IJDC | *Conference Paper*

The Transparency of an Honest Data Broker in Providing Electronic Health Record Data Sufficient for Reuse

Devan Ray Donaldson
Luddy School of Informatics, Computing, and Engineering

Titus Karl Ludwig Schleyer
Regenstrief Institute, Indiana University School of Medicine

Grace Gabriella Riordan
Luddy School of Informatics, Computing, and Engineering

Jamie Lian
Luddy School of Informatics, Computing, and Engineering

Abstract

Electronic Health Records (EHRs) offer a rich data source for clinical researchers to assess a wide variety of treatments and outcomes. Researchers can use Honest Data Brokers (HDBs) to gain access to EHR data for research. Unfortunately, HDBs' data analysts can inadvertently overlook nuances in clinical workflows when generating request code for data capture and cause data extraction errors that negatively impact EHR data quality. This paper presents findings from interviews with clinical researchers who have had experience with requesting EHR data from the Regenstrief Institute Data Core (RDC), an HDB with data analysts who provide authorised access to EHR data of nearly 25 million patients in the state of Indiana. Our participants wanted greater transparency when they had questions about the quality of the datasets they received. Participants wanted their data analysts to double check the data in their system, explain how they extracted the data, let them visit the data in their system, and/or view/edit the code used for data extraction. We offer a set of recommendations for HDBs on how to provide greater transparency to clinical researchers about the processes used to generate the EHR data they receive and discuss future directions for research.

Submitted 13 February 2024 ~ *Accepted* 22 February 2024

Correspondence should be addressed to Devan Ray Donaldson. Email: drdonald@iu.edu

This paper was presented at the International Digital Curation Conference IDCC24, 19-21 February 2024

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see
https://creativecommons.org/licenses/by/4.0/

Introduction

Electronic Health Records (EHRs) store patient charts digitally, containing information on findings, laboratory tests, medical history, diagnoses, treatments, such as medications and procedures, and outcomes. Researchers can use data from EHRs for many different purposes (Coorevits et al., 2013; Cowie et al., 2017; Thyvalikakath et al., 2020), ranging from recruiting eligible patients and healthcare providers into clinical research studies, to assessing whether new treatments result in improved outcomes for patients who suffer from chronic illnesses. The Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation (GDPR), Institutional Review Boards (IRBs), facility-specific regulations, and other legal and ethical frameworks typically specify how to protect patient identifiers and confidentiality. To access EHR data for research, researchers often use Honest Data Brokers (HDBs), who provide the data while complying with privacy and confidentiality regulations. However, issues of trust and transparency are key.

HDBs act as a neutral conduit for sharing EHR data between clinical and research settings, stripping HIPAA-protected health identifiers out of EHR datasets that researchers request (Dhir et al., 2008). Researchers have no choice but to trust HDBs' data analysts to extract the data they want out of clinical systems correctly when they are not allowed to do it themselves. Establishing this trust can be difficult since there are limits to how transparent HDBs' data analysts can be about the data and their data extraction processes (Choi et al., 2015). For example, data analysts may not be able to share the code they used to extract EHR data with researchers for multiple reasons, such as intellectual property constraints, lack of documentation or the simple fact that the code is typically not understandable by researchers. Furthermore, data analysts can inadvertently cause data extraction errors during pre-processing and de-identification (de Lusignan et al., 2011), or overlook nuances in clinical workflows when generating request code for data capture (Soares et al., 2021), both of which can negatively impact EHR data quality. In either case, greater transparency between clinical researchers and data analysts about what they do during data extraction can help build researchers' trust in an HDB's data analysts and the data they receive from them.

The purpose of the study is to explore trust when transparency in the context of EHR data reuse is limited. We examine clinical researchers' attitudes about their trust in an HDB's data analysts to correctly extract EHR data from clinical systems and provide them with access to those data. This paper presents findings from interviews with multiple clinical researchers who have had experience with requesting EHR data from the Regenstrief Institute Data Core (RDC), an HDB with data analysts who provide authorised access to EHR data of nearly 25 million patients in the state of Indiana through a health information exchange data repository (Dixon et al., 2016; McDonald et al., 2005; Overhage, 2016). This paper addresses the following research questions: When do clinical researchers require greater transparency about the datasets they receive from the RDC? What type(s) of transparency do these clinical researchers want? How do data analysts at the RDC provide the transparency that the clinical researchers want?

