Assessing the Quality of Electronic Health Record Data and Patient Self-Report Data

(Research In-progress)

Meredith N. Zozus, PhD,[1] Anita Walden,[1] Marcia Byers, PhD,[1] Thomas Powell MD,[1] Pei Wang, PhD,[1] Maryam Garza, MMCi,[1] Guilherme Del Fiol, MD, PhD,[2] Jessica Tenenbaum, PhD,[3] Matthew Nix, MD,[1] Carl Pieper, DrPH[3]

[1]University of Arkansas for Medical Sciences, [2]University of Utah
[3]Duke University

mzozus@uams.edu, ACWalden@uams.edu, MAByers@uams.edu, TPowell@uams.edu, PWang@uams.edu, MYGarza@uams.edu, guilherme.delfiol@utah.edu, jessie.tenenbaum@duke.edu, MWNix@uams.edu, carl.pieper@duke.edu

Abstract: Knowing the accuracy of self-reported medical data is critical to using the data in clinical decision-making and research. The same is true for data in Electronic Health Records (EHRs). For these data, accuracy reported in the literature varies widely leaving little to guide researchers in selection of the most accurate data source. This study addresses this gap by comparing patient self-report and EHR data and is the most extensive study to date in the accuracy of clinical data. The study design, data collection and preliminary results for race data are reported here. The initial comparison of race data in a small group of participating clinics showed a 33% discrepancy rate. Further, bias was evident in that all of the discrepant records were from patients reporting Hispanic ethnicity. Initial characterization of the results identified process differences among the clinics and lack of identification with the race categories among patients. The extent of variability in discrepancy rates across facilities and other data elements remains to be characterized but the necessity for accuracy assessment has been demonstrated.

Keywords: Data quality, Data accuracy, Data quality assessment, Electronic health record

INTRODUCTION

The importance of information quality in healthcare and health-related research is emphasized in Institute of Medicine (IOM) reports. (Davis, Nolan, et al. 1999; Dick, Steen et al. 1997; Stead and Lin 2009) The IOM (Davis, Nolan, et al. 1999) defines quality data as, "data that support the same conclusions as do error free data". Three major national efforts including the Patient Centered Outcomes Research Institute (PCORI, www.pcori.org), the Agency for Healthcare Research and Quality funded Electronic Data Management (EDM) Forum (www.edm-forum.org), and the National Institute of Health funded Healthcare Systems Research Collaboratory (http://www.rethinkingclinicaltrials.org) have all emphasized data quality in research either through policy or funding solicitations. The importance of data quality in research has long been recognized in federally funded clinical studies (Bagniewska, Black et al. 1986; DuChene, Hultgren, et al. 1986; Greenberg 1967; Kronmal, Davis et al. 1978; McBride and Singer 1995),

in industry trials conducted to support applications for marketing authorization, (Davis, Nolan, et al. 1999; SCDM 2013), in clinical registries (Arts, de Keizer, et al. 2002, Gliklich and Dreyer 2010), and recently in clinical studies relying on secondary use of healthcare data (NIH 2013; Zozus, Hammond, et al. 2014). The latter lists requirements for data quality in the solicitation with the goal of assuring that investigators demonstrate that data are capable of supporting research conclusions. The recent increased emphasis on research reproducibility and replication only heightens awareness and interest. (Baker 2015; Collins and Tabak 2014; Freedman, Cockburn, et al. 2015; Ioannidis 2005; NATURE 2014; Plant and Parker 2013; Young, Karr, et al. 2011; Young and Miller 2014)

With the almost ubiquitous adoption of Electronic Health Records (EHRs) in hospitals in the United States and office-based clinics not far behind, the aforementioned national efforts and the National Institutes of Health (NIH) funded Clinical and Translational Science Award (CTSA) program (www.ncats.nih.gov/ctsa), there is a large emphasis on secondary use of EHR data for research. Because routine care is selective in the information documented, interest has also increased in patient self-reported information as an alternate data source and to supplement routine care data. All of the initial seven trials conducted on the Healthcare Systems Research Collaboratory relied on EHR data and six of the seven augmented the EHR data with patient self-reported data. Together, these two data sources, patients and electronic health records, hold great promise for increasing the efficiency, generalizability and cost effectiveness of clinical research. These potential benefits, however, are dependent on the capability of the data to support research conclusions, and thus on data quality assessment. Unfortunately, there is little generalizable knowledge about the quality of EHR and patient self-reported data.

The preliminary results reported here are part of a large study to characterize the accuracy of EHR and patient self-report data across thirty-four medical conditions, eight procedures, hospitalizations, smoking status and class-level medications. The EHR Data Quality Study, currently underway contains three main aims. The first aim is to compare self-report and EHR data in the aforementioned areas and 1) estimate agreement rates between participant self-reported and EHR data, 2) estimate the Positive Predictive Value (PPV) and Negative Predictive Value (NPV) of self-reported and EHR data, and 3) characterize the reasons why the identified discrepancies exist. The second aim of the study investigates the likelihood that a participant reports data discrepancies to their healthcare provider and likelihood that reported discrepancies affect a change in the health record data. The third aim uses regressive methods to identify predictors of data discrepancies and to test hybrid algorithms leveraging both data sources to see if hybrid algorithms can outperform either data source alone in terms of predictive accuracy. Here we report the study design and preliminary results for the comparison of self-report and EHR data for race.

BACKGROUND

The medical record has long been considered the source or gold standard for comparison for research data under the assumption that if it is good enough for care, it is good enough for research. (Meads and Cooney 1982) Unfortunately the quality of data in medical records has been questioned for decades (Burnum 1989; Koran 1975a; Koran 1975b; van der Lei 1991). Hogan and Wagner (1997) and others reported the impact of variability and inaccuracies in electronic patient records, including treatment errors (Hogan and Wagner 1997; Leape, Bates, et al. 1995), underestimation of disease prevalence (Hogan and Wagner 1997; Johnson, Mant, et al. 1991), and underestimation of compliance with care standards (Hogan and Wagner 1997; Wilton and Pennisi 1994). Still others found under estimates of mortality from high risk procedures (Gallivan, Stark, et al. 2008), over estimates mortality for low-risk procedures (Gallivan, Stark, et al. 2008), false alerts to clinicians (Hogan and Wagner 1995; 1997), inaccurate decision support (Hasan and Padman 2006; Miettinen, and Korhonen 2008), inaccurate severity scores (Gibson, Haug, et al. 1996), and increased false negatives (Hogan and Wagner 1997) attributable to data error. Reported root causes of data quality problems in healthcare include failure to record information, lack of information consistency, lack of information exchange across care settings, and false commission

(Brown and Warmington 2002), pulling information forward (Hirschtick 2006), device artifacts (Benson, Junger et al. 2001), differences in data definition or recording of procedures, (Brown and Warmington 2002), and errors in data entry (Brown and Warmington 2002; Staes, Bennett et al. 2006).

