Enabling the Reuse of Electronic Health Record Data through Data Quality Assessment and Transparency

Nicole Gray Weiskopf

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Enabling the Reuse of Electronic Health Record Data through

Data Quality Assessment and Transparency

Nicole Gray Weiskopf

With the increasing adoption of health information technology and the growth in the resulting electronic repositories of clinical data, the secondary use of electronic health record data has become one of the most promising approaches to enabling and speeding clinical research. Unfortunately, electronic health record data are known to suffer from significant data quality problems. Awareness of the problem of electronic health record data quality is growing, but methods for measuring data quality remain ad hoc. Clinical researchers must handle this complicated problem without systematic or validated methods. The lack of appropriate or trustworthy electronic health record data quality assessment methodology limits the validity of research performed with electronic health record data.

This dissertation documents the development of a data quality assessment framework and guideline for clinical researchers engaged in the secondary use of electronic health record data for retrospective research. Through a systematic literature review and interviews with key stakeholders, we identified core constructs of data quality, as well as priorities for future approaches to electronic health record data quality assessment. We used a data-driven approach to demonstrate that data quality is task-dependent, indicating that appropriate data quality measures must be selected, applied, and interpreted within the context of a specific study. On the basis of these results, we developed and evaluated a dynamic guideline for data quality measures in order to help researchers choose data quality measures and methods appropriately within the context of reusing electronic health record data for research.

Table of Contents

List of Figures

Acknowledgements

I have a great many people to thank for their guidance, support, and assistance in the conception, development, and execution of the research described in this dissertation. First, I must thank my adviser, Dr. Chunhua Weng, who has been an excellent mentor over the past several years. She gave me the freedom to identify a topic that I found meaningful and compelling, while still ensuring that the work I did was timely and of good quality. This is a difficult balance to strike, and I very much appreciate it.

I would also like to thank all of my committee members, both at Columbia and at other institutions. Drs. Suzanne Bakken, George Hripcsak, Michael Kahn, Adam Wilcox, and Meredith Zozus shared with me an incredible amount of support and expertise. My work and this dissertation benefited greatly from their input.

My fellow students in the department have also been a tremendous source of support. I appreciate their willingness to sit through practice talks and exams, and their endless patience as sounding boards.

Lastly, I must thank my family. This has been a long and often stressful journey, and their support and patience has been invaluable. Thank you to my father, my in-laws, and my sister. Most of all, thank you to my mother, Joyce, and my husband, Lev. This would not have been possible without them. They were my constant sources of support, the first people I bounced ideas off of, and my go-to editors. I would not have made it through this degree without them. And thank you lastly to my daughter, Helen. I wouldn't say she makes me more productive, but she certainly makes all of this more worthwhile.

Chapter 1: Enabling the Reuse of Electronic Health Record Data through Data Quality Assessment and Transparency: An Overview

1.1 Introduction

With the increasing adoption of health information technology and the growth in the resulting electronic repositories of clinical data, the secondary use of electronic health record (EHR) data has become one of the most promising approaches to enabling and speeding clinical research. Unfortunately, EHR data are known to suffer from significant data quality problems exceeding those associated with data collected in traditional research settings. Awareness of the problem of EHR data quality is growing, but available methods for measuring data quality remain ad hoc. Clinical researchers are left to handle this complicated problem without systematic or

validated methods. The lack of appropriate or trustworthy EHR data quality assessment methodology limits the efficiency and validity of research performed with EHR data.

This dissertation documents the development of a data quality assessment framework and associated guidelines targeted at clinical researchers engaged in the secondary use of EHR data for retrospective research. Through a systematic literature review and interviews with key stakeholders, we identified core constructs of data quality relevant to the secondary use of EHR data, as well as priorities for future approaches to EHR data quality assessment. We used a data-driven approach to demonstrate that data quality is task-dependent, indicating that appropriate data quality measures must be selected, applied, and interpreted within the context of a specific study. On the basis of these results, we developed and evaluated a dynamic guideline for data quality measures in order to help researchers choose data quality measures and methods appropriately within the context of reusing EHR data for research.

1.2 Background

1.2.1 The Promise of EHR Data Secondary Use

The past few decades have seen growing adoption of electronic health records. Since the enactment of the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009, which includes incentives and penalties encouraging the adoption and meaningful use of EHRs,(1, 2) there has been an increase in the rate of EHR adoption. As of 2012, the rate of adoption of EHRs with at least basic functionality was 44.4% in non-federal acute care hospitals(3) and 39.6% in office-based physician practices.(4) Other estimates have placed rates of adoption amongst family physicians at 68% as of 2011.(5) The ubiquity of EHRs makes the secondary use of data collected during the course of clinical care a promising step towards decreasing research costs, increasing patient-centered research, and speeding the rate of medical discovery. Therefore, as the adoption of EHRs has made it easier to access and aggregate clinical

data, there has been growing interest among clinicians, researchers, and funding and regulatory agencies in conducting research with EHR data.(6-8)

The transition from paper records to EHRs has resulted in large repositories of computerized data, simplifying the process of reusing data recorded during the course of clinical care. The reuse of these data may diminish the costs and inefficiencies associated with clinical research. Many of the common processes involved in standard approaches to research, including participant recruitment and retention and the gathering of data, are expensive and time-consuming.(9) The identification and inclusion of certain populations in clinical research can be especially challenging. Overcoming the underrepresentation of ethnic minorities in research, for example, requires significant attention and effort.(10) Like other forms of retrospective research, studies that make use of EHR data do not require patient recruitment or data collection, eliminating much of the time and expense involved in prospective research. And, with the exception of large-scale research projects like the Framingham Heart Study and WICER, EHR data repositories contain more data than are collected for prospective studies, including more variables, longer periods of time, and greater numbers of patients.

The secondary use of EHR data may also address one of the primary limitations of randomized controlled trials: limited external validity.(11) The data from EHRs offer a window into the medical care, status, and outcomes of a diverse population that is representative of actual patients. This makes EHR data reuse particularly valuable for Comparative Effectiveness Research (CER)(12) and Patient Centered Outcomes Research (PCOR)(13) in their potential to enable comparisons of therapies that are meaningful for patients. As part of the American Recovery and Reinvestment Act of 2009, $1.1 billion were made available for CER.(14) The following year, the Patient Protection and Affordable Care Act established the PCOR Institute, which is dedicated to the guidance and funding of CER and PCOR.(15) It is clear that clinical research focusing on real life patient care and outcomes is currently a national priority, and EHR-derived datasets are ideally suited for such research.(16) Overall, the reuse of EHR data for

clinical research has the promise to bring us closer to a learning healthcare system, in which the generation of new medical knowledge and best practices arise naturally from the standard patient care process.(17)

1.2.2 EHR Data Quality

Despite these benefits, the reuse of EHR data has been limited by a number of factors, including concerns about the quality of the data and their suitability for research.(18, 19) It is generally accepted that clinical data are not recorded with the same care as research data.(20) Research data are often obtained and recorded under controlled conditions with carefully maintained workflows. The collection of high quality data is a primary focus in the traditional research approach. Data collected during the course of standard care, however, are in many ways a byproduct of the healthcare system. Clinical information is measured and recorded primarily to fulfill documentation requirements for patient care and administrative purposes. Moreover, it has been said that the adoption of health information technology like EHRs has led not to improvements in the quality of the data being recorded, but rather to the recording of a greater quantity of poor quality data.(21) Due in large part to concerns about DQ, van der Lei warned specifically against the reuse of clinical data for research and proposed what he called the first law of informatics: “[d]ata shall be used only for the purpose for which they were collected.”(22)

Although such concerns about data quality have existed since EHRs were first introduced, there remains no consensus as to the quality of electronic clinical data and its suitability for retrospective research. In fact, the quality of clinical data has been shown to be not only poor, but highly variable. Multiple studies have demonstrated that it is not possible to generalize data quality findings. Hogan and Wagner, in their 1997 literature review, found that the correctness of data ranged between 44% and 100%, and completeness between 1.1% and 100%, depending upon the clinical concepts being studied.(23) Similarly, Thiru et al., in calculating the sensitivity of different types of EHR data in the literature, found values ranging

between 0.26 and 1.00.(24) In a 2010 review, Chan et al. looked at the quality of the same clinical concepts across multiple institutions, and still found a great deal of variability. The completeness of blood pressure recordings, for example, fell anywhere between 0.1% and 51%.(25)

This variability in quality is due in part to differences in measurement, recording, information systems, clinical workflows, and clinical focus. Data quality variations can also be explained, however, by the concept of "fitness for use," which was first proposed by J. M. Juran.(26) Within the context of data quality and EHR data reuse, this means that an EHR-derived dataset is of sufficient quality when it allows a researcher to perform a study of interest and achieve results that are not invalidated by data quality problems. According to this definition, data quality is task-dependent, meaning that it is generally inadvisable to make assumptions about one EHR-derived dataset based upon another. Therefore, clinical researchers engaged in the secondary use of EHR data need systematic methods that will allow them to assess the quality of an EHR-derived dataset for a given research task.

1.3 Related Work

An ideal knowledge-based guideline for EHR data quality assessment should be all of the following: comprehensive, systematic, and dynamic. To be comprehensive it must include the major categories of data quality (e.g. correctness) that are relevant to the secondary use of EHR data. A systematic guideline would describe concrete, evidence-based, and, where possible, validated methods of assessing data quality. Finally, the guideline must be flexible in order to allow for fitness-for-use: different use cases will require different approaches to data quality assessment. Chapter 2 of this dissertation presents a thorough review of data quality categories and assessment methods from the informatics literature. A brief exploration of foundational work on data quality from informatics and related fields is included below.

1.3.1 Terminologies and Models for Data Quality Assessment

One of the most influential works on data quality is the industry model initially published by Wang and Strong in 1996.(27) Wang and Strong were heavily influenced by the concept "fitness for use," and were some of the first researchers to apply this definition of quality to data and databases from the data consumer viewpoint. Using a combination of literature review, user surveys, and user-based sorting tasks, Wang and Strong identified fifteen dimensions of data quality. The fifteen dimensions are organized in a hierarchy containing four primary categories, as shown in Figure 1.

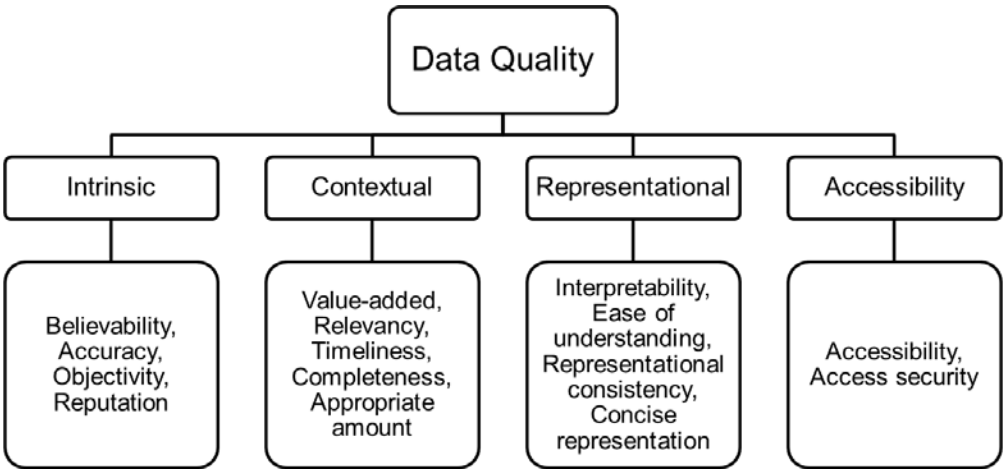


Figure 1.1 Wang and Strong's data quality model.

More recently, there have been increasing efforts to develop models of data quality specifically for EHRs. Liaw et al. performed a review of publications that used ontological approaches to data quality within the context of "routine data" use for chronic disease management, and found that the five most commonly included dimensions of data quality were accuracy, completeness, consistency, correctness, and timeliness.(28) A validated, comprehensive, accepted model or terminology of EHR data quality, however, does not currently exist.

1.3.2 Data Quality Assessment Approaches

A number of fields, particularly business, management, accounting, and information technology, have produced systematic approaches to data quality assessment. To understand the relevance of these approaches, however, it is important to understand what is meant by data quality assessment methods. “Methods” is meant to imply approaches for actually measuring data quality, which are rarely described in existing approaches.

Some approaches to data quality assessment, for example, present data quality metrics, rather than methods, where a metric is a system of measurement. Metrics are valuable because they allow standardized reporting and comparisons between results. They do not, however, give users direction on how to actually assess data quality in the first place. Hogan and Wagner, for example, present metrics for data completeness and correctness, but assume that a reference standard is available in order to calculate positive predictive value and sensitivity.(23) Pipino et al., in their heavily cited article on data quality assessment, describe three functional forms for reporting on data quality: ratios of problem elements to all elements, min or max reporting of data quality indicators, and weighted averages when there are multiple variables of interest.(29) Underlying all of these forms, however, is the assumption that users already have methods available to make subjective or objective assessments of data quality.

In other cases, data quality assessment refers to review of the processes surrounding data collection, storage, and governance, rather than methods that can be used to measure the quality of the data themselves. The International Monetary Fund Data Quality Assessment Framework, for example, essentially says that data are of good quality if regular steps are taken to ensure good data practices.(30) Similarly, Wang and Strong’s fifteen dimensions of data quality, described above, include concepts related to data representation and accessibility issues, which are not related to inherent issues of data quality.(27) In describing his model of data quality and data

error propagation, Ballou specifically stated that his work was process-oriented, rather than data-oriented.(31)

There are, however, a few existing methodological approaches to EHR data quality assessment. Falconer et al. proposed an eight-step process for examining the quality of diagnostic data that includes assessments of prevalence, completeness, accuracy, currency, consistency, sensitivity, and positive predictive value.(32) More recently, Kahn et al. developed a conceptual model for data quality assessment based upon fitness-for-use. The model incorporates a number of data quality categories from Wang and Strong's model, including accuracy, objectivity, believability, timeliness, and appropriate amount of data.(33) The Observational Medical Outcomes Partnership (OMOP), has developed automated data quality assessment "toolkits," called OSCAR and GROUCH, which can be used to measure, amongst other data quality dimensions, completeness, accuracy, and timeliness. These toolkits, however, are limited to use with the OMOP data model.(34)

EHR data quality assessment requires systematic approaches that allow the user to assess data quality without detailed knowledge of the processes that produced these data in the first place. Similarly, although standardized metrics are valuable for reporting, methods that give guidance on how to make measurements of data quality must be an area of focus. Overall, what is currently needed in order to enable secondary use of EHR data is a knowledge-based approach to EHR data quality assessment methods that can be applied systematically, though in a task-dependent manner.

1.4 Approach

1.4.1 Summary of Approach

This dissertation describes a mixed methods approach to the development and evaluation of a guideline for EHR data quality assessment within the context of secondary use. Our goal was to use knowledge- and data-driven approaches to combine the relevant categories of data quality

and proven methods of data quality assessment and present them in a way that considered user needs and perceptions. The studies described in Aims I, II, and III contributed necessary components and conceptual approaches to the guideline described in Aim V (Figure 2).

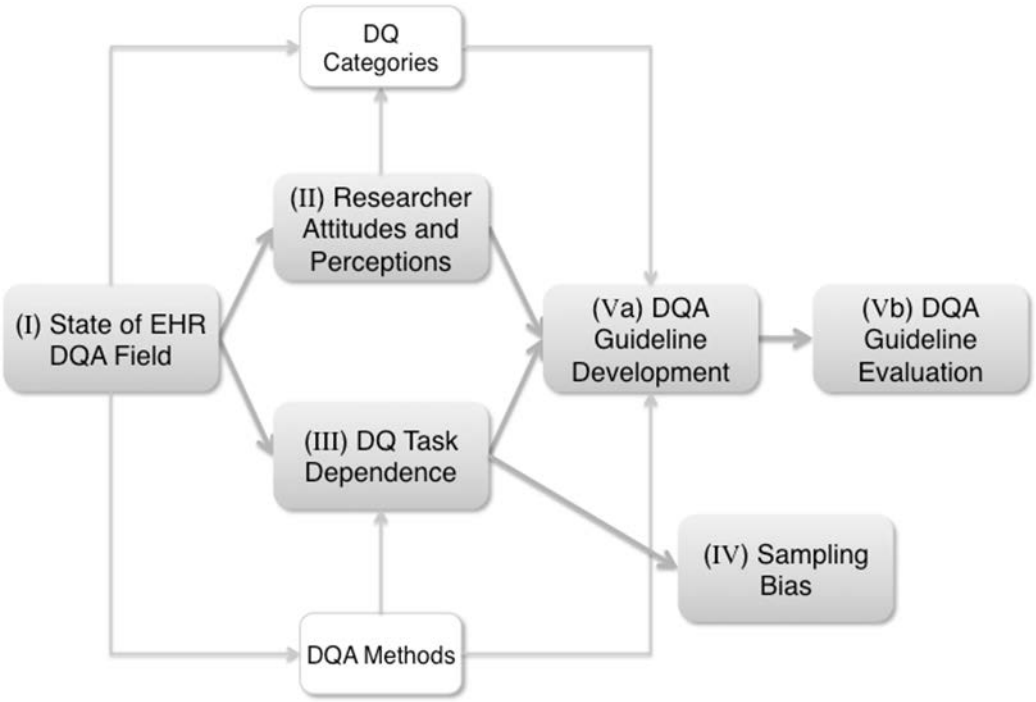


Figure 1.2 Dissertation flowchart.

Aim I: Identify the most commonly assessed categories of data quality and the most commonly used assessment methods in the secondary use of EHR data.

Aim II: Identify clinical researchers' attitudes toward and perceptions of EHR data quality as it relates to the secondary use of EHR.

Aim III: Determine if and how fitness-for-use applies to EHR data quality within the context of secondary use

Aim IV: Explore the relationship between patient health, data sufficiency, and the potential for sampling bias in the secondary use of EHR data.

Aim V: Develop and evaluate a data quality assessment guideline for clinical researchers engaged in the secondary use of EHR data.

1.4.2 Aim I: The state of the EHR data quality assessment field

Objective: Identify the most important categories of data quality and the most common data quality assessment methods from the informatics literature describing the secondary use of EHR data. Determine which methods are best suited to which categories.

Hypothesis: Different EHR data quality assessment methods are most commonly used for specific EHR data quality categories.

Research Questions:

- What categories of data quality are important to the secondary use of EHR data?
- What are common methods for EHR data quality assessment?
- Which methods are used for which categories?
- What are the current gaps in knowledge in the area of EHR data quality assessment?

Methods and Materials:

The primary methodology of this aim was a systematic literature review. The review focused on peer-reviewed articles describing original research that included a) secondary use of EHR data and b) data quality assessment. A two stage process was used to identify articles. First, a search was performed using PubMed's search feature and relevant MeSH headings. This search produced 230 results, which were then winnowed down to 44 relevant articles. Next, an ancestor search was used to find another 51 articles that were not identified during the initial search.

Two primary types of information were extracted from each of the 95 articles in the final sample: the features of data quality that were assessed, and the methods of assessment that were used. An iterative approach was used to group the data quality categories and methods into meaningful themes. Finally, the derived data quality methods and categories were applied to each of the 95 articles, which allowed us to determine how frequently each method was used to assess each category of data quality.

Primary Findings:

From the final pool of 95 articles, we identified five categories of EHR data quality and seven general approaches to EHR data quality assessment. . The data quality categories were derived from 27 unique data quality terms, and are as shown in Table 1.

Table 1.1 Categories of EHR data quality.

Data Quality Category	*Definition*
Completeness	Is a truth about a patient present in the EHR?
Correctness	Is an element that is present in the EHR true?
Concordance	Is there agreement between elements in the EHR, or between the EHR and another data source?
Plausibility	Does an element in the EHR makes sense in light of other knowledge about what that element is measuring?
Currency	Is an element in the EHR a relevant representation of the patient state at a given point in time?

The seven categories of data quality assessment approaches are detailed below in descending order of frequency use. . Most of the 95 articles did not explicitly label the approach used. Instead, these categories were derived from occasionally vague methodological descriptions.

Table 1.2 Categories of EHR data quality assessment approaches.

Data Quality Category	*Definition*
Gold standard	A dataset drawn from another source or multiple sources, with or without information from the EHR, is used as a gold standard.
Data element agreement	Two or more elements within an EHR are compared to see if they report the same or compatible information.
Element presence	A determination is made as to whether or not desired or expected data elements are present.
Data source agreement	Data from the EHR are compared to data from another source to determine if they are in agreement.
Distribution comparison	Distributions or summary statistics of aggregated data from the EHR are compared to the expected distributions for the clinical concepts of interest.
Validity check	Data in the EHR are assessed using various techniques that determine if values “make sense.”
Log review	Information on the actual data entry practices (e.g. dates, times, edits) is examined.

Certain methodological approaches are used more frequently for certain data quality categories. For example, gold standard-based assessment is the most commonly used approach for assessing correctness, while currency is assessed solely through the review of log data. . A complete mapping between the data quality categories and methodological approaches is shown in Figure 3.

The findings from the literature review indicate a few problems with current approaches to EHR data quality assessment within the context of secondary use. . First, the majority of researchers currently use ad hoc approaches, which limits the comparability and validity of data quality findings. . The lack of existing systematic approaches also increases the burden being put upon researchers who wish to use EHR data. . Second, there is currently an overreliance upon the use of gold standards for EHR data quality assessment. . This was by far the most common assessment approach in the articles included in the review. . In many—if not most—cases gold standards are not available, which means that this is not a generalizable approach.

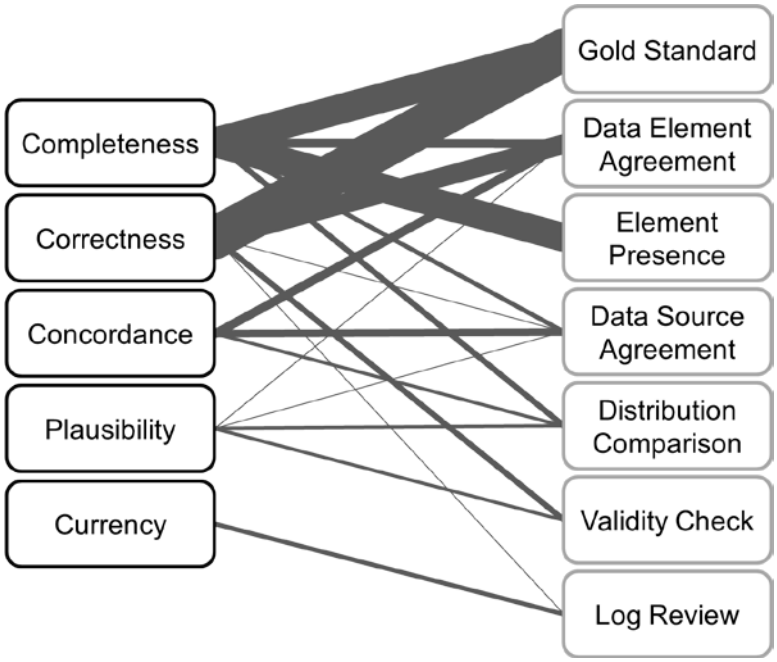


Figure 1.3 Mapping between data quality categories and methodological approaches.

1.4.3 Aim II: Clinical researcher attitudes and perceptions regarding EHR data quality

Objective: Gain an understanding of the factors that make clinical researchers more or less likely to engage in the secondary use of EHR data, especially as related to issues of data quality.

Hypothesis: Researcher awareness of and concerns about EHR data quality are a rate-limiting factor in the secondary use of EHR data.

Research Questions:

- How do clinical researchers define data quality?
- What do clinical researchers think about the quality of EHR data, specifically in the context of research?
- What factors make clinical researchers more or less likely to engage in the secondary use of EHR data?
- What sorts of research do clinical researchers prefer to perform with EHR data?

Methods and Materials:

Aim II relied upon a qualitative approach in order to develop an understanding of how clinical researchers view EHR data as a resource for secondary use. A series of ten unstructured interviews were conducted with clinical researchers in the Columbia University Medical Center community. Three broad categories of discussion were identified for the interviews: experience using EHR data for clinical research, perceptions of EHR data quality, and intent to use EHR data for research going forward. The interview guide was designed to explore researchers' actual experiences wherever possible, rather than hypotheticals. Following the interviews, the digital recordings were de-identified, transcribed, and then analyzed.

The interview data were analyzed using content analysis and a combined inductive-deductive approach. The inductive analysis was based upon the Precede-Proceed Model (shown in Figure 3), specifically the third phase of the model, which focuses on identifying factors that

make the target population (clinical researchers, in our scenario) more or less likely to engage in the desired behavior (the secondary use of EHR data).(35) The trustworthiness of the findings was ensured through two methods. Member checks were employed during the interviews and following analysis.(36, 37) An audit trail was also maintained over the course of the interviews and analysis and then reviewed by an expert in qualitative research.(36, 38, 39)

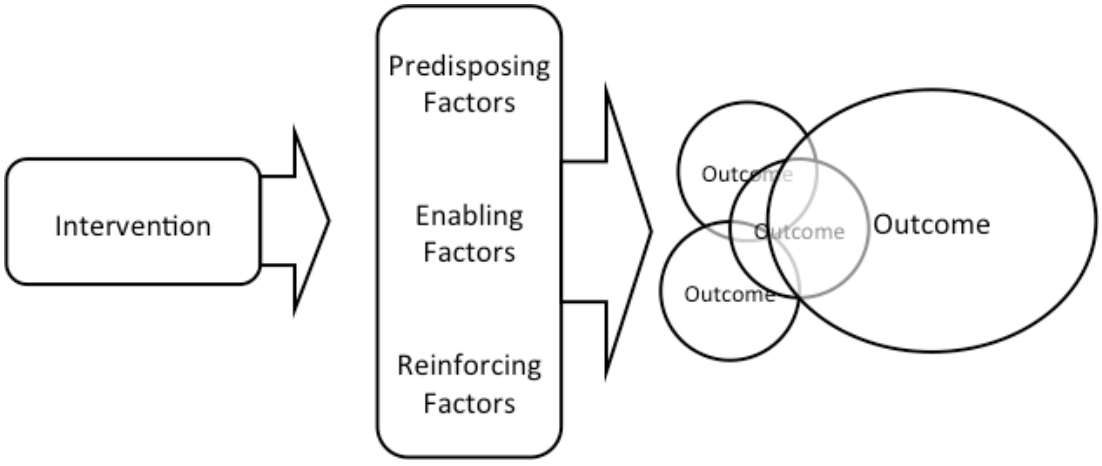


Figure 1.4 Precede-Proceed Model.

Primary Findings:

We identified seven categories of data quality that the participants believed were relevant to the secondary use of EHR data: completeness, correctness, fragmentation, granularity, concordance, structured, and signal-to-noise. . Three of these (completeness, correctness, and concordance) overlap with the five categories identified in the literature review described in Aim I (Chapter 2).

We also discovered that although the majority of the participants were aware of EHR data quality limitations and had conflicting or negative attitudes toward the secondary use of EHR data, they intended to begin or continue conducting research with EHR data. . For the most part, the perceived benefits and opportunities associated with the secondary use of EHR data outweighed their concerns about the limited validity of research conducted with these data. . This

finding drives home the importance of developing and disseminating systematic, valid, and feasible approaches to EHR data quality assessment.

Lastly, the participants emphasized the importance of involved domain experts and clinicians in the secondary use of EHR data. . Clinicians with experience using the EHR systems from which the data are extracted are uniquely suited to understanding the peculiarities and limitations of these data. . Therefore, the participant felt that data extraction, validation, and analysis should all be informed by individuals who are familiar with the data and data collection processes.

1.4.4 Aim III: The task-dependence of EHR data quality

Objective: Determine if the concept of fitness for use applies to EHR data quality within the context of secondary use.

Hypothesis: The use scenario for an EHR-derived dataset determines how data quality will be addressed, which in turn affects data quality findings.

Research Questions:

- Is EHR data quality task-dependent?
- How complete are EHR data?
- How do definitions and operationalizations of data quality concepts impact data quality findings?

Methods and Materials:

Aim III uses a data-driven approach to explore the concepts of EHR data completeness and fitness for use. Completeness was selected from the data quality categories described in Aims I and II because it is frequently described as being context-dependent.(27, 33) Four different operationalizations of EHR data completeness and associated methods of assessment were identified from the literature. The operationalizations were defined as follows:

Table 1.3 Operational definitions of EHR data completeness.

Operationalization	*Definition*
Documentation	A record contains all observations made about a patient.
Breadth	A record contains all desired types of data.
Density	A record contains a specified number or frequency of data points over time.
Predictive	A record contains sufficient information to predict a phenomenon of interest.

Each operationalization was tested against a dataset consisting of representative, frequently used types of data from the Columbia University Medical Center clinical data warehouse. These data included diagnoses, medication orders, laboratory results, visit information, and demographic information. Data from approximately 3.9 million unique patients records were used. Each operationalization was tested individually and plotted along increasing thresholds of completeness (i.e. fit for use). Simple versions of all four definitions were also tested against the data simultaneously to allow direct comparisons of which records were considered complete according to each definition.

Primary Findings:

The most basic take-home from this aim is that there are significant and unexpected limitations to the completeness of even the most basic EHR data elements. This was true for all four operational definitions. . Even with the most lenient cut-offs (i.e. at least one data point with the associated data of interest), only about half or fewer of the patient records can be considered complete according to the breadth, density, or documentation definitions.. . With more stringent cut-offs, where more data points are desired, these proportions drop precipitously. . Therefore, it is clear that EHR data completeness assessment should be a priority in the secondary use of EHR data.

Further complicating this issue is the fact that records that are complete according to one operational definition are not necessarily complete according to another. This is illustrated by Figure 4, which presents a visual comparison of the records deemed complete according to

illustrative examples of each of the four definitions. These results indicate that data completeness is dependent upon the use scenario and how completeness is operationalized, and that data quality assessment must be dynamic in order to capture task-dependence.

1.4.5 Aim IV: Sufficient EHR data quality and sampling bias

Objective: Explore the relationship between variation in data completeness and the potential for sampling bias in the secondary use of EHR data.

Hypothesis: Sicker patients have more complete records.

Research Questions:

- Is the quantity of data contained in a patient record random, or is it correlated with underlying patient health status?
- Is variation in EHR data quality a potential source of sampling bias?

Methods and Materials:

After uncovering extensive variation in EHR data completeness in Aim III, we wanted to further explore the issue of completeness and its possible determinants. Specifically, we hypothesized that completeness might be linked to patient health. To determine if this was the case, we needed to identify a variable to represent overall health, as well as a metric of completeness. We selected the American Society of Anesthesiologists (ASA) Physical Status Classification System as our health indicator, and conceptualized record completeness as counts of days with medication orders and days with laboratory results. A non-parametric test was used to test the relationship between patient health and record completeness.