Background

Van House (2003) asserted that use acts as a proxy for trust when she wrote, "using published sources, data collected by others, or even technology designed by others is an act of trust" (p. 278). Clinical researchers engage in this type of trusting behaviour when they use EHR data for

their research, also known as secondary use of EHR data. However, low EHR data quality can undermine trust, which in turn, can surface as non-use.

Research has shown that several factors can cause low EHR data quality (Kahn et al., 2016). Patients can provide incorrect information in the context of their clinical care (Wagner & Hogan, 1996), and healthcare professionals can input clinical data into operational systems incorrectly (Hammond et al., 2003). Data extraction errors can occur when migrating EHR data from the systems where the data were collected and de-identifying patient data for secondary use (de Lusignan et al., 2011). Data analysts can either include or exclude cases that they should not have when they do not have a clinical background and are unfamiliar with the clinical process where healthcare professionals generate and collect EHR data, which can adversely affect data quality (Soares et al., 2021). Various limitations in EHR data quality can also affect trust in EHR data (Weiskopf & Weng, 2013).

Cawsey (2022) provides definitions and a framework for data transparency that offers important links to trust and data quality. Cawsey (2022) defines data transparency as "the practice of making information about data collection, processing and usage easily accessible and understandable to the individuals whose data is being collected or used" (n.p.). Individuals can make informed decisions about whether they want to share their personal data with an organization if they have a clear understanding of its stance on data transparency (Cawsey, 2022). It is easier to identify errors or inconsistencies when data are transparent, which can be used to assess data quality. Data analysts analyse and interpret data which helps to ensure their accuracy, reliability and ethical use (Cawsey, 2022). Applied to the scenario of the secondary use of EHR data, the role of the data analyst in supporting data transparency is critical, since research has shown that EHR data have common data quality issues, such as incompleteness, inaccuracy, and inconsistency (Botsis et al., 2010).

Lamming et al. (2004) provides a metaphor from geology that is useful for understanding transparency. They use the behaviour of light on minerals and rocks as the analogy for the transfer of information and knowledge between individuals within organizations in supply management. For example, in geology, transparency refers to "light shining on or through a piece of mineral" (p. 294). This is analogous to the definition for transparency in supply management where information exists or is shared between two organizations (p. 294). In this metaphor transparency is on a continuum ranging from, opaqueness, "light cannot penetrate the surfaces nor pass through the structure of the substance," to translucency where "light can enter and exit the surfaces of the substance and pass through its structure, but is distorted or partly obscured in the passage," to clarity where "light enters and exits the surfaces or the substance and passes through its structure without alteration" (p. 294). In supply management, within a state of opaqueness, "information cannot be shared between the parties … but this constraint is acknowledged by both parties," within a state of translucency, "restricted information … may be shared", and within a state of clarity, "information … is shared candidly" (p. 294).

Research on EHR data quality assessment from the perspective of those who perform secondary use of EHR data, i.e., clinical researchers, is an understudied topic. Some clinical researchers have developed frameworks and tools to support EHR data quality assessment based on reviewing the literature on data quality definitions and dimensions and/or conducting their own case studies where they evaluate the data quality of one or more EHR datasets (Arts et al., 2002; Faulconer & de Lusignan, 2004; Kahn et al., 2012; Kahn et al., 2016). In contrast, some clinical researchers have asked other clinical researchers, besides themselves and their collaborators, about their "needs, attitudes, and perceptions as they relate to EHR data quality within the context of EHR data reuse" (Weiskopf, 2015, p. 54). This type of research has shown that many clinical researchers are concerned about quality in terms of completeness, correctness, fragmentation, granularity, concordance, structure, and signal-to-noise for the EHR data that they reuse (Weiskopf, 2015). The fact that this research has also shown that some clinical researchers believe that "EHR data require[s] substantial validation prior to reuse," and metadata should be attached to EHR data "so that researchers interested in using the data would have more information with which to inform the reuse process" could be an indication

that clinical researchers value and want greater transparency about EHR data that they want to reuse (Weiskopf, 2015, p. 64).

