2012. http://www.census.gov/prod/cen2010/briefs/c2010br-11.pdf.

[211] Islam NS, Khan S, Kwon S, Jang D, Ro M, and Trinh-Shevrin C. "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations: Historical Challenges and Potential Solutions." Journal of Health Care for the Poor and Underserved, 2010; 21(4): 1354–1381.

[212] Islam NS, Khan S, Kwon S, Jang D, Ro M, and Trinh-Shevrin C. "Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations: Historical Challenges and Potential Solutions." Journal of Health Care for the Poor and Underserved, 2010; 21(4): 1354–1381.

[213] Palo Alto Medical Foundation. "The Pan Asian Cohort Study." PAMF website, accessed July 15, 2013. http://www.pamf.org/pacs/. A similar pattern can be seen among women.

[214] American Cancer Society, "Cancer Facts & Figures 2013," accessed February 28, 2013, http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-036845.pdf.

[215] Institute of Medicine. "The Health of Lesbian, Gay, Bisexual, and Transgender People." March 31, 2011. http://www.iom.edu/Reports/2011/The-Health-of-Lesbian-Gay-Bisexual-and-Transgender-People.aspx

[216] National Institute of Mental Health. "A Parent's Guide to Autism Spectrum Disorder." 2011, http://www.nimh.nih.gov/health/publications/a-parents-guide-to-autism-spectrum-disorder/complete-index.shtml.

[217] Boyle CA and Boulet SL. "Health Care Use and Health and Functional Impact of Developmental Disabilities among US Children, 1997–2005." Archives of Pediatrics & Adolescent Medicine, 2009; 163(1): 19–26.

[218] Gurney JG, McPheeters ML, and Davis MM, "Parental Report of Health Conditions and Health Care Use among Children with and Without Autism: National Survey of Children's Health." Archives of Pediatrics & Adolescent Medicine, 2006; 160(8): 825–830.

[219] National Institute of Mental Health. "A Parent's Guide to Autism Spectrum Disorder."

[220] Gurney et al., "Parental Report of Health Conditions and Health Care Use among Children with and Without Autism."

[221] Croen LA, Najjar DV, Ray GT, Lotspeich L and Bernal P. "A Comparison of Health Care Utilization and Costs of Children with and Without Autism Spectrum Disorders in a Large Group-model Health Plan," Pediatrics, 2006; 118(4): e1203–1211.

[222] Langworthy-Lam KS, Aman MG, and Van Bourgondien ME. "Prevalence and Patterns of Use of Psychoactive Medicines in Individuals with Autism in the Autism Society of North Carolina." Journal of Child and Adolescent Psychopharmacology, 2002; 12(4): 311–321.

[223] Honeycutt T and Wittenburg T. "Identifying Transition-Age Youth with Disabilities Using Existing Surveys." Mathematica Policy Research, July 10, 2012. http://www.mathematica-mpr.com/publications/PDFs/disability/transition_age_youth_disabilities.pdf.

[224] More attention has been paid to the obstacles faced by youth with ASDs as they transition to college, work, and independent living. There is a body of literature around how schools should help students with ASD make this transition and federal measures in place to make sure schools plan for this transition because many students with ASDs receive special education services. Hendricks DR and Wehman P. "Transition from School to Adulthood for Youth with Autism Spectrum Disorders Review and Recommendations." Focus on Autism and Other Developmental Disabilities, 2009; 24(2):77-88.

[225] Bennett KJ, Olatsi B, and Probst J. "Health Disparities: A Rural-Urban Chartbook." South Carolina Rural Health Research Center, June 2008, http://rhr.sph.sc.edu/report/%287-3%29%20Health%20Disparities%20A%20Rural%20Urban%20Chartbook%20-%20Distribution%20Copy.pdf.

[226] Jones C, Parker T, Ahearn M, Mishra AK and Variyam J. "Health Status and Health Care Access of Farm and Rural Populations." USDA Economic Research Service. Economic Information Bulletin No. (EIB-57). August 2009. http://www.ers.usda.gov/publications/eib-economic-information-bulletin/eib57.aspx#.UeRQ5o2cfts

[227] Murphy J. "ONC Program Update." Presented at NCVHS Meeting, June 19, 2013. http://www.ncvhs.hhs.gov/130619p1.pdf

It should be noted that providers can receive payment through either the Medicare or the Medicaid payment meaningful use incentive program. To receive payment, providers must meet meaningful use (MU) criteria which are defined through the regulatory process and intended to facilitate improvement in quality and efficiency. There are three stages of meaningful use MU in these programs with increasingly challenging requirements and this data includes only the first payment, which is also the largest one. In the Medicaid program, providers may receive their first incentive payment for adoption, implementation, or upgrade (AIU) of an EHR system in recognition that Medicaid intensive providers are less likely to already have had an EHR due to resource limitations and may have a more difficult time raising the capital to finance or purchase a system on their own. It is not clear whether those who received payments for AIU or stage 1 will continue on to future stages. For more information these EHR incentive programs, see http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Basics.html.