Primary Findings:

On average, patients with a more severe health status have a higher number of days with medication orders and a higher number of days with laboratory results. A Kruskal-Wallice one-way analysis of variance revealed a significant effect of ASA Class on number of days with medication orders ($x2(3)=332.0$, $p<0.0001$) and on number of days with laboratory results

(x2(3)=202.2, p<0.0001). Post-hoc analyses using a Bonferonni correct Wilcoxon rank sum test showed significant differences between all ASA Classes except Class 1 and Class 2 for both variables. Figure 5 shows the average number of days with data broken down by ASA class for both medication orders and laboratory results.

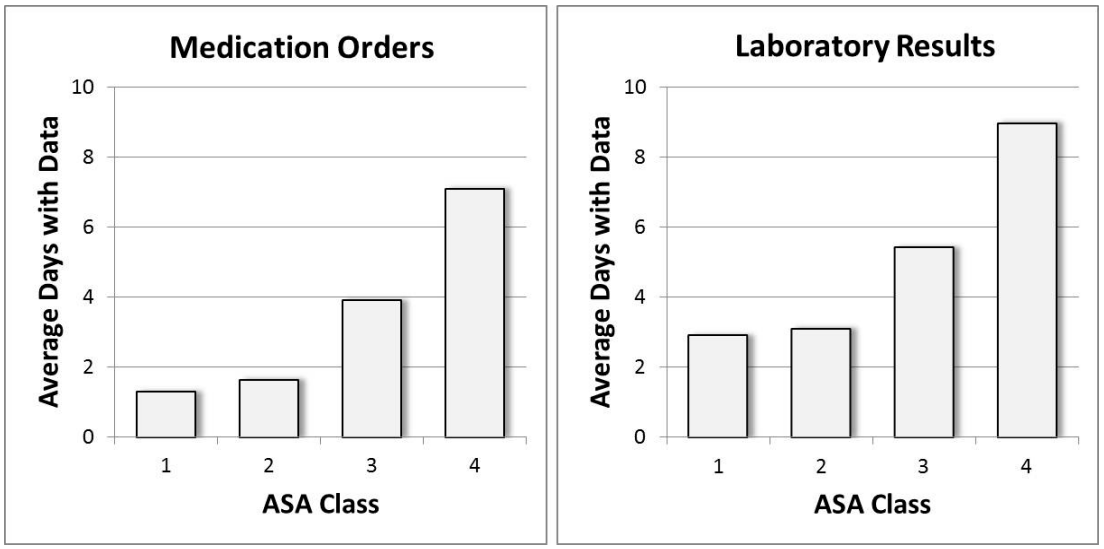


Figure 1.5 Average number of days with data per patient by ASA class.

These findings indicate two potential dangers to the secondary use of EHR data. First, a sample of patients drawn naively from the EHR-- assuming the records chosen are limited to those with sufficient data for the task at hand-- may be biased towards sicker patients. Second, researchers who wish to identify healthy controls for a study may have difficulty identifying patients with sufficient data. In both cases, the external validity of research conduct with EHR data may be compromised.

1.4.6 Aim V: Development and evaluation of EHR data quality assessment guideline

Objective: Use an iterative approach to the development of an EHR data quality assessment framework and guideline to be used by clinical researchers engaged in the secondary use of EHR data.

Methods and Materials:

The goal of this aim was to develop a data quality assessment framework and associated guidelines that took into account the findings of Aims I, II, and III. Specifically, the framework would combine the categories of data quality identified in Aims I and II and the methods of data quality assessment from Aim I and non-informatics literature. The concept of fitness for use, which was demonstrated in Aim III, also needed to be considered in the development process. The framework needed to be dynamic, since different secondary use scenarios would have different data requirements, and the determination of "sufficient" data quality had to be flexible. Finally, the framework and guideline were designed to allow users to bring their own expertise to bear on the data quality assessment process, as indicated by the results of Aim III.

The development of the framework and guidelines took place in two cycles. The initial development was followed by evaluation by content experts. Another round of development followed, as dictated by the experts' feedback. The evaluation used a basic psychometric approach, in which experts in the area of EHR data quality were asked to evaluate the framework and guidelines in terms of comprehensiveness, validity, clarity, and feasibility. Forced binary responses were used.

Primary Findings:

The resulting guideline is called 3x3 DQA, which refers to data quality assessment performed for three data quality constructs (correct, complete, current) across three data dimensions (patients, variables, time). The data quality constructs were selected and defined based upon the results from Aims I and II, and the data dimensions are taken from the basic dimensions of data flat files and data cubes. The full document is 26 pages long, and is divided into six primary sections: introduction, user guide, questions to determine the scope of the assessment, framework, assessment and reporting recommendations, and references. The framework, which is at the heart of the guideline and operationalizes each of the quality constructs across the data dimensions, is shown in Figure 5.

	A: COMPLETE	B: CORRECT	C: CURRENT
1: PATIENTS	**1A** Are there sufficient data points for each patient?	**1B** Is the distribution of values across patients plausible?	**1C** Were all data recorded during the timeframe of interest?
2: VARIABLES	**2A** Are there sufficient data points for each variable?	**2B** Is there concordance between variables?	**2C** Were variables recorded in the desired order?
3: TIME	**3A** Are there sufficient data points for each time?	**3B** Is the progression of data over time plausible?	**3C** Were data recorded with the desired regularity over time?

Figure 1.6 3x3 DQA framework.

The recommendations, which give actionable, knowledge- and evidence-based suggestions on how to measure and report data quality for each of the nine operationalizations in the framework, are divided into four parts. These include how to perform the measurement, how to report the findings, a simple example, and a list of relevant references. One of the two recommendations for assessing if data are correct across patients is shown in Figure 6.

The initial expert-based evaluation of 3x3 DQA included five participants. It is clear from the responses that certain aspects of the guideline require significant work. Although the participants liked the overall framework and approach, the wording of the constructs and operationalizations were subject to criticism regarding clarity and validity. Data quality assessment is an area in which clarity of meaning and intent is especially important. Responses to the recommendations were more positive, overall, though respondents requested further clarification regarding the role of the user. Specifically, there was concern about the frequency with which the user was asked to make judgments regarding which data were desired and how many data were sufficient. Some of the recommendations also require significant external knowledge, which was a cause of concern for some of the respondents.

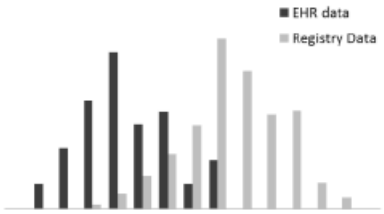
1B	The distribution of values across patients is plausible (approach II)
Measure	For variables that can be conceptualized as distributions of values across patients, identify reliable external sources of comparable data (e.g., national registry data). Select the appropriate statistical test given the variable type and distribution type. Also consider constructing graphical representations of the distributions to compare visually. Perform statistical and visual comparisons of your data and the external data source.
Report	Report the external data source, type of test(s) performed, and whether or not the results indicate that the distributions differ. How do the distributions differ (e.g. mean, skewness, etc.)? Are there possibly confounding differences between your population and the population for the external data source?
Example	Given a dataset of sixty patients and a continuous variable with a normal distribution and "research-quality" registry data for the same variable, a user can construct two normalized distributions, as shown below. A two-tailed t-test for this mock data results in $p < 0.01$. This specific example may indicate differences in population, or suggest systematic measurement error. 
References	Faulconer and de Lusignan 2004 Johnson, Mant et al. 1991 Iyen-Omofoman, Hubbard et al. 2011

Figure 1.7 One of two recommendations for data correctness across the patient dimension.

1.5 Contributions

The contributions of this dissertation fall into two categories. The most obvious—and primary—contribution of this dissertation is the initial iteration of 3x3 DQA, a guideline intended to assist clinical researchers in the assessment and reporting of EHR data quality within the context of secondary use. This guideline includes a framework and associated recommendations for data quality measurement and reporting, as well as a set of questions to help users determine

which elements are relevant for a given research task. While 3x3 DQA is not yet ready for dissemination, it has been well-received by evaluators, and is a promising step towards systematic, transparent, and evidence-based EHR data quality assessment.

The primary findings of each of the aims that led to the development of 3x3 DQA represent a set of "best practices" for EHR data quality assessment that can be used to direct informatics research independent from the guideline. These best practices include: literature-derived and user-derived sets of data quality categories that should be considered; knowledge-based and evidence-based data quality assessment methods that can be used; a need for a task-dependent approach that embraces fitness-for-use; the link between data quality and potential sampling bias in cohort selection; and the inclusion of user knowledge and user decision-making in the data quality assessment process (as opposed to a fully automated black-box approach).

1.6 Guide for the reader

Chapter 2 is an in-depth review of the literature on EHR data quality assessment within the context of secondary use. . I identified categories of data quality, methods of EHR data quality assessment, and several of the existing knowledge gaps in the current approaches to EHR data quality assessment.

Chapter 3 describes a qualitative study in which clinical researchers were asked about their experiences, beliefs, attitudes, and intentions as they relate to the secondary use of EHR data and EHR data quality.

Chapter 4 reports on a study of the completeness of EHR data, where completeness is operationalized for four different potential use cases. . This study demonstrates that EHR data quality is task-dependent.

Chapter 5 presents a follow-up study in which the relationship between completeness and underlying patient health status is explored, and demonstrates that selecting for patients with high quality EHR data may inadvertently bias samples towards sicker patients.

Chapter 6 details the development and evaluation of an EHR data quality assessment guideline.

Chapter 7 summarizes the contributions of the research described in this dissertation, as well as what I see as the next steps in EHR data quality assessment research and related areas of EHR data secondary use.

1.7 References

1. Blumenthal D. Stimulating the adoption of health information technology. N Engl J Med. 2009;360(15):1477-9.

2. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363(6):501-4.

3. Charles D, King J, Patel V, Furukawa M. ONC Data Brief No. 9: Adoption of Electronic Health record Systems among U.S. Non-federal Acute Care Hospitals: 2008-20012. The Office of the National Coordinator for Health Information Technology. 2013.

4. Hsiao C, Hing E. NCHS Data Brief No. 111: Use and characteristics of electronic health record systems among office-based physician practice: United States, 2001-2012. National Center for Health Statistics. 2012.

5. Xierali IM, Hsiao CJ, Puffer JC, Green LA, Rinaldo JC, Bazemore AW, et al. The rise of electronic health record adoption among family physicians. Annals of family medicine. 2013;11(1):14-9.

6. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1-9.

7. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care. 2007;13(6 Part 1):277-8.

8. National Center for Research Resources (U.S.). Widening the Use of Electronic Health Record Data for Research Bethesda, Md: National Institutes of Health; 2009 [cited 2011]. Available from: http://videocast.nih.gov/summary.asp?live=8062.

9. Sung NS, Crowley WF, Jr., Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA : the journal of the American Medical Association. 2003;289(10):1278-87.

10. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annual review of public health. 2006;27:1-28.

11. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005;365(9453):82-93.

12. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med. 2009;151(3):203-5.

13. Clancy C, Collins FS. Patient-Centered Outcomes Research Institute: the intersection of science and health care. Sci Transl Med. 2010;2(37):37cm18.

14. Text of the Recovery Act Related to Comparative Effectiveness Funding [updated March 2009June 6, 2012]. Available from: http://www.hhs.gov/recovery/programs/cer/recoveryacttext.html.

15. PCORI Establishing Legislation [June 6, 2012]. Available from: http://www.pcori.org/about/establishment/.

16. Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordyke RJ. Review: use of electronic medical records for health outcomes research: a literature review. Med Care Res Rev. 2009;66(6):611-38.

17. Institute of Medicine. The Learning Healthcare System. Washington, D.C.: The National Academies Press; 2007. Available from: http://www.nap.edu/catalog.php?record_id=11903.

18. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical care. 2013;51(8 Suppl 3):S30-7.

19. Weiner MG, Lyman JA, Murphy S, Weiner M. Electronic health records: high-quality electronic data for higher-quality clinical research. Inform Prim Care. 2007;15(2):121-7.

20. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? Ann Intern Med. 2009;151(5):359-60.

21. Burnum JF. The misinformation era: the fall of the medical record. Ann Intern Med. 1989;110(6):482-4.

22. van der Lei J. Use and abuse of computer-stored medical records. Methods of Information in Medicine. 1991;30(2).

23. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4(5):342-55.

24. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326(7398):1070.

25. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67(5):503-27.

26. Juran JM, Gryna FM. Juran's quality control handbook. 4th ed. New York: McGraw-Hill; 1988.

27. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems. 1996;12(4):5-34.

28. Liaw ST, Rahimi A, Ray P, Taggart J, Dennis S, de Lusignan S, et al. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. Int J Med Inform. 2013;82(1):10-24.

29. Pipino LL, Lee YW, Wang RY. Data quality assessment. Communications of the ACM. 2002;45(4):211-8.

30. International Monetary Fund. IMF Data Quality Assessment Framework (DQAF) Factsheet. 2003.

31. Ballou D, Pazer H. Modeling data and process quality in multi-input, multi-output information systems. Management Science. 1985;31(2):150-62.

32. Faulconer ER, de Lusignan S. An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar. Inform Prim Care. 2004;12(4):243-54.

33. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical care. 2012;50 Suppl:S21-9.

34. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60.

35. Green L, Kreuter M. Health promotion planning: an educational and ecological approach. 4th ed. Boston, MA: McGraw Hill; 2005.

36. Carlson J. Avoiding traps in member checking. The Qualitative Report. 2010;15(5):1102-13.

37. Hoffart N. A member check procedure to enhance rigor in naturalistic research. Western Journal of Nursing Research. 1991;13(4):522-35.

38. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educational Communication and Technology Journal. 1981;29(2):75-91.

39. Krefting L. Rigor in qualitative research: the assessment of trustworthiness. Am J Occup Ther. 1991;45(3):214-22.

Chapter 2: The State of the Electronic Health Record Data Quality Assessment Field[1]

2.1 Introduction

As the adoption of electronic health records (EHR) has made it easier to access and aggregate clinical data, there has been growing interest in conducting research with data collected during the course of clinical care.(1, 2) The NIH has called for increasing the reuse of electronic records for research, and the clinical research community has been actively seeking methods to enable secondary use of clinical data.(3) EHRs surpass many existing registries and data repositories in volume, and the reuse of these data may diminish the costs and inefficiencies associated with clinical research. Like other forms of retrospective research, studies that make

[1] This chapter was originally published in the Journal of the American Medical Informatics Association. The full citation for this publication is: Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51.

use of EHR data do not require patient recruitment or data collection, both of which are expensive and time-consuming processes. The data from EHRs also offer a window into the medical care, status, and outcomes of a diverse population that is representative of actual patients. The secondary use of data collected in EHRs is a promising step towards decreasing research costs, increasing patient-centered research, and speeding the rate of new medical discoveries.

Despite these benefits, reuse of EHR data has been limited by a number of factors, including concerns about the quality of the data and their suitability for research. It is generally accepted that, as a result of differences in priorities between clinical and research settings, clinical data are not recorded with the same care as research data.(4) Moreover, Burnum stated that the introduction of health information technology like EHRs has led not to improvements in the quality of the data being recorded, but rather to the recording of a greater quantity of bad data.(5) Due to such concerns about data quality, van der Lei warned specifically against the reuse of clinical data for research and proposed what he called the first law of informatics: "[d]ata shall be used only for the purpose for which they were collected."(6)

Although such concerns about data quality have existed since EHRs were first introduced, there remains no consensus as to the quality of electronic clinical data or even agreement as to what "data quality" actually means in the context of EHRs. One of the most broadly adopted conceptualizations of quality comes from Juran, who said that quality is defined through "fitness for use."(7) In the context of data quality, this means that data are of sufficient quality when they serve the needs of a given user pursuing specific goals.

Past study of EHR data quality has revealed highly variable results. Hogan and Wagner, in their 1997 literature review, found that the correctness of data ranged between 44% and 100%, and completeness between 1.1% and 100%, depending upon the clinical concepts being studied.(8) Similarly, Thiru et al., in calculating the sensitivity of different types of EHR data in the literature, found values ranging between 0.26 and 1.00.(9) In a 2010 review, Chan et al. looked at the quality of the same clinical concepts across multiple institutions, and still found a

great deal of variability. The completeness of blood pressure recordings, for example, fell anywhere between 0.1% and 51%.(10) Due to differences in measurement, recording, information systems, and clinical focus, the quality of EHR data is highly variable. Therefore, it is generally inadvisable to make assumptions about one EHR-derived dataset based on another. We need systematic methods that will allow us to assess the quality of an EHR-derived dataset for a given research task.

Our review primarily differs from those highlighted above in its focus. The previous reviews looked at data quality findings, while ours instead focuses on the methods that have been used to assess data quality. In fact, the earlier reviews were explicitly limited to studies that relied upon the use of a reference standard, while we instead explore a range of data quality assessment methods. The contributions of this literature review are an empirically based conceptual model of the dimensions of EHR data quality studied by clinical researchers and a summary and critique of the methods that have been used to assess EHR data quality, specifically within the context of reusing clinical data for research. Our goal is to develop a systematic understanding of the approaches that may be used to determine the suitability of EHR data for a specific research goal.

2.2 Methods

We identified articles in the literature by performing a search of the literature using standard electronic bibliographic tools. The literature search was performed by the first author on PubMed in February of 2012. As observed by Hogan and Wagner in their literature review, there is no Medical Subheadings (MeSH) term for data quality,(8) so a brief exploratory review was performed to identify relevant keywords. The final list included “data quality,” “data accuracy,” “data reliability,” “data validity,” “data consistency,” “data completeness,” and “data error.” The MeSH heading for EHRs was not introduced until 2010, so the older and more general MeSH heading “medical record systems, computerized” was used instead. The phrases “electronic

health record," "electronic medical record," and "computerized medical record" were also included in order to capture articles that may not have been tagged correctly. We searched for articles including at least one of the quality terms and at least one of the EHR terms. Results were limited to English language articles. The full query is shown below.

("data quality" OR "data accuracy" OR "data reliability" OR "data validity" OR "data consistency" OR "data completeness" OR "data errors" OR "data error") AND (electronic health record OR electronic medical record OR computerized medical record OR medical records systems, computerized [mh]) AND English[lang]

This search produced 230 articles, all of which were manually reviewed by the first author to determine if they met selection criteria. Specifically, the articles retained for further review 1) included original research using data quality assessment methods, 2) focused on data derived from an EHR or related system, and 3) were published in a peer-reviewed journal. Articles dealing with data from purely administrative systems (e.g. claims databases) were not included. These inclusion criteria resulted in 44 relevant articles. Next, we performed an in-depth ancestor search, reviewing the references of all of the articles in the original pool of 44. This allowed us to identify an additional 51 articles, resulting in a final pool of 95 articles meeting our inclusion criteria that then were used to derive results in this study.

From each article we abstracted the features of data quality examined, the methods of assessment used, and basic descriptive information including about the article and the type of data being studied. Through iterative review of the abstracted data, we derived broad dimensions of data quality and general categories of assessment strategies commonly described in the literature. Finally, we reviewed the 95 articles again, categorizing every article based upon the dimension or dimensions being assessed, as well as the assessment strategies used for each of those dimensions.

Prior to beginning this analysis, we searched for preexisting models of EHR data quality, but were unable to find any. We decided that the potential benefits of adapting a data quality

model from another field were outweighed by the risks of approaching our analysis through the lens of a model that had not been validated in the area of EHR data quality. Furthermore, using an existing model to guide analysis has the potential to obscure information contained in the data.(11) By imposing an existing model from a different discipline, we would have run the risk of missing important findings. Therefore, we decided to use an inductive, data-driven coding approach. This approach provides advantages over the deductive approach by allowing us better coverage of the dimensions and methods of data quality assessment.

2.3 Results

The majority of papers reviewed (73%) looked at structured data only, or at a combination of structured and unstructured data (22%). For our purposes, unstructured data types include free-entry text, while structured data types include coded data, values from pre-populated lists, or data entered into fields requiring specific alpha-numeric formats.

2.3.1 Summary of Data Quality Dimensions

Ignoring variations due to lexical categories and negation, the articles contained 27 unique terms describing dimensions of data quality. Features of data quality that were mentioned or described but not assessed were not included in our analysis. We grouped the terms together based on shared definitions. A few features of good data described in the literature, including sufficient granularity and the use of standards, were not included in our analyses. This decision was made due to the limited discussion of these features, the fact that they could be considered traits of good data practice instead of data quality, and because no assessment methods were described. Overall, we empirically derived five substantively different dimensions of data quality from the literature. The dimensions are defined below.

- *Completeness*: Is a truth about a patient present in the EHR?
- *Correctness*: Is an element that is present in the EHR true?

- *Concordance*: Is there agreement between elements in the EHR, or between the EHR and another data source?
- *Plausibility*: Does an element in the EHR makes sense in light of other knowledge about what that element is measuring?
- *Currency*: Is an element in the EHR a relevant representation of the patient state at a given point in time?

The list of data quality terms and their mappings to the five dimensions described above are shown in Table 1. The terms chosen to denote each of the dimensions were the clearest and least ambiguous from each of the groups. There was a great deal of variability and overlap in the terms used to describe each of these dimensions. "Accuracy," for example, was sometimes used as a synonym for correctness, but in other articles meant both correctness and completeness. The dimensions themselves, however, were abstracted in such a way as to be exhaustive and mutually exclusive based upon their definitions. Every article identified could be matched to one or more of the dimensions.

Table 2.1 Terms from literature to denote five common dimensions of EHR data quality.

Completeness	*Correctness*	*Concordance*	*Plausibility*	*Currency*
Accessibility	Accuracy	Agreement	Accuracy	Recency
Accuracy	Corrections made	Consistency	Believability	Timeliness
Availability	Errors	Reliability	Trustworthiness	
Missingness	Misleading	Variation	Validity	
Omission	Positive predictive value			
Presence	Quality			
Quality	Validity			
Rate of recording				
Sensitivity				
Validity				

2.3.2 Summary of Data Quality Assessment Methods

A similar process was used to identify the most common methods of data quality assessment. The strategies used to assess the dimensions of data quality fell into seven broad categories of methods, many of which were used to assess multiple dimensions. These general methods are listed and defined below.

- *Gold standard*: A dataset drawn from another source or multiple sources, with or without information from the EHR, is used as a gold standard.
- *Data element agreement*: Two or more elements within an EHR are compared to see if they report the same or compatible information.
- *Element presence*: A determination is made as to whether or not desired or expected data elements are present.
- *Data source agreement*: Data from the EHR are compared to data from another source to determine if they are in agreement.
- *Distribution comparison*: Distributions or summary statistics of aggregated data from the EHR are compared to the expected distributions for the clinical concepts of interest.
- *Validity check*: Data in the EHR are assessed using various techniques that determine if values "make sense."
- *Log review*: Information on the actual data entry practices (e.g. dates, times, edits) is examined.

A summary of which methods were used to assess which dimensions is shown in Table 2. The graph in Figure 1 shows the strength of the pairwise relationships between the dimensions and methods. Some of the methods were used to assess only certain dimensions of data quality, which other methods were applied more broadly. Element presence, for example, was used to

assess completeness, but none of the other dimensions. Data element agreement and data source agreement, however, were applied more broadly. Most of the dimensions were assessed using an assortment of methods, but currency was only measured using a single approach.

Table 2.2 Dimensions of data quality and methods of data quality assessment. In decreasing order of frequency, the dimensions are listed from left to right, and the methods from top to bottom. The numbers in the cells correspond to the article counts.

Dimension / *Method*	*Completeness*	*Correctness*	*Concordance*	*Plausibility*	*Currency*	
Gold standard	12-35	12-23, 25, 26, 28-35, 42-44, 58, 73-81				*37*
Data element agreement	57, 59-65	46, 50, 51, 57, 59, 61-63, 65, 82-89	38, 40, 51, 60, 95-97	44, 93		*26*
Element presence	36-58					*23*
Data source agreement	47, 66-68	94	98-103	105		*11*
Distribution comparison	69-72		31, 46, 69	59, 104, 106		*10*
Validity checks		32, 84, 90-92		68, 93		*7*
Log review		93			42, 54, 61, 79	*5*
	61	*57*	*16*	*7*	*4*	

2.3.2.1 Completeness

Completeness was the most commonly assessed dimension of data quality and was an area of focus in 61 (64%) of the articles. Generally speaking, completeness referred to whether or not a truth about a patient was present in the EHR. Most of the articles used the term completeness to describe this dimension, but some also referred to data availability or missing data. In others, completeness was subsumed into more general concepts like accuracy or quality.

Some articles cited the statistical definition of completeness suggested by Hogan and Wagner, in which completeness is equivalent to sensitivity.(8)

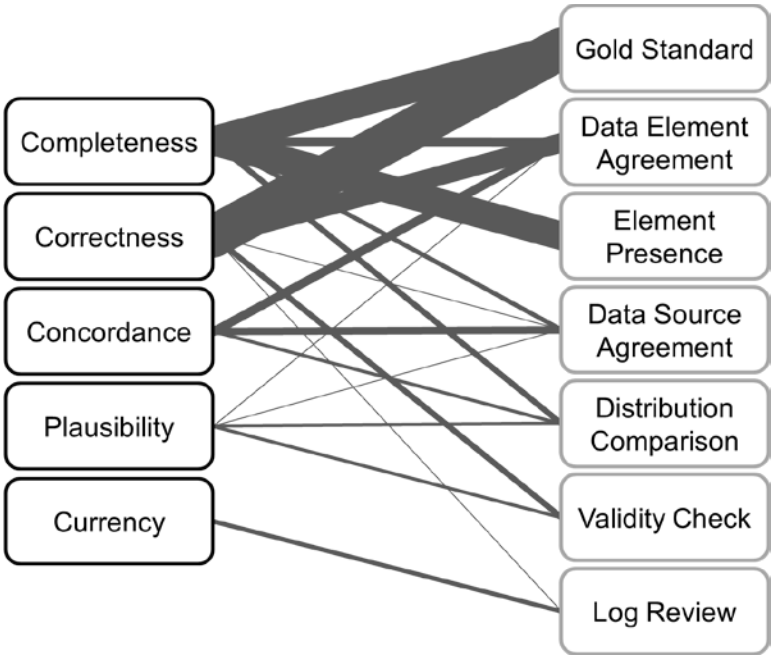


Figure 2.1 Mapping between dimensions of data quality and data quality assessment. The weight of the edge connecting a dimension and method indicates the relative frequency of that combination.

Many articles assessed EHR data completeness by using another source of data as a gold standard. The gold standards used included concurrently kept paper records,(12-18) information supplied by patients,(19-21) review of data by patients,(22-25) clinical encounters with patients,(26-28) information presented by trained standard patients,(29, 30), information requested from the treating physician,(31) and alternate data sources from which EHR elements were abstracted.(32, 33) A similar approach involved triangulating data from multiple sources within the EHR to create a gold standard.(34, 35)

Other researchers simply looked at the presence or absence of elements in the EHR. In some cases, these were elements that were expected to be present, even if they were not needed for any specific task.(36-50) In other situations, the elements examined were dependent upon the task at hand, meaning that the researchers determined whether or not the EHR data were complete

enough for a specific purpose.(51-58) Other methods for assessing completeness included looking at agreement between elements from the same source,(57, 59-65) agreement between the EHR and paper records,(47, 66, 67) agreement between the EHR and another electronic source of data,(47, 68) and comparing distributions of occurrences of certain elements between practices(69) or with nationally recorded rates.(70-72)

2.3.2.2 Correctness

The second most commonly assessed dimension of data quality was correctness, which was included in 57 (60%) of the articles. EHR data were considered correct when the information they contained was true. Other terms that were commonly used to describe this concept included accuracy, error, and quality. Occasionally, correctness included completeness, due to the fact that some researchers consider missing data to be incorrect (i.e., errors of omission). The definition of correctness suggested by Hogan and Wagner states that data correctness is the proportion of data elements present that are correct, which is equivalent to positive predictive value.(8)

Comparison of EHR data to a gold standard was by far the most frequently used method for assessing correctness. These gold standards included: paper records;(12-18, 44, 73, 74) information supplied by patients through interviews(19, 20, 42, 75), questionnaires,(21, 76) data review,(22, 23, 25, 77) or direct data entry;(43) clinical encounters with patients;(26, 28, 78) information presented by trained standard patients;(29, 30) automatically recorded data;(79) contact with the treating physician;(31, 58, 80) and alternate data sources from which information matching EHR elements were abstracted.(32, 33) Some researchers developed gold standards by extracting and triangulating data from within the EHR.(34, 35, 76, 81)

The second most common approach to assessing correctness was to look at agreement between elements within the EHR. Usually this involved verifying a diagnosis by looking at associated procedures, medications, or lab values.(46, 50, 57, 59, 61-63, 82, 83) Similarly, some

articles reported on agreement between related elements(65, 84) and errors identified through the examination of the use copy and paste practices(85, 86). Other researchers looked specifically at agreement between structured elements and unstructured data within EHRs.(51, 87) One of the more formal approaches described for assessing correctness was the data quality probe, proposed by Brown and Warmington, which is a query that, when run against an EHR database, only returns cases with some disagreement between data elements.(88, 89)

A few articles described the use of validity checks to assess correctness. These included review of changes of sequential data over time(90), identifying end digit preferences in blood pressure values(91, 92), and comparing elements to their expected value ranges. (32, 84) Two other approaches to were using corrections seen in log files as a proxy for correctness,(93) and comparing data on the same patients from a registry and an EHR.(94)

2.3.2.3 Concordance

Sixteen (17%) of the articles reviewed assessed concordance. Data were considered concordant when there was agreement or compatibility between data elements. This may mean that two elements recording the same information for a single patient have the same value, or that elements recording different information have values that make sense when considered together (e.g. biological sex is recorded as female, and procedure is recorded as gynecologic exam). Measurement of concordance is generally based upon elements contained within the EHR, but some researchers also included information from other data sources. Common terms used in the literature to describe data concordance include agreement and consistency.

The most common approach to assessing concordance was to look at agreement between elements within the EHR,(38, 40, 51, 95) especially diagnoses and associated information like medications or procedures.(60, 96, 97) The second most common method used to assess concordance was to look at the agreement of EHR data with data from other sources. These other sources included billing information,(98) paper records,(99-101) patient-reported data,(102) and

physician-reported data.(103) Another approach was to compare distributions of data within the EHR to distributions of the same information from similar medical practices (46, 69) or with national rates.(31)

2.3.2.4 Plausibility

Seven (7%) of the articles assessed the plausibility of EHR data. In this context, data were plausible if they were in agreement with general medical knowledge or information and were therefore feasible. In other words, assessments of plausibility were intended to determine whether or not data could be trusted or if they were of suspect quality. Other terms that were used to discuss and describe EHR data plausibility include data validity and integrity.