The reported data quality problems and root causes are not surprising given the information flow and operations performed on the data in healthcare settings (Figure 1). This list does not include the considerable work in assessing sensitivity and specificity of tests, medical coding, or inter- and intra-clinician reliability in interpreting test results and making diagnoses. The model in Figure 1 was synthesized from four manuscripts reporting data or information flow in healthcare and between healthcare and secondary uses. (Hogan and Wagner 1997; Wyatt 1995; Gilbert, Lowenstein et al. 1996; Nagurney, Brown et al. 2005). Each operation performed on data has an associated error rate. While some data errors can be detected and corrected, others cannot. Thus in the absence of significant planned activities to monitor and control the error rates, the more operations performed on a data value, the higher the likelihood of error.

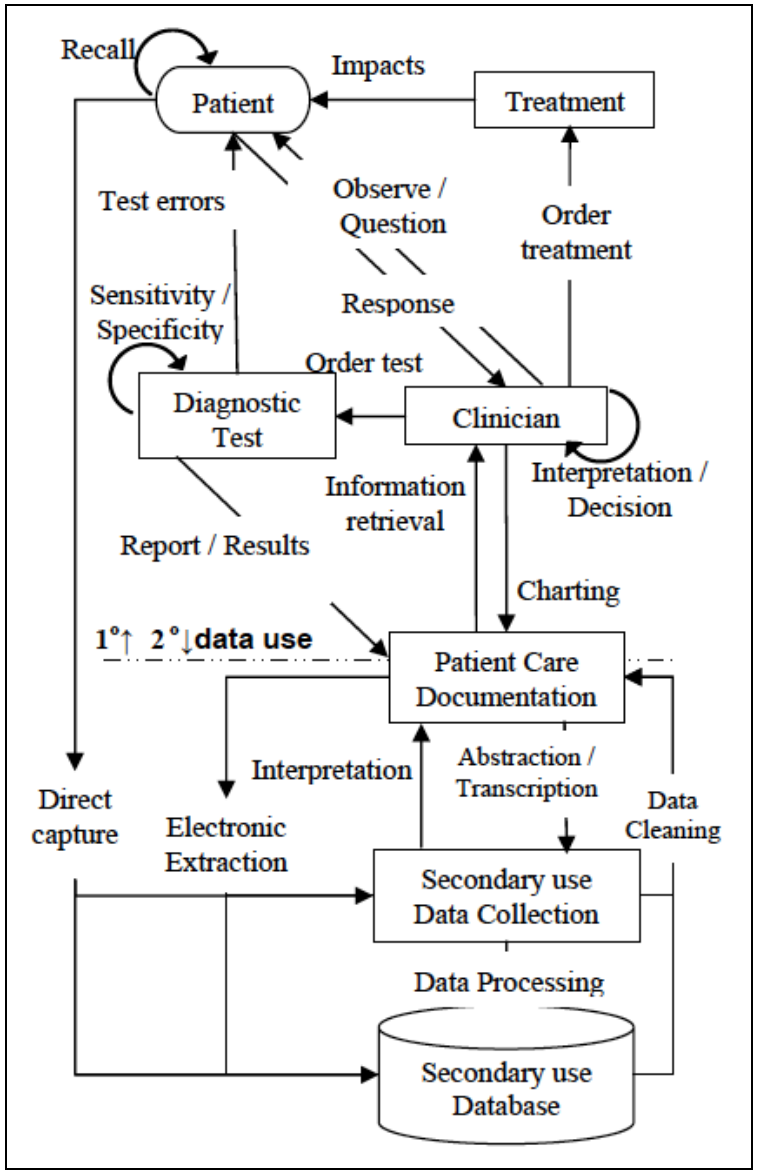


Figure 1: Information Flow in Health Care and to Secondary Uses

Two large reviews of EHR data quality have been conducted. Thiru, Hassey, and Sullivan (2003) conducted a systematic review of studies on the quality of EHR data for primary care. In a more recent study, Chan, Fowles et al. (2010) conducted a review focused on suitability of EHR data quality for healthcare quality measurement. They concluded that, "Issues related to data accuracy, completeness, and comparability must be addressed before routine EHR-based quality of care measurement can be done with confidence". For example, notable articles report differences in quality measures depending on the source of clinical data used for the measure, i.e., differences between results obtained using structured EHR data, abstracted data, and administrative data sources (Braun, Kritchevsky et al. 2006; Williams, Watt et al. 2006; Watt, Williams, et al. 2003).

In many studies today investigators have the opportunity to use either self-reported data or EHR data or both. Knowing which provides the greatest certainty in the presence of a patient characteristic, diagnosis for a given condition or occurrence of an event is necessary.

Establishing the accuracy of data depends on existence of a gold standard for comparison (Zozus, Hammond, et al. 2014). Establishing the accuracy of self-reported data including self-reported medical conditions, procedures, and events such as hospitalizations is complicated by the lack of such a gold standard in health care. As a result, investigators wishing to validate data electronically extracted or manually abstracted from medical records have used patient self-report as the source for comparison, while others wishing to validate patient reported data have used data electronically extracted or manually abstracted from medical records as the source for comparison. A wide variety of sources for comparison exist including second or repeated self-report at a later time point, data manually abstracted from medical records, structured data electronically extracted from medical records or other databases. The latter extracted structured data often include diagnoses, medications, lab data, vital signs, diagnostic test results, claims data, birth, death, and reportable conditions. Zozus, Hammond, et al. (2014) report a hierarchy of sources for comparison based on closeness to the truth and level of independence from the data to be assessed.