Typically, research on EHR data quality assessment is a means to an end, often culminating in the development of a framework, tool, or guidelines for EHR data quality assessment. In this paper, we take a different approach where we ask about EHR data quality assessment to pinpoint when and where issues of transparency and trust arise for clinical researchers who reuse EHR data from an HDB. We argue this type of research is important for two reasons. First, it can be used to inform a feedback loop between data analysts and clinical researchers where data analysts can better understand the needs of clinical researchers. Second, it can be used to inform the development of frameworks, tools, and guidelines to enhance data delivery and improve clinical researchers' EHR data reuse experiences.

Methods

The Indiana University Human Subjects Office reviewed our study and classified it as exempt per 45 CFR 46 104 (d)(2) (IRB Study# 1910559646). From October to November 2020, we conducted semi-structured interviews with seven clinical researchers who had published peer-reviewed journal articles citing the use of the RDC as a/the source of their data. Additionally, participants selected for our interviews were evaluated based on the available history and records found of them requesting EHR data from the RDC. The techniques used to generate the most suitable participants for the study consisted of using Google Scholar and PubMed along with key search terms relevant to the RDC. Confirmed authors of publications using data from the RDC from 2015 - 2019 were sent invitations to participate in the research via email. No incentives for participation were provided.

During the interviews, we asked participants about their experience with datasets they received from the RDC. Follow up questions centered on how well the data matched their requirements and perceived data quality. Interviews were video-recorded using Zoom and transcribed. An anonymised version of our dataset is publicly available at the Harvard Dataverse (Donaldson, 2024).

We analysed the interview transcripts for circumstances where participants required or mentioned wanting greater transparency from their data analysts to evaluate the quality of the datasets they received. Additionally, we analysed the interview transcripts for discussion of trust in data analysts to extract EHR data that were sufficient for reuse. We used NVivo to code the transcripts when definitions and methods of trust and transparency were discussed. We calculated inter-rater reliability for the third and fourth author using Cohen's kappa. The coders achieved a score of 0.78.

Findings

Study Participants

Of the seven participants, two were early-career (one assistant professor and one PhD student), one was mid-career (an associate professor), and four were late-career researchers (full professors or institute executives). Most (i.e., five out of seven) had requested data from the RDC more than ten times in the past five years prior to data collection, with two being heavy users (e.g., requesting data multiple times per month). Of the other two, one had requested data three times and the other had never requested data from the RDC; she was a PhD student who had permission to use a dataset her advisor had requested. All had terminal degrees (e.g., MD, DMD, or PhD) in fields ranging from biostatistics and biomedical informatics to chemistry,

dentistry, medicine, pharmacy, and public health or were pursuing one. Five were women; two were men.

Interview Results

Participants required or mentioned wanting greater transparency from their data analysts when they had questions about the datasets they received from the RDC. Specifically, participants wanted their data analysts to double check the data in their system, explain how they extracted the data, let them visit the data in their system, and/or view/edit the code used for data extraction.

Transparency at the initial stages helps

Several participants reported that they are able to acquire EHR data that are sufficient for their reuse when they communicate early with their data analysts. It helps when data analysts share details about how they pulled the data during the initial pull. As P3 points out:

> 'Once we get that first pull, we've gone through it together with the data analysts. They'll present what they have and where they got it, and then we'll ask some questions. We go back and forth as necessary. And then usually by the second pull, we come up with a model that everybody understands. We know where the data is coming from. We know what we're getting. And then our data analysts are able to continuously run the code to pull the data we need unless there's been a change.'