The most common approach to assessing the plausibility of EHR data was to perform some sort of validity check to determine if specific elements within the EHR were likely to be true or not. This included looking for elements with values that were outside biologically plausible ranges or that changed implausibly over time(68) or zero-valued elements.(93) Other researchers compared distributions of data values between practices(59, 104) or with national rates,(105, 106) or looked at agreement between related elements.(44, 93)

2.3.2.5 Currency

Currency of EHR data was assessed in four (4%) of the 95 articles. Currency was often referred to in the literature as timeliness or recency. Data were considered current if they were recorded in the EHR within a reasonable period of time following measurement or, alternatively, if they were representative of the patient state at a desired time of interest. In all four articles, currency was assessed through the review of data entry logs. In three of the four, researchers reviewed whether desired data were entered into the EHR within a set time limit.(42, 61, 79) In the fourth, researchers considered whether each type of data element was measured recently enough to be considered medically relevant.(54)

2.4 Discussion

We identified five dimensions of data quality and seven categories of data quality assessment methods. Examination of the types of methods used, as well as overlap of the methods between dimensions, reveals significant patterns and gaps in knowledge. Below, we explore the major findings of the literature review, specifically highlighting areas that require further attention, and make suggestions for future research.

2.4.1 Terminology and dimensions of data quality

One of the biggest difficulties in conducting this review resulted from the inconsistent terminology used to discuss data quality. We had not expected, for example, the overlap of terms between dimensions, or the fact that the language within a single article was sometimes inconsistent. The clinical research community has largely failed to develop or adopt a consistent taxonomy of data quality.

There is, however, overlap between the dimensions of data quality identified during this review and those described in pre-existing taxonomies and models of data quality. Wang and Strong's conceptual framework of data quality, for example, contains fifteen dimensions, grouped into four categories: intrinsic, contextual, representational, and accessibility.(107) Our review focused on intrinsic (inherent to the data) and contextual (task-dependent) data quality issues. The dimensions we identified overlapped with two of the intrinsic features (accuracy and believability, which are equivalent to correctness and plausibility) and two of the contextual features (timeliness, which is equivalent to currency, and completeness). The only dimension we identified that does not appear in Wang and Strong's framework is concordance.

The Institute of Medicine (IOM) identified four attributes of data quality relevant to patient records: completeness, accuracy, legibility, and meaning (related to comprehensibility).(108) As the IOM points out, electronic records by their nature negate many of the concerns regarding legibility, so we are left with three relevant attributes, two of which we

identified through our review. Meaning is a more abstract concept and is likely difficult to measure objectively, which may be why we did not observe assessments of this dimension in the literature.

Although the five dimensions of data quality derived during our review were treated as mutually exclusive within the literature, we feel that only three can be considered fundamental: correctness, completeness, and currency. By this we mean that these dimensions are non-reducible, and describe core concepts of data quality as it relates to EHR data reuse. Concordance and plausibility, on the other hand, while discussed as separate features of data quality, appear to serve as proxies for the fundamental dimensions when it is not possible to assess them directly. This supposition is supported by the overlap observed in the methods used to assess concordance and plausibility with those used to assess correctness and completeness. A lack of concordance between two data sources, for example, indicates error in one or both of those sources: an error of omission, resulting in a lack of completeness, or an error of commission, resulting in a lack of correctness. Similarly, data that do not appear to be plausible may be incorrect, as in the case of a measurement that fails a range check, or incomplete, such as aggregated diagnosis rates within a practice that do not match the expected population rates. It may be that correctness, completeness, and currency are properties of data quality, while plausibility and concordance are methodological approaches to assessing data quality. Additionally, researchers may refer to plausibility or concordance when they believe that there are problems with completeness or correctness, but have no way to be certain that errors exist or which data elements might be wrong.

2.4.2 Data quality assessment methodology

We observed a number of noteworthy patterns within the literature in terms of the types of data quality assessments used and the manner in which data quality assessment was discussed. For example, 37 of the 95 articles in our sample relied upon a gold standard to assess data quality.

There are a few problems with this approach. First, the data sources used could rarely be considered true gold standards. Paper records, for example, may sometimes be more trusted than electronic records, but they should not be considered entirely correct or complete. Perhaps more importantly, a gold standard for EHR data is simply not available in most cases. This will become more problematic as the use of de-identified datasets for research becomes more common. A "fitness for purpose" approach, which suggests that the quality of each dataset compiled for a specific task must be assessed, necessitates the adoption of alternatives to gold standard-based methods.

In addition to the overreliance upon gold standards, the majority of the studies we identified relied upon an "intuitive" understanding data quality and used ad hoc methods to assess data quality. This tendency has been observed in other fields as well.(107, 109) Most of the studies included in this review presented assessment methodologies that were developed with minimal empirical or theoretical basis. Only a few researchers made the effort to develop generalizable approaches that could be used as a step towards a standard methodology. Faulconer and de Lusignan, for example, proposed a multi-step, statistically driven approach to data quality assessment.(61) Hogan and Wagner suggested specific statistical measures of the correctness and completeness of EHR elements that have been adopted by other researchers.(8) Certain methods, including comparing distributions of data from the EHR to expected distributions or looking for agreement between elements within the EHR, lend themselves more readily to generalization. Brown and Warmington's data quality probes, for example, could be extended to various data elements, though they require detailed clinical knowledge to implement.(88, 89) Some researchers looking at the quality of research databases pulled from general practices in the United Kingdom have adopted relatively consistent approaches to comparing the distributions of data concerning specific clinical phenomena to information from registries and surveys.(31, 70, 71, 106) In most cases, however, the specific assessment methods described in the literature would be difficult to apply to other datasets or research questions. If the reuse of EHR data for

clinical research is to become common and feasible, development of standardized, systematic data quality assessment methods is vital.

Additionally, if as a field we intend to adopt the concept of "fitness for purpose", it is important to consider the intended research use of EHR data when determining if they are of sufficient quality.(110) Some dimensions may prove to be more task-dependent, or subjective, while others are essentially task-independent, or objective.(109) It will be important to develop a full understanding of the interrelationships of research tasks and data characteristics as they relate to data quality. For example, the completeness of a set of data elements required by one research protocol may differ from the completeness required for a different protocol. Many factors, including clinical focus, required resolution of clinical information, and desired effect size, can affect the suitability of a dataset for a specific research task.(26)

2.4.3 Future Directions

We believe that efforts to reuse EHR data for clinical research would benefit most from work in a few specific areas: adopting a consistent taxonomy of EHR data quality; increasing awareness of task-dependence; integrating work on data quality assessment from other fields; and adopting systematic, statistically-based methods of data quality assessment. A taxonomy of data quality would enable a structured discourse and contextualize assessment methodologies. The findings in this review regarding the dimensions of data quality may serve as a stepping-stone towards this goal. Task-dependence is likely to become a growing issue as efforts to reuse EHR data for research increase, particularly since data quality assessment does not have a one-size-fits-all solution. One approach to addressing the problem of EHR data quality and suitability for reuse in research would be to look at what has been done outside of clinical research, since data quality has been an area of study in fields ranging from finance to industrial engineering. Finally, it is important that the clinical researchers begin to move away from ad hoc approaches to data

quality assessment. Validated methods that can be adapted for different research questions are the ideal goal.

2.4.4 Limitations

There were a number of limitations to this review. First, the search was limited. Due to the lack of MeSH term for data quality and the variation in terminology used to discuss data quality, it is possible that our original search may have missed some relevant articles. We believe that our decision to review the references of each article improved the saturation of our sample.

It is also important to note that our classification process was largely subjective and was performed by only one of the authors. It is possible that the original researchers might disagree with our interpretations. We chose to use an iterative process to label and categorize the dimensions of data quality and methods of assessment described in each article in an effort to develop a consistent coding scheme.

Finally, it is likely that the dearth of literature discussing data quality in the reuse of EHR data for clinical research is due in part to underreporting. A common first step in analyzing any dataset is to review distributions, summary statistics, and histograms, but this process is rarely described in publications. Such methods are therefore likely to be underrepresented in this review. Greater transparency regarding data cleaning or checking steps would be advisable, as it could help to establish acceptable reporting standards for the reuse of EHR data in research.

2.5 Conclusion

The secondary use of EHR data is a promising area of research. However, the problems with EHR data quality necessitate the use of quality assessment methodologies to determine the suitability of these data for given research tasks. In this review of the literature we have identified the major dimensions of data quality that are of interest to researchers, as well as the general assessment techniques that have been utilized. Data quality is not a simple problem, and if the reuse of EHR data is to become an accepted approach to medical research, the clinical research

community needs to develop validated, systematic methods of EHR data quality assessment. We encourage researchers to be consistent in their discussion of the dimensions of data quality, systematic in their approaches to measuring data quality, and to develop and share best practices for the assessment of EHR data quality in the context of reuse for clinical research.

2.6 References

1. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1-9.

2. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care. 2007;13(6 Part 1):277-8.

3. National Center for Research Resources (U.S.). Widening the Use of Electronic Health Record Data for Research Bethesda, Md: National Institutes of Health; 2009 [cited 2011]. Available from: http://videocast.nih.gov/summary.asp?live=8062.

4. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? Ann Intern Med. 2009;151(5):359-60.

5. Burnum JF. The misinformation era: the fall of the medical record. Ann Intern Med. 1989;110(6):482-4.

6. van der Lei J. Use and abuse of computer-stored medical records. Methods of Information in Medicine. 1991;30(2).

7. Juran JM, Gryna FM. Juran's quality control handbook. 4th ed. New York: McGraw-Hill; 1988.

8. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4(5):342-55.

9. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326(7398):1070.

10. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67(5):503-27.

11. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. American Journal of Evaluation. 2005;27(2):237-46.

12. Ayoub L, Fu L, Pena A, Sierra JM, Dominguez PC, Pui CH, et al. Implementation of a data management program in a pediatric cancer unit in a low income country. Pediatr Blood Cancer. 2007;49(1):23-7.

13. Barrie JL, Marsh DR. Quality of data in the Manchester orthopaedic database. BMJ (Clinical Research Ed). 1992;304(6820):159-62.

14. Hohnloser JH, Fischer MR, Konig A, Emmerich B. Data quality in computerized patient records. Analysis of a haematology biopsy report database. Int J Clin Monit Comput. 1994;11(4):233-40.

15. Mikkelsen G, Aasly J. Consequences of impaired data quality on information retrieval in electronic patient records. Int J Med Inform. 2005;74(5):387-94.

16. Ricketts D, Newey M, Patterson M, Hitchin D, Fowler S. Markers of data quality in computer audit: the Manchester Orthopaedic Database. Annals of the Royal College of Surgeons of England. 1993;75(6):393-6.

17. Roukema J, Los RK, Bleeker SE, van Ginneken AM, van der Lei J, Moll HA. Paper versus computer: feasibility of an electronic medical record in general pediatrics. Pediatrics. 2006;117(1):15-21.

18. Wallace CJ, Stansfield D, Gibb Ellis KA, Clemmer TP. Implementation of an electronic logbook for intensive care units. Proc AMIA Symp. 2002:840-4.

19. Kaboli PJ, McClimon BJ, Hoth AB, Barnett MJ. Assessing the accuracy of computerized medication histories. Am J Manag Care. 2004;10(11 Pt 2):872-7.

20. Porter SC, Mandl KD. Data quality and the electronic medical record: a role for direct parental data entry. Proc AMIA Symp. 1999:354-8.

21. Whitelaw FG, Nevin SL, Milne RM, Taylor RJ, Taylor MW, Watt AH. Completeness and accuracy of morbidity and repeat prescribing records held on general practice computers in Scotland. The British Journal of General Practice: The Journal of the Royal College of General Practitioners. 1996;46(404):181-6.

22. Pyper C, Amery J, Watson M, Crook C. Patients' experiences when accessing their on-line electronic patient records in primary care. British Journal of General Practice. 2004;54(498):38-43.

23. Staroselsky M, Volk LA, Tsurikova R, Newmark LP, Lippincott M, Litvak I, et al. An effort to improve electronic health record medication list accuracy between visits: patients' and physicians' response. Int J Med Inform. 2008;77(3):153-60.

24. Staroselsky M, Volk LA, Tsurikova R, Pizziferri L, Lippincott M, Wald J, et al. Improving electronic health record (EHR) accuracy and increasing compliance with health maintenance clinical guidelines through patient access and input. Int J Med Inform. 2006;75(10-11):693-700.

25. Weingart SN, Cleary A, Seger A, Eng TK, Saadeh M, Gross A, et al. Medication reconciliation in ambulatory oncology. Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources. 2007;33(12):750-7.

26. Logan JR, Gorman PN, Middleton B. Measuring the quality of medical records: a method for comparing completeness and correctness of clinical encounter data. Proc AMIA Symp. 2001:408-12.

27. Smith PC, Araya-Guerra R, Bublitz C, Parnes B, Dickinson LM, Van Vorst R, et al. Missing clinical information during primary care visits. JAMA : the journal of the American Medical Association. 2005;293(5):565-71.

28. Bentsen BG. The accuracy of recording patient problems in family practice. Journal of Medical Education. 1976;51(4):311-6.

29. Berner ES, Kasiraman RK, Yu F, Ray MN, Houston TK. Data quality in the outpatient setting: impact on clinical decision support systems. AMIA Annu Symp Proc. 2005:41-5.

30. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. Medical care. 2004;42(11):1066-72.

31. Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. Pharmacoepidemiol Drug Saf. 2004;13(7):437-41.

32. Staes CJ, Bennett ST, Evans RS, Narus SP, Huff SM, Sorensen JB. A case for manual entry of structured, coded laboratory data from multiple sources into an ambulatory electronic health record. J Am Med Inform Assoc. 2006;13(1):12-5.

33. Madsen KM, Schonheyder HC, Kristensen B, Nielsen GL, Sorensen HT. Can hospital discharge diagnosis be used for surveillance of bacteremia? A data quality study of a Danish hospital discharge registry. Infect Control Hosp Epidemiol. 1998;19(3):175-80.

34. Aronsky D, Haug PJ. Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index. J Am Med Inform Assoc. 2000;7(1):55-65.

35. Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. J Am Med Inform Assoc. 1996;3(3):234-44.

36. Einbinder JS, Rury C, Safran C. Outcomes research using the electronic patient record: Beth Israel Hospital's experience with anticoagulation. Proceedings / the Annual Symposium on Computer Application [sic] in Medical Care Symposium on Computer Applications in Medical Care. 1995:819-23.

37. Forster M, Bailey C, Brinkhof MW, Graber C, Boulle A, Spohr M, et al. Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings. Bulletin of the World Health Organization. 2008;86(12):939-47.

38. Goulet JL, Erdos J, Kancir S, Levin FL, Wright SM, Daniels SM, et al. Measuring performance directly using the veterans health administration electronic medical record: a comparison with external peer review. Medical Care. 2007;45(1):73-9.

39. Hahn KA, Ohman-Strickland PA, Cohen DJ, Piasecki AK, Crosson JC, Clark EC, et al. Electronic Medical Records Are Not Associated With Improved Documentation in Community Primary Care Practices. American Journal of Medical Quality. 2011;26(4):272-7.

40. Jelovsek F, Hammond W. Formal error rate in a computerized obstetric medical record. Methods Inf Med. 1978;17(3):151-7.

41. Jones RB, Hedley AJ. A computer in the diabetic clinic.Completeness of data in a clinical information system for diabetes. Practical Diabetes International. 1986;3(6):295-6.

42. Ndira SP, Rosenberger KD, Wetter T. Assessment of data quality of and staff satisfaction with an electronic health record system in a developing country (Uganda): a qualitative and quantitative comparative study. Methods Inf Med. 2008;47(6):489-98.

43. Olola CH, Narus S, Poynton M, Nebeker J, Hales J, Rowan B, et al. Patient-perceived usefulness of an emergency medical card and a continuity-of-care report in enhancing the quality of care. Int J Qual Health Care. 2011;23(1):60-7.

44. Pearson N, O'Brien J, Thomas H, Ewings P, Gallier L, Bussey A. Collecting morbidity data in general practice: the Somerset morbidity project. BMJ. 1996;312(7045):1517-20.

45. Porcheret M, Hughes R, Evans D, Jordan K, Whitehurst T, Ogden H, et al. Data quality of general practice electronic health records: the impact of a program of assessments, feedback, and training. J Am Med Inform Assoc. 2004;11(1):78-86.

46. Pringle M, Ward P, Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. Br J Gen Pract. 1995;45(399):537-41.

47. Scobie S, Basnett I, McCartney P. Can general practice data be used for needs assessment and health care planning in an inner-London district? Journal of Public Health Medicine. 1995;17(4):475-83.

48. Soto CM, Kleinman KP, Simon SR. Quality and correlates of medical record documentation in the ambulatory care setting. BMC Health Services Research. 2002;2(1):22-.

49. Tang PC, LaRosa MP, Gorden SM. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. Journal of the American Medical Informatics Association: JAMIA. 1999;6(3):245-51.

50. Thiru K, De Lusignan S, Hague N. Have the completeness and accuracy of computer medical records in general practice improved in the last five years? The report of a two-practice pilot study Health Informatics Journal. 1999;5(4):224-32.

51. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. AMIA Summits Transl Sci Proc. 2010;2010:1-5.

52. de Lusignan S, Hague N, Brown A, Majeed A. An educational intervention to improve data recording in the management of ischaemic heart disease in primary care. J Public Health (Oxf). 2004;26(1):34-7.

53. Jensen RE, Chan KS, Weiner JP, Fowles JB, Neale SM. Implementing electronic health record-based quality measures for developmental screening. Pediatrics. 2009;124(4):e648-54.

54. Williams JG. Measuring the completeness and currency of codified clinical information. Methods Inf Med. 2003;42(4):482-8.

55. Agnew-Blais JC, Coblyn JS, Katz JN, Anderson RJ, Mehta J, Solomon DH. Measuring quality of care for rheumatic diseases using an electronic medical record. Annals of the Rheumatic Diseases. 2009;68(5):680-4.

56. Asche C, Said Q, Joish V, Hall CO, Brixner D. Assessment of COPD-related outcomes via a national electronic medical record database. International Journal of Chronic Obstructive Pulmonary Disease. 2008;3(2):323-6.

57. Linder JA, Kaleba EO, Kmetik KS. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. Medical Care. 2009;47(2):208-16.

58. Lo Re V, 3rd, Haynes K, Forde KA, Localio AR, Schinnar R, Lewis JD. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. Pharmacoepidemiol Drug Saf. 2009;18(9):807-14.

59. de Lusignan S, Chan T, Wood O, Hague N, Valentin T, Van Vlymen J. Quality and variability of osteoporosis data in general practice computer records: implications for disease registers. Public Health. 2005;119(9):771-80.

60. de Lusignan S, Valentin T, Chan T, Hague N, Wood O, van Vlymen J, et al. Problems with primary care data quality: osteoporosis as an exemplar. Inform Prim Care. 2004;12(3):147-56.

61. Faulconer ER, de Lusignan S. An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar. Inform Prim Care. 2004;12(4):243-54.

62. Hassey A, Gerrett D, Wilson A. A survey of validity and utility of electronic patient records in a general practice. BMJ. 2001;322(7299):1401-5.

63. Tang PC, Ralston M, Arrigotti MF, Qureshi L, Graham J. Comparison of Methodologies for Calculating Quality Measures Based on Administrative Data versus Clinical Data from an Electronic Health Record System: Implications for Performance Measures. Journal of the American Medical Informatics Association : JAMIA. 2007;14(1):10-5.

64. Hohnloser JH, Puerner F, Soltanian H. Improving coded data entry by an electronic patient record system. Methods of Information in Medicine. 1996;35(2):108-11.

65. Horsfield P. Trends in data recording by general practice teams: an analysis of data extracted from clinical computer systems by the PRIMIS project. Informatics in Primary Care. 2002;10(4).

66. Jick H, Terris BZ, Derby LE, Jick SS. Further validation of information recorded on a general practitioner based computerized data resource in the united kingdom. Pharmacoepidemiology and Drug Safety. 1992;1(6):347-9.

67. Neal RD, Heywood PL, Morley S. Real world data--retrieval and validation of consultation data from four general practices. Family Practice. 1996;13(5):455-61.

68. Noel PH, Copeland LA, Perrin RA, Lancaster AE, Pugh MJ, Wang CP, et al. VHA Corporate Data Warehouse height and weight data: opportunities and challenges for health services research. J Rehabil Res Dev. 2010;47(8):739-50.

69. de Lusignan S, Chan T, Wells S, Cooper A, Harvey M, Brew S, et al. Can patients with osteoporosis, who should benefit from implementation of the national service framework for older people, be identified from general practice computer records? A pilot study that illustrates the variability of computerized medical records and problems with searching them. Public Health. 2003;117(6):438-45.

70. Haynes K, Forde KA, Schinnar R, Wong P, Strom BL, Lewis JD. Cancer incidence in The Health Improvement Network. Pharmacoepidemiol Drug Saf. 2009;18(8):730-6.

71. Iyen-Omofoman B, Hubbard RB, Smith CJ, Sparks E, Bradley E, Bourke A, et al. The distribution of lung cancer across sectors of society in the United Kingdom: A study using national primary care data. BMC Public Health. 2011;11(1):857.

72. Johnson N, Mant D, Jones L, Randall T. Use of computerised general practice data for population surveillance: comparative study of influenza data. BMJ (Clinical Research Ed). 1991;302(6779):763-5.

73. Dambro MR, Weiss BD. Assessing the quality of data entry in a computerized medical records system. J Med Syst. 1988;12(3):181-7.

74. Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis of psychosis on general practice computer system. BMJ : British Medical Journal. 1993;307(6895):32-4.

75. Dawson C, Perkins M, Draper E, Johnson A, Field D. Are outcome data regarding the survivors of neonatal care available from routine sources? Arch Dis Child Fetal Neonatal Ed. 1997;77(3):F206-10.

76. Van Weel C. Validating long term morbidity recording. Journal of Epidemiology and Community Health. 1995;49 Suppl 1:29-32.

77. Powell J, Fitton R, Fitton C. Sharing electronic health records: the patient view. Inform Prim Care. 2006;14(1):55-7.

78. Meara, Bhowmick BK, Hobson P. Accuracy of diagnosis in patients with presumed Parkinson's disease. Age and Ageing. 1999;28(2):99-102.

79. Vawdrey DK, Gardner RM, Evans RS, Orme JF, Jr., Clemmer TP, Greenway L, et al. Assessing data quality in manual entry of ventilator settings. J Am Med Inform Assoc. 2007;14(3):295-303.

80. Van Staa TP, Abenhaim L, Cooper C, Zhang B, Leufkens HG. The use of a large pharmacoepidemiological database to study exposure to oral corticosteroids and risk of fractures: validation of study population and results. Pharmacoepidemiology and Drug Safety. 2000;9(5):359-66.

81. Margulis AV, Garcia Rodriguez LA, Hernandez-Diaz S. Positive predictive value of computerized medical records for uncomplicated and complicated upper gastrointestinal ulcer. Pharmacoepidemiol Drug Saf. 2009;18(10):900-9.

82. de Lusignan S, Khunti K, Belsey J, Hattersley A, van Vlymen J, Gallagher H, et al. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. Diabet Med. 2010;27(2):203-9.

83. de Burgos-Lunar C, Salinero-Fort MA, Cardenas-Valladolid J, Soto-Diaz S, Fuentes-Rodriguez CY, Abanades-Herranz JC, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. BMC Med Res Methodol. 2011;11:146.

84. Basden A, Clark EM. Data integrity in a general practice computer system (Clinics). International Journal of Bio-Medical Computing. 1980;11(6):511-9.

85. Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are Electronic Medical Records Trustworthy? Observations on Copying, Pasting and Duplication. AMIA Annual Symposium Proceedings. 2003;2003:269-73.

86. Weir CR, Hurdle JF, Felgar MA, Hoffman JM, Roth B, Nebeker JR. Direct text entry in electronic progress notes. An evaluation of input errors. Methods of Information in Medicine. 2003;42(1):61-7.

87. Hogan WR, Wagner MM. Free-text fields change the meaning of coded data. Proc AMIA Annu Fall Symp. 1996:517-21.

88. Brown PJ, Warmington V. Data quality probes-exploiting and improving the quality of electronic patient record data and patient care. Int J Med Inform. 2002;68(1-3):91-8.

89. Brown PJ, Warmington V. Info-tsunami: surviving the storm with data quality probes. Inform Prim Care. 2003;11(4):229-33; discussion 34-7.

90. Haerian K, McKeeby J, Dipatrizio G, Cimino JJ. Use of clinical alerting to improve the collection of clinical research data. AMIA Annu Symp Proc. 2009;2009:218-22.

91. Alsanjari ON, de Lusignan S, van Vlymen J, Gallagher H, Millett C, Harris K, et al. Trends and transient change in end-digit preference in blood pressure recording: studies of sequential and longitudinal collected primary care data. Int J Clin Pract. 2012;66(1):37-43.

92. de Lusignan S, Belsey J, Hague N, Dzregah B. End-digit preference in blood pressure recordings of patients with ischaemic heart disease in primary care. J Hum Hypertens. 2004;18(4):261-5.

93. Benson M, Junger A, Quinzio L, Fuchs C, Michel A, Sciuk G, et al. Influence of the method of data collection on the documentation of blood-pressure readings with an Anesthesia Information Management System (AIMS). Methods Inf Med. 2001;40(3):190-5.

94. Lawrenson R, Todd JC, Leydon GM, Williams TJ, Farmer RD. Validation of the diagnosis of venous thromboembolism in general practice database studies. British Journal of Clinical Pharmacology. 2000;49(6):591-6.

95. Owen RR, Thrush CR, Cannon D, Sloan KL, Curran G, Hudson T, et al. Use of electronic medical record data for quality improvement in schizophrenia treatment. Journal of the American Medical Informatics Association: JAMIA. 2004;11(5):351-7.

96. Stein HD, Nadkarni P, Erdos J, Miller PL. Exploring the degree of concordance of coded and textual data in answering clinical queries from a clinical data repository. J Am Med Inform Assoc. 2000;7(1):42-54.

97. Terry AL, Chevendra V, Thind A, Stewart M, Marshall JN, Cejic S. Using your electronic medical record for research: a primer for avoiding pitfalls. Fam Pract. 2010;27(1):121-6.

98. Roos LL, Sharp SM, Wajda A. Assessing data quality: a computerized approach. Soc Sci Med. 1989;28(2):175-82.

99. Mikkelsen G, Aasly J. Concordance of information in parallel electronic and paper based patient records. Int J Med Inform. 2001;63(3):123-31.

100. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. BMJ (Clinical Research Ed). 1991;302(6779):766-8.

101. Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing Paper-based with Electronic Patient Records: Lessons Learned during a Study on Diagnosis and Procedure Codes. Journal of the American Medical Informatics Association : JAMIA. 2003;10(5):470-7.

102. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. 2008;14(8):530-9.

103. Conroy MB, Majchrzak NE, Silverman CB, Chang Y, Regan S, Schneider LI, et al. Measuring provider adherence to tobacco treatment guidelines: a comparison of electronic medical record review, patient survey, and provider survey. Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco. 2005;7 Suppl 1:S35-43-S35-43.

104. Haynes K, Bilker WB, Tenhave TR, Strom BL, Lewis JD. Temporal and within practice variability in the health improvement network. Pharmacoepidemiol Drug Saf. 2011;20(9):948-55.

105. Hansell A, Hollowell J, Nichols T, McNiece R, Strachan D. Use of the General Practice Research Database (GPRD) for respiratory epidemiology: a comparison with the 4th Morbidity Survey in General Practice (MSGP4). Thorax. 1999;54(5):413-9.

106. Kaye JA, Derby LE, del Mar Melero-Montes M, Quinn M, Jick H. The incidence of breast cancer in the General Practice Research Database compared with national cancer registration data. British Journal of Cancer. 2000;83(11):1556-8.

107. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems. 1996;12(4):5-34.

108. Committee on Improving the Patient Record Institute of Medicine. The Computer-Based Patient Record:An Essential Technology for Health Care, Revised Edition. Dick RS, Steen EB, Detmer DE, editors: The National Academies Press; 1997.

109. Pipino LL, Lee YW, Wang RY. Data quality assessment. Communications of the ACM. 2002;45(4):211-8.

110. de Lusignan S. The optimum granularity for coding diagnostic data in primary care: report of a workshop of the EFMI Primary Care Informatics Working Group at MIE 2005. Inform Prim Care. 2006;14(2):133-7.

Chapter 3: Clinical Research Attitudes and Perceptions Regarding Electronic Health Record Data Quality[1]

3.1 Introduction

The reuse of electronic health record (EHR) data in research is one of the most promising approaches to comparative effectiveness research, patient-centered outcomes research, and the development of a learning healthcare system.(1-5) These promises include increased efficiency, decreased cost, and improved representation of target patient populations. A number of stakeholders, including patients, clinicians, researchers, and policy-makers, stand to influence and benefit from the secondary use of EHR data. Clinical researchers, in particular, are likely to be

[1] This chapter is a manuscript in preparation to be submitted to a peer-reviewed journal. Nicole Weiskopf will be listed as the first author, Suzanne Bakken as the second author, and Chunhua Weng as the senior author.

responsible for the translation of information acquired during clinical care processes into actionable knowledge.(6)

Although significant work has been done on the acceptance of EHRs as clinical tools,(7-10) little attention has been paid to the factors influencing EHR data reuse. This is especially important in light of the growing focus on EHR data quality. EHR data quality has been a topic of interest since EHRs and other health information technology systems were first introduced,(11-13) but it is only in past decade that there has been attention paid to the necessity for comprehensive and systematic approaches to EHR data quality within the context of EHR data reuse.(14-16)

In order to enable the reuse of EHR data, it is important to understand researcher perceptions of and attitudes towards EHR data and their suitability for research endeavors.. What are researcher perceptions of EHR data quality? Are these perceptions a barrier to the reuse of the data? An assessment of the needs of researchers in regards to EHR data quality and reuse is necessary to guide ongoing and future research efforts.

The most direct way to gain an understanding of researcher needs, attitudes, and perceptions as they relate to EHR data quality within the context of EHR data reuse in research is to ask them. Therefore, we conducted a series of semi-structured interviews with clinical researchers within the Columbia University Medical Center community. Our goal was to gain an understanding of actual researcher experiences conducting research with EHR data. We hypothesized that clinical researchers, inasmuch as they were aware of EHR data quality problems, saw data quality as a barrier to research based on EHR data.