[228] For an overview of HITECH programs and how they are designed to work together see: Gold M, McLaughlin C, Devers K, Berenson B, and Bovbjerg RR. "Obtaining Providers' "Buy-In" and Establishing Effective Means of Health Information Exchange Will Be Critical to HITECH's Success." Health Affairs, 2012; 31(3):514-526.

[229] DesRoches CM, Campbell EG, Rao SR, et al. "Electronic health records in ambulatory care—a national survey of physicians." *New England Journal of Medicine,* 359(1):50-60, 2008.

[230] Hsiao CJ, Jha AK, King J, Patel V, Furukawa MF, and Mostashari F. "Office-Based Physicians Are Responding To Incentives And Assistance By Adopting And Using Electronic Health Records." Health Affairs. July 2013; Epub ahead of print.

[231] DesRoches CM, Campbell EG, Rao SR, et al. "Electronic health records in ambulatory care—a national survey of physicians." New England Journal of Medicine, 359(1):50-60, 2008.

[232] The expert panel disagreed on the need to include physician notes and nursing assessments to classify a hospital as having a basic system, so two definitions of Basic EHR were developed. Since Meaningful Use does not require clinician notes, adoption of at least Basic EHR is based on the definition of Basic without clinician notes.

[233] DesRoches CM, Charles D, Furukawa MF, Joshi MS, Kralovec P, Mostashari F, Worzala C, and Jha AK. "Adoption Of Electronic Health Records Grows Rapidly, But Fewer Than Half Of US Hospitals Had At Least A Basic System In 2012." Health Affairs, 2013; 32(8):1-8.

[234] Charles D, Furukawa M, and Hufstader M. "Electronic Health Record Systems and Intent to Attest to Meaningful Use among Non-Federal Acute Care Hospitals in the United States: 2008-2011." ONC Data Brief no. 1 Washington, DC: Office of the National Coordinator for Health IT. February 2012.

[235] Nakamura MM, Ferris TG, DesRoches CM, and Jha AK. "Electronic health record adoption by children's hospitals in the United States." Archives of Pediatrics and Adolescent Medicine, 2010; 164(12): 1145-51.

[236] Jha AK, DesRoches CM, Kralovec PD, and Joshi MS. "A Progress Report on Electronic Health Records in U.S. hospitals." *Health Affairs*, 2010; 29(10): 1951–7.

[237] Furukawa MF, Patel V, Charles D et al. "Hospital Electronic Health Information Exchange Grew Substantially in 2008–2012." *Health Affairs*, August 2013; 32(8):1346-1354.

[238] Nakamura, et al, "Electronic Health Record Adoption by Children's Hospitals in the United States."

[239] McGinn CA, Grenier S, Duplantie J, et al. "Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review." *BMC Medicine*, 9(46): 2011.

[240] Moiduddin A and Stromberg S. *Health information technology in California's rural practices: assessing the benefits and barriers*. Oakland, CA: California Healthcare Foundation, 2009; Bahensky JA, Jaana M, and Ward MM. "Health care information technology in rural America: electronic medical record adoption status in meeting the national agenda." *Journal of Rural Health*, 24(2): 101–5, 2008; Bahensky JA, Ward MM, Nyarko K, and Li P. "HIT Implementation in Critical Access Hospitals: Extent of Implementation and Business Strategies Supporting IT Use." *Journal of Medical Systems*, 35(4): 599-607, 2011.

[241] Institute of Medicine. "Key Capabilities of an Electronic Health Record System: Letter Report." July 31, 2003. http://www.iom.edu/Reports/2003/Key-Capabilities-of-an-Electronic-Health-Record-System.aspx.

[242] Ford EW, Menachemi N, Huerta TR, and Yu F. "Hospital IT Adoption Strategies Associated with Implementation Success: Implications for Achieving Meaningful Use." *Journal of Healthcare Management / American College of Healthcare Executives,* 2010; 55(3): 175–88; discussion 188–89.