In this example, transparency is key because the data analysts are upfront about what data they have and where they got it. By meeting with her data analysts face-to-face, she was able to ask clarifying questions that helped her understand the data and helped the data analysts understand what she wanted. The type of transparency she valued the most was knowing where the data were coming from and knowing what data she could expect to receive.

Double-checking the data

Several participants mentioned the power of asking data analysts to double-check the data they received in their system. For example, this was an effective way for P7 to figure out why she was getting a different number of patients with a certain diagnosis code than she expected in one of her feasibility studies:

> 'We had requested patient records with certain International Classification of Diseases (ICD) codes. At the time, the data analyst pulled the data. Then there were some discrepancies later on. When we got the final record, we had fewer patients with the ICD codes. So, we were trying to figure out why. And then they said, the first data analyst also included customised codes which should not have been there.'

This example shows that when data analysts double-check the code they use to pull data, it can be useful for helping researchers understand why they receive data that are different from what they expect.

Re-extraction

Some participants mentioned having to ask the data analysts to re-extract data because they encountered some inconsistencies. P2 provides an example of this based on his experience with a dataset:

> 'For example, when we saw something inconsistent happening in the demographics file, that's when we went back and asked the analyst what's going on here? They looked back and figured it out that one of them was a matching issue. But to get to that state, they re-extracted multiple columns so we could look at the data.'

In this example, P2 was better positioned to assess the quality of his dataset after his data analyst re-extracted the data.

Data visitation

Some participants enjoyed direct benefits from being able to visit data at the RDC. For example, P2 encountered some inconsistencies in a dataset that he could not figure out. Consequently, he travelled to his data analyst's office to review the data jointly to figure out what was wrong with his dataset:

> 'I was there at Regenstrief with the data analyst looking at the data so we could try and figure it out on their system before they delivered it, because they deal with it in the identified format, and we receive the de-identified format.'

Other participants who had never visited data at the RDC commented on the value that they thought it would add. For example, P6 said:

> 'I wasn't able to go where the data is stored. That has to be through the data analyst. Sometimes I feel like if I can see the data for myself, then it would be easier to understand. But I think I couldn't because of how the RDC is set up, and because of regulations and things. So, when I have questions, I have to ask the data analyst. Sometimes I wish I could see more things.'

It is not clear from the interviews why some participants were given the opportunity for data visitation and others were not. Comparing P2 and P6, perhaps P2 was allowed to visit the data because he was the original data requestor and P6 was not allowed data visitation because she was not the original data requestor, her advisor was. In any case, P6's comments illustrate that she thought data visitation would have increased her understanding of the data.

View/edit data extraction code

Participants shared different views on the importance of data analysts providing the code they used for data extraction to offer greater transparency regarding the datasets they received from the RDC. For example, P3 was not in favour of this:

> 'I don't see what they see. I don't know how to program and extract from the tables. Ultimately, we are at the mercy of the data analyst truly understanding what the data look like. I don't have a lot of ways to double check that the datasets I receive from the RDC are 100% accurate, especially for these large pulls.'

In contrast, P5 demanded trust through transparency by repeatedly providing feedback to the RDC about wanting to receive the code her data analysts use to extract the datasets she receives:

> 'Whenever they pull data, they need to provide documentation for me because I also do programming. In my mind, it is crucial how you pull these data. They cannot just give me a dataset using a black box approach and say, "trust me, I pulled the data right." So, we had asked for this for many years because myself and my group are programmers. We can edit their code. A lot of their data pulls use the same software we use. Why can't they just give us a program that they use to pull data? So, this way, one more pair of

> eyes can make sure you don't forget one ICD code, or you don't forget to restrict the time window. If I want to restrict it to only this one year, I can make sure you don't forget that. To me, it would be very reassuring if I can see everything you programmed and see if it lines up with what we envisioned. And this would help us know, in cases where we don't get the data we envisioned, whether it's the programmer's fault.'