3.2 Methods

3.2.1 Sample

The target population for this study was clinical researchers within the Columbia University Medical Center community. We used purposive sampling in order to identify the most

representative range of views on EHR data quality and EHR data reuse.(17) Specifically, we wished to identify researchers who represented various dimensions of experience: seniority, gender, clinical training, research experience, and experience using EHR data in research.

Though recommendations regarding sample sizes vary widely, previous research has indicated that the majority of relevant themes can be identified using semi-structured interviews with six participants.(18) We set an initial goal of ten interviews, with the understanding that further interviews would be conducted if saturation was not achieved.(19)

3.2.2 Procedure

Institutional Review Board and Consent

The study was determined to be exempt from the Institutional Review Board approval process by the Columbia University Medical Center Institutional Review Board. All participants, however, were asked to provide informed consent prior to being interviewed. Compensation in the form of a twenty-five dollar pre-paid credit card was provided to all participants upon completion of the interview.

Recruitment

Potential participants were contacted by email. The email contained basic information on the study and included the information sheet and consent form (see Appendix A) as an attachment.

Instrument

We used Phase 3 of the Precede-Proceed model as the basis for our instrument design.(20) Although the Precede-Proceed model was developed for guiding the development, implementation, and evaluation of interventions to alter health-related behaviors, the model can be extended and adapted to other behaviors of interest, including use of health-related technologies.(21, 22)

Phase 3 of the Precede-Proceed model focuses on potentially modifiable factors that may influence target behaviors—specifically predisposing, enabling, and reinforcing factors. Within the context of researcher re-use of EHR data in research, this means that Phase 3 can be used to guide an exploration of researcher perceptions of EHR data quality, their attitudes towards EHR data reuse, the ease or difficulty of reusing EHR data, and the rewards or costs of reusing EHR data.

We prepared a set of experience-based questions to guide semi-structured interviews, each of which was intended to last between 30 and 60 minutes. Rather than asking directly about predisposing, enabling, and reinforcing factors, the questions instead invite the participants to describe their experience using EHR data in research (if they have such experience), their perceptions of EHR data quality, and their intention to use EHR data in research going forward. A wrap-up question was also included at the end of all interviews. By asking broad, open-ended questions, our hope was to elicit a broad range of experiences and perceptions from the participants.(23, 24) The complete interview instrument is included in Appendix B.

Data Collection

All interviews were recorded with the participants' permission. Immediately following each interview, the digital recordings were transferred to a secure computer and the interviewer's initial thoughts and impressions were written down. Each interview recording was reviewed and de-identified using the Audacity audio editor (http://audacity.sourceforge.net/). Following de-identification, the digital recordings were transcribed by a professional transcription service. The transcripts were compared to the digital recordings and edited for accuracy. All electronic files were stored in an encrypted file on a secure server. Paper consent forms were saved in a locked file in a secure office.

Analysis

The interview transcripts were analyzed through content analysis, using a combined top-down and bottom-up approach.(25, 26) Coding was performed iteratively,(27) and began with the

top-down step. An initial coding scheme was developed based upon the three major categories of Phase 3 of the Precede-Proceed model. Additional codes, related to data quality categories and assessment methods, were identified from a previous literature review on EHR data quality assessment.(28) A first pass through the transcripts was then performed, during which existing codes were assigned where possible, and important excerpts that did not fit into the existing scheme were identified.

During the deductive half of the analysis process, the coding scheme was updated based upon the first round of analysis. New themes and categories were identified and old ones were eliminated or edited. During a second pass through the transcripts, the updated coding scheme was applied, and final refinements were made to the coding scheme.

3.2.3 Ensuring Trustworthiness

Member Checks

Member checks were performed at two stages in the data collection and analysis process. First, during the interviews the interviewer reflected phrases and concepts back to the participants to clarify and confirm understanding.

Following the completion of the final coding scheme and coding of data, approximately 28 months after the initial interviews were conducted, all participants were contacted by email and asked to confirm the major themes identified from their individual interviews.(29, 30) Each participant received an email with a brief description of the member check process. An encrypted document, containing a selection of the major themes identified from that participant's interview, was attached. A second email was sent containing a password to decrypt the attachment. Each participant was asked to respond indicating whether the themes faithfully captured their beliefs, attitudes, and experiences with regards to electronic health record data and data quality at the time of the interview.

Audit Trail

Documentation of the interview and analysis processes was maintained over the course of the study, creating an audit trail. The trail included the interview guide, the initial coding scheme, the first pass of coded data, the final coding scheme, the second pass of coded data, and detailed notes on the analysis and coding process. These documents were audited by Suzanne Bakken, a faculty member, to assess the reliability of the analysis and resulting findings.(30-32)

3.3 Results

3.3.1 Participants

Ten clinical researchers were interviewed, though one interview was dropped because the interview guide could not be completed in the time allotted. The nine remaining participants had an average age of 43, with a standard deviation of 17 years. Six of the nine participants were in their thirties. Four of the participants were female and five were male. Eight of the participants were medical doctors, though not all had practiced clinically, and three held doctorates besides medical degrees.

3.3.2 Coding Scheme

Overall, 38 unique thematic codes were identified and used. The complete coding scheme is presented in Figure 1. The coding scheme was mapped to the three primary factor groups from the Precede-Proceed model (predisposing, enabling, and reinforcing factors). The majority of themes derived during the deductive analysis phase fell into predisposing category.

3.3.3 Predisposing Factors

Most findings fell into the predisposing factors category, which includes researcher knowledge, perceptions, and attitudes regarding data quality as a general concept, the quality of

PREDISPOSING FACTORS	***knowledge of DQ problems***	completeness correctness fragmentation granularity concordance structured signal-to-noise	
	beliefs about state of EHR DQ	overall trust in EHR data	the data can't be trusted conflicted feelings about the data
		quality of different data types	lab results diagnoses meds billing data demographic data notes
		quality of data from different institutions	
	beliefs about factors impacting EHR DQ	data traits	optional vs mandatory automatic vs manual recording objective vs subjective structured vs unstructured
		system features	not designed for clinical care not designed for research was designed for billing system fragmentation
		user behaviors	impacted by knowledge of data use clinical care process
	others' attitudes to EHR data		
	attitudes to EHR data reuse	positive attitudes negative attitudes conflicted attitudes	
	recommendations for reuse	institutional knowledge is required EHR data require validation attach metadata to EHR data	
ENABLING FACTORS	***availability and accessibility of data***		
	identification of relevant data		
	skills required for data reuse		
REINFORCING FACTORS	***reuse of EHR data can facilitate research***		
	reuse of EHR data can facilitate publication		

Figure 3.1 The coding scheme developed during the analysis process.

EHR data specifically, and the reuse of EHR data in research. Seven thematic categories were identified: knowledge of relevant categories of data quality, beliefs about the state of EHR data quality, beliefs about factors that impact EHR data quality, beliefs about others' attitudes towards EHR data quality and data reuse, general attitudes towards EHR data reuse, and beliefs about how EHR data should be reused.

3.3.3.1 Knowledge of relevant data quality problems

The participants described a number of different data quality problems and concerns that they knew of that they thought were relevant to EHR data. Seven basic dimensions of data quality were derived from the problems and concerns: completeness, correctness, fragmentation, granularity, concordance, non-structured, and signal-to-noise.

3.3.3.2 Beliefs about the state of EHR data quality

In terms of participants beliefs about the current state of EHR data quality, most reported either that they did not trust the data, or that they had mixed feelings regarding the trustworthiness of the data (see Table 1 for examples). They also believed that there was variation in quality depending upon institutes, as well as the types of data recorded (e.g. medication orders, laboratory results, notes, billing data, etc.).

Table 3.1 Predisposing factors related to beliefs about the state of EHR data quality: overall researcher trust in EHR data.

Theme	*Representative Participant Statements*
EHR data cannot be trusted	It's legendary among trialists that what's in a chart—it's an electronic health record now but it was a chart before—the kind of rules about what you put in there, who puts it in, and how careful they are…is pretty sloppy. I think it's much better than it probably used to be, but I think it does have a long way to go before it's something I feel entirely comfortable without having to go through a lot of internal validation steps.
Conflicted feelings about EHR data	If [data quality] was like, one to ten, maybe a six. I think—we tend to be pretty comprehensive, I think accuracy—we tend to be overall pretty accurate for a lot of the important things. I think reliability is a big issue.

3.3.3.3 Beliefs about factors that influence EHR data quality

The participants identified three primary types of factors that they believed influenced the quality of EHR data: traits of the data, system features, and user (e.g. clinician) behaviors (see Table 2 for examples). Data traits include mechanisms and requirements at the variable level. Whether an element is required or optional, automatically recorded or manually recorded, objective (e.g. lab result) or subjective (e.g. diagnosis), and structured or unstructured were all seen as potentially influencing the quality of the relevant data. Some participants, for example, thought that manually recorded data were likely to be of better quality than automatically recorded data, since they invited some degree of human validation, while other participants believed the opposite, and expected that automatically recorded information was likely to be more trustworthy. Those who mentioned optional versus required elements believed that optional variables were more likely to be of good quality than required elements, since the latter might be entered without care simply to progress to the next data entry step. Objective variables were generally believed to be of better quality than subjective elements. Beliefs regarding structured versus unstructured variables were split. Some thought that clinicians were more likely to faithfully capture nuanced information in a note, while others thought that the very nature of unstructured data made it unsuitable for reuse.

Most participants cited similar system features or issues that they believed negatively influenced EHR data quality: EHRs are designed for billing, rather than clinical care or research, and EHRs and institutional health information technology systems tend to be fragmented (either within an EHR, or between multiple EHRs and related systems that are meant to function as a single system). Participants expressed significant frustration regarding their perceptions that EHR design, implementation, and use are guided by administrative and billing needs, which they believed resulted in data that was of insufficient quality for research. Fragmentation was believed to make it difficult to identify, collect, and validate relevant data.

In terms of user behaviors, some participants believed that clinicians were so focused on clinical practice that they paid little attention to the quality of the data they entered beyond what they felt was strictly necessary for caring for a patient. Not all participants who mentioned this issue described it as a negative, but rather as a simple fact of the care process. Similarly, some participants believed that more explicit knowledge of how or if EHR data might be reused might influence clinicians to pay more attention to data quality at the time of data entry.

Table 3.2 Predisposing factors related to beliefs about factors that influence EHR data quality: data traits, system features, and user behaviors that participants believed might influence the quality of EHR data.

Theme	*Representative Participant Statements*
Traits of data	If data collection can be automated, I think you can have a higher confidence that the data are accurate. So for instance, if somebody came to me and they were very interested in blood pressure in the ICU…there are tags that nurses make to blood pressure measurements that they've verified and validated. And so in the data, I'm sure that there is an embedded tag that indicates that a human being has seen and verified this result. And I might suggest that, if possible, only those values be used…. Anytime a human being hasn't laid eyes on a result and had a chance to flag it as valid or invalid, there's obviously a risk of a mistake.
System features	When you just go with people who with blood pressure measurements and then you like, wait, you want those people who are high blood pressure but they are treated so their blood pressure is lower now, you need to code them as hypertensives as well.
User behaviors	The data was extracted from a team that is employed by the hospital. And they feel very strongly that I need to work with the data in a particular environment supported by the hospital, meaning they're not just going to give me a bunch of information from patients. And working within that environment is sort of a new concept…that has been the biggest limitation.

3.3.3.4 Beliefs about other's attitudes towards EHR data quality and reuse

Some of the participants believed that their perceptions of and attitudes towards EHR data quality and EHR data reuse were different from those of their fellow researchers or clinicians. Specifically, some participants believed that they were more aware of the limitations

of EHR data than others in their position. One of the more senior participants believed that younger clinicians, particularly those who had always worked with EHRs instead of paper records, would be likely to trust the data coming from EHRs.

3.3.3.5 Attitudes towards the reuse of EHR data

The participants had varying attitudes towards the reuse of EHR data. Some felt positively towards it, some conflicted, and others negatively (see Table 3 for examples). Some participants expressed inconsistent attitudes over the course of a single interview. Most of the few positive comments regarding EHR data reuse were with qualification. Simply having the data available was seen as a positive, when the converse was not having the data at all, and electronic data were seen as more accessible and less fragmented than paper records. The majority of negative attitudes towards the reuse of EHR data stemmed from concerns about retrospective research broadly (e.g. unknown confounders), the potential to intentionally or unintentionally alter or influence findings through the data transformation and cleaning processes, and the suspected poor quality of EHR data. The participants who expressed conflicted attitudes felt that there were both significant limitations to and major opportunities in the reuse of EHR data in research.

Table 3.3 Predisposing factors related to attitudes towards the reuse of EHR data in research.

Theme	*Representative Participant Statements*
Positive attitudes	I think it's definitely a step forward from where things used to be. It's much better than hand inputting data in separate silos…over and over every time you want to get a project done, like having some basic stuff that's available for everybody to use. It definitely gives you data are that suitable for research….
Negative attitudes	But even just looking at like 10 charts, like there were so many holes that we like discarded using the data whatsoever and we don't have the manpower to go back, to review like, five of those charts. For me it took like a couple of hours. So we wanted to do this for something like a thousand patients and we don't have like staffing. So it's sort of like we are waiting on that project until we know the data's better. There're a lot of hidden choices about how do I define the date of

	diagnosis, which is not as well-defined as I thought it would be. How do I define the presentation of a disease that can be different for a lot of people? And in order for you to talk about all of that, the methods section would be twenty pages long.
Conflicted attitudes	I'm just saying, you know what? [EHR data] is what it is. And people have spoken about problems with it, but that's systemic error, and this is some additional pieces of the puzzle that probably give us more insight into important disease processes. And I think that's sort of fine, it's sort of saying that I'm not agreeing with the entire record…let's sort of all accept the problems with it. I mean, I think it's not even—starting at this point in time, it's not even a question of positive or negative, because I think it's the way of the future, or the way of the present.

Beliefs about how to reuse EHR data

The participants had a number of beliefs regarding how EHR data should be reused, both in terms of the processes used and the types of research being performed (see Table 4 for examples). In terms of the latter, some participants believed that EHR data were best-suited to preliminary research and hypothesis generation, and that results from EHR-based research should be confirmed using more traditional studies or data. Many shared the opinion that EHR data required substantial validation prior to reuse, and that the processes of identifying, validating, and reusing the data should be guided by individuals with relevant clinical and institutional expertise (e.g. nurses or physicians). Others suggested that metadata, which could include substantial information about the recording of a piece of data (e.g. clinician, location, etc.) as well as previous validation efforts, be attached to EHR data, so that researchers interested in using the data would have more information with which to inform the reuse process.

Table 3.4 Predisposing factors related to beliefs about how to reuse EHR data in research.

Theme	*Representative Participant Statements*
Only use EHR data in preliminary research and hypothesis generation	I think, you at least initially, until you know that, for your set of questions, that the de-identified dataset is pretty accurate, at least initially the dataset should probably just be used for generating hypotheses.

	Or you're putting a grant in and you need like some preliminary data that looks encouraging or interesting or whatever. And if you don't have it there, people reading the grant are much less enthusiastic about giving a high score if you don't show them something that can encourage them to believe in you.
Clinical and institutional expertise should inform reuse	I think it's hard to use [EHR data] when it's pulled straight from a database without having someone clinically or -- with clinical knowledge or background to review it, to use it and think that the data quality is good. I don't think that we feel comfortable enough like that to do it. And I think it's hard like even, you know, even one person reviewing a chart, even if you know, like two different people review the same chart and pull on what they think is the relevant data for a patient, like labs, or meds or diagnosis. Like, they may come with different answers as well. So it's not saying that humans are perfect and the chart review that like looks at everything over a time chronologically is necessarily more accurate. It probably can account for some of the variations, the computers that doesn't sort of check for little things right -- like the one data that the lab got the wrong value for potassium. So like, I think that overall, we feel like it's probably more reliable and we still need to check all of it manually. Finding very good clinicians and sitting down. So that's sort of my philosophy right now and we're writing this sort of lessons learned piece. It's basically, if you need to do any research on this, just have to be a multidisciplinary team… there's no way around it basically and you just need physicians who are dealing with this on a daily basis, informaticians and -- of course the traditional people like statisticians and epidemiologists, that kind of thing is fine but you just can't do it with just one person.
EHR data require validation prior to reuse	I haven't actually worked with data points that I know are going to be filled with unreliable things, but the way that I would handle that is the way that anybody would handle a large data set that they weren't responsible for collecting. They would run frequencies, they would look for clear outliers, one blood glucose of 500 in a patient who has blood glucoses in the normal range is more likely to be a statistical outlier or mistake or whatever. And those patients, if you have the luxury, you flag those and you either throw out the data entirely and call it missing or you go back to the chart and verify that it was real.
Metadata should be attached to EHR data	There needs to be a data dictionary with a clear delineation of what the source of the information is plus indicators if that element came from different sources…what the source of that information came from so you can be selective about how you use it. There should be, whether the researcher does it or whether it's published—and it would be great

	if it were pre-published—indicators of quality for those select variables that people felt were important and prone to misinterpretation, misclassification or whatever. I guess I can find out how many missing values there are in a particular variable, but in terms of the dataset as a whole, I would like to know for each of these key variables, how many are missing and how many not, where they…came from and information about other people's experiences…. That kind of thing, so that a researcher can make a choice. If that is their major outcome, they—or just a covariate—they can make a choice as to whether that is a good use of the data.

3.3.4 Enabling Factors

The enabling factors cited by participants included the availability and accessibility of EHR data, identifying relevant patients for data reuse, and the skills required to reuse the data (see Table 5 for examples). Many of the participants complained of difficulty gaining access to the data they wanted, largely because of the institutionally-mandated process researchers must go through when requesting EHR data. Some participants also said that the cohort identification process was deceptively difficult, in part because of the limited reliability of diagnoses and related data. Initial queries may have limited sensitivity, specificity, or both, leading to an extended, iterative query process, as well as manual review of results. Lastly, some described lacking the specific skills needed to work with the data after gaining access. This may be due to the format the data are extracted to, or special environments where researchers must conduct their work.

Table 3.5 Enabling factors: factors influencing ease or difficulty of researcher reuse of EHR data in research.

Theme	*Representative Participant Statements*
Availability and accessibility of data	That's a big barrier here. I mean, you can wait a long time to get results because there's really just two people writing programs to pull the stuff out. But if you start just letting anybody do it, I suspect you would have a lot of misunderstanding about where things were and whatnot. So it just seems like that's an awfully small—they probably ought to charge a fair price for it, put it in the grants as an expense, have a little more infrastructure to deal with it.

	It would be a lot easier if you could use the information that's already in the electronic record and database that -- just download that into the database so that correlates with all the other clinical information. That's a major disadvantage but that's I guess, technical or political or, you know, commercial, I'm not sure what -- I mean there's lot of reasons why it's not being done.
Identification of relevant data	When you just go with people who with blood pressure measurements and then you like, wait, you want those people who are high blood pressure but they are treated so their blood pressure is lower now, you need to code them as hypertensives as well.
Skills required for data reuse	The data was extracted from a team that is employed by the hospital. And they feel very strongly that I need to work with the data in a particular environment supported by the hospital, meaning they're not just going to give me a bunch of information from patients. And working within that environment is sort of a new concept…that has been the biggest limitation.

3.3.5 Reinforcing Factors

The participants reported two primary factors rewarding the choice to reuse EHR data: the facilitation of research and the facilitation of publication (see Table 6 for examples). They felt that reusing EHR data could increase the efficiency of research and increase the number of research questions that could be answered. These improvements in the research process, in turn, could lead to an increase in papers published.

Table 3.6 Reinforcing factors: costs and benefits to researcher reuse of EHR data in research.

Theme	*Representative Participant Statements*
Facilitation of research	EHR [data] is the promise for all the comparative effectiveness research, right? You have to have a little bit of skepticism, don't trust the data implicitly, use it for what it is and certainly, it's better than not having it and it facilitates a tremendous amount of research that wouldn't be possible otherwise without exhaustive chart reviews or manual processing, so it has huge benefits, but it's not a panacea.
Facilitation of publication	…the NYP database could be a paper-generating monster…

3.4 Discussion

The enabling and reinforcing factors identified from the interviews were mostly in agreement with what is already known about the secondary use of EHR data. Some of the predisposing factors, however, were not in line with commonly accepted wisdom about secondary use. Clinical researchers have certain beliefs regarding "best practices" in EHR data reuse that are not currently addressed, and should be incorporated into workflows, models, and methodological approaches to EHR data quality assessment and reuse.

3.4.1 Reinforcing and Enabling Factors

The primary reinforcing factors identified from the interviews are very much in agreement with the factors currently driving the secondary use of EHR data: data reuse can improve efficiency and facilitate the research process and publication.(33, 34) The enabling factors identified were similarly unsurprising. Data accessibility and ease-of-use have been highlighted as an important (and problematic) issues within the informatics and broader researcher communities. Data security, governance, and availability are all ongoing topics of research.(35-37) Cohort identification has also been identified as a deceptively complicated task,(38) and is the central focus of the eMERGE Network.(39)

3.4.2 Conceptualizations of Data Quality

Learning how clinical researchers view data quality is an important step in encouraging and improving the secondary use of EHR data. The participants in this study identified seven categories of data quality: completeness, correctness, fragmentation, granularity, concordance, structured, and signal-to-noise. There is a certain degree of overlap with Wang and Strong's popular model of data quality, which includes accuracy, completeness, and concise representation among its 15 dimensions.(40)

Fragmentation, concordance, granularity, and signal-to-noise may represent the issues of data quality that arise specifically from the use of EHRs and other health information technology. The requirement that multiple information systems work in concert, which can lead to problems of fragmentation and lack of concordance, is common within large medical centers and systems, where there are often different EHRs for inpatient care, office-based practices, emergent care, etc. Data may also be recorded multiple times within the same system, leading to the same problems. The longitudinal nature of patient care, along with copy-and-paste problems,(41) can make it difficult to identify and locate the relevant data for a patient of interest. Finally, the granularity of data is extremely important in healthcare. Data that are too granular can be difficult to group together meaningfully, while data that are not granular enough may not provide sufficient information. It is therefore unsurprising that these four categories of data quality are more likely to be found within the informatics literature. Hersh, et al., for example, describe unstructured notes, multiple data sources, and granularity as potential causes of problems with the secondary use of EHR data in comparative effectiveness research, in addition to the more common data quality issues of accuracy and completeness.(35)

3.4.3 Attitudes toward Secondary Use of EHR Data

We had hypothesized that, to the extent clinical researchers were aware of EHR data quality problems, these concerns would be a barrier to secondary use. There were a few findings in this area that were both important and surprising. First, we had not anticipated the fact that every participant was at least somewhat concerned about EHR data quality. This may have been partially a function of the fact that all participants were members of large medical center with a robust EHR system, and most had practiced clinically at the center. Whatever the reason, the majority of the participants expressed conflicting beliefs about EHR data quality, and conflicting attitudes regarding the use of EHR data in research.

The most important finding in this area is summarized by a quotation from one of the participants: "I mean, I think—starting at this point in time—it's not even a question of positive or negative, because I think it's the way of the future, or the way of the present." This sentiment was echoed, though not so explicitly, throughout the interviews. Most participants were aware of EHR data quality problems, concerned about the validity of research conducted using EHR data, but nevertheless planned on using EHR data for research. Our hypothesis, therefore, was only half right. Concerns about EHR data quality existed, but were *not* a barrier to research. If this attitude can be generalized to the broader population of clinical researchers, then it emphasizes the importance of and immediate need for the development of methodological approaches to the assessment and improvement of EHR data quality. If secondary use is going to continue, it is important to make the research conducted using EHR data as trustworthy and as valid as possible.

3.4.4 Involving Experts in the Secondary Use Process

One point that many of the participants made is that an expert—either on the clinical topic of interest or the clinical workflows that generated the data—should be involved in the identification, validation, and analysis of EHR data for clinical research. In our previous research we suggested that a systematic approach for EHR data quality assessment was necessary,(28) and other groups, particularly the Observational Medical Outcomes Partnership, are working not only systematic, but automated methods.(42, 43) Our findings from the current study, however, suggest that clinical researchers desire a certain degree of human influence in the process of deciding how and if EHR data should be used in research. They believe that clinicians have a better understanding of what the data are actually measuring, how they are recorded, and what they mean than someone who approaches the data as an outsider. It may be that a "black box" approach to data quality assessment would have limited acceptance amongst clinical researchers. It is also possible that the participants were right, and that personal experience and clinical expertise is still vitally important in the reuse of EHR data, particularly as it relates to the data

quality assessment process. Unfortunately, this kind of expert input has the potential to decrease the efficiency and the consistency of EHR data secondary use. It may also pose a problem when researchers work with clinically-generated datasets that have been fully deidentified or drawn from multiple institutions, since in these scenarios it may no longer possible to draw conclusions about data quality based upon the provenance of the data (though the use of metadata could alleviate this problem).

3.4.5 Possible Approaches to the Data Quality Problem

Three primary approaches for handling the problem of questionable EHR data quality were mentioned during the interviews. The first, and currently most common, is to assess and, if possible, "fix" the data after it has been collected. Studies that involve the secondary use of EHR data take this approach when they involve the validation of the data or the imputation of missing values. The advantage of this approach is that it does not interfere with the clinical workflows already in place in the data collection processes. The disadvantage is that there are limited options for actually improving the data at this stage in the data lifecycle.

A second option would be to introduce an intervention before the point of data extraction. This could mean changing clinical workflows or documentation practices. Alternatively, the development and introduction of new EHRs and other health information technology systems that encourage good documentation and otherwise improve efficiency could improve the quality of the data collected. Although some of the participants were hesitant to suggest that research should be a priority in clinical settings, most felt that clinical care itself is not enough of a priority (versus billing or administrative services), and that EHRs that embrace clinical workflows and patient care could only improve data quality.

A final option would be to simply limit the types of research being conducted with EHR data. Specifically, using EHR data only for preliminary research or hypothesis generation would mean that if the results are invalidated by data quality problems, this lack of validity will be

uncovered during further study using more controlled data. Given the eagerness most of the participants expressed to use EHR data for research, however, this seems unlikely.

3.4.6 Limitations and Future Work

There are a number of limitations to this work, mostly centered around the lack of generalizability of our findings. Our sample size was small and from only a single institution. Also, the majority of our sample had extensive experience using an EHR, either as clinicians or informaticians. It is unclear whether the broader clinical researcher community would be as family with EHR data, or problems of EHR data quality. Alternatively, it may be possible that individuals who are less familiar with EHRs would be less comfortable using EHR data in their research. Ideally, this study should be repeated with a more representative sample of clinical researchers. We plan to replicate this study in other institutions using a larger sample in the future.

Another limitation is that we have no way to link the perceptions and attitudes of our participants to concrete behaviors. In other words, do the views espoused during these interviews actually reflect participant behaviors as they relate to EHR data reuse and data quality assessment? Would a participant who says the data should always be validated by a clinical domain expert actually secure the assistance of a domain expert when participating in the secondary use of EHR data? Further work would be needed to determine how the predisposing, enabling, and reinforcing factors identified in this study actually impact the research process using EHR data.

3.5 Conclusion

Our findings indicate that clinical researchers are aware of data quality issues, but that data quality issues are not a barrier to the secondary use of EHR data. The major barriers and rewards for engaging in the secondary use of EHR data are largely unsurprising. Improved

efficiency of research and publication were both seen as benefits, while issues like limited data accessibility and availability were identified as hindering the research process.

Clinical researchers identified a number of potential causes of data quality problems, and were especially critical of EHRs themselves. They also felt that EHR data quality was variable, and largely dependent upon the clinical variables and data types being used. Most believed that domain experts should be involved in the process of validating EHR data prior to reuse, which indicates that a black box approach to EHR data quality assessment or the lack of transparency of EHR data quality may not be readily accepted by clinical researchers. Future work on EHR data reuse and data quality assessment should take researcher perceptions and attitudes into account.

3.6 References

1. Hersh WR, Cimino JJ, Payne PR, Embi PJ, Logan JR, Weiner MG, et al. Recommendations for the use of operational electronic health record data in comparative effectiveness research. eGEMs (Generating Evidence & Methods to improve patient outcomes). 2013;1(1).

2. Olsen L, Aisner D, McGinnis JM, editors. The Learning Healthcare System. Washington, D.C.: National Academies Press; 2007.

3. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1-9.

4. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? Ann Intern Med. 2009;151(5):359-60.

5. Steiner J, Pace WD, Motsashari F, Coye MJ. Using evidence to build a learning health care system. Washington, D.C.: AcademyHealth, 2012.

6. Westfall JM, Mold J, Fagnan L. Practice-based research--"Blue Highways" on the NIH roadmap. JAMA : the journal of the American Medical Association. 2007;297(4):403-6.

7. Moody LE, Slocumb E, Berg B, Jackson D. Electronic health records documentation in nursing: nurses' perceptions, attitudes, and preferences. Comput Inform Nurs. 2004;22(6):337-44.

8. Huryk LA. Factors influencing nurses' attitudes towards healthcare information technology. J Nurs Manag. 2010;18(5):606-12.

9. Ward R, Stevens C, Brentnall P, Briddon J. The attitudes of health care staff to information technology: a comprehensive review of the research literature. Health Info Libr J. 2008;25(2):81-97.

10. Menachemi N. Barriers to ambulatory EHR: who are 'imminent adopters' and how do they differ from other physicians? Inform Prim Care. 2006;14(2):101-8.

11. Burnum JF. The misinformation era: the fall of the medical record. Ann Intern Med. 1989;110(6):482-4.

12. Roos LL, Sharp SM, Wajda A. Assessing data quality: a computerized approach. Soc Sci Med. 1989;28(2):175-82.

13. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. BMJ (Clinical Research Ed). 1991;302(6779):766-8.

14. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical care. 2012;50 Suppl:S21-9.

15. Faulconer ER, de Lusignan S. An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar. Inform Prim Care. 2004;12(4):243-54.

16. Devine EB, Capurro D, van Eaton E, Alfonso-Cristanco R, Devlin A. Preparing electronic clinical data for quality improvement and comparative effectiveness research: the SCOAP CERTAIN automation and validation project. EGEMs (Generating Evidence & Methods to improve patient outcomes). 2013;1(1):16.

17. Teddlie C, Yu F. Mixed methods sampling: a typology with examples. Journal of Mixed Methods Research. 2007;1(1):77-100.

18. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods. 2006;18(1):59-82.

19. Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychol Health. 2010;25(10):1229-45.

20. Green L, Kreuter M. Health promotion planning: an educational and ecological approach. 4th ed. Boston, MA: McGraw Hill; 2005.

21. Bakken S, Lantigua RA, Busacca LV, Bigger JT. Barriers, enablers, and incentives for research participation: a report from the Ambulatory Care Research Network (ACRN). J Am Board Fam Med. 2009;22(4):436-45.

22. Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform. 2003;36(3):218-27.

23. Tappen R. Advanced Nursing Research: From Theory to Practice. Sudbury, MA: Jones & Bartlett Learning; 2011.

24. Waltz C, Strickland O, Lenz E. Measurement in nursing research. Philadelphia, PA: F.A. Davis Company; 1984.

25. Thomas D. A general inductive approach for analyzing qualitative evaluation data. American Journal of Evaluation. 2006;27(2):237-46.

26. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. International journal of Qualitative Methods. 2006;5(1):80-92.

27. Weston C, Gandell T, Beauchamp J, McAlpine L, Wiseman C, Beauchamp C. Analyzing interview data: the development and evolution of a coding system. Qualitative Sociology. 2001;24(3):381-400.

28. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51.

29. Hoffart N. A member check procedure to enhance rigor in naturalistic research. Western Journal of Nursing Research. 1991;13(4):522-35.

30. Carlson J. Avoiding traps in member checking. The Qualitative Report. 2010;15(5):1102-13.

31. Krefting L. Rigor in qualitative research: the assessment of trustworthiness. Am J Occup Ther. 1991;45(3):214-22.

32. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educational Communication and Technology Journal. 1981;29(2):75-91.

33. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care. 2007;13(6 Part 1):277-8.

34. Sung NS, Crowley WF, Jr., Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA : the journal of the American Medical Association. 2003;289(10):1278-87.

35. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical care. 2013;51(8 Suppl 3):S30-7.

36. Hruby GW, McKiernan J, Bakken S, Weng C. A centralized research data repository enhances retrospective outcomes research capacity: a case report. J Am Med Inform Assoc. 2013;20(3):563-7.

37. Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(34):4243-8.

38. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc. 2014;21(2):221-30.

39. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC medical genomics. 2011;4:13.

40. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems. 1996;12(4):5-34.

41. Wrenn JO, Stein DM, Bakken S, Stetson PD. Quantifying clinical narrative redundancy in an electronic health record. J Am Med Inform Assoc. 2010;17(1):49-53.

42. Hartzema AG, Reich CG, Ryan PB, Stang PE, Madigan D, Welebob E, et al. Managing data quality for a drug safety surveillance system. Drug Saf. 2013;36 Suppl 1:S49-58.

43. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60.

Chapter 4: The Task-Dependence of Electronic Health Record Data Quality[1]

4.1 Introduction

With the growing availability of large electronic health record (EHR) databases, clinical researchers are increasingly interested in the secondary use of clinical data.(1, 2) While the prospective collection of data is notoriously expensive and time-consuming, the use of an EHR may allow a medical institution to develop a clinical data repository containing extensive records for large numbers of patients, thereby enabling more efficient retrospective research. These data are a promising resource for comparative effectiveness research, outcomes research, epidemiology, drug surveillance, and public health research.

[1] This chapter was originally published in the Journal of Biomedical Informatics. The full citation for this publication is:
Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform. 2013;46(5):830-6.

Unfortunately, EHR data are known to suffer from a variety of limitations and quality problems. The presence of incomplete records has been especially well documented.(3-6) The availability of an electronic record for a given patient does not mean that the record contains sufficient information for a given research task.

Data completeness has been explored in some depth. The statistics community has focused extensively on determining in what manner data are missing. Specifically, data may be considered to be missing at random, missing completely at random, or missing not at random.(7, 8) Datasets that meet these descriptions require different methods of imputation and inference.

The statistical view of missing or incomplete data, however, is not sufficient for capturing the complexities of EHR data. EHR records are different from research data in their methods of collection, storage, and structure. A clinical record is likely to contain extensive narrative text, redundancies (i.e., the same information is recorded in multiple places within a record), and complex longitudinal information. While traditional research datasets may suffer from some degree of incompleteness, they are unlikely to reflect the broad systematic biases that can be introduced by the clinical care process.

There are several dimensions to EHR data completeness. First, the object of interest can be seen as the patient or as the health care process through which the patient was treated; there is a difference between complete information about the patient versus complete information about the patient's encounters. A patient with no health care encounters and an empty record has a complete record with respect to the health care process, but a blank one with respect to the patient. Furthermore, one can measure completeness at different granularities: the record as a whole or of logical components of the record, each of which may have its own requirements or expectations (e.g., demographic patient information versus the physician thought process). (9, 10) Another dimension of completeness emerges from the distinction between intrinsic and extrinsic data requirements. One can imagine defining minimum information requirements necessary to consider a record complete (which could be with respect to either the patient or the health care

process), or one can tailor the measurement of completeness to the intended use. Put another way, we can see completeness in terms of intrinsic expectations (i.e., based a priori upon the content) or extrinsic requirements (based upon the use). (11, 12)

The EHR data consumers who define these extrinsic requirements will have different data needs, which will in turn dictate different conceptualizations of a complete patient record. Here, Juran's definition of quality becomes valuable: "fitness for use."(12) It may be that data completeness does not have a simple, objective definition, but is instead task-dependent. Wang and Strong, for example, in their work developing a model of data quality, define completeness as "[t]he extent to which data are of sufficient breadth, depth, and scope for the task at hand."(13) In other words, whether a dataset is complete or not depends upon that dataset's intended use or desired characteristics. In order to determine the number of complete records available for analysis one must first determine what it means to have a complete patient record. The quality of a dataset can only be assessed once the data quality features of interest have been identified and the concept of data quality itself has been defined.(11)

Multiple interpretations of EHR completeness, in turn, may result in different subsets of records that are determined to be complete. The relationships between research task, completeness definition, and completeness findings, however, are rarely made explicit. Hogan and Wagner offer one of the most widely used definitions: "the proportion of observations that are actually recorded in the system."(5) This definition does not, however, offer specific measures for determining whether a record is complete. Neither does it account for the possibility that completeness may be task-dependent. What proportion of observations should be present? Which observations are desired? Are there any other considerations beyond simple proportion? Furthermore, observations are complex, nested concepts, and it must be determined what level of detail or granularity is needed or expected. In order of increasing detail, one could record a visit that occurred, the diagnoses, all the symptoms, a detailed accounting of the timing of all the symptoms, the clinician's thought process in making a diagnosis, etc.

In the sections below, we enumerate four specific operational and measurable definitions of completeness. These definitions are not exhaustive, but they illustrate the diversity of possible meanings of EHR data completeness. We ran the definitions against our clinical database in order to demonstrate the magnitude of completeness in the database and to illustrate the degree of overlap among the definitions.

4.2 Methods

Previously, we conducted a systematic review of the literature on EHR data quality in which we identified five dimensions of data quality that are of interest to clinical researchers engaged in the secondary use of EHR data. Completeness was the most commonly assessed dimension of data quality in the set of articles we reviewed.(3) Based upon this exploration of the literature on EHR data quality, consideration of potential EHR data reuse scenarios, and discussion with stakeholders and domain experts, we describe four prototypical definitions of completeness that represent a conceptual model of EHR completeness. Further definitions of completeness are possible and may become apparent as the reuse of EHR data becomes more common and more use cases and user needs are identified.

Figure 1 presents a visual model of the four definitions of completeness, which are described further in section 2.1. In this model of EHR data, every potential data point represents some aspect of the patient state at a specific time that may be observed or unobserved as well as recorded or unrecorded. The longitudinal patient course, therefore, can be represented as a series of points over time that may or may not appear in the EHR.

4.2.1 Definitions

Documentation: A record contains all observations made about a patient

The most basic definition of a complete patient record described in the literature is one where all observations made during a clinical encounter are recorded.(5) This is an objective, task-independent view of completeness that is, in essence, a measure of the fidelity of the

documentation process. Assessments of documentation completeness rely upon the presence of a reference standard, which may be drawn from contacting the treating physician (14), observations of the clinical encounter,(15) or comparing the EHR data to an alternate trusted data source—often a concurrently maintained paper record.(16-19) Documentation completeness is also relevant to the quality measurements employed by the Centers for Medicare & Medicaid Services.(20)

In secondary use cases, however, the data consumer may be uninterested in the documentation process. Instead, completeness is determined according to how well the available data match the specific requirements of the task at hand, meaning that completeness in these situations is more often subjective and task-dependent. While documentation completeness is intrinsic, the following three definitions of completeness are extrinsic and can only be applied once a research task has been identified.

Breadth: A record contains all desired types of data

Some secondary use scenarios require the availability of multiple types of data. EHR-based cohort identification and phenotyping, for example, often utilize some combination of diagnoses, laboratory results, medications, and procedure codes.(21-23) Quality of care and clinician performance assessment also rely upon the presence of multiple data types within the EHR (the relevant data types vary depending upon clinical area).(20, 24-27) More broadly, researchers interested in clinical outcomes may require more than one type of data to properly capture the clinical state of patients.(28, 29) In the above cases, therefore, a complete record may be one where a breadth of desired data types is present. It is important to note that the absence of a desired data type in a record does not necessarily indicate a failure in the clinical care process or in the recording process. Rather, it may be that a data type that is desired for research was not relevant from a clinical standpoint, and therefore was not observed.

Density: A record contains a specified number or frequency of data points over time

In many secondary use scenarios, EHR data consumers require not only a breadth of data types, but also sufficient numbers and density of data points over time.(30) Some of the phenotyping algorithms developed by the eMERGE Network, for example, rely upon the presence of multiple instances of the same laboratory tests, diagnoses, or medications,(31) and sometimes specify desired time periods between the recording of these data within the EHR.(32, 33) Clinical trial eligibility criteria, which can be compared to patient records to identify relevant cohorts, also contain complex temporal data specifications,(34) as do EHR data requests submitted by clinical researchers.(35) Breadth and density can be considered complementary, orthogonal dimensions of completeness. A single point of patient data, for example, has breadth and density of one.

Predictive: A record contains sufficient information to predict a phenomenon of interest

Our final and most complex definition of EHR data completeness arises when one considers that the overall goal of much research is the ability to predict an outcome(13). It is possible to train various computational models, some of which being more tolerant of missing data than others, using EHR-derived datasets. Researchers may be interested in predicting, amongst other clinical phenomena, disease status and risk,(36-38) readmission,(39, 40) or mortality.(41, 42) Depending upon the model employed, data needs may be implicit, rather than explicit. The metric for completeness is performance on the task, rather than counts of data points. The data that are required are those that are sufficient to make a prediction. Therefore, it may that two records with different data profiles are both complete according to this definition.

4.2.2 Data

NewYork-Presbyterian Hospital (NYPH) is a not-for-profit hospital in New York City consisting of five locations. For the purposes of this research, we included data from: the Milstein Hospital, a tertiary care hospital, and its associated ambulatory areas; Allen Hospital, a community hospital; and Morgan Stanley Children's Hospital. All are in upper Manhattan. These

locations and their affiliated offices treat close to 300,000 unique patients per year. The patient population is 56% female, with an average age of 51 years. The population is 32% Hispanic, 10% Asian, 19% Black, and 39% White.

A number of different health information technology systems are in place at NYPH. In this study, we used data from Allscripts's Sunrise Clinical Manager for clinical care, Cerner Millennium for ancillary services, and Eagle Registration for administrative transactions.

4.2.3 Experiments

Four experiments were designed to demonstrate applications of each of the above definitions to EHR data. A fifth experiment was used to compare the datasets deemed complete according to each of the four definitions. We sampled representative data types for each definition. Specifically, we selected data types that are expected to be present in most EHRs, and which are commonly required in research use cases. These data types include, but are not limited to, admission and discharge information, laboratory results, medication orders, and basic demographic information. Each day a patient was present in the hospital or an affiliated medical office represents an opportunity to observe and record data on the patient state. Each data observation and recording opportunity, in turn, includes multiple data types (e.g., diagnosis, laboratory result, etc.).

EHR completeness can be measured at different levels of granularity. One might examine, for example, the completeness of a full patient record (e.g. each patient represents a potential subject or case), or of specific data types (e.g., lab values are extracted and aggregated across patients). It can also be argued that at any granularity, EHR data never have total completeness. For the purposes of this demonstration, however, we have chosen to measure completeness at the patient record level, and have categorized records as either complete or incomplete according to each definition of completeness. Completeness according to each definition, therefore, is reported in number of patient records that meet the relevant criteria.

Rather than provide generalizable completeness findings for EHR data, our goal is to explicitly define and measure completeness from various perspectives and to illustrate the misalignment and intersections among different definitions of completeness.

4.2.3.1 Documentation

If a complete record must contain all information that was gathered during a clinical encounter—a potential data collection point-- a record is incomplete if there was a failure in the recording process. Determining when there was a failure to record data, however, is difficult without a reference standard. NYPH policy dictates that every day that a patient is present in the hospital or one of its affiliated offices, a narrative note should be entered into their record. Therefore, to illustrate this definition, we considered a record without a note on any day that a patient was present for treatment to be incomplete. Inherent in this approach is the assumption that visits are themselves appropriately recorded.

We extracted visit data on all patients in the NYPH clinical data warehouse and determined on which days they were present. Each day was considered to be a potential data collection point. We then identified all days where a patient had a narrative note or report recorded. Every day a patient was present without an associated note or report was said to be a data point that did not meet the definition of documentation completeness.

4.2.3.2 Breadth

When researchers require a breadth of information about patients, a record is considered complete if certain desired types of information are present. The information required for a record to be deemed complete will vary according to the research task at hand. For this experiment, we chose to look for the presence of five data types frequently found in patient records: laboratory results, medication orders, diagnoses, sex, and date of birth. In this example, a patient with all five data types present would be said to have a complete record. Given the multiplicity of laboratory tests, we also looked specifically at two common laboratory results: blood glucose and

hemoglobin measurements. For all patients, we measured the coverage of laboratory results, medication orders, and diagnoses for each day that they were present in the hospital or an affiliated office. The presence of sex and date of birth were assessed once for each patient.

4.2.3.3 Density

Some research tasks require the availability of multiple data points over time. Moreover, these data points may be required with some degree of regularity or covering a desired period of time. A complete record, therefore, would be one with a desired number of data points over a set period of time, spaced at sufficiently even intervals. For this experiment, we looked at the quantity and temporal distribution of patient visits, medication orders, and laboratory results. We approached this view of completeness in two ways. First, we looked at the number of clinical data points over the course of a patient record. Second, we applied an adjustment described by Sperrin et al. that accounts for the temporal irregularity of data.(30)

$$I = 2/n + \frac{(n-2)}{n}\left[1 - \sqrt{(n-1)Var\{g_t; i = 1, \ldots, n-1\}}\right]$$

where

$$g_t = \frac{x_{i+1} - x_i}{x_n - x_1}$$

I gives the average amount of information provided by each data point by accounting for the variability between those points. In the ideal situation, where all points are evenly spaced, *I* = 1. Multiplying *I* by *n* gives the number of effective data points. Sperrin et al. also proposed a linear adjustment that may be used to determine not only how evenly spaced data are, but to what extent a period of interest is covered by those data points. A set of points evenly spaced over a month may give sufficient information about that month, but if the period of interest is a full year, that information becomes insufficient. The adjustment, given a period of interest [a, b], is shown below.

$$I^* = I \times \frac{\min\{b, x_n\} - \max\{a, x_1\}}{b - a}$$

4.2.3.4 Predictive

One goal of reusing EHR data is to predict something or to find associations. Therefore, a record that contains sufficient information to predict successfully can be considered to be sufficiently complete for the stated purpose. We illustrated the definition for predictive completeness by assessing our ability to predict return visits. Such prediction is important in the context of health care reform, because institutions are striving to reduce readmission rates, and predicting who is likely to return allows institutions to target resources to prevent readmissions. We employed a logistic regression model using type and number of visits, number of medications, and number and value of common laboratory tests as the independent variables and using the presence of a gap of 180 days or more in future visits as the dependent variable.

4.2.3.5 Comparison of Completeness Definition Results

Further analysis was performed in order to compare records considered to be complete according to the four definitions of completeness. A documentation complete record was one with at least one visit accompanied by a narrative note. Records with breadth completeness were those that included a patient's date of birth, sex, and at least one medication order, laboratory test, and diagnosis. For density, we considered the presence of medication orders and laboratory tests over time, since these data types represent common clinical actions. Temporal resolution was considered down to the second. Sperrin's *I* was used to calculate the number of effective data points. Finally, we determined the predictive completeness of records using a simplified version of the logistic regression model described in section 2.3.4. The dependent variable was a gap in each patient record of at least 180 days, and the independent variables were counts of medication orders, laboratory results, and visits in the three, six, and twelve months preceding a potential gap.

4.3 Results

4.3.1 Documentation Completeness

Of the approximately 3.9 million patients with data in the clinical data warehouse, 48.3% have at least one visit recorded where a free-text note or report would be expected. Due to the gradual process of EHR adoption within NYPH, the percentage of missing notes has dropped drastically over the years (Figure 2). The overall rates of non-missing notes compared to the rates of visits are shown in Figure 3. Of all the patients with data in the clinical data warehouse, 18.5% have at least one visit with an associated note or report, 7.1% have five or more, and 4% have ten or more. Since 1986, 23.6% of all recorded visits have been accompanied by notes or reports. Over the most recent calendar year, however, the rate of completeness according to this definition has been significantly higher: 98.6% of inpatient visits, 73.8% of outpatient visits, and 95.0% of emergency visits have same day notes or reports recorded.

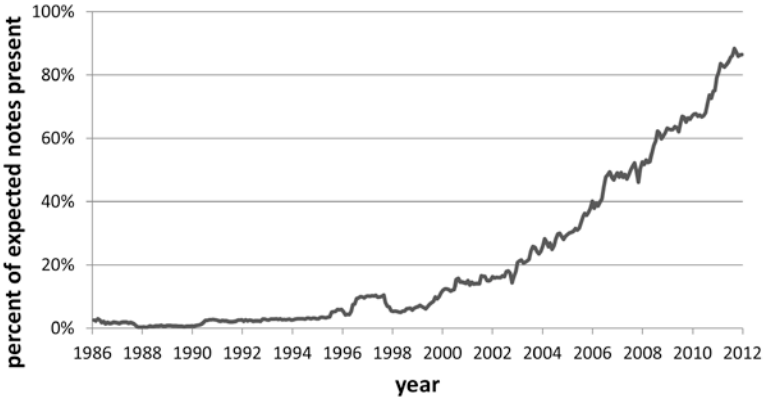


Figure 4.1 Documentation completeness over time. The documentation completeness of records has improved as documentation practices have changed and EHR adoption has increased.

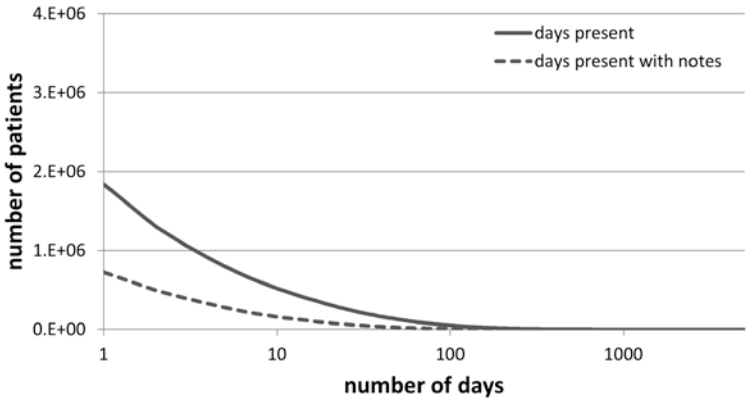


Figure 4.2 Documentation completeness. Shows the number of patients who have been present in the hospital for a certain number of days, as well as the number of patients whose records have notes associated with a certain number of days that they have been present.

4.3.2 Breadth

Of the patients with data in the clinical data warehouse, 29.3% had at least one visit with a recorded laboratory result (20.0% glucose, 23.0% hemoglobin), 12.6% had at least one with a medication order, and 44.5% had at least one with a diagnosis. The vast majority of patient records included basic demographic information: 97.8% had a valid date of birth recorded, and 99.6% had sex recorded.

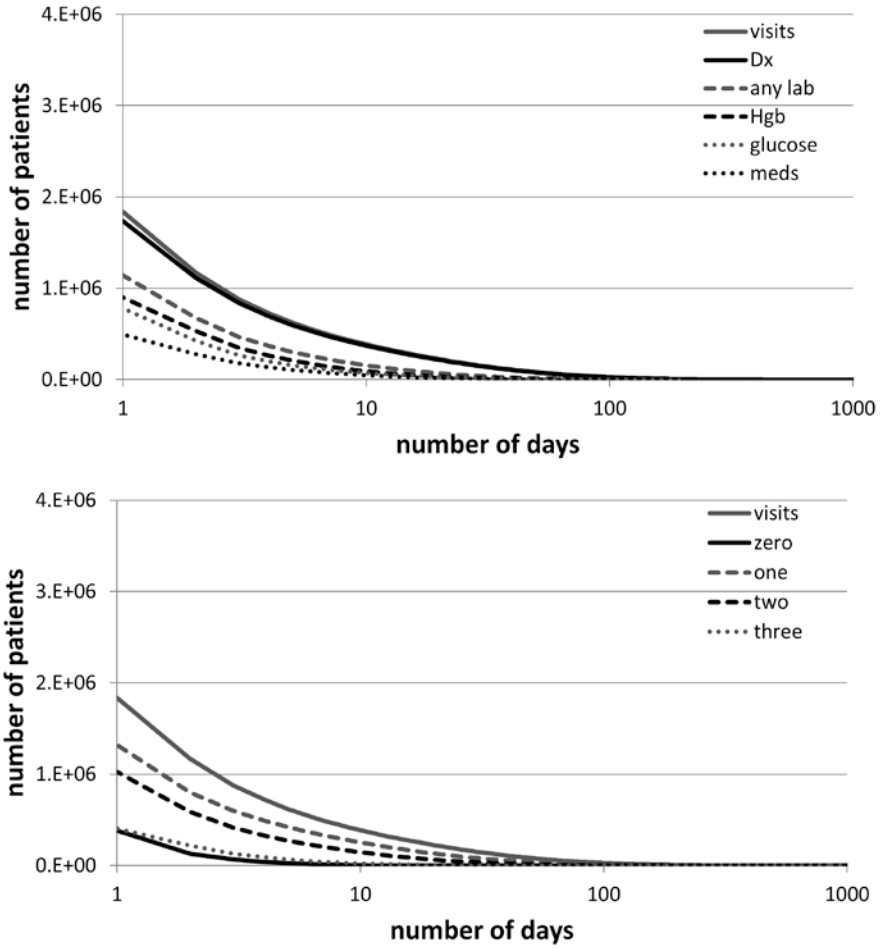


Figure 4.3 Breadth completeness. The number of patients with laboratory results, medication orders, and diagnoses on the same day as compared to the number of days when they were present in the hospital. Below, the number of patients with zero, one, two, or all of these data types present in their record on the same day.

Figure 4 shows the rates of visits with associated medications, laboratory tests, and diagnoses, as well as the rates of visits with none, one, two, or all three types of information. Of the patients with records in the clinical data warehouse, 10.4% had at least one visit with all three data types, 26.2% had at least one visit with exactly two, and 33.8% had at least one visit with exactly one.

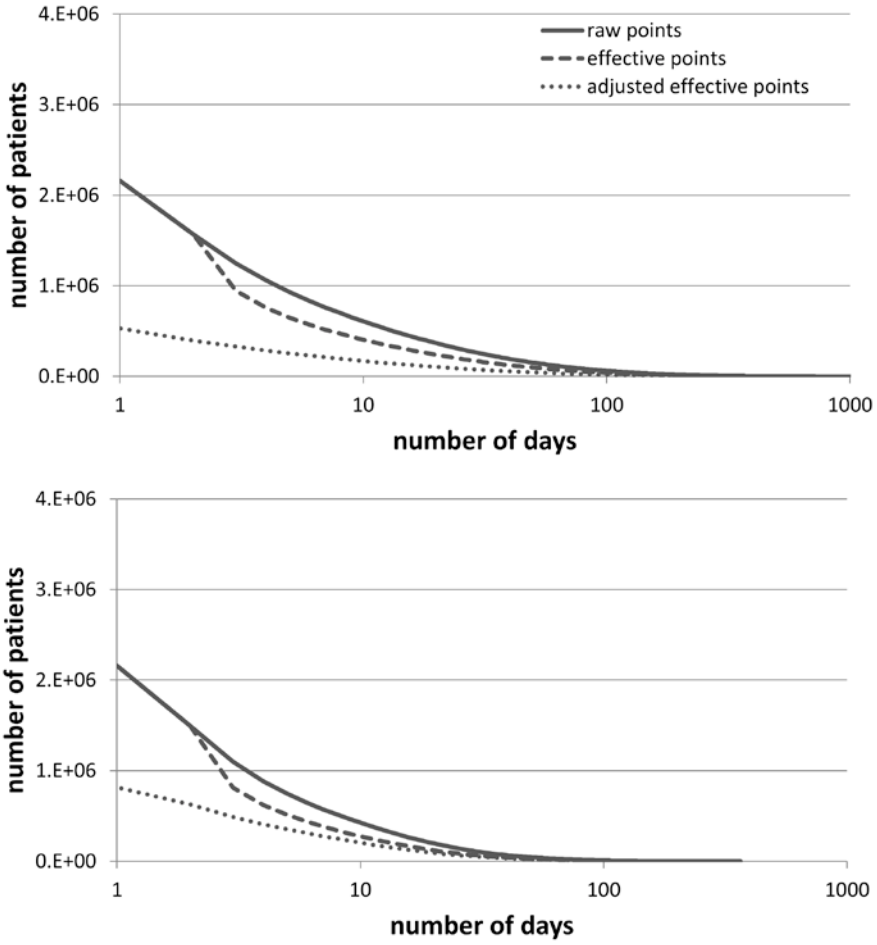


Figure 4.4 Density completeness of records. The number of patients with a given number of days with recorded visit events, laboratory results, or medication orders. The raw number of days, the number of days adjusted for variance, and the number of days adjusted for variance and time period are shown.

4.3.3 Density

Overall, 55.4% of the patients with records in the clinical data warehouse had at least one day with a recorded admission event, discharge event, laboratory result, or medication order. Twenty-three point eight percent had at least five, and 15.6% had at least ten. With Sperrin's *I* applied, 16.6% had at least five, and 10.4% had at least ten. With Sperrin's *I* and the linear adjustment, these figures dropped even further: 13.6% had at least one, 6.5% had at least five, and 4.4% had at least ten. If the time span of interest is limited to the year in which each patient spent the most days at the hospital, the rates of raw visits and effective visits meeting criteria are lower,

but the rates of adjusted visits are higher. Figure 5 shows the rates of raw, effective, and adjusted counts of days that patients were present in the hospital.

4.3.4 Predictive

We were able to predict 180-day-or-greater gaps in visits and data with an accuracy of 0.89. The area under the receiver operating characteristic curve was 0.79. The nature of the visits and duration of the records were predictive of gaps. Based upon this conceptualization of completeness, unlike breadth or density completeness, individual cases are predicted either correctly or incorrectly, so there is no sense of an intermediate completeness on an individual case.

4.3.5 Comparison of Completeness Definition Results

A comparison of the records satisfying the breadth, density, documentation, and predictive definitions of completeness is shown in Figure 6.(43) Overall, 55.7% of patients in the CDW have at least one point of clinical data, and 26.9% meet the criteria for at least one definition of completeness. In terms of density, only 11.8% have a complete record when completeness is defined as at least 15 laboratory results or medication orders adjusted for temporal variance. When completeness is defined as a breadth of five data types of interest (date of birth, sex, medication order, laboratory test, and diagnosis), 11.4% of patients have complete records. Patients with documentation complete records —meaning they had at least one visit with an associated note-- accounted for 18.5% of all patients. Finally, the presence or absence of a gap of 180 days or more could be correctly predicted for 8.4% of patients. Only 0.6% of patient records could be considered complete according to the implementations of all four definitions.

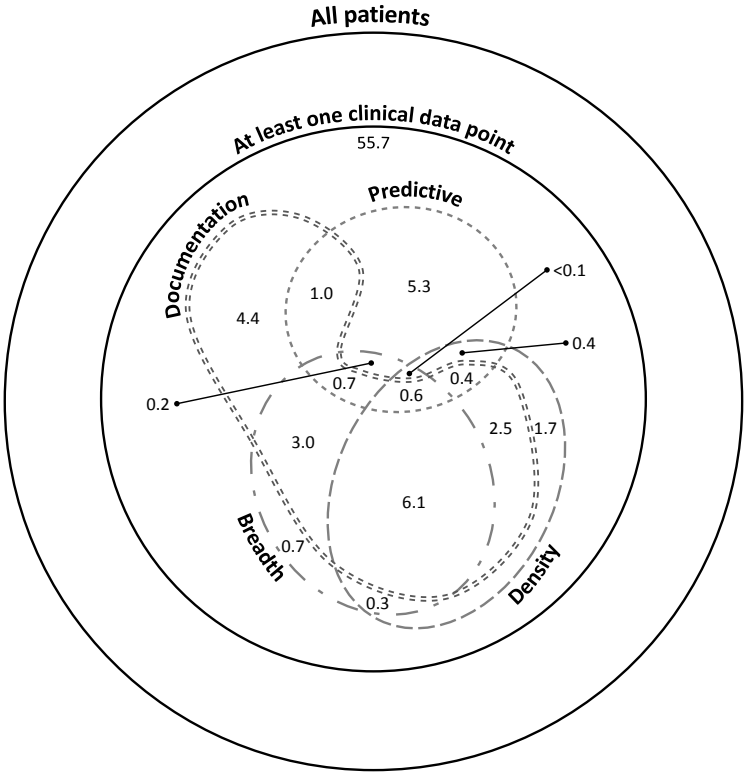


Figure 4.5 Comparison of completeness definition results. Subsets of patients with complete records according to the density (medication orders and laboratory tests over time with Sperrin's adjustment), breadth (record includes date of birth, sex, and at least one medication order, laboratory test, and diagnosis), documentation (at least one visit accompanied by a note), and predictive (a gap of 180 days can be correctly predicted) completeness definitions.

4.4 Discussion

At the time of this study, the clinical data warehouse contained the electronic records of approximately 3.9 million patients, but the number of records with sufficient information for various analyses is likely much lower. Only about half would be considered complete according to any of the four definitions using the least stringent cutoffs (e.g., at least one data point, at least one visit, or at least one medication or laboratory result). Only about a quarter would be considered complete with more detailed data requirements (e.g., at least one visit with an associated note or laboratory result, at least five visits over the course of a record). When limited not only to complete records, but also to a relevant cohort, the amount of useful information will drop even further. By any definition only a fraction of all the records are complete and suitable for reuse.