243 HealthIT.gov FAQs, http://www.healthit.gov/providers-professionals/faqs/what-information-does-electronic-health-record-ehr-contain . Accessed May 24, 2013

244 NORC at the University of Chicago. "Howard University Hospital Diabetes Treatment Center—Using Multi-modal Health IT Tools to Improve Quality and Delivery of Care in an Urban Setting." June 2012, http://www.healthit.gov/sites/default/files/pdf/HowardCaseStudyReport.pdf.

245 Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, and Baker DW. "Disparities in Enrollment and Use of an Electronic Patient Portal." Journal of General Internal Medicine, 2011; 26(10): 1112–1116.

246 Interviews with Savitz, Callahan, and Ehrenfeld

247 Technical Expert Panel Discussion, July 24, 2013

248 Interviews with Callahan, Ehrenfeld

249 Interview with Walker

250 Webster PS and Sampangi S. "Report on Data Improvement Pilot on Patient Ethnicity and Race (DIPPER): Pilot Design and Proposed Voluntary Standard." Rhode Island Medical Journal, January 2013.

251 Mearian L. "How Big Data Will Save Your Life," Computer World, April 25, 2013, http://www.computerworld.com/s/article/9238593/How_big_data_will_save_your_life.

252 Interviews with Savitz, Elliott, and Hornbrook

253 Decker SL, Jamoom EW, Sisk JE. Physicians in non-primary care and small practices and those age 55 and older lag in adopting electronic health record systems. Health Affairs. April 2012. 10.1377/hlthaff.2011.1121; DesRoches CM, et al. Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. Health Affairs2012;31;10.1377/hlthaff.2012.0153; Stronger Economies Together (SET), a USDA Rural Development program found 63 percent of rural health providers did not have EHR as of 2012 (USDA website)

254 DesRoches CM, Worzala C, and Bates S. "Some Hospitals Are Falling Behind in Meeting 'Meaningful Use' Criteria and Could Be Vulnerable to Penalties in 2015." *Health Affairs*. 2013; 32(8): 1355–60.

255 Interview with Savitz.

256 Interview with Hornbrook.

257 Interview with Savitz.

258 Interview with Luft.

259 Interview with Croen.

260 Interview with Hornbrook.

261 Fiks AG, Grundmeier RW, Margolis B et al. "Comparative Effectiveness Research Using the Electronic Medical Record: An Emerging Area of Investigation in Pediatric Primary Care." *Journal of Pediatrics* 2012; 160(5): 719–24.

262 American Academy of Pediatrics Website, "About ePROS." Accessed September 19, 2013. http://www2.aap.org/pros/epros/; and interview with Wasserman and Fiks.

263 Interview with Savitz.

264 Interview with Hornbrook.

265 Interview with Wasserman and Fiks.

266 University of California, Davis, Health System. No date. "Patient Questions for Demographics."

267 Institute of Medicine. "Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records - Workshop Summary - Institute of Medicine." December 20, 2012. http://iom.edu/Reports/2012/Collecting-Sexual-Orientation-and-Gender-Identity-Data-in-Electronic-Health-Records.aspx.

268 Technical Expert Panel discussion, July 24, 2013

269 Technical Expert Panel discussion, July 24, 2013; and Kaelber D. "Clinical Research Informatics." Presentation to Case Western Reserve University, 2013.

270 Powell J and Buchan I. "Electronic Health Records Should Support Clinical Research," Journal of Medical Internet Research 7, no. 1 (March 14, 2005), doi:10.2196/jmir.7.1.e4.

271 Lohr K, and Steinwachs D. Health services research: An evolving definition of the field. Health Services Research, 2002; 37(1):15-17.

272 Agency for Healthcare Research and Quality. "What is Comparative Effectiveness Research?" AHRQ website, accessed July 10, 2013. http://effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/

273 Mearian L. "How Big Data Will Save Your Life," Computer World, April 25, 2013, http://www.computerworld.com/s/article/9238593/How_big_data_will_save_your_life.

[274] Kaelber DC, Foster W, Gilder J, Lover TE, and Jain AK. "Patient Characteristics Associated with Venous Thromboembolic Events: a Cohort Study Using Pooled Electronic Health Record Data." Journal of the American Medical Informatics Association. 2012; 19(6):965-72.