Discussion and Future Directions

Our findings reflect the importance of trust through transparency in EHR data reuse because some of our participants wanted their data analysts to share as much information as possible about their data extraction processes rather than trusting that they pulled the data they wanted correctly. This sentiment causes us to recommend that HDBs share the code they use for data extraction with researchers, or at least allow them to view it, of course, in accordance with regulatory guidelines. Additionally, we recommend that HDBs clearly explain their data visitation policies to their customers and offer data visitation as an option for them, especially those who feel that it would increase their understanding of the data.

Based on our findings, we extend the metaphor of transparency where the behaviour of light acts as an analogy for the transfer of information and knowledge in supply management to the secondary use of EHR data received from an HDB. Our definition of transparency refers to an HDB's data analysts sharing EHR data with clinical researchers. Opaqueness is present when, for whatever reason, an HDB's data analyst cannot share EHR data with a clinical researcher. Our participants did not experience opaqueness. Translucency is when an HDB's data analyst shares de-identified data with a clinical researcher. Our participants cited several examples of translucency. Sometimes the translucency was sufficient for data reuse, and at other times, usually when participants had concerns about data quality, the translucency was insufficient for data reuse. Clarity is when an HDB's data analyst freely shares patient identifying EHR data and the code they used to pull the data with a clinical researcher. Similar to when applying this metaphor in supply management (Lamming et al., 2004), it is unlikely that the HDB we studied (or any other HDB for that matter) can provide total clarity. Notwithstanding, HDBs may be able to provide 'fissures' of clarity that can benefit clinical researchers by clarifying concerns they may have about data quality. Examples of fissures include sharing the code that data analysts use to pull the data and allowing clinical researchers data visitation.

Clinical researchers need to understand that a lot of interpersonal work beyond simply submitting a request is needed to ensure that EHR data is being pulled correctly. Trust in a HDB depends on bridging the knowledge gap between researchers and analysts by creating a mutual understanding and including individuals on research teams with common languages they can use to communicate. HDBs can improve their delivery of data requests by engaging with the requesting researcher early on in the process so they can better understand their needs while making it clear what they can and cannot pull upfront. As one participant (P3) suggested, "the process could be improved by having there be a required conversation between the investigator and the analyst after the submission of the request." The consensus among our participants was that establishing a relationship with the data analyst led to significantly better data quality because they built an understanding of the researcher's needs and their knowledge on the backend of the data through engaging in long-term communication through involvement in projects over time.

Going forward, we suggest two directions for future research. First, conduct semi-structured interviews with data analysts at HDBs to identify current opportunities and challenges to providing greater transparency about the processes they use to extract EHR data and support reuse. Second, develop and administer a large-scale survey to clinical researchers to understand how they evaluate EHR data quality. Both research activities could inform the future development of a framework for communication between clinical researchers and data analysts

that takes issues of trust and transparency into account with respect to EHR data quality and reuse.

Acknowledgements

The authors thank several past and present administrators, research scientists, and staff at the Regenstrief Institute for their support of this project (Peter Embi, Eneida Mendonca, and Faye Smith). Additionally, the authors thank several past members of the Donaldson lab group who worked on this project (Emily Grover, Jacob Samm, Huixin Tian, and Julie Wasserman). We thank Kaitlin L. Costello for reading and providing feedback on previous drafts of this manuscript. This research is supported by a grant from the Indiana University Enhanced Mentoring Program with Opportunities for Ways to Excel in Research (EMPOWER).

References

Arts, D. G., De Keizer, N. F., & Scheffer, G. J. (2002). Defining and improving data quality in medical registries: a literature review, case study, and generic framework. Journal of the American Medical Informatics Association : JAMIA, 9(6), 600–611. https://doi.org/10.1197/jamia.m1087

Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. Summit on translational bioinformatics, 2010, 1–5.