Moreover, the number of records in the relevant dataset varies depending upon the definition of completeness being used, which is in turn dependent upon user needs. Someone who is interested in patient care or outcomes over the longitudinal patient record will require very different data from someone looking at a cross-section of a patient population or someone studying the quality of care delivered at a medical institution. These users might identify complete records through, respectively, the density, breadth, and documentation completeness definitions described in this paper. As we have shown, each of these definitions results in a different number of complete records. Before making a determination of how many complete records are available for analysis, therefore, a researcher should first determine and specify what their data needs are, and then select the appropriate definition of completeness and provide it together with the completeness analysis result.

Further complicating the issue of completeness is the fact that not only do different definitions of completeness result in different numbers of useable records, these definitions may also point to different sets of relevant records. One might expect that a record that satisfies one definition of completeness is likely to satisfy another, but this is not necessarily the case. As shown in the comparison of the four definitions of completeness, the resulting sets of useable records share only partial overlap (Figure 6). In this study, documentation completeness suggests breadth or density completeness, possibly because our method of determining documentation completeness (section 2.3.1) requires the presence of at least one recorded visit. Predictive completeness, on the other hand, has little overlap with the other three result sets. Although 26.9% of the records in our CDW meet the criteria for at least one of the definitions of completeness, only 0.6% meet the criteria for all four. Therefore, explicitly selecting a relevant definition of EHR completeness is necessary to identify not only how many records are complete, but also which records are complete.

It is important to note that a range of defined completeness is possible and will depend in part upon the complexity of the task for which the data will be used. Taking a trivial example

based on the concept of predictive completeness, predicting the patient's age next year requires only the current age, implying most of the patients' records are complete, but predicting the age at which a patient will die is very difficult. Patients with rapidly fatal diseases may be predicted from their diagnoses, but others would be more difficult. Similarly, simple research tasks are likely to require less breadth or density of data than more complex tasks.

There may be analytic ways to address or avoid incompleteness. For example, the algorithm to predict gaps could be used to decide if an individual record is complete. If a patient has a gap and a gap was predicted from preceding data, then perhaps the gap was real; for example, the patient may have been healthy during the period. If, however, the patient has a gap and a gap was not predicted, then perhaps some data are missing. For example, perhaps the patient did have visits but the patient went to a different health provider. Thus the prediction may indicate the likelihood of completeness in the sense of the first definition (i.e., were the data that should have been there present). One could then potentially filter out cases with apparently missing visits.

4.4.1 Limitations

The rates of complete records identified in this study are not generalizable to other institutions. Differences in populations served, settings, workflows, HIT, and data procedures result in unique data profiles. The definitions of completeness described in this study, however, are not specific to our institution. The idea that information quantity can only be determined following the identification of a relevant definition of EHR completeness and the selection of an appropriate method of measurement is generalizable.

The definitions of completeness described in this study are primarily illustrative and are not exhaustive, as we may have failed to take into account all the needs of potential data consumers. We did not study, for example, the relationship between record completeness and underlying patient status. That is, a healthy patient's record would be expected to look very

different from a sick patient's. Further work is needed to more thoroughly and rigorously model the concept of completeness as it relates to the secondary use of EHR data.

The four definitions of completeness described in this study also require further exploration. In the case of predictive completeness, for example, it is unclear how to interpret the result: what level of prediction is sufficient to consider the EHR to be complete? Complicating this is the difficulty distinguishing the cause of low predictive accuracy. It could be because of lack of data, tackling a problem that is hard to solve, or the difficulty of developing an appropriate model.

Finally, completeness is closely tied to other dimensions of data quality. In examining completeness, we made no assumptions regarding the correctness of the data. The fact that data are present does not mean that they are necessarily trustworthy. A full assessment of an EHR-derived dataset prior to reuse should go beyond completeness.

4.5 Conclusion

We have illustrated that multiple definitions of completeness may be used, that they lead to different degrees of measured completeness for the same dataset, and that the number of complete records in a typical clinical database may be far lower than the nominal total. As researchers and clinicians continue the trend of repurposing EHR data for secondary use, it is important to bear in mind that these clinical data may not satisfy completeness requirements. Completeness, however, is contextual and is determined through an understanding of specific data needs. The number of complete records available for analysis is dependent upon the definition of completeness being used. Each definition results in a different set of complete records. We urge EHR data consumers to be mindful of the potential limitations of a dataset prior to committing to its use, explicit in their choice of completeness definition, and transparent about completeness findings when reporting results.

4.6 References

1. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care. 2007;13(6 Part 1):277-8.

2. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1-9.

3. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51.

4. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326(7398):1070.

5. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4(5):342-55.

6. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67(5):503-27.

7. Rubin D. Inference and missing data. Biometrika. 1976;63(3):581-92.

8. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods. 2002;7(2):147-77.

9. Blobel B. Advanced and secure architectural EHR approaches. Int J Med Inform. 2006;75(3-4):185-90.

10. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. AMIA Summits Transl Sci Proc. 2010;2010:1-5.

11. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002;9(6):600-11.

12. Juran JM, Gryna FM. Juran's quality control handbook. 4th ed. New York: McGraw-Hill; 1988.

13. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems. 1996;12(4):5-34.

14. Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. Pharmacoepidemiol Drug Saf. 2004;13(7):437-41.

15. Logan JR, Gorman PN, Middleton B. Measuring the quality of medical records: a method for comparing completeness and correctness of clinical encounter data. Proc AMIA Symp. 2001:408-12.

16. Barrie JL, Marsh DR. Quality of data in the Manchester orthopaedic database. BMJ (Clinical Research Ed). 1992;304(6820):159-62.

17. Hohnloser JH, Fischer MR, Konig A, Emmerich B. Data quality in computerized patient records. Analysis of a haematology biopsy report database. Int J Clin Monit Comput. 1994;11(4):233-40.

18. Ricketts D, Newey M, Patterson M, Hitchin D, Fowler S. Markers of data quality in computer audit: the Manchester Orthopaedic Database. Annals of the Royal College of Surgeons of England. 1993;75(6):393-6.

19. Roukema J, Los RK, Bleeker SE, van Ginneken AM, van der Lei J, Moll HA. Paper versus computer: feasibility of an electronic medical record in general pediatrics. Pediatrics. 2006;117(1):15-21.

20. Centers for Medicare & Medicaid Services. Quality Measures [updated 4/9/20135/16/13]. Available from: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/index.html.

21. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet. 2010;86(4):560-72.

22. Thadani SR, Weng C, Bigger JT, Ennever JF, Wajngurt D. Electronic screening improves efficiency in clinical trial recruitment. J Am Med Inform Assoc. 2009;16(6):869-73.

23. Weng C, Batres C, Borda T, Weiskopf NG, Wilcox AB, Bigger JT, et al. A real-time screening alert improves patient recruitment efficiency. AMIA Annu Symp Proc. 2011;2011:1489-98.

24. Agnew-Blais JC, Coblyn JS, Katz JN, Anderson RJ, Mehta J, Solomon DH. Measuring quality of care for rheumatic diseases using an electronic medical record. Annals of the Rheumatic Diseases. 2009;68(5):680-4.

25. Goulet JL, Erdos J, Kancir S, Levin FL, Wright SM, Daniels SM, et al. Measuring performance directly using the veterans health administration electronic medical record: a comparison with external peer review. Medical Care. 2007;45(1):73-9.

26. Jensen RE, Chan KS, Weiner JP, Fowles JB, Neale SM. Implementing electronic health record-based quality measures for developmental screening. Pediatrics. 2009;124(4):e648-54.

27. Linder JA, Kaleba EO, Kmetik KS. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. Medical Care. 2009;47(2):208-16.

28. Asche C, Said Q, Joish V, Hall CO, Brixner D. Assessment of COPD-related outcomes via a national electronic medical record database. International Journal of Chronic Obstructive Pulmonary Disease. 2008;3(2):323-6.

29. Einbinder JS, Rury C, Safran C. Outcomes research using the electronic patient record: Beth Israel Hospital's experience with anticoagulation. Proceedings / the Annual Symposium on Computer Application [sic] in Medical Care Symposium on Computer Applications in Medical Care. 1995:819-23.

30. Sperrin M, Thew S, Weatherall J, Dixon W, Buchan I. Quantifying the longitudinal value of healthcare record collections for pharmacoepidemiology. AMIA Annu Symp Proc. 2011;2011:1318-25.

31. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. J Am Med Inform Assoc. 2012;19(2):212-8.

32. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am J Hum Genet. 2011;89(4):529-42.

33. Overby C, Pathak J, Haerian K, Perotte A, Murphy S, Bruce K, et al. A collaborative approach to develop an electronic health record phenotyping algorithm for drug-induced liver injury. J Am Med Inform Assoc. 2013;In press.

34. Boland MR, Tu SW, Carini S, Sim I, Weng C. EliXR-TIME: A Temporal Knowledge Representation for Clinical Research Eligibility Criteria. AMIA Summits Transl Sci Proc. 2012;2012:71-80.

35. Hao T, Rusanov A, Weng C. Extracting and normalizing temporal expressions in clinical data requests from researchers. International Health Informatics Conference; Beijing, China2013.

36. Zhao D, Weng C. Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. J Biomed Inform. 2011;44(5):859-68.

37. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al. A predictive model for progression of chronic kidney disease to kidney failure. JAMA : the journal of the American Medical Association. 2011;305(15):1553-9.

38. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Medical care. 2010;48(6 Suppl):S106-13.

39. Cholleti S, Post A, Gao J, Lin X, Bornstein W, Cantrell D, et al. Leveraging derived data elements in data analytic models for understanding and predicting hospital readmissions. AMIA Annu Symp Proc. 2012;2012:103-11.

40. Nijhawan AE, Clark C, Kaplan R, Moore B, Halm EA, Amarasingham R. An electronic medical record-based model to predict 30-day risk of readmission and death among HIV-infected inpatients. J Acquir Immune Defic Syndr. 2012;61(3):349-58.

41. Wells BJ, Jain A, Arrigain S, Yu C, Rosenkrans WA, Jr., Kattan MW. Predicting 6-year mortality risk in patients with type 2 diabetes. Diabetes care. 2008;31(12):2301-6.

42. Kennedy EH, Wiitala WL, Hayward RA, Sussman JB. Improved cardiovascular risk prediction using nonparametric regression and electronic health record data. Medical care. 2013;51(3):251-8.

43. Micallef L, Rodgers P. Drawing area-proportional venn-3 diagrams using ellipses. 12th Annual Grace Hopper Celebration of Women in Computing, ACM Student Research Competition and Poster Session; Baltimore, MD2012.

Chapter 5: Sufficient Electronic Health Record Data and Potential Sampling Bias[1]

5.1 Introduction

There is great promise in the reuse of electronic health record (EHR) data for clinical research purposes. Retrospective research that reuses existing datasets is generally faster and less costly than prospective research. EHR data have the added benefit of being representative of actual healthcare consumers. For these reasons, amongst others, there is growing interest in the secondary use of EHR data.(1-3)

When working with EHR data, however, it is important to be mindful of potential data quality caveats.(4) A number of studies have established that poor data quality may challenge the

[1] The majority of this chapter was originally published in the 2013 AMIA Annual Proceedings. Nicole Weiskopf and Alex Rusanov were listed as equally contributing first authors. The full citation for this publication is:
Weiskopf NG, Rusanov A, Weng C: Sick patients have more data: the non-random completeness of electronic health records. AMIA Annu Symp Proc 2013, 2013:1472-1477.

suitability of EHR data for research.(5-7) In our previous work, we identified three core dimensions of data quality in which researchers engaged in the secondary use of EHR data are interested: completeness, correctness, and currency.(8) We further illustrated that EHR data completeness is a task-dependent phenomenon, and that how record completeness is defined depends upon the intended use for the data.(9)

EHR data completeness may be understood differently in the clinical setting versus in the research setting. Clinically, completeness is most commonly understood to mean fidelity of documentation.(6) In other words, a record is considered complete if it contains all information that was observed. When repurposed for secondary use, however, the concept of "fitness for use" can be applied.(10, 11) In secondary use settings, EHR data completeness becomes extrinsic, and is dependent upon whether or not there are sufficient types and quantities of data to perform a research task of interest. It should be noted that a record deemed incomplete for a given research use might be complete from a clinical standpoint. As an example, the clinical requirement for frequency of blood pressure and heart rate measurement during provision of anesthetic care is every five minutes.(12) A record documenting these two variables at five-minute intervals would thus be complete from a clinical standpoint. A researcher, however, might be interested in beat-to-beat variations. A record documenting these variables at five-minute intervals would thus be incomplete from a research standpoint. Similarly, the record would also be incomplete if a researcher was interested in pulmonary artery pressures and these were not documented.

When considering the problem of incomplete data, it is important to understand the manner in which data are incomplete. In statistics, data are understood to be missing at random or missing not at random.(13, 14) When incompleteness is a random phenomenon, the signal may become noisier, but is otherwise unchanged. Non-random incompleteness, however, may introduce a spurious signal into a dataset. Therefore, it is important to understand not only how much data is missing, but also if those data are missing according to a certain pattern that might have clinical relevance. The same could be said of correctness as well: random error is less

problematic than systematic error. Similarly, consider data currency: if a random subset of the data are out of date it is less likely to cause a problem than if, for example, only data from less healthy patients are out of date. When engaged in the secondary use of EHR data for research, it is important to know whether the records identified as complete for a given research task are representative of all patients of interest, or if the act of selecting for completeness actually results in a non-representative group of records. The literature on this topic is limited.

Researchers who use EHR-derived data often apply implicit data quality requirements to their datasets. In selecting a cohort for study, eligibility criteria often include requirements that patients have at least *x* number of laboratory values, or have been followed for *y* number of months, or that they must have records containing *y* different points of data over time. The implication of these criteria is that to be eligible a patient must have a record that meets a certain level of data quality, specifically completeness in these examples. In the future, researchers may even have explicit criteria for record inclusion. Perhaps all records included will need to be free from detectable error. Any type of cohort restriction, however, introduces the possibility of selection bias. If a study only included patients under the age of twenty, the researchers would likely acknowledge that the results might not be generalizable to older adults. But what does it mean to say that the results of a study cannot be generalized to patients with less complete records? Is there a clinically meaningful difference between patients with good quality data and those with worse quality data?

In the study reported here, we hypothesize the existence of a relationship between record completeness and the underlying health status of the patient of interest. Through the use of representative data types and a broadly applicable definition of complete EHR data, our study proves that the selection of only complete records has the potential to create a biased, non-representative dataset.

5.2 Methods

5.2.1 Identification of a Health Status Indicator

In order to look at the relationship between patient health and record completeness from the secondary use perspective, we required the use of a reliable measure of health status that does not rely entirely on the presence of these data types in the record. The American Society of Anesthesiology (ASA) Physical Classification score(15, 16) is a subjective assessment of overall health or illness severity assigned by an anesthesia provider to every patient requiring anesthetic services. The ASA score is based not only on information present in the EHR but also on interviews with the patient, family and the patient's other healthcare providers well as information (i.e. laboratory results, imaging and other diagnostic testing results, medical records, prescriptions, etc.) from sources outside the institution and thus not present in the EHR. It is a severity of illness score that is prospectively assigned to each patient by a scoring expert (anesthesia provider) and is thus much less reliant on EHR data than retrospectively assigned scores (e.g Charlson Comorbidity Index) which rely solely on information present in the EHR. We also sought a measure that is routinely recorded for a broad spectrum of the patient population, as opposed to those that focus on a specific disease (e.g., New York Functional Status Classification for heart failure patients) or segment of the population (e.g., children). The score separates patients into one of six categories (Table 1), with the letter "E" appended for emergency procedures, and has been shown to be strongly correlated with other clinical risk predictors(17) as well as outcomes.(18-20) Although the requirement of an ASA score limits our cohort to those who have received anesthesia, we felt that it allowed us to capture a more broad population than the other illness measures and have an severity of index measure that was mostly independent of presence of data in the EHR.

5.2.2 Data

Table 5.1 . American Society of Anesthesiology (ASA) Physical Status Classification.

ASA Class	*Definition*
1	A normal healthy patient
2	A patient with mild systemic disease
3	A patient with severe systemic disease
4	A patient with severe systemic disease that is a constant threat to life
5	A moribund patient who is not expected to survive without the operation
6	A declared brain-dead patient whose organs are being removed for donor purposes

The Columbia University Department of Anesthesiology provides anesthetic services for operating rooms at three hospitals – a tertiary-care academic medical center (Milstein), a dedicated children's hospital (CHONY) and a small community hospital (Allen). Additionally, anesthetic services are also provided for the labor and delivery floors at CHONY and Allen as well as several "off-site" locations such as endoscopy, radiology, neuroradiology, cystoscopy, cardiac catherization and electrophysiology suites, and ophthalmologic surgery suites. For the majority of patients for whom we provide anesthetic services an anesthesia information management system, CompuRecord (Philips Healthcare, Andover, MA) is used for electronic recordkeeping. The ASA Class score is recorded for all valid procedures. The data collected by this system is periodically migrated to a research database, with the last update occurring in October 2012.

The Columbia University Medical Center Institute Review Board approved this study. We began by identifying all patients in the CompuRecord research database with a recorded ICD-9 and CPT code. These codes have been recorded in the database since April 2012, so we identified patients with anesthesia procedures recorded between April, 2012 and the end of September, 2012. To minimize bias introduced by having multiple procedures requiring anesthesia within our time period of interest (one year preceding the procedure), we focused only on the earliest procedure for each patient in our dataset and then excluded all patients who had a

procedure in the preceding year, or who were less than one year of age at the time of the procedure. From the remaining set of patients we pulled the ASA Class for a randomly selected set of 5000 patients for the current analysis. Due to the infrequent occurrence of ASA Classes 5 and 6 we dropped the patients in these classes from our analyses.

For demonstration purposes, we selected two common clinical data types through which to assess record completeness: laboratory results and medication orders. Because we did not have a specific use case in mind, we did not use a hard cut-off for defining a complete record. Instead, we assumed that each record existed on a continuum of less or more complete, as implied by counts of days where data were present. The raw counts of days with medication orders and days with laboratory results were calculated for each of the 5,000 patients during the year preceding their procedures using data obtained from the Clinical Data Warehouse (CDW). These data are drawn from a combination of Allscripts's Sunrise Clinical Manager for clinical care and Cerner Millennium for ancillary services.

5.2.3 Data Analysis Methods

We compared the completeness of records between ASA Classes for both data types using the Kruskal-Wallis one-way analysis of variance. Further post-hoc analyses were performed using the Wilcoxon rank-sum test with a Bonferroni correction.

5.3 Results

The average age of patients in our sample at the time of surgery was 45 with a standard deviation of 24 years, and 61% were female. The ASA Class distribution for our sample is summarized in Figure 1.

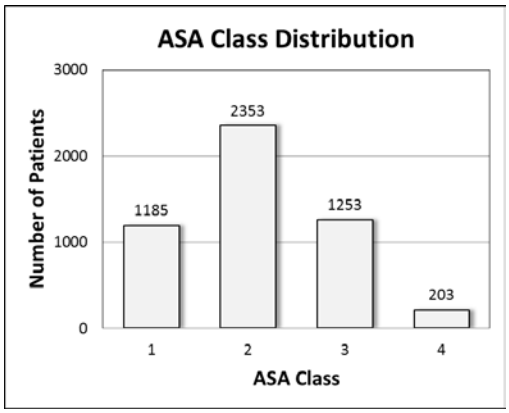


Figure 5.1 Distribution of ASA Class in study population.

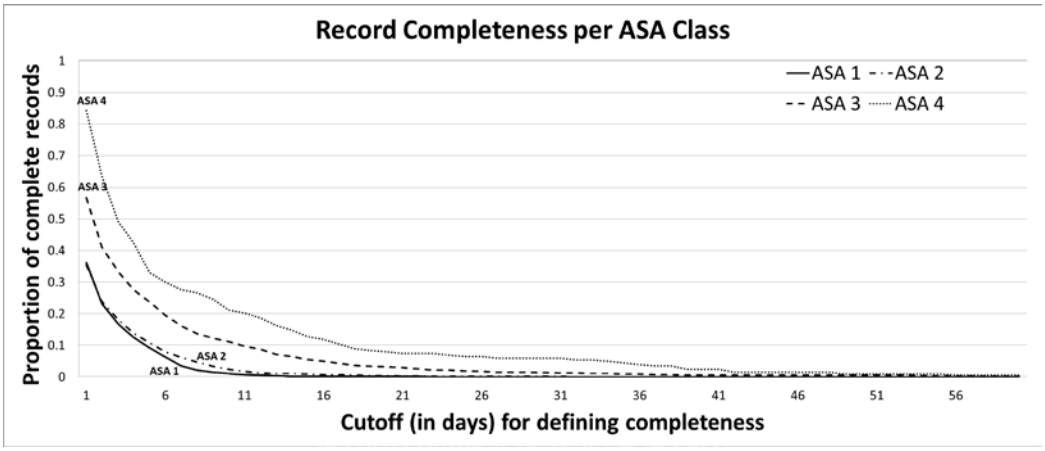


Figure 5.2 Record completeness, per ASA class, over a range of cutoffs (1-60), where cutoffs are the minimum number of values required in each of two categories (medication orders and laboratory results) to make a record complete. For cutoffs >60 all proportions are zero, and thus not shown.

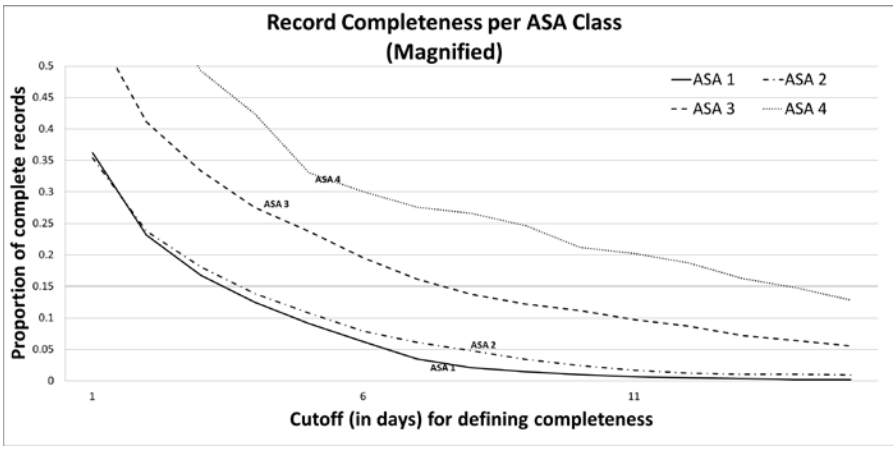


Figure 5.3 Magnified view of record completeness, per ASA class, over a range of cutoffs (1-15), where cutoffs are the minimum number of values required in each of two categories (medication orders and laboratory results) to make a record complete.

We defined as complete any record having at least N recorded values in each of the two categories (medication orders and laboratory results), where N is the cutoff value, in days. Using this definition, the ratio of complete records over total records in each ASA Class was plotted as N was varied from 1 to 60 days (Figure 2). A magnified view of the plot is shown (Figure 3) to facilitate visualization by improving resolution in the range where N is between 1 and 15. As a specific example, we show the distribution of the proportion of records that are complete across ASA Classes given an arbitrary cut-off of N=7 (Figure 4). As the cutoff of number of desired points increases, the distribution of patients skews further towards the records of patients with an ASA status of 3 or 4.

The average number of days with data for patients in each ASA Class are shown for medication orders and laboratory results in Figure 5. Patients with a more severe health status have a higher number of days with data, on average. A Kruskal-Wallice one-way analysis of variance revealed a significant effect of ASA Class on number of days with medication orders ($x^2(3)=332.0$, $p<0.0001$) and on number of days with laboratory results ($x^2(3)=202.2$, $p<0.0001$). Post-hoc analysis using a Wilcoxon rank sum test with Bonferroni correction showed significant

differences between all ASA Classes except Class 1 and Class 2 for both data types, all with p-values <0.0001.

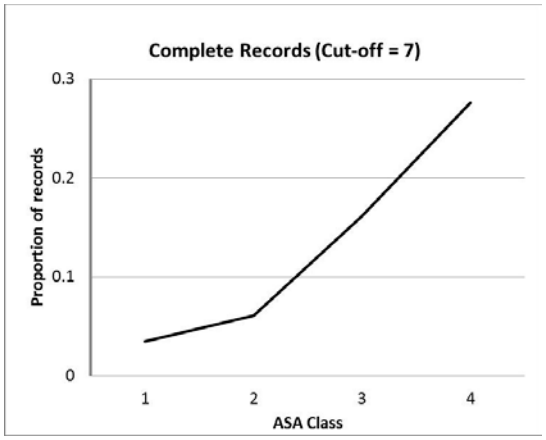


Figure 5.4 Complete records by ASA Class where complete records are those having at least seven values in each of the two categories (medication orders and laboratory results).

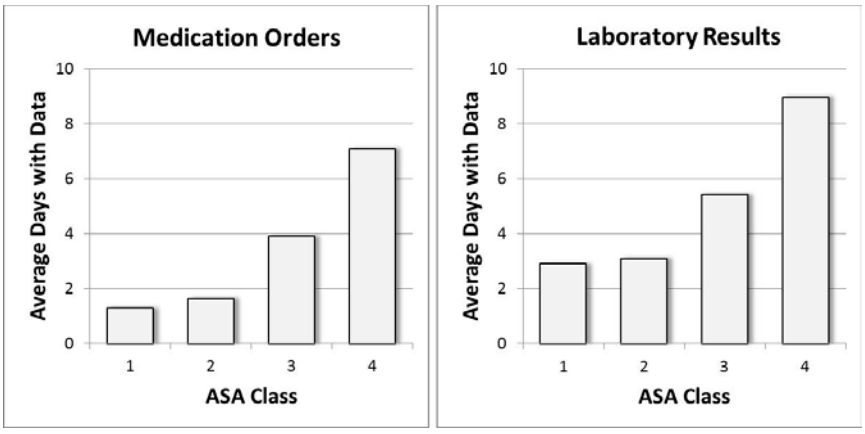


Figure 5.5 Average number of days with data per patient by ASA class. For both medication orders and laboratory results, all ASA Classes are significantly different except for Classes 1 and 2.

5.4 Discussion

The results indicate that our hypothesis that there is a relationship between patient health status and EHR completeness is correct. EHR data are not missing at random. Both medication orders and laboratory results show a significant increase in data points per patient as ASA Class

increases. Moreover, as the cutoff for what constitutes a complete record increases, the proportion of records considered complete concentrates towards the patients in higher ASA Classes.

This indicates that a sample of patients naively drawn from the EHR-- assuming the records chosen are limited to those with sufficient data to be considered complete for the task at hand-- will almost invariably be biased towards sicker patients. In other words, those patients with records deemed complete for a given study are unlikely to be representative of the population of interest. These findings are in line with previous work exploring the complex relationships among data quality, bias, and health status. In one example, Wennberg et al. used insurance claims data to demonstrate bias in comorbidity measurement by showing that Charlson Comorbity Index scores are associated with the frequency of physician visits,{Wennberg, 2013 #231} suggesting that data quality is compromised by differences in healthcare utilization. Similarly, Collins et al. identified a relationship between patient mortality and increased rates of nursing documentation, suggesting that more acutely ill patients are likely to have more thoroughly documented records.{Collins, 2013 #236;Collins, 2012 #237} In a study of a pneumonia severity index based upon EHR data, Hripcsak et al. found that the addition of cohort selection criteria that required the presence of sufficient data to make a reliable diagnosis substantially limited sample size and significantly altered mortality rates.{Hripcsak, 2011 #423} They note that the addition of simple sample restraints, while beneficial in their case, has the potential to significantly narrow the sample, leading to the possibility of bias. At least in the case of completeness, data quality problems introduce not only noise into EHR data, but bias as well.

A secondary finding suggested by our results is that data consumers may have a difficult time identifying healthy patients with sufficient data for secondary use. This is problematic for those seeking healthy controls or comparison cohorts for research purposes.

5.4.1 Limitations

Because of the reliance of our analysis on the presence of ASA Class for all patients in our sample, only anesthesia patients were included in this study. These patients may not be representative of the overall population with data in the CDW. By extension, the CDW itself may not be representative of typical EHR databases due to the tertiary academic medical setting. Nevertheless, we believe that the essential concept proven by this study—that there is a relationship between record completeness and patient health—will hold true across a broad range of clinical database and institutions.

5.4.2 Future Directions

We intend to continue this research by investigating factors that might be driving the observed trend. For example, we could consider the procedure performed, the diagnosis, the emergency status of the surgery, or inpatient/outpatient status prior to surgery. All are potential confounders.

It would also be helpful to focus on special populations, such as children or pregnant women. One might expect, for example, that pregnant women would be mostly ASA Classes 1 and 2, but with good chart completeness prior to their anesthetic due to pre-natal care. These special populations may require further consideration.

Finally, we hope in the future to identify or develop methodological approaches to selecting complete patient records from EHR databases while avoiding the introduction of bias in the form of patient health status.

5.5 Conclusion

In this sample of 5,000 patients the percentage of complete records varied with ASA Class. This relationship held true for laboratory result and medication order data. Using ASA Class as a surrogate for patient health, the data confirm our initial hypothesis that there exists a

statistically significant difference between the completeness of electronic health records for sick patients and the completeness of electronic health records for healthy patients. Sicker patients tend to have more complete records and healthier patients tend to have records that are less complete.

These results should serve as a word of caution to researchers wishing to use EHR data for research. Investigators wishing to reuse EHR data must be aware that blind sampling of complete records within an EHR database may skew the sampled population towards sicker patients.

5.6 References

1. Bloomrosen M, Detmer DE. Advancing the Framework: Use of Health Data—A Report of a Working Conference of the American Medical Informatics Association. J Am Med Inform Assoc. 2008;15(6):715-22.

2. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1-9.

3. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care. 2007;13(6 Part 1):277-8.

4. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EB, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical Care. 2013;51(8 Suppl 3):S30-7.

5. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67(5):503-27.

6. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4(5):342-55.

7. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326(7398):1070.

8. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51.

9. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform. 2013.

10. Juran JM, Gryna FM. Juran's quality control handbook. 4th ed. New York: McGraw-Hill; 1988.

11. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems. 1996;12(4):5-34.

12. American Society of Anesthesiologists. Standards for basic anesthetic monitoring 1986 [8/20/13]. Available from: http://www.asahq.org/For-Members/~/media/For%20Members/documents/Standards%20Guidelines%20Stmts/Basic%20Anesthetic%20Monitoring%202011.ashx.

13. Rubin D. Inference and missing data. Biometrika. 1976;63(3):581-92.

14. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods. 2002;7(2):147-77.