[275] Weissberg J. "Use of Large System Databases: Cox-2 Inhibitors." Kaiser Permanente, Presentation, The Learning Healthcare System, Institute of Medicine—Roundtable on EBM. May, 2006. http://www.iom.edu/~/media/Files/Activity%20Files/Quality/VSRT/S1bWeissbergReadOnly.pdf.

[276] Technical Expert Panel discussion, July 24, 2013

[277] Interviews with Croen and Savitz.

[278] Interview with Croen.

[279] Interview with Holve.

[280] Interview with Wasserman and Fiks.

[281] Interviews with Glasgow and West and Schilling.

[282] Interview with West and Schilling.

[283] Interview with Croen.

[284] Interview with Califf.

[285] Interview with Hornbrook.

[286] Interviews with Croen and Hornbrook.

[287] Technical Expert Panel discussion, July 24, 2013.

[288] Interviews with Savitz and Luft.

[289] Interview with Walker.

[290] Interview with Capponi.

[291] Technical Expert Panel discussion, July 24, 2013.

[292] Interview with Savitz.

[293] Interview with Walker.

[294] Interview with Capponi.

[295] Interview with Hornbrook.

[296] Bellin E, Fletcher DD, Geberer N, Islam S, and Srivastava N. "Democratizing Information Creation from Health Care Data for Quality Improvement, Research, and Education-the Montefiore Medical Center Experience," Academic Medicine: Journal of the Association of American Medical Colleges, 2010; 85(8): 1362–1368.

[297] Kern LM, Malhotra S, Barron Y, Quaresimo J, Dhopeshwarkar R, Pichardo M, Edwards AM, and Kaushal R. "Accuracy of Electronically Reported 'Meaningful Use' Clinical Quality Measures: a Cross-sectional Study." Annals of Internal Medicine, 2013; 158(2): 77–83.

[298] Parsons A, McCullough C, Wang J and Shih S. "Validity of Electronic Health Record-derived Quality Measurement for Performance Monitoring." Journal of the American Medical Informatics Association, 2012; 19(4): 604–609.

[299] Hornbrook MC, Whitlock EP, Berg CJ, Callaghan WM, Bachman DJ, Gold R, Bruce FC, Dietz PM, and Williams SB. "Development of an Algorithm to Identify Pregnancy Episodes in an Integrated Health Care Delivery System." Health Services Research, 2007; 42(2): 908–927.

[300] Interview with Luft.

[301] Interview with Savitz.

[302] Jensen PB, Jensen LJ, and Brunak S. "Mining electronic health records: towards better research applications and clinical care." Nature Reviews, Genetics, 2012; 13:295-405.

[303] Ehrenfeld J. "Identification of LGBT Patients and Health Disparities: Using Electronic Health Records." Presented at the Sexual Orientation and Gender Identity Data Collection in Electronic Health Records: A Workshop, Institute of Medicine, October 12, 2012.

[304] Stanfill MH, Williams M, Fenton SH, Jenders RA, and Hersh WR. "A Systematic Literature Review of Automated Clinical Coding and Classification Systems." Journal of the American Medical Informatics Association: JAMIA, 2010; 17(6): 646–651.

[305] Chan KS, Fowles JB, and Weiner JP. "Review: Electronic Health Records and the Reliability and Validity of Quality Measures: a Review of the Literature." Medical Care Research and Review, 2010: 67(5): 503–527.

[306] Stanfill MH, Williams M, Fenton SH, Jenders RA, and Hersh WR. "A Systematic Literature Review of Automated Clinical Coding and Classification Systems." Journal of the American Medical Informatics Association: JAMIA, 2010; 17(6): 646–651.

[307] Interview with Walker.

[308] Interview with Franklin.

[309] Interview with Hornbrook.

[310] Chan KS, Fowles JB, and Weiner JP. "Review: Electronic Health Records and the Reliability and Validity of Quality Measures: a Review of the Literature." *Medical Care Research and Review*, 2010: 67(5): 503–527.

[311] Interview with Franklin.

[312] Jaret P. "Mining Electronic Records for Revealing Health Data." New York Times, January 14, 2013. http://www.nytimes.com/2013/01/15/health/mining-electronic-records-for-revealing-health-data.html.

[313] Interview with Hornbrook.

[314] Technical Expert Panel discussion, July 24, 2013.

[315] Interviews with Hornbrook and Califf.

[316] Technical Expert Panel discussion, July 24, 2013.

[317] Interview with Elliott.

[318] Interview with Wasserman and Fiks.

[319] Technical Expert Panel discussion, July 24, 2013.