Although the published literature contains empirical agreement rates for many medical conditions designation of a gold standard is somewhat arbitrary and varied sources are used for comparison. Some compare self-report data to health record data while others compare health record data to other databases of varying degrees of independence. Further, the agreement rates are usually reported for only one or a small number of medical conditions and the reasons for discrepancies are rarely reported. Also complicating interpretation of reported measures of data accuracy, sometimes the source of comparison is treated as a gold standard supporting measures such as sensitivity, specificity, positive predictive value and negative predictive value. While in other cases the source of comparison is treated as an independent source equally likely to be in error, in which case, measures such as agreement rate or chance adjusted agreement rate are used. Lastly both the numerator (counts of discrepancies or errors) and the denominator (counts of data values assessed) of data accuracy have been shown to vary by a factor of four or more depending on the counting rules used by the assessor (Nahm, Dzeim, et al. 2004).

Due to the aforementioned variability, reports of data accuracy or surrogates thereof in the literature cannot be synthesized in a meaningful way. Table 1 presents results from a small sample of the recent literature reporting measures of data accuracy for seven self-reported medical conditions. In this brief review the sensitivity, specificity, PPV, NPV, Kappa, and overall agreement varied considerably (Table 1). The variability in measures reported and in measures themselves is indicative of the literature for accuracy of self-reported and EHR reported patient characteristics, medical conditions and events.

Table 1: Example Literature Reports of Relevant Data Accuracy Measures for Select Medical Conditions

Condition	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Kappa	Overall % Agreement
Diabetes[1, 5, 6, 9, 11,13]	32.0-86.7	97-99.7	70.0-94.3	97.4-99.46	0.76-0.87	66-97.2
Hypertension[1, 5, 6, 9]	49.4 - 91	81.4 – 95.3	49.4 – 86.4	89.4 – 95.7	0.41 – 0.75	88.4 – 97
MI[1, 10]	89.5	98.2	73.4	99.4	0.80	81 – 97.8
Stroke[1, 2, 3, 10]	78.4	98.6	67.4	99.2	0.71	65 - 100
Heart failure[1, 11]	68.6 – 88.5	97.0 – 98.3	36.8 – 65.1	99.2 – 99.6	0.46 – 0.74	96.3
Cancer[6, 11]	71 – 93.1	89 – 99.7	92.3	99.73	0.92	
High cholesterol[9, 10]	59.1	84.2	62.7	82.1		95

1: Okura Y, Urban, et al. 2004. 2: Tretli, Lund-Larsen, et al. 1982. 3: Paganini-Hill, Chao, et al. 1993. 4: Walker, Whincup, et al. 1998. 5: Goldman, Lin, et al. 2003. 6: Kehoe, Wu, et al. 1994. 7: Ferraro and Farmer 1999. 8: Ferraro and Su 2000. 9: Martin, Leff, et al. 2000. 10: Wada, Yatsuya, et al. 2009. 11: Baumeister, Kriston, et al. 2010. 12: Dowd and Zajacova 2010. 13: Fort, Wilcox, et al. 2014. NOTE: References 4 ,7, 8 and 12 in the sample did not report the statistics used in the table, and thus do not appear as the source for any of the measures. References 4, 7, 8 and 12 are included for completeness sake.

Increasing evidence shows that patients may be a valuable source of information about medical conditions, procedures, hospitalizations and medications. Practices in healthcare such as history taking and medication reconciliation commonly use patients as a source of this information already. Thus, for healthcare data, patients may provide the missing source for comparison (Hanauer, Prieb et al. 2014). For example, Dave deBronkart, known as, "e-Patient Dave", transferred his personal health data from his healthcare facility into Google Health, an early web based personal health record system. The transferred data contained errors including a false medication warning, exaggerated diagnoses, and conditions that Dave didn't have. Dave knew these were wrong and reported the discrepancies publicly (Wangsness 2009). In a recent study, patient prompted amendment requests to health records were reviewed (Hanauer, Prieb, et al. 2014). In total 77.8% of the amendment requests resulted in rectification of incorrect information in the health record. Hanauer et al. concluded that increased access to medical records could encourage patient participation in improving the accuracy of healthcare data. In diabetes, the answer to that question in a recent study (Fort, Wilcox et al. 2014) was yes. Thus, evidence is mounting that patients may play an important role in improving data quality in healthcare.

This PCORI-funded EHR Data Quality Study, has the opportunity to significantly add to what is known about data quality of patient self-report data and EHR data through assessment of multiple conditions and procedures as well as other clinical information commonly used in research such as medications, hospitalizations and smoking status. This study is unique in (1) the broad inclusion of commonly reused clinical data, (2) interviewing of patients to discuss the data discrepancies and (3) categorization of discrepancies according to root cause and most likely correct data source through double independent coding and third party adjudication. This approach provides an improved dataset for comparison. As such, the study provides not only agreement-based measures of accuracy, but also measures of sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV).

METHODS

The methodology used for aim one of the EHR Data Quality Study compares self-report and EHR data for medical conditions, procedures, hospitalizations, smoking status and class-level medications. The study is being conducted in Arkansas and North Carolina. In North Carolina, where the study originated, participants from a longitudinal community registry and biorepository, The Measurement to Understand Reclassification of Disease of Cabarrus/Kannapolis (MURDOCK) Study, have been followed for eight years and have provided self-report data for medical conditions, procedures, hospitalizations, smoking status and medications with annual updates as well as consent to access their EHR data (Bhattacharya, Dunham et al. 2012). In Arkansas, the study is conducted within the University of Aransas for Medical Sciences (UAMS) central Arkansas Family Medicine clinics and outlying Regional Programs. Participants will be recruited in the clinic waiting rooms; consented participants will provide self-report data for the aforementioned areas as well as authorized access to their EHR data. In addition to having been consulted in the design of the study, and participating in the conduct through interviews, to better understand the impact of health literacy on data quality, patients will also participate in the data collection and analysis.

A mixed methods approach is used to 1) estimate agreement rates between participants' self-reported and EHR data by computational methods, 2) estimate the Positive Predictive Value (PPV) and Negative Predictive Value (NPV) of EHR and self-reported data by computational methods and 3) understand how the discrepancies are distributed and the reasons why the identified discrepancies exist through qualitative analysis of recorded interviews. Discrepancies identified through the comparison of EHR and self-reported data are reported to and discussed with study participants during semi-structured telephone interviews. An interview guide prompts a dialog about each discrepancy between the participant and interviewer. The recorded interviews are subsequently coded. Two independent coders, one of which will

be a local patient, will code each discrepancy. Two codes will be assigned to each discrepancy to indicate (1) the root cause for the discrepancy and (2) the coder's impression of which data source is most likely correct. Root cause codes will be developed from the data as the study progresses to accrue a controlled terminology for root causes of data discrepancies in self-report data and EHR data. A third person will adjudicate coding differences with the coders. The interview-adjudicated dataset serves as the gold standard to which the self-report and EHR data are compared. Further, the interview-adjudicated dataset remains associated with the discrepant data values creating a source of EHR data and patient self-report data with known discrepancies.