15. Saklad M. Grading of Patients for Surgical Procedures. Anesthesiology. 1941;2(3):281-4.

16. Dripps R. New classification of physical status. Anesthesiology. 1963;24:111.

17. Davenport D, Bowe E, Henderson W, Khuri S, Mentzer RJ. National Surgical Quality Improvement Program (NSQIP) risk factors can be used to validate American Society of Anesthesiologists physical status classification (ASA PS) levels. Ann Surg. 2006;243(5):636-41.

18. Dripps R, Lamont A, Eckenhoff J. The role of anesthesia in surgical mortality. JAMA. 1961;178(3):261-6.

19. Vacanti C, VanHouten R, Hill R. A statistical analysis of the relationship of physical status to postoperative mortality in 68,388 cases. Anesth Analg. 1970;49(4):564-6.

20. Wolters U, Wolf T, Stützer H, Schröder T. ASA classification and perioperative variables as predictors of postoperative outcome. Br J Anaesth. 1996;77(2):217-22.

21. Collins SA, Vawdrey DK. "Reading between the lines" of flow sheet data: nurses' optional documentation associated with cardiac arrest outcomes. Appl Nurs Res. 2012;25(4):251-7.

22. Collins SA, Cato K, Albers D, Scott K, Stetson P, Bakken S, et al. Relationship between nursing documentation and mortality. American Journal of Critical Care. 2013;22(4):306-13.

Chapter 6: Development and Evaluation of Electronic Health Record Data Quality Assessment Guideline

6.1 Introduction

The previous aims have highlighted the need for systematic methodology for EHR data quality assessment within the context of secondary use. Existing reviews of the informatics literature have demonstrated that EHR data tend to be of poor and variable quality.(1-3) Researchers interested in using EHR data must be given the tools necessary to assess data quality. Unfortunately, current approaches to EHR data quality assessment tend to be ad hoc, which limits the transparency of data quality findings and the validity and comparability of results from research studies that use these data.(4) This is especially disappointing in light of the fact that a potential advantage of the secondary use of EHR data is improved generalizability due to the increased representativeness of the subjects when compared to more traditional research.(5)

A guideline-based approach to EHR data quality assessment has the potential to improve the validity, comparability, and transparency of research conducted using EHR data. Guidelines also have the potential to decrease the burden placed upon researchers, who must otherwise find or develop methods for data quality assessment. Substantial research on clinical guidelines, which are also intended to simplify and standardize complex processes, have been shown to improve the consistency and quality of care.(6) Ideally, a data quality assessment guideline would do the same for the secondary use of EHR data.

An EHR data quality assessment guideline, however, must be more than just systematic. It must also be evidence-based where possible, or at least expert knowledge-based. It must also consider the needs and preferences of clinical researchers, who would be the target users. Lastly, because EHR data quality is task-dependent, the guideline must be flexible enough to accommodate a fitness-for-use approach to assessment; not all data quality measurements and measurement methods will be relevant for all research or all data.(7, 8)

This chapter describes the design and evaluation of 3x3 DQA: Dynamic, evidence-based guidelines to enable electronic health record data quality assessment and reporting for retrospective research.

6.2 Methods

3x3 DQA is the cumulative product of three different studies, each of which identified concepts, methods, and priorities that were incorporated into the guidelines. From the beginning, the overall goal was for the guidelines to be clear, concrete, knowledge- and evidence-based, and actionable. The specifics, however, including the data quality constructs addressed, the methodology described, and the manner in which the entire process of EHR data quality assessment was approached, were determined by a combination of literature-based, qualitative, and data-driven research.

6.2.1 Initial Guideline Development

6.2.1.1 Guideline Format Considerations

Before beginning the process of designing 3x3 DQA, we knew that it would need to contain a great deal of information, which can be overwhelming and confusing if not presented properly. Initially, we had intended to develop the guidelines using the SAGE guideline model, which can be presented using a simple flowchart approach that includes three types of nodes: context, decision, and action.(9) After gaining a deeper understanding of the complexities of EHR data quality assessment, however, it became clear that SAGE and other similar models were not suitable. The goal of SAGE and other clinical guideline models is to identify a specific end action. In EHR data quality assessment, however, there are likely to be multiple end actions (i.e. multiple methods to assess the quality of the data). We needed to use a knowledge representation model that would allow users to identify multiple relevant constructs of data quality and multiple associated methods of data quality assessment.

Successful representations of complex research processes often present information using at least two levels of information. The first level, which is intended to be easily accessible, may be a visual or simple textual representation of the framework. More information, in substantial detail, is then available if the user wants to proceed with the model or framework. Two examples of such an approach are the RE-AIM framework, which can be succinctly summarized by its full name (**R**each **E**ffectiveness **A**doption **I**mplementation **M**aintenance), and the Precede-Proceed model used in Chapter 3 of this dissertation, which is presented as a visual model.(10, 11) Both are used to guide the complex planning, implementation, and/or evaluation of public health interventions.

Therefore, we decided that the best approach to the complexities of EHR data quality assessment was to use a simple, visual framework that summarizes the process in a conceptually clear fashion, coupled with in-depth guidelines describing the assessment methods dictated by the

framework. As indicated by the findings of Aim III (Chapter 4), we knew that parts of the framework and associated guidelines would not be needed in all use cases, and therefore also decided to develop a series of questions that would help users determine the scope of their research and identify the relevant sections of the framework and guidelines.

6.2.1.2 Data Quality Construct Identification

The constructs of data quality included in the framework were derived from two sources: the literature review described in Chapter 2, and the interviews with clinical researchers in Chapter 3. The literature review identified five categories of data quality that are described and assessed in the informatics literature: complete, correct, concordant, plausible, and current data. It was determined, however, that data concordance and plausibility, however, are used as methodological proxies for the other categories of data quality, especially correctness. A lack of concordance between two variables, for example, indicates that one or both of those variables are incorrect. Similarly, when one states that a specific value is implausible, the implication is that it is not correct. The connection between concordance, plausibility, and correctness is illustrated in a data accuracy assessment comparison hierarchy for healthcare data proposed by Zozus et al.{Zozus, Assessing Data Quality for Healthcare Systems Data Used in Clinical Research: A Healthcare Systems Research Collaboratory Core White Paper #466} Their hierarchy suggests that there are a number of ways to assess the accuracy of healthcare data, ranging from gold standard data (rarely available) to known standards and other data values.

We also decided to set limitations on the scope of the guideline based upon previous work on data quality. Wang and Strong defined four broad categories in their hierarchical model: intrinsic, contextual, representational, and accessibility data quality.{Wang, 1996 #405} They state that the intrinsic constructs are those that capture the quality of data in their own right, without knowledge of how the data will be used, while contextual data quality refers to the fact that some constructs of data quality are task-dependent. Both are concerned with the quality of

the data themselves. Representational data quality and accessibility, however, largely concern the systems and formats in which the data are measured, recorded, and stored. EHRs and other health information technology systems are extraordinarily complex, which is already the subject of a substantial field of research. We decided to limit the scope of our data quality assessment guidelines to those aspects of data quality that concern the data themselves, rather than those that evaluate systems.

Table 6.1 EHR data quality constructs, their sources, their categories, and their inclusion status.

Construct	*Literature*	*Researchers*	*Proxy*	*Category*	*Included*
Completeness	X	X		contextual	X
Concordance	X	X	X	intrinsic	
Correctness	X	X		intrinsic	X
Currency	X			contextual	X
Fragmentation		X		representational / accessibility	
Granularity		X		contextual	
Plausibility	X		X	intrinsic	
Signal-to-noise		X		representational / accessibility	
Structuredness		X		representational / accessibility	

Of the seven constructs of data quality identified through interviews with clinical researchers, three would be considered aspects of accessibility or representational data quality: fragmentation, structuredness, and signal-to-noise. All are concerned with how data are recorded and stored, which in turn influences how easy they are to retrieve. Of the four remaining constructs, one is concordance, which, as described above, we determined to be a proxy of correctness. Another, granularity, though clearly important to researchers, is very rarely addressed in the informatics literature or literature on data quality more broadly. We did not feel that there was sufficient information available to make methodological recommendations regarding the assessment and reporting of granularity. The remaining two constructs, correctness

and completeness, overlap with the three core constructs from the literature review. Table 1 summarizes the sources and considerations for each of the nine total constructs.

6.2.1.3 Framework Development

The primary goal of the framework was to give users a conceptual overview of what data quality means, how applies to the secondary use of EHR data, and how data quality assessment can be approached from a practical standpoint. Aim III (Chapter 4) demonstrated that EHR data completeness can be operationalized in different meaningful ways depending upon what data will be needed and how they will be used. Therefore, we determined that the framework should extend this concept of multiple operationalizations of the most important data quality constructs, which were identified in Aims I (Chapter 2) and II (Chapter 3), and link these operationalizations to different possible data "needs" that would in turn be dictated by the user.

6.2.1.4 Recommendation Development

The detailed methodological recommendations were dictated by the different operationalizations from the framework. Therefore, each data quality construct, having multiple operationalizations, would also have multiple recommendations. A primary goal of 3x3 DQA is transparency of EHR data quality findings, so the recommendations had to include not only methodological approaches, but suggestions for reporting as well. Ideal recommendations are knowledge- and evidence-based, so the methodological approaches were pulled from literature on EHR data quality assessment and data quality more broadly. Many of these sources were initially identified in Aim I (Chapter 2).

6.2.1.5 Scope Identification Questions

Once the 3x3 DQA framework had been developed, all combinations of the operationalizations were calculated. Next, various assumptions and logical rules were applied, and a subset of realistically possible combinations was identified. A series of yes or no questions was then developed to help users identify which data quality constructs and operationalizations

were relevant for their intended research. The possible combinations of answers to these questions corresponded to the subset of operationalizations deemed realistic.

6.2.2 Evaluation

Following the initial draft of 3x3 DQA, an expert-based evaluation was developed to guide further iterative development of the document. Other evaluation approaches, including scenario-based testing and pilot testing with actual users, were determined to be more appropriate later in the development process. The evaluation instrument was based largely upon the literature on psychometric instrument development and testing. Unfortunately, although this kind of expert-based testing is often used, it is rarely reported in detail.(12)

An expert-based evaluation was conducted instead of a user-based evaluation for multiple reasons. First, an expert-based evaluation seemed more appropriate at this stage in the development of 3x3 DQA. User-based evaluations are more complex and require greater time and resources, and it was therefore determined that a lighter-weight expert evaluation and subsequent round of iterative design should be performed prior to user-based evaluation. Second, the primary goal of this evaluation was to assess the validity, clarity, and comprehensiveness of 3x3 DQA, which is especially well-suited to experts. Users, in contrast, are better-suited to evaluating features like usefulness and ease-of-use.

Finally, this was the first opportunity to incorporate expert knowledge and opinions into the development of 3x3 DQA. Up to this point, 3x3 DQA had been the product of a literature review (Chapter 2), a data-driven exploration of EHR data quality (Chapters 4 and 5), and feedback from target users (Chapter 3). Input from actual experts on EHR data quality within the context of secondary use, however, had not been elicited in any direct way. This initial evaluation, therefore, served not only as a way to determine the overall quality and value of 3x3 DQA, but also as a way to incorporate expert knowledge and feedback into the further development of the framework and guideline.

6.2.2.1 Evaluation Design

The evaluation instrument was designed so as to mirror the levels of the guideline: framework and constructs, operationalizations, recommendations, and the scope identification questions. The content experts were asked to consider the clarity, validity, comprehensiveness, feasibility, and usefulness of the different sections. Each question had a forced binary response option, both to make the questionnaire quicker to fill out and to make the results easier to analyze.(13)

At the highest level is the framework, which includes the three constructs of EHR data quality. Respondents were asked to assess the clarity and comprehensiveness of the overall framework, as well as the clarity and validity of each of the constructs. Next are the nine operationalizations of the data quality constructs, wherein each construct is projected onto the three data dimensions. The evaluation includes questions about the clarity and validity of each operationalization, as well as the comprehensiveness of the operationalizations in capturing each construct. Each operationalization is then detailed in actionable terms in the recommendations for measurement and reporting. Respondents were asked to assess the clarity and feasibility of each of the recommendations. Finally, two general questions were included for the scope identification questions, asking the content experts to evaluate the questions in terms of clarity and usefulness. Each questionnaire item or set of items also included a space for optional free-form response. The complete evaluation instrument is included in Appendix C.

6.2.2.2 Expert Identification and Recruitment

Content experts in the area of EHR data quality assessment were identified in one of two ways. Either they were collaborators on the PCORI Methods project entitled *Building PCOR Value and Integrity with Data Quality and Transparency Standards*, led by Michael Kahn of the University of Colorado Denver, or they were required to have published at least two peer-reviewed articles on EHR data quality assessment. Ten content experts were identified based

upon these criteria, and invited by email to participate in the evaluation of 3x3 DQA. No compensation was offered.

6.3 Results

6.3.1 Initial Guideline Document

Version 1.0 of 3x3 DQA can be viewed in its entirety in Appendix D. The guideline document is 25 pages long, and includes an introduction, a user guide, scope identification questions, the 3x3 DQA framework, data quality assessment and reporting recommendations, and relevant references.

6.3.1.1 The 3x3 DQA Framework

The 3x3 DQA framework, which is at the heart of the guideline document, is shown in Figure 2. Based on the data quality categories identified in Aims I (Chapter 2) and II (Chapter 3), three core constructs of data quality were selected as being essential for the secondary use of EHR data. To be fit for use, EHR data must be complete, correct, and current. These constructs are defined in Table 2:

Table 6.2 Core data quality constructs and definitions from 3x3 DQA.

Construct	*Definition*
Complete	Data are sufficient in quantity for the task at hand.
Correct	Data are free from error.
Current	Data were recorded at the desired relative or absolute time(s).

These three constructs are then operationalized across the three potential dimensions of EHR data: patients, variables, and time. These data dimensions correspond to the standard dimensions of data flat files and cubes. Patients are equivalent to rows, variables to fields, and time to the longitudinal dimension that is present in some datasets. These data dimensions were originally identified in Aim III (Chapter 4), and were used as the basis of three of the definitions of completeness described therein. It is theoretically possible that a fourth data dimension could

be added to represent the possibility of including data from multiple institutions. The inclusions of multiple data sources would dictate the operationalization of the three data quality constructs across institutions. Comparisons of variables between institutions, for example, could indicate problems with correctness. 3x3 DQA could easily accommodate this fourth data dimension, but because the majority of the foundational research in Aim II (Chapter 3) and Aim III (Chapter 4) were conducted at a single institution, it was determined that it was more appropriate to limit the scope of the framework and guideline to a single institution as well.

	A: COMPLETE	B: CORRECT	C: CURRENT
1: PATIENTS	**1A** Are there sufficient data points for each patient?	**1B** Is the distribution of values across patients plausible?	**1C** Were all data recorded during the timeframe of interest?
2: VARIABLES	**2A** Are there sufficient data points for each variable?	**2B** Is there concordance between variables?	**2C** Were variables recorded in the desired order?
3: TIME	**3A** Are there sufficient data points for each time?	**3B** Is the progression of data over time plausible?	**3C** Were data recorded with the desired regularity over time?

Figure 6.1 3x3 DQA framework. Data quality constructs are at the top, data dimensions along the side, and cells contain corresponding operationalizations.

Each cell corresponds to the operationalization of one of the three data quality constructs across one of the three data dimensions. Correctness, which is especially difficult to measure without a reference standard, is operationalized through concordance (agreement between values) and plausibility, which are often viewed as independent categories of data quality in the informatics literature, but are generally proxies for correctness.(4)

6.3.1.2 Recommendations for Assessment and Reporting

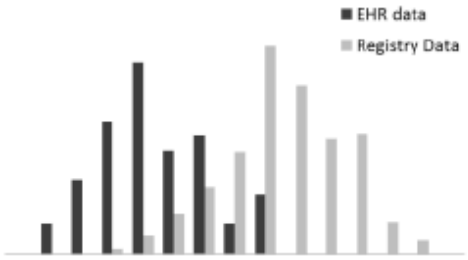
1B	The distribution of values across patients is plausible (approach II)
Measure	For variables that can be conceptualized as distributions of values across patients, identify reliable external sources of comparable data (e.g., national registry data). Select the appropriate statistical test given the variable type and distribution type. Also consider constructing graphical representations of the distributions to compare visually. Perform statistical and visual comparisons of your data and the external data source.
Report	Report the external data source, type of test(s) performed, and whether or not the results indicate that the distributions differ. How do the distributions differ (e.g. mean, skewness, etc.)? Are there possibly confounding differences between your population and the population for the external data source?
Example	Given a dataset of sixty patients and a continuous variable with a normal distribution and "research-quality" registry data for the same variable, a user can construct two normalized distributions, as shown below. A two-tailed t-test for this mock data results in $p < 0.01$. This specific example may indicate differences in population, or suggest systematic measurement error. 
References	Faulconer and de Lusignan 2004 Johnson, Mant et al. 1991 Iyen-Omofoman, Hubbard et al. 2011

Figure 6.2 One of two recommendations for measuring and reporting if a dataset is correct across the patient dimension, which is operationalized as "Is the distribution of values across patients plausible?"

Each operationalization within the framework has an associated recommendation for assessment and reporting, except for the assessment of correctness across patients (cell 1B),

which has two recommendations. Notably, the recommendations do not address implementation, which is highly dependent upon the data of interest, and often require the user to call upon relevant domain knowledge. The recommendations also do not include "cut-offs," or points at which a dataset would be deemed to be of sufficient quality. Such a determination must be made by the user depending upon their understand of the data being used, the clinical phenomena being examined, and the methods of analysis utilized.

Each recommendation has four sections: measure, report, example, and references (where available). "Measure" describes how to actually assess the data quality construct as described in the relevant operationalization, "Report" indicates what needs to be included when describing data quality findings. This includes not only the numerical results of the measurement process, but also any rules that were put in place for purposes of assessment. "Example" presents a simple demonstration of how the measure would be implemented, and what would be reported, sometimes with relevant visualizations. Finally, "references" supplies source and example publications for the specific measures described in the recommendation, allowing users to read further if they so choose. See Figure 3 for an example of a recommendation page from the guideline document.

6.3.1.3 Scope Identification Questions

There exist 512 combinations of the nine data quality assessment operationalizations. By applying a number of logical constraints to these combinations (e.g. all use cases will require correct and complete data, but not necessarily current data), we identified 26 realistically possible combinations. The possible combinations depend upon which data dimensions (patients, variables, time) were included in the target dataset, and which aspects of currency were important for the use case. All 26 combinations are summarized in Table 3.

Table 6.3 Possible combinations of 3x3 DQA operationalizations.

Data dimensions to assess	*Operationalizations of currency*
Multiple patients, variables, and time points.	Check timeframe, relative order, and frequency of points. Check timeframe and relative order, but not frequency of points. Check timeframe and frequency of points, but not relative order. Check timeframe, but not frequency of points or the relative order. Check frequency of points and relative order, but not timeframe. Check relative order, but not timeframe or frequency of points. Check frequency of points, but not timeframe or relative order. Don't check timeframe, relative order, or frequency of points.
Multiple patients and variables, but only one time point.	Check timeframe and relative order, but not frequency of points. Check timeframe, but not relative order or frequency of points. Check relative order, but not timeframe or frequency of points. Don't check timeframe, relative order, or frequency of points.
Multiple patients and time points, but only one variable.	Check timeframe and frequency of points, but not relative order. Check timeframe, but not relative order or frequency of points. Check frequency of points, but not timeframe or relative order. Don't check timeframe, relative order or frequency of points.
Multiple variables and time points, but only one patient.	Check relative order and frequency of points, but not timeframe. Check relative order, but not timeframe or frequency of points. Check frequency of points, but not timeframe or relative order. Don't check timeframe, relative order, or frequency of points.
Multiple patients, but only one variable and one time point.	Check timeframe, but not relative order, or frequency of points. Don't check timeframe, relative order, or frequency of points.
Multiple variables, but only one patient and one time point.	Don't check timeframe, relative order, or frequency of points. Check relative order, but not timeframe or frequency of points.
Multiple time points, but only one patient and one variable.	Check frequency of points, but not timeframe or relative order. Don't check timeframe, relative order, or frequency of points.

The scope identification questions are organized in three phases. The first includes three questions that ask users to consider whether they have more than one variable, patient, and time point in their study, in order to determine which data dimensions (the framework rows) are relevant to the intended study (see Figure 4). Phase two includes a total of eight questions, though users will only be asked to respond to one of them. These questions check to make sure the user has identified the correct data dimensions. The third and final phase includes three questions, and again the user will determine which one to answer based upon their responses to the phase one questions. These questions help the user determine which operationalizations of currency are

relevant to the intended study. Upon completing the scope identification questions, the user should know which operationalizations and, by extension, which recommendations are necessary for the intended study.

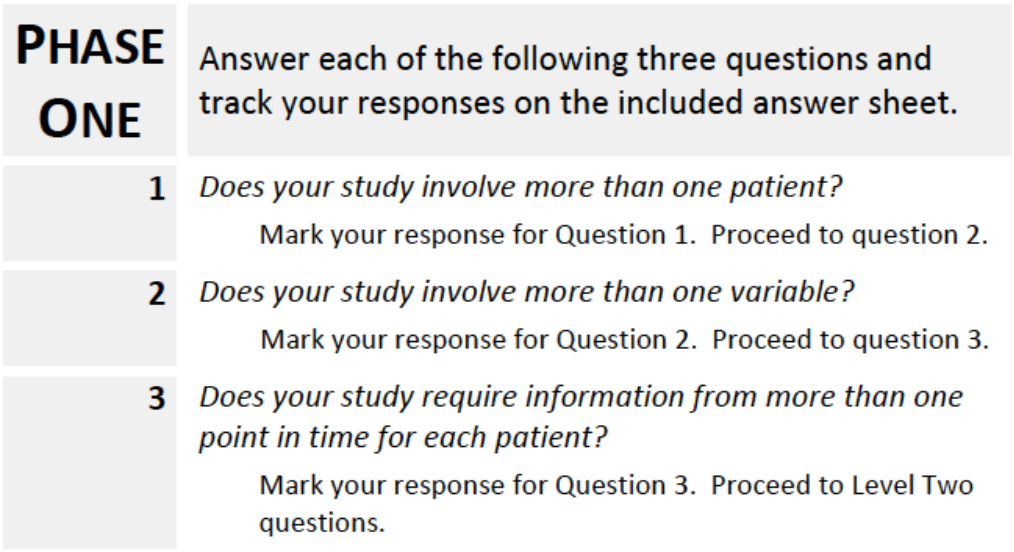


Figure 6.3 The phase one scope identification questions. These questions help the user determine which data dimensions are relevant to their study, which in turn dictates which data quality operationalizations and recommendations are necessary.

6.3.2 Quantitative Evaluation Results

Of the ten content experts who were asked to participate in the evaluation, nine expressed interest in the project, and five completed and returned the evaluations before May 28, 2014. The quantitative and qualitative results for the five responses are summarized below. Only two participants responded to the questions regarding the questions to identify scope, so those questions and their responses were not included in the analysis.

Of the five participants who completed the evaluation, two thought the framework was valid, and two thought it was clear. The three constructs of data quality also received low marks, which are summarized in Figure 5. The responses to the operationalizations, shown in Figure 6, were more mixed. The recommendations received the most positive responses overall, as shown in Figure 7.

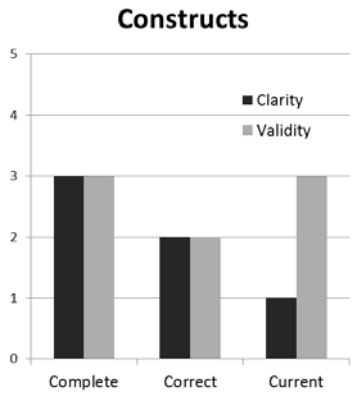


Figure 6.4 Number of content experts who thought the three data quality constructs were clear and valid.

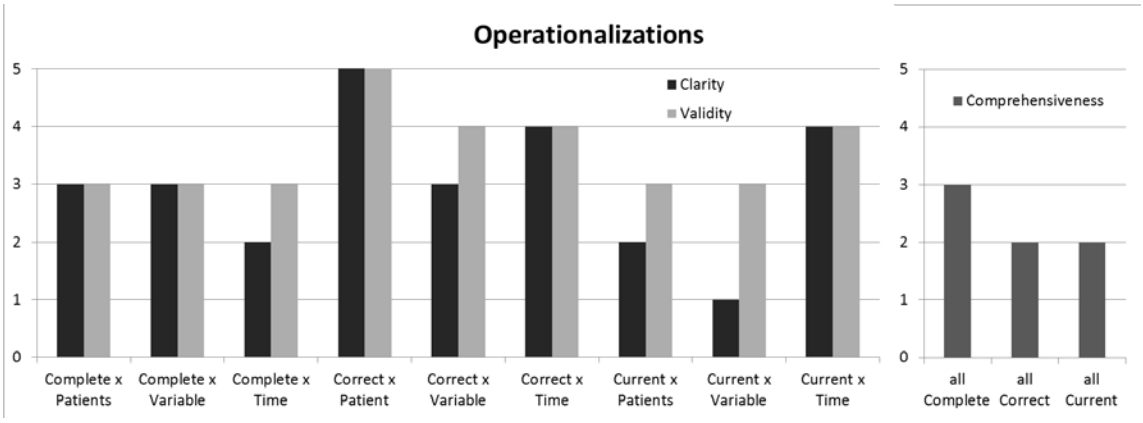


Figure 6.5 Number of content experts who thought the nine operationalizations were clear and valid, as well as the number who thought that the three operationalizations for each construct were comprehensive.

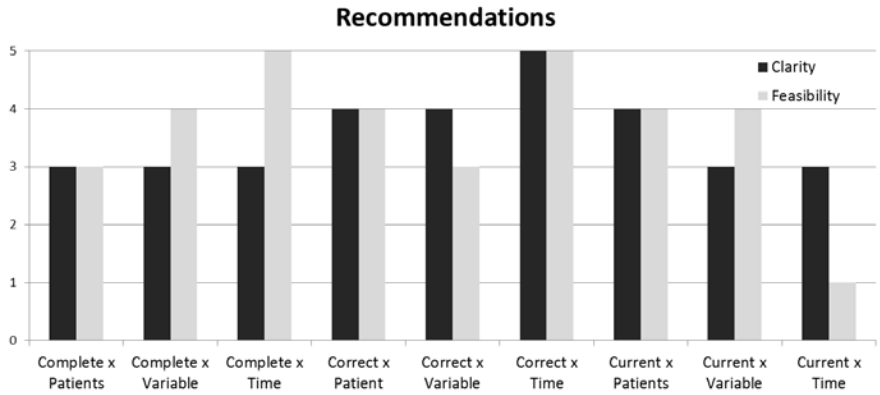


Figure 6.6 Number of content experts who thought the nine recommendations were clear and feasible.

6.3.3 Qualitative Evaluation Results

The overall response to 3x3 DQA was positive. One content expert called the guideline "a coherent and thoughtful piece that should support research activities in an important way." Another said, "This is really great and very nicely presented." All five respondents, however, had significant critiques and suggestions.

6.3.3.1 Framework Responses

Based on their comments, none of the respondents had problems with the clarity of the framework structure, which features the quality constructs projected across the data dimensions. Four out of the five, however, had issues with at least some of the wording used in the operationalizations, which are featured in the framework cells. One respondent also felt that there needed to be more clarity regarding when in the research process the framework would be applied. In terms of comprehensiveness, three respondents had suggestions for features of data quality they felt should be included: ensuring proper mapping and matching between different data elements (patients, variables, etc.), consistency of representation, inclusion of the right patients (i.e. cohort selection), and methods for identifying and removing duplicate records.

Construct Responses

Of the three constructs, the definition of complete data had the highest overall ratings. One evaluator, however, felt that the inclusion of the phrase "task at hand" in the definition made the construct both unclear and potentially invalid. Another felt that the construct was both unclear and invalid because it was likely that a research scenario would be designed around a dataset, not that the dataset would be evaluated for a given research scenario: "Typically, a researcher will evaluate data quality/availability and develop a research design appropriate for the data. One question is whether the data are sufficient/appropriate given the research design. Without understanding how the data will be used it will be difficult to understand the questions or answers."

Responses to the complete data definition were mixed, with three complaints being about the phrase “free from error.” One content expert thought the definition was valid, but unclear because the meaning of “error” was unspecified, while the other two felt the phrase made the definition both unclear and invalid. All three emphasized that the source of error must be considered (e.g. incorrectly measure, incorrectly recorded, incorrectly retrieved, etc.). In contrast, another respondent stated, “This is simple—‘free from error’—which is good.”

The definition of current data had the lowest rating for clarity of the three constructs. Three of the respondents requested clarification regarding whether the time of interest was when the data point was measured or when it was recorded, suggesting that the latter is not informative, and also that it is not always available. One respondent wrote, “…the time of recording [a vaccination] may not be meaningful and is often not transferred as metadata with [the] dataset, but the date I put down as the receipt date is the time that is important for data users.” A different content expert also pointed out that the term “current” usually means “up to date,” and that “timely” was perhaps a better word for this construct.

6.3.3.2 Operationalization Responses

Two of the respondents had significant concerns about the clarity and validity of the operationalizations of complete data. One disliked the use of the term “sufficiency,” and seemed to want to move towards a less task-dependent definition of complete data and have some type of denominator for patient counts. Another felt the concept of time as it was used to characterize a denominator was unclear and unhelpful. There was also a concern that the three operationalizations are all ways of measuring the same thing, and do not give different information: “To me, these are all the same question. I want to know whether my cohort has the right data available for a study, during the study period. It is one question to me, not three.”

The operationalizations of complete data were the most popular, though two respondents felt that the use of the word “concordance” was unclear. One content expert also disliked that all

three operationalizations relied upon external knowledge, and also wished that the data quality assessment considered the data lifecycle (i.e. measurement to recording to extraction). Another respondent, however, highlighted these as being the strongest part of the guideline overall.

As with the overall construct of current data, the individual operationalizations of current data had the lowest overall ratings of clarity. Three respondents disliked the phrase "desired order," and were not sure what it meant. Two disliked the emphasis on user decision-making (i.e. "timeframe of interest," "desired order," and "desired regularity"). One of the two suggested that if we wanted to emphasize task-dependence, it might be better to use consistent phrasing across all operationalizes (e.g. "sufficient," as in the complete data operationalizations, instead of "desired"). One of the content experts suggested that we include a measurement of the time between when an event occurred and when it was recorded.

6.3.3.3 Recommendation Responses

One of the content experts did not write comments for the individual recommendations, but instead stated, "These recommendations…are where this comes to life. I'd be hesitant to make it too formal…but there are key recommendations that everyone should do before using a data source."