[320] Benson B. "Legacy EHR System and Data Lookup a Thing of the Past." HITECH Answers, accessed May 25, 2013, http://www.hitechanswers.net/legacy-ehr-system-data-lookup/.

[321] Interview with Elliott.

[322] Interview with Califf.

[323] Technical Expert Panel discussion, July 24, 2013.

[324] McGraw D and Leiter A. "Legal and Policy Challenges to Secondary Uses of Information from Electronic Clinical Health Records." AcademyHealth, 2012.
Kern LM, Malhotra S, Barron Y, Quaresimo J, Dhopeshwarkar R, Pichardo M, Edwards AM, and Kaushal R. "Accuracy of Electronically Reported 'Meaningful Use' Clinical Quality Measures: a Cross-sectional Study." Annals of Internal Medicine, 2013; 158(2): 77–83.

[325] Selker H, Grossman C, Adams A, et al. "The Common Rule and Continuous Improvement in Health Care: A Learning Health System Perspective." Institute of Medicine Discussion Paper, October 1, 2011. http://www.iom.edu/Global/Perspectives/2012/CommonRule.aspx; Nass SJ, Levit LA, Gostin LO. *Beyond the HIPAA Privacy Rule*. Washington DC, National Academies Press, 2009.

[326] HHS Press Office. "New rule protects patient privacy, secures health information." News Release, January 17, 2013. http://www.hhs.gov/news/press/2013pres/01/20130117b.html

[327] DeGraw D and Leiter A. "Legal and Policy Challenges to Secondary Uses of Information from Electronic Clinical Health Records." AcademyHealth, 2012.

[328] Jensen PB, Jensen LJ, and Brunak S. "Mining electronic health records: towards better research applications and clinical care." Nature Reviews, Genetics, 2012; 13:295-405.

[329] Obel N, Omland LH, Kronborg G, Larsen CS, Pedersen C, Pedersen G, Sørensen HT, Gerstoft J. "Impact of non-HIV and HIV Risk Factors on Survival in HIV-infected Patients on HAART: a Population-based Nationwide Cohort Study." PloS One, 2011; 6(7).

[330] Clark S and Weale A. "Information Governance in Health: An Analysis of the Social Values Involved in Data Linkage Studies." Economic and Social Research Council, 2011.

[331] Hoffman S and Podgurski A. "Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research." Social Science Research Network Scholarly Paper, September 7, 2011. http://papers.ssrn.com/abstract=1923187.

[332] Ingelfinger JR and Drazen JM. "Registry Research and Medical Privacy." The New England Journal of Medicine, 2004; 350(14): 1452–1453.

[333] Felt U, Bister MD, Strassnig M, and Wagner U. "Refusing the Information Paradigm: Informed Consent, Medical Research, and Patient Participation." Health (London, England: 1997), 2009; 13 (1): 87–106.

[334] Federal Trade Commission. "Fair Information Practices Principles." http://www.ftc.gov/reports/privacy3/fairinfo.shtm, accessed August 25, 2013.

[335] Hoffman S and Podgurski A. "Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research." Social Science Research Network Scholarly Paper, September 7, 2011. http://papers.ssrn.com/abstract=1923187.

[336] Noble S, Donovan J, Turner E, Metcalfe C, Lane A, Rowlands MA, Neal D, Hamdy F, Ben-Shlomo Y, and Martin R. "Feasibility and Cost of Obtaining Informed Consent for Essential Review of Medical Records in Large-scale Health Services Research." Journal of Health Services Research & Policy, 2009; 14(2): 77–81.

[337] Grande D, Mitra N, Shah A, Wan F, and Asch D. "A National Survey of Patient Preferences about Secondary Uses of Electronic Health Information." Presented June 25, 2013 at AcademyHealth Annual Research Meeting, Baltimore, Maryland.

[338] Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. "Written Informed Consent and Selection Bias in Observational Studies Using Medical Records: Systematic Review." BMJ, 2009; 338: b866.

[339] U.S. Food and Drug Administration. "Should Your Child be in a Clinical Trial?" http://www.fda.gov/forconsumers/consumerupdates/ucm048699.htm, accessed August 25, 2013.

[340] McGraw D and Leiter A. "A Policy and Technology Framework for Using Clinical Data to Improve Quality." Houston Journal of Law & Policy. 2012; 137-167.

[341] Interview with Elliott.

[342] Interview with Walker.

[343] Interview with Croen.

[344] Interview with Ehrenfeld.

[345] Interview with Capponi.

[346] Interviews with Savitz and Callahan.