The initial computational comparisons provide agreement rates between self-report and EHR data (bottom relationship in Figure 2). An interview-improved data source (top box in Figure 2) is created after the coding process providing a gold standard dataset. The self-report data and EHR data are then each compared to the improved data source to calculate sensitivity, specificity, positive predictive value and negative predictive value of each data source relative to the gold standard (Figure 2).

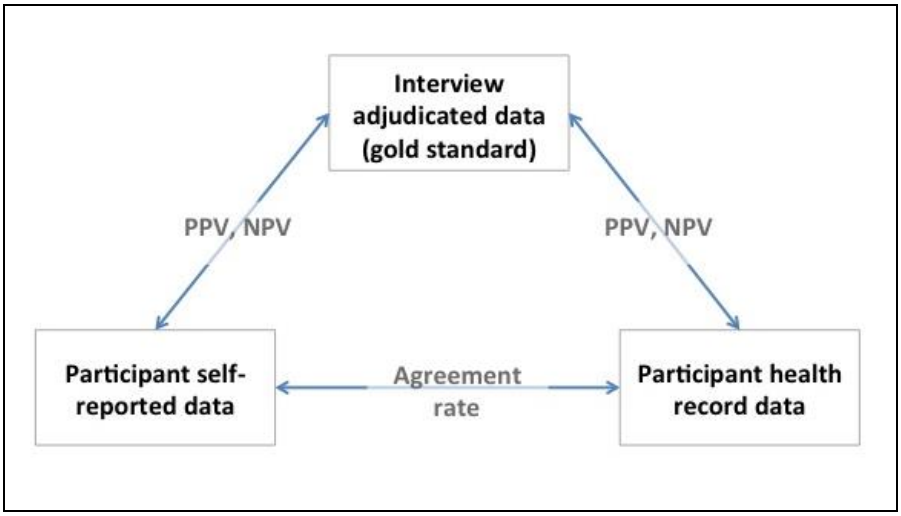


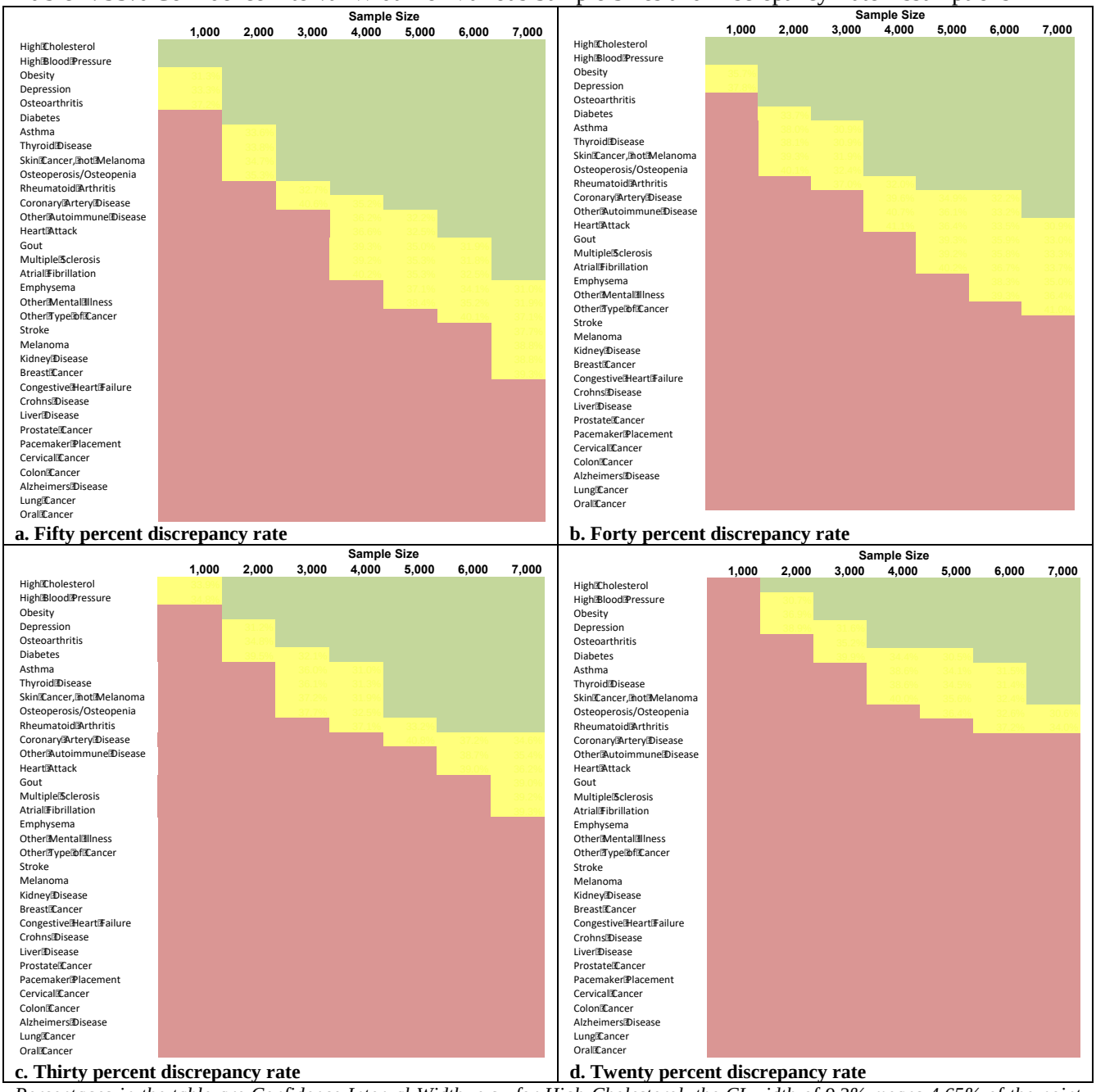
Figure 2: Structure of Study Comparisons