Two of the content experts had significant concerns regarding the clarity and feasibility of the recommendations for measuring and reporting on complete data. One would have preferred that the recommendations be clearer regarding how judgments would be made regarding if a dataset met the necessary threshold for quality. "[The recommendations] are about availability. The recommendations say nothing about the judgment that is required to turn availability into sufficiency…" Another respondent said that the term "variable" had to be clearly defined in order to be applied, and also said that these measures might be difficult to execute for large datasets.

For the most part, the content experts were satisfied with the feasibility and clarity of the recommendations for measuring and reporting data correctness. There were a few small

suggestions, including being more explicit about how many variables had to be assessed and defining concordance. The majority of the comments came from one respondent, who had significant suggestions for all three recommendations. For the first approach for assessing correctness across patients, this content expert suggested distinguishing between plausible values within an individual record versus across all patient records, and for the second approach pointed out that the identification and use of an external benchmark is not always feasible. This respondent was also concerned about the burden of having users come up with their own rules to assess concordance between variables and the plausibility of sequences of data. Lastly, this respondent also pointed out that it could also be useful to assess trends of data at the population level, not just the patient level.

One major concern regarding the feasibility of the recommendations for current data was that the assessments relied upon the availability of certain log data or metadata (e.g. time when the data were recorded), which may not always be included in a dataset. Two of the content experts disliked the phrase “within range” in the recommendation for assessing currency across patients, and felt it was misleading. Three of the respondents disliked the use of formal calculations described for assessing currency across time, or at least felt it was out of line with the rest of the recommendations.

6.3.3.4 Questions to Identify Scope Responses

Only two of the content experts included comments on the scope questions. One thought the questions were a good start, but that their clarity would be improved by emphasizing the importance of context (i.e. the intended research). Another felt that many of the questions were unhelpful, since most research would have the same requirements, and suggested that some illustrative examples might help.

Table 6.4 Triangulation of experts' quantitative and qualitative feedback.

Framework		Constructs		Operationalizations			Recommend.		
Clear	*Comp*	*Clear (3)*	*Valid (3)*	*Clear (9)*	*Valid (9)*	*Comp. (3)*	*Clear (9)*	*Feas. (9)*	**Sample Expert Statements**
	yes	0	0	4	3	1	8	8	Typically, a researcher will evaluate data quality/availability and develop a research design appropriate for the data. Recording a diagnosis code for "diabetes" in an EHR during a visit in which the clinician orders a test for diabetes is not necessarily an error, it might be a local policy that all diabetes tests ordered get coded with that diagnosis. I want to know whether my cohort has the right data available for a study, during the study period. It is one question for me, not three. These recommendations…are where this comes to life.
		0	0	1	6	0	5	6	I think missing from the framework is actually the frame—when in the research process are we supposed to use this? It seems to be aimed at the analysis of a data set—after the data-collection process has been specified. [T]he concept missing for me is my data-quality workflow, as a quality assessor or researcher. You saw my earlier comment before, the "pathway" of data from physical event to recording in the dataset. These three operationalizations don't cover all of them, so I presume you are making a choice based on some sort of tradeoff, having to do with ease of checking.
yes		3	2	6	6	0	7	9	Again it is contextual—fitness for purpose definition. But overall the logic of self-assessment and self-determination of what "sufficient" is makes sense. Progression on data over time reflects clinical course and will vary depending on a number of diagnostic, management and prognostic factors. So need constraints in framing the research question(s). I think the realist approach should be emphasized, i.e. the importance of context. The issue of "actors" is another important scope question as EHR-based research can be used for research about the care provider and interventions as well as impact on patients.
yes		0	2	9	9	3	4	3	I don't see anything to address the quality issue of, "Are the right patients included in the data?" Perhaps this is more of a research question…but it seems to cross into the data quality boundary when someone attempts to use the data for something that's not fit for purpose. [I]t feels as if [completeness] depends on the 'task at hand.' If the goal is to estimate an effect, then completeness requires that the estimate can be generated without bias due to confounding. [Using an] external reference is a good idea, but practically is quite difficult, both in terms of logistics and methodologically ensuring that the external reference should be comparable to the source population.
	yes	2	3	7	8	3	8	7	Each construct seems like it should be followed by the term "for the task at hand." [The completeness across patients recommendation] may be difficult for larger datasets, composite variables, and deciding when to do this…where does this get represented? [For the current across patients recommendation,] this is clear—I'm not sure how feasible it is. [For the current across time recommendation,] without metadata for recording data I'm not sure how feasible this is.

6.3.4 Data Triangulation

As a complement to the separate qualitative and quantitative results for each section of 3x3 DQA, we also looked at the responses for each expert. Table 4 shows a summary of each expert's quantitative scores and a selection of their key qualitative comments. The purpose of this table is to demonstrate how each expert's feedback hangs together throughout the evaluation.

6.4 Discussion

While the overall response to 3x3 DQA was positive, the quantitative and qualitative results both indicated that there is significant work to be done to improve validity, clarity, comprehensiveness, and feasibility. The content experts, for example, largely disregarded the scope identification questions. This result suggests that the scope identification questions either need to be significantly improved, or not included at all. The recommendations, in contrast, had relatively high levels of approval, and the problems mentioned by the content experts should be relatively easy to fix.

Each expert had at least one overarching concern or guiding principle in their evaluation (see Table 4). The first expert felt strongly that the context in which the data was originally collected and maintained was vitally important to understand the meaning of the data, and also disliked the more abstract parts of the guideline, particularly the construct definitions and, to a lesser extent, the operationalizations. This expert, however, liked the concrete suggestions contained within the recommendations. The second expert believed that in order to understand data quality it was necessary to understand the underlying processes and lifecycle of the data, and felt that a weakness of the entire guideline was that not explicitly stated *when* in the data lifecycle 3x3 DQA was intended to be used. The third expert comes from a strong background in ontological science, and therefore felt that the guideline needed to be more ontologically coherent. This expert also felt that the importance of context and the concept of fitness-for-use needed to feature more centrally in the guideline. The fourth expert also emphasized the

importance of context in the meaning of the three core constructs of the guideline, and believed that the identifying a research question was necessary to understand how any of the constructs would really be operationalized and implemented. The fourth reviewer, along with the fifth reviewer, was also very concerned about the feasibility of some of the recommendations, and believed that some of them would be difficult to carry out in practice, particularly with large datasets or unique patient populations.

There were a few major lessons that can be applied to the entirety of 3x3 DQA. One is the importance of phrasing and terminology. The three data quality construct definitions (and the word "current") all received major criticism. The operationalizations also received low ratings on clarity, especially for current data. The way in which data quality is defined and operationalized is clearly both important and complex. Part of this complexity stems from the lack of consistency in thought on how best to describe and represent data quality. The disagreement over the definition of correct data is a good example: three respondents disliked the term "free from error," while another singled it out as the best definition. Going forward, there is significantly more work to be done on the definitions and operationalizations of the constructs.

Another major lesson is the major goals and guiding principles of the development of 3x3 DQA need to be more clearly laid out in the actual document. Specifically, we wanted the guideline to embrace the task-dependent nature of EHR data quality and data quality assessment, as well as the central role of the target users (clinical researchers) in determining how to apply the recommendations and whether or not data were of good enough quality for their purposes. The comments, however, suggested that these two principles were sometimes unclear, with one content expert explicitly disliking the use of terms like "sufficient" or "desired." Therefore, 3x3 DQA may need to have a more clear "mission statement" that highlights the guiding principles and explains how they were derived.

One of the limitations of this evaluation was the choice to use a forced binary choice for the survey items, rather than a Likert scale. Forced binary items are ideal for when one wants to

determine if an individual completely agrees with a survey item. Therefore, from an iterative design perspective, forced binary items were useful. If one of the experts was at all unsatisfied with some aspect of 3x3 DQA, we wanted to know this, because it meant that further work was required. From an analysis standpoint, however, a forced binary item results in limited information. With a Likert scale a tester can convey a range of satisfaction or dissatisfaction, which would have allowed for more in-depth analysis of the responses to 3x3 DQA. As it is, there is not sufficient data to determine if some aspect of the guideline is "close to comprehensive" or "nearly valid." Rather, if a respondent interpreted the binary items strictly to mean "comprehensive" versus "not comprehensive," we can only determine if they felt some work was still needed, not how much work was still needed.

A final point is that there is always a trade-off between clarity and complexity. One of the goals of 3x3 DQA was to impose a coherent conceptual framework onto the lengthy and difficult process of EHR data quality assessment. In improving the clarity and usability, however, there was a loss of scope and complexity. There are only three data quality constructs included in the guideline, as opposed to the fifteen defined by Wang and Strong or the five identified in the literature review and the seven identified in the interviews described in Chapters 2 and 3 of this dissertation. The decision to only assess complete, correct, and current data was made in part because these were identified as the core constructs of EHR data quality, but also because including all possible constructs could easily become unmanageable. Likewise, the decision to project these three constructs across the three data dimensions was intended to operationalize EHR data quality in a logical and actionable fashion, but it is also a design choice. It is important to determine if the losses in scope and complexity are outweighed by the gains in coherence and clarity, as well as if there are ways to improve the balance.

6.4.1 Future Work

The next step in the iterative development of 3x3 DQA will be another round of design and improvement, based upon the feedback of the five content experts. After that, it is vital that the guideline undergo scenario testing and pilot testing. While experts are important in establishing the underlying conceptual basis and validity of the guideline, user experiences and feedback are necessary for further development. Finally, the actual impact of 3x3 DQA will need to be evaluated. Does using 3x3 DQA increase the quality of research conducted using EHR data? Does it improve the awareness and knowledge of researchers for appropriately selecting and applying pertinent data quality measures? Does it improve the transparency of research and also the interpretability of the research results? And most importantly, does it improve the validity of research conducted with EHR data? This last question is especially hard to answer, since it requires the establishment of a difficult baseline: what is the validity of research conducted with EHR data without data quality assessment, or with ad hoc approaches to EHR data quality assessment?

6.5 Conclusions

This chapter describes the first two phases in the iterative design of a guideline for EHR data quality assessment: the initial design and an expert evaluation. 3x3 DQA is a standalone document targeted at clinical researchers engaged in the secondary use of EHR data. It is meant to embrace a fitness-for-use approach to data quality that is flexible enough to accommodate different study designs and data requirements. Rather than relying upon the availability of a reference standard, the 3x3 DQA guides users in utilizing external sources of medical information and knowledge to evaluate data quality. The guideline results from qualitative, data-driven, and literature-based investigation to understand and assess EHR data quality issues.

Based on an evaluation of the validity, clarity, comprehensiveness, and feasibility of 3x3 DQA by EHR data quality content experts, the guideline appears to be a promising start, though it

requires significant continual development. Specifically, the constructs and operationalizations of EHR data quality need to be improved, and primary goals and principles of the guideline to be explicitly stated and explained to intended users. Automated execution of the guideline should also be explored to reduce cognitive overhead for potential users to interpret the complex guideline logic. Further iterations of 3x3 DQA will require extensive testing and evaluation to demonstrate real world usefulness and impact.

6.6 References

1. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67(5):503-27.

2. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4(5):342-55.

3. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003;326(7398):1070.

4. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51.

5. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005;365(9453):82-93.

6. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet. 1993;342(8883):1317-22.

7. Juran JM, Gryna FM. Juran's quality control handbook. 4th ed. New York: McGraw-Hill; 1988.

8. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform. 2013;46(5):830-6.

9. The SAGE Project Home Page [June 6, 2012]. Available from: http://sage.wherever.org/model/components.html.

10. Green L, Kreuter M. Health promotion planning: an educational and ecological approach. 4th ed. Boston, MA: McGraw Hill; 2005.

11. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. American journal of public health. 1999;89(9):1322-7.

12. Grant J, Davis L. Selection and use of content experts for instrument development. Research in Nursing & Health. 1997;20(3):269-74.

13. Dolnicar S, Grun B, Leisch F. Quick, simple and reliable: forced binary survey questions. International Journal of Market Research. 2011;53(2):231-52.

Chapter 7: Conclusions and Future Work

7.1 Introduction

This dissertation investigates the problems and priorities of electronic health record (EHR) data quality and data quality assessment within the context of secondary use. The first study (Chapter 2) was a literature review that established the need for a systematic approach to EHR data quality assessment. The second (Chapter 3) was a qualitative study exploring clinical researchers' perceptions of EHR data and attitudes towards their reuse for research, and demonstrated the prominent need for EHR data quality assessment methods, as well as the desire for clinician involvement in the validation and reuse process. The third study (Chapter 4) featured a data-driven approach to EHR data completeness, which demonstrated that EHR data quality is a task-dependent phenomenon. The fourth study (Chapter 5), inspired by the third, explored the connection between patient health and EHR data completeness, and established that there is a risk of sampling bias in the reuse of EHR data when imposing data sufficiency requirements. The last study (Chapter 6) describes the development and expert-based evaluation of 3x3 DQA, a dynamic, evidence-based guideline to enable EHR data quality assessment and reporting for

retrospective research that incorporates the major lessons and principles demonstrated by the previous chapters of this dissertation.

7.2 Contributions

This dissertation makes two primary contributions to the improvement of the secondary use of EHR data. Aims I, II, III, and IV led to the discovery and development of guiding principles and best practices in the reuse of EHR data and data quality assessment. These best practices, in turn, were used to guide the design of 3x3 DQA, the other primary contribution of this dissertation.

7.2.1 Best Practices for EHR Data Quality Assessment

From Aim I, the literature review, we learned that current approaches to EHR data quality assessment are ad hoc and overly reliant upon arbitrarily chosen reference standards, which are not always available. Aim I also resulted in five categories of data quality and seven approaches to data quality assessment that may be applied to EHR data quality reuse or, in the future, other sources electronic health data (e.g. registry data or social media data).

In Aim II, the interviews, we learned that clinical researchers are aware of EHR data quality problems, but intend to engage in the secondary use of EHR data despite these problems, indicating that there is an immediate need for data quality assessment methodology is. The interviews also indicated that to be accepted by clinical researchers, an EHR data quality assessment tool should engage the user in the evaluation and decision making process. The interviews also led to the identification of seven possible categories of EHR data quality.

Aim III, which was the data-driven exploration of completeness, proved that EHR data quality is a task-dependent phenomenon, which means that any approach to EHR data quality assessment should abide by the principle of fitness-for-use. This aim also demonstrated that EHR data quality can be operationalized across the data dimensions of variables, patients, and time.

In the course of Aim IV, which looked at the relationship between patient record completeness and patient health status, we proved that by naively selecting patient records with sufficient data for a given study, researchers may inadvertently bias their cohort towards sicker patients, thereby limiting the external generalizability of their research.

7.2.2 EHR Data Quality Assessment Guideline

The best practices described above were used as guiding principles in the development of 3x3 DQA, which is the primary contribution of this dissertation. 3x3 DQA is a dynamic, evidence-based guideline meant to enable EHR data quality assessment and reporting for retrospective research. Though not yet ready for dissemination (further iterative development, evaluation, and testing are required), the guideline is a promising approach for improving the efficiency, validity, and transparency of research relying on the secondary use of EHR data.

7.3 Future Work

7.3.1 Designing and Evaluating 3x3 DQA

Although feedback from the content experts who participated in the initial evaluation of 3x3 DQA was largely positive, it is clear that further improvements and testing are needed. The design process behind the guideline is iterative. Following redesigns based upon the evaluation described in Chapter 6, the guideline will need to undergo further testing, likely followed by further redesign.

Future testing efforts will be needed to establish the validity and reliability of 3x3 DQA, possibly using techniques for psychometric instrument development.(1) Ideally, evaluation will also include scenario-based testing with representative users, and eventually pilot testing with real data and real research. It is important that the actual real world impact of the guideline be assessed. Does it simply improve data quality, or does it actually improve the validity of research?

It is also hoped that further iterations of 3x3 DQA will include a partially automated data quality assessment "toolkit," which will eliminate some of the complex implementation processes that currently interfere with the guideline's comprehensibility and use. Some degree of user guidance and training will always be required—by necessity and by design—but some of the burden could and should be decreased, improving efficiency and consistency.

7.3.2 Putting 3x3 DQA into Practice

Following another round of iterative design based upon the expert evaluation described in Chapter 6, the next phase of testing will be to use 3x3 DQA with actual EHR data. The goal of this work will be to determine the usability and usefulness of 3x3 DQA when used by target users in real-world settings with the intended sources of data. This implementation should reveal the usefulness and usability of the guideline, as well as its effectiveness in revealing data quality problems and the potential of the guideline to alleviate the burden of ad hoc approaches to data quality assessment.

A number of opportunities to implement 3x3 DQA in applicable research situations exist. One such opportunity exists at the University of Colorado, Denver, where Dr. Lisa Schilling has proposed piloting of 3x3 DQA on an EHR-derived dataset that has been compiled for a planned study. Another opportunity exists at the Oregon Health and Science University, where I will be working beginning in the fall of 2014. 3x3 DQA will likely serve as the basis upon which much of my future research efforts will be based, including an exploration of the impact of data quality upon the findings of research conducted using EHR-derived data.

Certain features of 3x3 DQA will also be incorporated into other models of EHR data quality assessment. Dr. Meredith Zozus at Duke's Translational Medicine Institute has adapted the operationalizations of completeness from 3x3 DQA as part of her work as a member of the NIH Research Collaboratory on pragmatic clinical trials and the use of healthcare data.{Zozus, 2014 #466} Certain features of 3x3 DQA are also being incorporated into the work of the PCORI

methods group led by Dr. Michael Kahn at University of Colorado, Denver. The parts of 3x3 DQA that focus on EHR data completeness have been especially well-received, and will probably be included within the PCORI unified model. In both cases, the elements derived from 3x3 DQA will eventually be used with EHR-derived datasets, providing an opportunity where features of the guideline can be piloted in real-world situations. Such work will be ideal for evaluating patient-centered datasets like those being compiled as part of the PCORnet project.

7.3.3 Exploring the Data Lifecycle

All of the work described in this dissertation either uses or acts upon data after the point of extraction from the Columbia University Medical Center's clinical data warehouse. There exists significant interest, however, in the assessment and improvement of EHR data quality earlier in the data lifecycle. Assessing data quality after extraction has the potential to improve the validity of research conducted using the data. Assessing the data within the clinical system can improve patient care, clinical workflow, and quality of care.

The EHR data lifecycle is constantly and iteratively evolving. Each time a patient has contact with the healthcare system their record may receive additional notes, medication actions, laboratory results, etc. The patient record is an incredibly complex, living document, and the processes by which data are entered already place extensive time requirements upon clinicians. Therefore, efforts to assess or improve data quality prior to extraction must be developed and implemented in such a way as to prevent additional burden. There may be specific points in the EHR data lifecycle where data quality is already a focus, though perhaps not overtly. Medication reconciliation is an example where this already occurs to some extent.(2, 3) Although the purpose of medication reconciliation is to improve patient care by preventing medication errors, the process also leads to improved data quality, and can be used to assess data quality. Finding other such points in the data lifecycle may be a promising approach to assessing and improving data quality.

7.3.4 Patient Engagement

Another possible approach to EHR data quality assessment and improvement is through increasing patient engagement. The growing presence of patient communities like PatientsLikeMe and individuals like Dave deBronkart (e-Patient Dave) indicate that there are significant sub-populations of patients who are eager to be participants in their own healthcare and the clinical research process. Improving patient access to and awareness of their electronic clinical data may be one untapped opportunity to encourage patients to assess and improve EHR data quality. This is most commonly done through patient portals or personal health record (PHR) systems,(4) which can be accessed by patients at home. More recently, however, there have been efforts to engage patients in the inpatient setting,(5) where they may be more likely to review their own data.

It may also be possible to triangulate data from the EHR and other health information technology (e.g. pharmacy data), as well as other entirely disparate sources of health data (e.g. social media) and demographic data (e.g. census data) in order to improve overall health-related data quality.(6)

7.4 Conclusions

This dissertation and the work described herein represent a promising step towards the development of best practices for and systematic approaches to the assessment of EHR data quality within the context of secondary use. This is an incredibly complex area, and there is still a great deal of work to be done. The promise of EHR data reuse, however, as well as the eagerness of researchers to use the data, elucidates the need to make this work an immediate and ongoing priority in order to increase the efficiency and validity of medical research.

7.5 References

1. Norbeck JS. What constitutes a publishable report of instrument development? Nursing research. 1985;34(6):380-2.

2. Pippins JR, Gandhi TK, Hamann C, Ndumele CD, Labonville SA, Diedrichsen EK, et al. Classifying and predicting errors of inpatient medication reconciliation. Journal of general internal medicine. 2008;23(9):1414-22.

3. Pronovost P, Weast B, Schwarz M, Wyskiel RM, Prow D, Milanovich SN, et al. Medication reconciliation: a practical tool to reduce the risk of medication errors. Journal of critical care. 2003;18(4):201-5.

4. Hanauer DA, Preib R, Zheng K, Choi SW. Patient-initiated electronic health record amendment requests. J Am Med Inform Assoc. 2014.

5. Prey JE, Woollen J, Wilcox L, Sackeim AD, Hripcsak G, Bakken S, et al. Patient engagement in the inpatient setting: a systematic review. J Am Med Inform Assoc. 2013.

6. Weber GM, Mandl KD, Kohane IS. Finding the Missing Link for Big Biomedical Data. JAMA : the journal of the American Medical Association. 2014.

Information Sheet and Consent Form for Researcher Interviews

Research Purpose
The purpose of this interview is to explore researcher perceptions of the quality and suitability of EHR data for clinical research, as well as how these attitudes translate to actual research intentions and practices. To accomplish this goal, the aims of this study are to: 1) develop a theoretical understanding of researcher beliefs and concerns regarding the quality of EHR data and their reuse for clinical research and 2) learn how researchers define data quality and its dimensions.

Information on Research
We invite you to participate in an approximately 60 minute interview during which we will ask you to share your opinions about Electronic Health Record data and their quality and suitability for reuse in clinical research. With your permission, we will audio record and transcribe the interview to summarize your perceptions. You may be contacted with follow-up questions in the future.

Potential Risks
A potential risk is loss of confidentiality regarding your participation in the interview. However, the audio recording of the interview will be stripped of identifying information and stored in a locked office. The transcription of the interview will be also be de-identified, and will be stored in a password-protected computer in a secure area. Your opinions will not be linked with your name in reports of research findings.

Potential Benefits
There are no direct benefits to interview participants. We may share the study results with you at a later time, as well as invite you to our informatics workshop on "secondary use of clinical data for research."

Alternative Procedures
The alternative is to not to participate in the interview.

Compensation
You will receive $25 per hour in compensation for your participation in the interview.

Additional Costs
There are no costs to you for participating in the interview.

Additional Information
If you have any questions or concerns about the study, you may contact Nicole Weiskopf at 347-216-0520 or via email at nicole.weiskopf@dbmi.columbia.edu. If you have any questions about your rights as a research subject you may contact the Columbia University Medical Center Institutional Review Board at 212-305-5883.

Confidentiality
Although every effort will be made to protect the confidentiality of your interview data, absolute confidentiality cannot be guaranteed. By participating in the interview you grant permission for your interview data to be made available to:
- The investigator and study staff who may be evaluating the study;
- Columbia University;
- Authorized representatives of the Office of Human Research Protections ("OHRP"); and
- The Columbia University Institutional Review Board ("IRB") that independently reviews the study to assure adequate protection of research participants, as required by federal regulations.

The investigator, regulatory authorities, and IRB may keep the research records indefinitely. If the results of the study are published or presented at a medical or scientific meeting you will not be identified.

Voluntary Participation
Your participation in the interview is voluntary. You may decide not to participate in the interview. If you decide to participate, you are free to withdraw from the interview at any time. Your refusal to participate, or your early withdrawal, in the interview will not affect any benefits to which you are otherwise entitled from Columbia University.

_____	_____
Signature	Date
_____	_____
Signature of Person Obtaining Consent	Date

Interview Guide

I. Actual Use of EHR data for research
   a. *Why don't you start by briefly telling me about the type of clinical research you do?*
   b. *Have you ever used EHR data for research purposes?*
      i. Yes
         1. *Could you tell me how you've used EHR data in clinical research? E.g. what you were studying, the research design, etc.*
         2. *Can you think of some specific positives or negatives that arose from using EHR data?*
      ii. No
         1. *Why not?*

II. Perceptions of EHR DQ
   a. Defining data quality
      i. *What does the term data quality mean to you?*
      ii. *Specifically in the context of reusing EHR data for clinical research, what do you think are the most important aspects of data quality?*
      iii. *Are there any special considerations related to data quality in this situation?*
   b. Opinions of EHR DQ
      i. *Overall, what are your thoughts on the quality and suitability of EHR data for research?*
      ii. *Do you think EHR data quality differs by type of element? E.g. notes, labs, diagnoses, medications, etc.*

III. Intent to use EHR data for research
   a. *Are there types of research that you would be more likely to conduct using EHR data?*
   b. *Given the opportunity, would you use EHR data to for clinical research?*
      i. *Do you intend to conduct clinical research using EHR data?*

IV. Wrap-up
   a. *Is there anything else related to the topic of EHR data quality and clinical research that you'd like to discuss?*

Introduction

Thank you for agreeing to evaluate 3x3 DQA, a data quality assessment guideline. Please consider the context of secondary use of electronic health record (EHR) data in retrospective research when responding to the evaluation questions. Your responses will be used to guide further development of 3x3 DQA.

If you have any questions or comments, please contact Nicole Weiskopf at ngw2105@columbia.edu.

What to expect:

This evaluation is formatted to mirror to the framework and recommendations, which have four levels:

1. The framework

2. The three data quality constructs: complete, correct, and current

3. The three operationalizations of each construct, which correspond to the nine cells of the framework

4. The recommendations for each operationalization, which describe how to perform and report data quality measurements

The evaluation ends with two questions regarding the scope identification questions.

You will be asked to evaluate the clarity, validity, comprehensiveness, feasibility, and usefulness of the guideline components. ***Please check all statements with which you agree.***

Framework (page 11)

☐ **Clarity**: The framework is easy to interpret and understand.

If not, do you have any suggestions to improve the clarity of the framework?

☐ **Comprehensiveness:** The framework captures all the constructs of data quality that are relevant to EHR data quality assessment within the context of secondary use.

If not, what other constructs should be included?

Constructs (page 10)

Definition: "**Complete** Data are sufficient in quantity for the task at hand."

- ☐ ***Clarity***: The definition of complete data is easy to interpret and understand
- ☐ ***Validity***: The definition is representative of data completeness

Do you have any suggestions to improve the construct definition?

Definition: "**Correct** data are free from error."

- ☐ ***Clarity***: The definition of correct data is easy to interpret and understand
- ☐ ***Validity***: The definition is representative of data completeness

Do you have any suggestions to improve the construct definition?

Definition: "**Current** data were recorded at desired relative or absolute time(s)."

- ☐ ***Clarity***: The definition of current data is easy to interpret and understand
- ☐ ***Validity***: The definition is representative of data currency

Do you have any suggestions to improve the construct definition?

Operationalizations of Complete Data (Framework cells, page 11)

Clarity: Which of the following are easy to interpret and understand?

- ☐ "Are there sufficient data points for each patient?"
- ☐ "Are there sufficient data points for each variable?"
- ☐ "Are there sufficient data points for each time?"

Do you have any suggestions to improve the clarity of the above operationalizations?

Validity: Which of the following are representative of the construct of data completeness?

- ☐ "Are there sufficient data points for each patient?"
- ☐ "Are there sufficient data points for each variable?"
- ☐ "Are there sufficient data points for each time?"

Do you have any suggestions to improve the validity of the above operationalizations?

☐ **Comprehensiveness**: Taken together, the three operationalizations of complete data capture the entire construct of data completeness.

If not, what other operationalizations should be included?

Operationalizations of Correct Data (Framework cells, page 11)

Clarity: Which of the following are easy to interpret and understand?

- ☐ "Is the distribution of values across patients plausible?"
- ☐ "Is there concordance between variables?"
- ☐ "Is the progression of data over time plausible?"

Do you have any suggestions to improve the clarity of the above operationalizations?

Validity: Which of the following are representative of the construct of data correctness?

- ☐ "Is the distribution of values across patients plausible?"
- ☐ "Is there concordance between variables?"
- ☐ "Is the progression of data over time plausible?"

Do you have any suggestions to improve the validity of the above operationalizations?

☐ **Comprehensiveness**: Taken together, the three operationalizations of correct data capture the entire construct of data correctness.

If not, what other operationalizations should be included?

Operationalizations of Current Data (Framework cells, page 11)

Clarity: Which of the following are easy to interpret and understand?

- ☐ "Were all data recorded during the timeframe of interest?"
- ☐ "Were variables recorded in the desired order?"
- ☐ "Were data recorded with the desired regularity over time?"

Do you have any suggestions to improve the clarity of the above operationalizations?

Validity: Which of the following are representative of the construct of data currency?

- ☐ "Were all data recorded during the timeframe of interest?"
- ☐ "Were variables recorded in the desired order?"
- ☐ "Were data recorded with the desired regularity over time?"

Do you have any suggestions to improve the validity of the above operationalizations?

☐ **Comprehensiveness**: Taken together, the three operationalizations of current data capture the entire construct of data currency.

If not, what other operationalizations should be included?

Recommendations for Complete Data Assessment (pages 13-15)

Complete across **Patients**: "Are there sufficient data points for each patient?"

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Complete across **Variables**: "Are there sufficient data points for each variable?"

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Complete across **Time**: "Are there sufficient data points for each time?"

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Recommendations for Correct Data Assessment (pages 16-19)

Correct across **Patients**: “Is the distribution of values across patients plausible?”

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Correct across **Variables**: “Is there concordance between variables?”

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Correct across **Time**: “Is the progression of data over time plausible?”

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Recommendations for Current Data Assessment (pages 20-22)

Current across **Patients**: "Were all data recorded during the timeframe of interest?"

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Current across **Variables**: "Were variables recorded in the desired order?"

☐ ***Clarity***: The recommendation is easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Current across **Time**: "Were data recorded with the desired regularity over time?"

☐ ***Clarity***: The recommendation easy to interpret and understand.

☐ ***Feasibility***: The implementation and execution of the recommendation would be manageable.

Do you have any suggestions to improve the recommendation?

Questions to Identify Scope (pages 5-9)

☐ **Clarity**: The questions are easy to interpret, follow, and understand.

Do you have any suggestions to improve the clarity of the questions?

☐ **Usefulness:** The questions are useful in identifying the scope of a research task and the relevant data quality assessment operationalizations and recommendations.

Do you have any suggestions to improve the usefulness of the questions? Alternatively, do you think the questions should not be included in the guidelines?

Other Comments or Suggestions