[347] McGraw D and Leiter A. "A Policy and Technology Framework for Using Clinical Data to Improve Quality." Houston Journal of Law & Policy. 2012; 137-167.

[348] Luft H. "Embedded Research: Doing Research on the Organization Within Which You Work." Presented June 2013 at AcademyHealth Annual Research Meeting, Baltimore, Maryland.

[349] Snyder C. "Considerations for Using Patient-Reported Outcomes in Clinical Practice: A Case Study." Presented June 2013 at AcademyHealth Annual Research Meeting, Baltimore, Maryland.

[350] Sabharwal R, Holve E, Rein A, and Segal C. "Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy." Issue Briefs and Reports, March 1, 2012, http://repository.academyhealth.org/edm_briefs/1.

[351] Multicenter Perioperative Outcomes Group website. http://mpog.med.umich.edu/, accessed August 26, 2013.

[352] McGraw D. "Data Governance Challenges & Opportunities in Health Services Research." Presented June 24, 2013 at AcademyHealth Annual Research Meeting, Baltimore, Maryland.

[353] Rosenbaum S. "Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access." Health Services Research, 2010; 45(5 Pt 2): 1442–1455.

[354] Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, and Detmer DE. "Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper." Journal of the American Medical Informatics Association: JAMIA, 2007; 14(1): 1–9.

[355] Interview with Capponi.

[356] Interview with Elliott.

[357] Interview with Walker.

[358] McGraw D and Leiter A. "A Policy and Technology Framework for Using Clinical Data to Improve Quality." Houston Journal of Law & Policy. 2012; 137-167.

[359] Sabharwal R, Holve E, Rein A, and Segal C. "Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy." Issue Briefs and Reports, March 1, 2012, http://repository.academyhealth.org/edm_briefs/1.

[360] Technical Expert Panel discussion, July 24, 2013.

[361] Sabharwal R, Holve E, Rein A, and Segal C. "Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy." Issue Briefs and Reports, March 1, 2012, http://repository.academyhealth.org/edm_briefs/1.

[362] Pace WD, Cifuentes M, Valuck RJ, Staton EW, Brandt EC, and West DR. "An Electronic Practice-Based Network for Observational Comparative Effectiveness Research." Annals of Internal Medicine, 2009; 151(5): 338–340.

[363] American Academy of Pediatrics Website, "About ePROS." Accessed September 19, 2013. http://www2.aap.org/pros/epros/

[364] Sabharwal R, Holve E, Rein A, and Segal C. "Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy." Issue Briefs and Reports, March 1, 2012, http://repository.academyhealth.org/edm_briefs/1.

[365] Brown J, Syat B, Lane K and Platt R. "Blueprint for a Distributed Research Network to Conduct Population Studies and Safety Surveillance." Effective Health Care Program Research Reports Number 27. Agency for Healthcare Research and Quality, June 2010. http://effectivehealthcare.ahrq.gov/reports/final.cfm.

[366] Merrill M. "Pilot Project for Distributed Research Network Will Use EHRs," August 12, 2008, http://www.healthcareitnews.com/news/pilot-project-distributed-research-network-will-use-ehrs.

[367] Randhawa GS and Slutsky JR. "Building Sustainable Multi-functional Prospective Electronic Clinical Data Systems." Medical Care, 2012; 50 Suppl: S3–6.

[368] Technical Expert Panel discussion, July 24, 2013.

[369] Massachusetts eHealth Institute. "PopMedNet: Distributed Data Network." http://mehi.masstech.org/what-we-do/hie/mdphnet/popmednet, accessed August 25, 2013.

[370] Holve E, Segal C, and Lopez MH. "Opportunities and Challenges for Comparative Effectiveness Research (CER) With Electronic Clinical Data." Medical Care, 2012; 50 Suppl: S11–S18.

[371] McGraw D. "Data Governance Challenges & Opportunities in Health Services Research." Presented June 24, 2013 at AcademyHealth Annual Research Meeting, Baltimore, Maryland.

[372] Technical Expert Panel discussion, July 24, 2013.

[373] Segal C and Holve E. "Emerging Data Resources, Tools, and Publications from the ARRA-CER Infrastructure Awards." Presented June 2013 at Academy Health Annual Research Meeting, Baltimore, Maryland.