Identification of the discrepancies between self-report and EHR data represents a large portion of the work in this study. We define "discrepancy between self-report and EHR data" as a meaningful difference between the self-report data versus the participant's EHR data. An example of a discrepancy is, a participant indicating "no" to diabetes where the EHR data was confirmative for a diagnosis of diabetes over the same time period. In order to computationally identify such discrepancies, we need to 1) ascertain the self-report "answer" for each medical condition, procedure, medication, hospitalization and smoking status, 2) ascertain the EHR "answer" for each medical condition, procedure, medication, hospitalization and smoking status for each participant, and 3) compare them and note all differences. Obtaining the EHR answer requires developing an algorithm for each, a total of 45 algorithms (34 medical conditions, 8 procedures plus one each for hospitalizations, smoking status and class-level medications). The study applies a three-way classification of the self-report and EHR answers, "yes," "no" and "uncertain". The uncertain category for the self-report data is a response option on the data collection form. The uncertain category for the EHR data is obtained by relaxing the criteria for the algorithms bringing the total algorithm count up to ninety. Candidate algorithms were identified from multiple authoritative sources including the EMERGE Network (www.emerge.mc.vanderbilt.edu), PheKb (www.phekb.org/phenotypes), and the Centers for Medicare and Medicaid (CMS). Where authoritative phenotypes were not available, the study team searched the literature and worked with clinicians to develop candidate algorithms. Multiple algorithm variants have been developed for each algorithm and are being assessed using relative set membership. For the confirmatory algorithm, we require high specificity and sensitivity whereas for the uncertain category, we relax the specificity in a manner that preserves mutual exclusivity of the two sets.

The sample size for the EHR Data Quality Study is based on medical condition reporting rates in the North Carolina cohort. The sensitivity analysis for 95% confidence intervals is shown in Table 2. Due to the wide range of condition frequency in the population, ranging from 43% of participants reporting high

cholesterol to 0.2% reporting oral cancer. The actual confidence intervals will be reported with the measured results and will determine the conclusions that can be drawn from the data. Based on the sample size of 5,500 total participants and the discrepancy rates reported in the literature, reasonable confidence intervals will likely be obtainable for 25 percent of the conditions. Hospitalizations, medications and smoking status occur for most participants significantly narrowing the confidence intervals for hospitalizations, medications and smoking status.

Table 2: 95% Confidence Interval Width for Various Sample Sizes and Discrepancy Rate Assumptions



a. Fifty percent discrepancy rate

b. Forty percent discrepancy rate

c. Thirty percent discrepancy rate

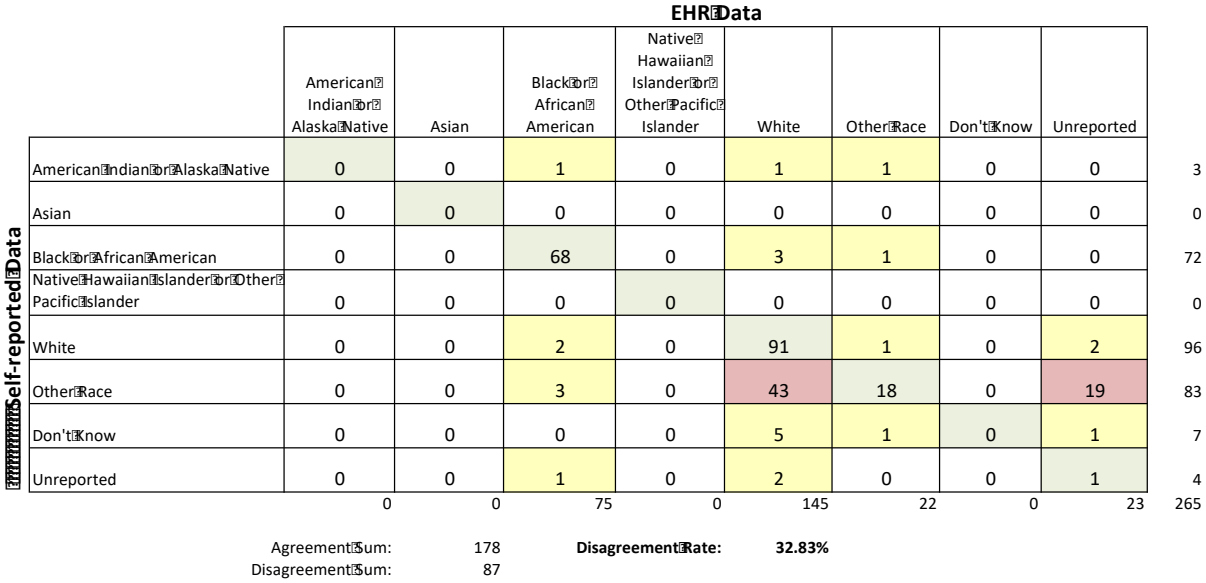
d. Twenty percent discrepancy rate

Percentages in the table are Confidence Interval Width, e.g., for High Cholesterol, the CI width of 9.3% means 4.65% of the point estimate of the discrepancy rate on either side of the point estimate of the discrepancy rate. Green: CI width of 30% (+/- 15% on either side of the point estimate of the discrepancy rate); Orange: CI width greater than 30% and up to 40% (between 15% and 20% on either side of the point estimate of the discrepancy rate); Red: CI width greater than 40% (+/- 20% or more on either side of the point estimate of the discrepancy rate)

PRELIMINARY RESULTS

Early Discrepancy Analysis

An early agreement rate analysis was done for the race data using the first set of clinics from which we received EHR data. The clinics are federally qualified clinics in North Carolina. Federally qualified clinics are required to report race and ethnicity data. EHR data were linked with the participant self-report data and compared (Figure 3). There were 265 participants with both self-report and EHR data. Figure 3 shows self-reported race (left-most column) as compared to race obtained from the EHR (top row). Agreements between self-reported and EHR race appear on the diagonal. The preliminary analysis identified a 32.8% (95% CI 28–39) discrepancy rate between self-reported race and race documented in the EHR. Further, all eighty-seven of the discrepant values were from participants also reporting Hispanic ethnicity. Thus, in addition to having a high discrepancy rate, the race data are also biased in that individuals with Hispanic ethnicity are more likely to have discrepant race data.