[374] Patient-Centered Outcomes Research Institute. "Improving Our National Infrastructure to Conduct Comparative Effectiveness Research." PCORI website, accessed July 10, 2013. http://www.pcori.org/funding-opportunities/improving-our-national-infrastructure-to-conduct-comparative-effectiveness-research/

[375] Luft HS. "Commentary: Protecting Human Subjects and Their Data in Multi-site Research." Medical Care, 2012; 50 Suppl: S74–76.

[376] Interview with Holve.

[377] Kahn MG, Raebel MA, Glanz JM, Riedlinger K, and Stein JF. "A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research." Medical Care, 2012; 50 Suppl: S21–29.

[378] Conn J. "More Than 300 Vendors Share Ambulatory Care EHR Market." ModernHealthcare, October 24, 2012, http://www.modernhealthcare.com/article/20121024/NEWS/310249954.

[379] Dreyer N. "Interfacing Registries with EHRs." Presented at AHRQ Annual Conference. September 14, 2009. http://www.ahrq.gov/news/events/conference/2009/dreyer/index.html.

[380] Eastwood B. "6 Big Data Analytics Use Cases for Healthcare IT." CIO.com, April 23, 2013. http://www.cio.com/article/732160/6_Big_Data_Analytics_Use_Cases_for_Healthcare_IT.

[381] HL7 website: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=6

[382] Allen T. "Better Care through Sharing Electronic Medical Records," Health Affairs Blog, September 4, 2012, http://healthaffairs.org/blog/2012/09/04/better-care-through-sharing-electronic-medical-records/.

[383] Interviews with Callahan and Walker.

[384] Interview with Glasgow.

[385] Interview with Hornbrook.

[386] Interview with Califf.

[387] Interview with Capponi.

[388] Interview with Holve.

[389] Interview with Hornbrook.

[390] Interview with West and Schilling.

[391] Interview with McBurnie.

[392] Technical Expert Panel discussion, July 24, 2013.

[393] Adler-Milstein J, Bates DW, and Jha AK. "U.S. Regional Health Information Organizations: Progress and Challenges." Health Affairs, 2009; 28(2):483–492.

[394] Aligning Forces for Quality. "Reform In Action: Can Publicly Reporting the Performance of Health Care Providers Spur Quality Improvement?" April 2012. http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2012/rwjf400299

[395] Bradley CJ, Penberthy L, Devers KJ, and Holden DJ. "Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future." Health Services Research, 2010; 45(5p2): 1468–1488.

[396] Centers for Disease Control. "CDC Features - Providing Quality Cancer Data," accessed May 25, 2013, http://www.cdc.gov/Features/CancerRegistries/.

[397] Hornbrook MC, Fishman PA, Ritzwoller DP, Elston-Lafata J, O'Keeffe-Rosetti MC, and Salloum RG. "When Does an Episode of Care for Cancer Begin?" Medical Care, 2013; 51(4): 324–329.

[398] Field K, Kosmider S, Johns J, Farrugia H, Hastie I, Croxford M, Chapman M, Harold M, Murigu N, and Gibbs P. "Linking Data from Hospital and Cancer Registry Databases: Should This Be Standard Practice?" Internal Medicine Journal, 2010; 40(8): 566–573.

[399] Jain SH, Conway PH, and Berwick DM, "A Public-private Strategy to Advance the Use of Clinical Registries." Anesthesiology, 2012; 117(2): 227–229.

[400] Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I and Brand CA. "Data Linkage: A Powerful Research Tool with Potential Problems." BMC Health Services Research, 2010; 10(1): 346.

[401] Belmont J and McGuire AL. "The Futility of Genomic Counseling: Essential Role of Electronic Health Records." Genome Medicine, 2009; 1(5): 48.

[402] Belmont J and McGuire AL. "The Futility of Genomic Counseling: Essential Role of Electronic Health Records." Genome Medicine, 2009; 1(5): 48.

[403] Jensen PB, Jensen LJ, and Brunak S. "Mining electronic health records: towards better research applications and clinical care." Nature Reviews, Genetics, 2012; 13:295-405.

[404] Jensen PB, Jensen LJ, and Brunak S. "Mining electronic health records: towards better research applications and clinical care." Nature Reviews, Genetics, 2012; 13:295-405.

[405] Centers for Disease Control and Prevention Website, "National Health and Nutrition Examination Survey: How to Access the Genetic Data Sets." Access September 19, 2013. http://www.cdc.gov/nchs/nhanes/genetics/genetic_access.htm

[406] Hamilton J. "Matching DNA with Medical Records to Crack Disease and Aging." NPR, All Things Considered, November 19, 2012. http://www.npr.org/blogs/health/2012/11/19/165498842/matching-dna-with-medical-records-to-crack-disease-and-aging.

[407] Trinidad SB, Fullerton SM, Ludman EJ, Jarvik GP, Larson EB, and Burke W. "Research Ethics. Research Practice and Participant Preferences: The Growing Gulf." Science, 2011; 331(6015): 287–288.

[408] Lunshof JD, Chadwick R, Vorhaus DB, and Church GM. "From Genetic Privacy to Open Consent," Nature Review Genetics, 2008; 9(5): 406–411.

[409] Steiner C, "The Healthcare Cost and Utilization Project (HCUP): Linked Data Enhancements and Improved Analytic Capacity," April 10, 2013.

[410] Andrews R. "Clinically-Enhanced Statewide Hospital Discharge Data: Practical Experience and Potential Value." Presented June 23, 2013 at AcademyHealth Annual Research Meeting; Baltimore, Maryland.

[411] Technical Expert Panel discussion, July 24, 2013.

[412] NORC at the University of Chicago. "Patient Care Management and Rewards Program—Promoting and Tracking Wellness Behaviors within the Context of an Existing Case-management Program." June 2012, http://www.healthit.gov/sites/default/files/pdf/AEH_CaseStudyReport.pdf.

[413] Vayena E, Mastroianni A, and Kahn J. "Caught in the Web: Informed Consent for Online Health Research." Science Translational Medicine, 2013; 5(173): 173fs6–173fs6.

[414] Interview with Savitz.

[415] Interview with Croen.

[416] Interview with Hornbrook.

[417] Interview with Savitz.

[418] Waidmann TA, Ormond BA, and Spillman BC. Potential Savings through Prevention of Avoidable Chronic Illness among CalPERS State Active Members. Urban Institute, April 2012. http://www.urban.org/publications/412550.html

[419] Katz N, Andrews R, Zingmond D, and Weiser T. "Statewide Initiatives to Improve Race Ethnicity and Language Data: Three Unique Approaches." Presented at Council of State and Territorial Epidemiologists annual conference, Pasadena, CA, June 10, 2013. https://cste.confex.com/cste/2013/webprogram/Paper1519.html

[420] Center for Medicare and Medicaid Services, "Social and Behavioral Domains and Measures for Domains for Electronic Clinical Quality Measures (eCOM)." Accessed September 19, 2013. https://www.fbo.gov/index?s=opportunity&mode=form&id=77b0f00d5508ca8cef522072de3c5b0a&tab=core&_cview=0; and "CMS Orders Study on Including Social, Behavioral Health Data in EHRs." iHealthBeat, September 16, 2013. http://www.ihealthbeat.org/articles/2013/9/16/cms-commissions-study-on-including-social-behavioral-health-in-ehrs

[421] Arispe IE. "The National Center for Health Statistics: Adapting to meet new data needs." Presented June 2013 at AcademyHealth Annual Research Meeting; Baltimore, Maryland.

[422] Interview with Luft.

[423] Interview with Felix.

[424] Interview with Califf.

[425] Interview with Walker.

[426] Interview with McBurnie.

[427] Interview with Ehrenfeld.

[428] Interview with Hornbrook.

[429] Interview with Luft.

[430] Interview with Luft.

[431] Interview with Luft.

[432] Technical Expert Panel discussion, July 24, 2013.

[433] Technical Expert Panel discussion, July 24, 2013.

[434] Interview with Croen.

[435] Interview with Chang Weir.

[436] Interview with Chang Weir.

[437] Interview with Croen.

[438] Interview with Hornbrook.

[439] Interview with Hornbrook.

[440] Interview with Chang Weir.

[441] Interviews with Chang Weir, Hornbrook, and Ehrenfeld.

[442] Interview with Ehrenfeld.

[443] Interview with Capponi.

[444] Interview with McBurnie.

[445] Interviews with Walker and Elliott.

[446] Interview with Hornbrook.

[447] Interviews with Ehrenfeld, Croen, and Hornbrook.

[448] Interviews with Ehrenfeld and Chang Weir.

[449] Interview with McBurnie.

[450] Interview with West and Schilling.

[451] Interview with McBurnie.

[452] Interview with Savitz.

[453] Interview with West and Schilling.

[454] Technical Expert Panel discussion, July 24, 2013.

[455] McGraw D and Leiter A. "A Policy and Technology Framework for Using Clinical Data to Improve Quality." *Houston Journal of Law & Policy*. 2012; 137-167.