Self-reported Data \ EHR Data	American Indian or Alaska Native	Asian	Black or African American	Native Hawaiian or Other Pacific Islander	White	Other Race	Don't Know	Unreported	
American Indian or Alaska Native	0	0	1	0	1	1	0	0	3
Asian	0	0	0	0	0	0	0	0	0
Black or African American	0	0	68	0	3	1	0	0	72
Native Hawaiian or Other Pacific Islander	0	0	0	0	0	0	0	0	0
White	0	0	2	0	91	1	0	2	96
Other Race	0	0	3	0	43	18	0	19	83
Don't Know	0	0	0	0	5	1	0	1	7
Unreported	0	0	1	0	2	0	0	1	4
	0	0	75	0	145	22	0	23	265

Agreement Sum: 178 **Disagreement Rate: 32.83%**
Disagreement Sum: 87

Figure 3: Preliminary results indicating agreement between EHR and self-reported data for race. Cells on the diagonal represent agreement. Off-diagonal cells categorize the different ways in which the data were discrepant. Red cells note instances of particularly high disagreement.

Because the clinics were federally qualified clinics, required to report race and ethnicity, we expected a high rate of agreement between self-reported and EHR documented race data. However, this was not the case; the disagreement rate was higher than expected. When we discussed the results with the clinics and with patients, two factors came to light. First, different clinics use different methods for collecting race and ethnicity data over time. Some clinics use the "look see" method where the registration staff enter their impression of the patient's race and ethnicity even though the Office of Management and Budget (OMB) standard federal race and ethnicity categories are to be self reported by the individual to whom the data pertain. Secondly, we discovered that many patients of Hispanic ethnicity do not identify with the OMB categories. These preliminary results from a small subset of clinics illustrate the mixed methods approach of computationally identifying the discrepancies and contextualizing them through the involvement of the clinics and participants for more in-depth understanding.

Implementation Related Results

From our experience with start-up of this study, we have many implementation-related learnings. The first pertains to acquisition of EHR data. Many organizations are not yet accustomed to working within the Health Insurance Portability and Accountability Act (HIPAA) to exchange health information for research. Doing so at the facility level even with HIPAA authorization in the consent form for the study requires a contractual agreement with the providing healthcare facility. These, in some cases took over a year to finalize. Further, in the case of the North Carolina cohort even though the parent study had verbal agreement from facilities to provide data given patient consent, one large facility, when it came time to finalize the data use agreement, had undergone changes in leadership and after two years of discussion, refused to participate. Though we were not able to discern directly, their refusal may have been due to perceived risk of exposure of data problems (providing EHR data is one thing, having discrepancies in provided data studied and reported to patients is yet another). The refusal may have also been due to concerns about providing data to researchers from a neighboring health system, or due to concerns about lost opportunity cost for the research the health system itself could undertake using the data. While disclosure could have been forced with patient release forms under the HIPAA rules, the study opted not to do this. Thus, we recommend to researchers and funders alike that applications for research support that include acquisition of EHR data include executed data sharing agreements.

Another significant learning concerned the record linkage. Initially we assumed that facilities would want to link their own data, meaning that the study sends a list of contact information and other identifiers for consented participants and the health system links the data and provides data for those participants. None of the participating facilities were able to do this. The office-based practices had very limited technical computer support for extraction and use of data from their EHR system. All participating facilities provided the study a complete set of all of their EHR data and the study team linked the data. Several facilities granted study personnel access to their EHR to set up and execute the EHR data transfer. These arrangements required broader language in contracts between the sites and the study. The large health systems were similar but for a different reason. Their record linkage software is bundled in with Master Patient Index (MPI) or EHR software and they cannot use that to link external data such as the consented study participant list. Thus, the study team performed all of the record linkage.

A third significant learning was in interactions with the clinical facilities themselves. While the study was designed to be conducted external to the facilities, providing the participants a list of their data discrepancies necessarily involved the facilities in the study. Because any participant could show up to an office visit and report a discrepancy, all clinicians and staff at each facility had to be briefed on the study. The facilities handled HIPAA change request compliance differently. Under HIPAA, patients have the right to request an amendment to their health record and the facility is required to respond to such requests. All facilities in the North Carolina cohort preferred the study to instruct participants to report the data discrepancies during the encounter. The large health system in the Arkansas cohort preferred an escalation process for reporting discrepancies to a study-supported nurse case manager, and assisting the patient with a HIPAA amendment request after discussion with the case manager. This is because the nurse case manager can make amendments like those routinely made in the medication reconciliation process, e.g., changing an antibiotic that was completed a year ago to inactive. The smaller facilities in the Arkansas cohort initially preferred the discrepancies to be reported in the encounter, but also plan to use the nurse case manager escalation process. Primary care encounters are short, many occurring within a fifteen-minute window, thus impinging on encounter time in any way required significant discussions with facilities and working out a workflow for each clinic that resulted in the least impact on wait times and encounter times. For his reason, even the in-clinic recruiting for the Arkansas cohort will occur differently for each clinic with some having study staff call patients with scheduled appointments, soliciting interest and asking them to come to their appointment an hour early, while others have the registration staff ascertain interest and refer interested patients to a study-supported research assistant in

the clinic. In all cases, we needed to institute a process where the participant, after consent, could leave the clinic with the study self-report forms and the study-supported research assistant would follow-up with the patient to retrieve the forms.

DISCUSSION

Though the study is still ongoing, the preliminary findings concerning race and ethnicity data already have implications for secondary use of health record data. Personalized medicine requires studying disparities and heterogeneity of treatment effect across race and ethnicity as well as other patient characteristics. The medical community is looking to the promise of large EHR datasets for these analyses. Discrepancy rates of 33% and bias of the type found here are extensive enough to undermine these analyses. Even if the clinics studied here are not representative, and as federally qualified clinics in a rural area they may not be, variability between healthcare facilities may be substantial based on the results here. Such variability further jeopardizes analyses based on EHR data. Thus data quality assessment and data quality intervention should be planned as part of studies depending on EHR data.

The preliminary findings clearly demonstrate the necessity of assessing accuracy of clinical data prior to use. In information systems and in subsequent cleaning prior secondary use discrete data such as race categories are easily standardized so that semantically distinct data values have only one representation. Data quality dimensions such as standardization, column completeness and timeliness as commonly operationalized are considerably easier to measure than accuracy. These and other easily measured dimensions may be helpful but without accuracy assessment are not sufficient to support secondary use of clinical data. Further, use of common surrogates for data accuracy such as conformance to valid values and consistency with other values within the database should not be considered equal to or substituted for accuracy assessment unless they are validated as indicative of data quality.

From the work required thus far to start the study in North Carolina and Arkansas, it is clear that clinic workflow in study implementation is a major consideration and that study operationalization and design is best done in partnership with patients and participating healthcare facilities – both have significantly impacted the design and operations of this study. Continued work with the facilities during and after the study will likely be fruitful in understanding which data discrepancies are important to facilities, what interventions for data quality improvement facilities choose and of those which are effective and sustained. The race discrepancies were of interest to the federally qualified clinics in North Carolina and the clinic group planed to take the results into a quality improvement cycle. Thus, the data will likely improve. However, data discrepancies that aren't important to clinical facilities will not likely garner interest sufficient to initiate or sustain data quality improvements. Such low-interest data elements will likely continue to have high discrepancy rates that will likely preclude some secondary uses of the data. Any generalizable findings in this regard will inform choice of data sources for clinical research.

CONCLUSION

From the preliminary race analysis, it is clear that significant discrepancies will exist in at least some data elements and for some facilities. The extent of variability in discrepancy rates across facilities and data elements remains to be characterized but the necessity for accuracy assessment has been demonstrated. This is the most extensive study of self-report and EHR data quality to date and, based on the preliminary results, we expect to learn an amazing amount about the rate and distribution of data discrepancies and about their impact on facilities and on research results.

ACKNOWLEDGEMENTS

This research is funded through a contract (ME-1409-22573) from the Patient Centered Outcomes Research Institute (PCORI). Early work on the acquisition and linkage of the EHR data was supported by a gift to Duke University from the David H. Murdock Institute for Business and Culture and by Duke University's institutional commitment to grants K99LM011128 and R00LM011128 from the National Institutes of Health (NIH) National Library of Medicine (NLM). The ideas and opinions expressed here are those of the authors and not necessarily those of PCORI, the National Institutes of Health or any other organization.

REFERENCES

Arts D.G.T., de Keizer N.F., Scheffer G., Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. Journal of the American Medical Informatics Association Vol. 9 No. 6, Nov / Dec 2002.

Baker M, Irreproducible biology research costs put at $28 billion per year. NATURE News 9 June 2015.

Bagniewska A, Black D, Molvig K, Fox C, Ireland C, Smith J, Hulley S, Data quality in a distributed data processing system: the SHEP Pilot Study. Control Clin Trials 1986; 7: 27–37.

Baumeister, H., Kriston L, Bengel J, Härter M., High agreement of self-report and physician-diagnosed somatic conditions yields limited bias in examining mental-physical comorbidity. Journal of clinical epidemiology, 2010. 63(5): p. 558-565.

Benson M, Junger A, Quinzio L, 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. Jul 2001;40(3):190-195.

Bhattacharya S, Dunham AA, Cornish MA, Christian VA, Ginsburg GS, Tenenbaum JD, Nahm ML, Miranda ML, Califf RM, Dolor RJ, Newby LK, The Measurement to Understand Reclassification of Disease of Cabarrus/Kannapolis (MURDOCK) Study Community Registry and Biorepository. Am J Transl Res. 2012;4(4):458-70.

Braun BI, Kritchevsky SB, Kusek L, Wong ES, Solomon SL, Steele L, Richards CL, Gaynes RP, Simmons B, Comparing bloodstream infection rates: The effect of indicator specifications in the evaluation of processes and indicators in infection control (EPIC) Study. Infect Control Hosp Epidemiol 2006; 27:14-22.

Brown PJ, Warmington V. Data quality probes-exploiting and improving the quality of electronic patient record data and patient care. Int J Med Inform. Dec 18 2002;68(1-3):91-98.

Burnum JF. The misinformation era: the fall of the medical record. Ann Intern Med. Mar 15 1989;110(6):482-484.

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. Oct 2010;67(5):503-527.

Collins FS and Tabak LA, NIH plans to enhance reproducibility, NATURE Vol 505, 2014, 612-3.

Davis JR, Nolan VP, Woodcock J, Estabrook RW, Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making. Roundtable on Research and Development of Drugs, Biologics, and Medical Devices, Institute of Medicine. Washington DC: National Academy Press; 1999.

Dick RS, Steen EB, Detmer DE, eds. The computer-based patient record: an essential technology for healthcare. 2nd ed. Washington DC: National Academy Press; 1997.

Dowd, J. and A. Zajacova, Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data. Annals of epidemiology, 2010. 20(10): p. 743-749.

DuChene AG, Hultgren DH, Neaton JD, Grambsch PV, Broste SK, Aus BM, Rasmussen WL, Forms control and error detection procedures used at the Coordinating Center of the Multiple Risk Factor Intervention Trial (MRFIT). Control Clin Trials 1986; 7: 34S–45S.

Ferraro, K. and M. Farmer, Utility of health data from social surveys: Is there a gold Standard for measuring morbidity? American Sociological Review, 1999.

Ferraro, K. and Y. Su, Physician-evaluated and self-reported morbidity for predicting disability. American journal of public health, 2000. 90(1): p. 103-108.

Fort D, Wilcox AB, Weng C., Could Patient Self-reported Health Data Complement EHR for Phenotyping? AMIA Annu Symp Proc. 2014 Nov 14;2014:1738-47.

Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165, 2015.

Gallivan S, Stark J, Pagel C, Williams G, Williams WG. Dead reckoning: can we trust estimates of mortality rates in clinical databases? Eur J Cardiothorac Surg. Mar 2008;33(3):334-340.

Gibson RF, Haug PJ, Horn SD. Lessons from evaluating an automated patient severity index. J Am Med Inform Assoc. Sep-Oct 1996;3(5):349-357.

Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: Where are the methods? *Ann Emerg Med.* Mar 1996;27(3):305-308.

Gliklich RE, Dreyer NA, eds. Registries for evaluating patient outcomes: A user's guide. 2nd ed. Prepared by Outcome DEcIDE Center [Outcomes Sciences, Inc. d/b/a Outcome] under contract no. HHSA29020050035I TO3. AHRQ Publication No. 10-ECH049. Rockville MD: Agency for Healthcare Research and Quality. September 2010.

Goldman N, Lin IF, Weinstein M, Lin YH, Evaluating the quality of self-reports of hypertension and diabetes. J Clin Epidemiol. 2003 Feb;56(2):148-54.

"Greenberg Report" Organization, Review, and Administration of Cooperative Studies: A Report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967. Controlled Clinical Trials 9:137-148, 1988.

Hanauer DA, Preib R, Zheng K, Choi SW, J Am Med Inform Assoc 2014; 21:992–1000.

Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annu Symp Proc. 2006:324-328.

Hirschtick RE. A piece of my mind. Copy-and-paste. JAMA. May 24 2006;295(20):2335-2336.

Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. Sep-Oct 1997;4(5):342-355.

Hogan WR, Wagner MM. Using belief networks to enhance sharing of medical knowledge between sites with variations in data accuracy. Proc Annu Symp Comput Appl Med Care. 1995:218-222.

Ioannidis JPA, Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. JAMA, July 13, 2005—Vol 294, No. 2.

Johnson N, Mant D, Jones L, Randall T. Use of computerised general practice data for population surveillance: comparative study of influenza data. BMJ. Mar 30 1991;302(6779):763-765.

Kehoe, R., Wu SY, Leske MC, Chylack LT Jr., Comparing self-reported and physician-reported medical history. American journal of epidemiology, 1994. 139(8): p. 813-818.

Koran LM. a The reliability of clinical methods, data and judgments (first of two parts). N Engl J Med. Sep 25 1975;293(13):642-646.

Koran LM. b The reliability of clinical methods, data and judgments (second of two parts). N Engl J Med. Oct 2 1975;293(14):695-701.

Kronmal RA, Davis K, Fisher LD, Jones RA, Gillespie MJ. Data management for a large collaborative clinical trial (CASS: Coronary Artery Surgery Study). Comput Biomed Res 1978; 11: 553–566.

Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, Hallisey R, Ives J, Laird N, Laffel G, Nemeskal R, Petersen LA, Porter K, Servi D, Shea BF, Small SD, Sweitzer BJ, Thompson BT, Vander Vliet M, Bates D, Hojnowski-Diaz P, Petrycki S, Cotugno M, Patterson H, Hickey M, Kleefield S, Kinneally E, Dempsey Clapp M, Hackman JR, Edmondson A, Systems analysis of adverse drug events. ADE Prevention Study Group. JAMA. Jul 5 1995;274(1):35-43.

Martin, LM, Leff M, Calonge N, Garrett C, Nelson DE., Validation of self-reported chronic conditions and health services in a managed care population. Am J Prev Med, 2000. 18(3): p. 215-8.

McBride R and Singer SW, Introduction [to the 1995 Clinical Data Management Special Issue of Controlled Clinical Trials]. Controlled Clinical Trials 16:1S-3s, 1995.

Meads S and Cooney JP, The medical record as a data source: use and abuse. Top Health Rec Manage 2: 23–32, 1982.

Miettinen M, Korhonen M. Information quality in healthcare: coherence of data compared between organization's electronic patient records. Paper presented at: 21st IEEE International Sympossium on Computer-based Medical Systems 2008.

Nagurney JT, Brown DF, Sane S, Weiner JB, Wang AC, Chang Y. The accuracy and completeness of data collected by prospective and retrospective methods. *Acad Emerg Med.* Sep 2005;12(9):884-895.

Nahm M, Dziem G, Fendt K, Freeman L, Masi J, Ponce Z, Data Quality Survey Results. Data Basics, 2004;10(2):13-19.

National Institutes of Health (NIH), RFA-RM-13-012: NIH Health Care Systems Research Collaboratory - Demonstration Projects for Pragmatic Clinical Trials Focusing on Multiple Chronic Conditions (UH2/UH3). Posted on August 7, 2013

NATURE, Journals unite for reproducibility. NATURE Vol 515, 2014. Available from http://www.nature.com/news/journals-unite-for-reproducibility-1.16259?WT.ec_id=NATURE-20141106.

Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ., Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. J Clin Epidemiol. 2004 Oct;57(10):1096-103.

Paganini-Hill A, Chao A., Accuracy of recall of hip fracture, heart attack, and cancer: a comparison of postal survey data and medical records. Am J Epidemiol. 1993 Jul 15;138(2):101-6.

Plant A, Parker GC, Translating Stem Cell Research from the Bench to the Clinic: A Need for Better Quality Data. Stem cells and development, Vol. 22, No 18, 2013.

Society for Clinical Data Management, Good Clinical Data Management Practices (GCDMP), October 2013. Available from http://www.scdm.org

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. Jan-Feb 2006;13(1):12-15.

Stead W, Lin H, eds. Computational technology for effective health care: immediate steps and strategic directions. Washington DC: National Academy Press; 2009.

Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. May 17 2003;326(7398):1070.

Tretli S, Lund-Larsen PG, Foss OP., Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. J Epidemiol Community Health. 1982 Dec;36(4):269-73.

van der Lei, J., Use and Abuse of Computer- Stored Medical Records. Meth. Inform. Med.. Vol. 30. No.2. 1991.

Wada, K., Yatsuya H, Ouyang P, Otsuka R, Mitsuhashi H, Takefuji S, Matsushita K, Sugiura K, Hotta Y, Toyoshima H, Tamakoshi K., Self-reported medical history was generally accurate among Japanese workplace population. Journal of clinical epidemiology, 2009. 62(3): p. 306-313.

Walker MK, Whincup PH, Shaper AG, Lennon LT, Thomson AG., Validation of patient recall of doctor-diagnosed heart attack and stroke: a postal questionnaire and record review comparison. Am J Epidemiol. 1998 Aug 15;148(4):355-61.

Wangsness L, Electronic Health Records Raise Doubt, Boston Globe, April 13, 2009

Watt, Ann, Williams S, Lee K, Robertson J, Koss RG, Loeb JM, "Keen Eye on Core Measures: Joint Commission Data Quality Study Offers Insights into Data Collection, Abstracting Processes." Journal of AHIMA 74, no.10 (November 2003): 20-25.

Williams SC, Watt A, Schumaltz SP, Koss RG, Loeb JM, Assessing the reliability of standardized performance indicators, International Journal for Quality in Health Care 2006; Vol 18, No 3: pp. 246–55

Wilton R, Pennisi AJ. Evaluating the accuracy of transcribed computer-stored immunization data. Pediatrics. Dec 1994;94(6 Pt 1):902-906.

Wyatt J. Acquisition and use of clinical data for audit and research. *J Eval Clin Pract.* Sep 1995;1(1):15-27.

Young S, Karr A, Deming, data and observational studies: A process out of control and needing fixing. Significance, September 2011, 116-120.

Young SS and Miller HI, Are Medical Articles True on Health, Disease? Sadly, Not as Often as You Might Think. Genetic Engineering & Biotechnology News May 1, 2014 (Vol. 34, No. 9)

Zozus MN, Hammond WE, Green BB, MD, Kahn MG, Richesson RL, Rusincovitch SA; Simon GE, SmerekMM, Assessing Data Quality for Healthcare Systems Data Used in Clinical Research. Available from www.nihcollaboratory.org.