Cawsey, M. (2022). The ultimate guide to data transparency. Stibo Systems, Master Data Management. Retrieved from https://www.stibosystems.com/blog/data-transparency#:~:text=Data%20transparency%20refers%20to%20the,is%20being%20collected%20or%20used

Choi, H. J., Lee, M. J., Choi, C.-M., Lee, J., Shin, S.-Y., Lyu, Y., Park, Y. R., & Yoo, S. (2015). Establishing the role of honest broker: Bridging the gap between protecting personal health data and clinical research efficiency. PeerJ, 3. https://doi.org/10.7717/peerj.1506

Coorevits, P., Sundgren, M., Klein, G. O., Bahr, A., Claerhout, B., Daniel, C., Dugas, M., Dupont, D., Schmidt, A., Singleton, P., De Moor, G., & Kalra, D. (2013). Electronic health records: new opportunities for clinical research. Journal of internal medicine, 274(6), 547–560.

Cowie, M. R., Blomster, J. I., Curtis, L. H., Duclaux, S., Ford, I., Fritz, F., ... & Zalewski, A. (2017). Electronic health records to facilitate clinical research. Clinical Research in Cardiology, 106, 1-9.

de Lusignan, S., Liaw, S. T., Krause, P., Curcin, V., Vicente, M. T., Michalakidis, G., Agreus, L., Leysen, P., Shaw, N., & Mendis, K. (2011). Key concepts to assess the readiness of data for international research: data quality, lineage and provenance, extraction and processing errors, traceability, and curation. Contribution of the IMIA Primary Health Care Informatics Working Group. Yearbook of medical informatics, 6, 112–120.

Dhir, R., Patel, A.A., Winters, S., Bisceglia, M., Swanson, D., Aamodt, R., Becich, M.J. (2008). A multidisciplinary approach to honest broker services for tissue banks and clinical data: A pragmatic and practical model. Cancer. 113, 1705–1715. https://doi.org/10.1002/cncr.23768

Dixon, B. E., Whipple, E. C., Lajiness, J. M., & Murray, M. D. (2016). Utilizing an integrated infrastructure for outcomes research: a systematic review. Health information and libraries journal, 33(1), 7–32.

Donaldson, D. (2024). "Replication Data for: "The Transparency of an Honest Data Broker in Providing Electronic Health Record Data Sufficient for Reuse." Harvard Dataverse. Retrieved from https://doi.org/10.7910/DVN/OVMREQ

Faulconer, E. R., & de Lusignan, S. (2004). An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar. Informatics in primary care, 12(4), 243–254.

Hammond, K. W., Helbig, S. T., Benson, C. C., & Brathwaite-Sketoe, B. M. (2003). Are electronic medical records trustworthy? Observations on copying, pasting and duplication. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2003, 269–273.

Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S. T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., & Schilling, L. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Washington, DC), 4(1), 1244. https://doi.org/10.13063/2327-9214.1244

Kahn, M. G., Raebel, M. A., Glanz, J. M., Riedlinger, K., & Steiner, J. F. (2012). A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical care, 50 Suppl(0), S21–S29. https://doi.org/10.1097/MLR.0b013e318257dd67

Lamming, R., Caldwell, N., & Harrison, D. (2004). Developing the concept of transparency for use in supply relationships. British Journal of Management, 15(4), 291-302.

McDonald, C. J., Overhage, J. M., Barnes, M., Schadow, G., Blevins, L., Dexter, P. R., Mamlin, B., & INPC Management Committee (2005). The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. Health affairs (Project Hope), 24(5), 1214–1220.

Overhage JM. (2016). The Indiana Health Information Exchange. In B. E. Dixon (Ed.), Health Information Exchange: Navigating and managing a network of health information systems (1st ed., pp. 267-279). Academic Press.

Soares, N., Singhal, S., Kloosterman, C., & Bailey, T. (2021). An interdisciplinary approach to reducing errors in extracted electronic health record data for research. Perspectives in Health Information Management, 18, 1-9.

Thyvalikakath, T. P., Duncan, W. D., Siddiqui, Z., LaPradd, M., Eckert, G., Schleyer, T., Rindal, D. B., Jurkovich, M., Shea, T., Gilbert, G. H., & National Dental PBRN Collaborative Group (2020). Leveraging Electronic Dental Record Data for Clinical Research in the National Dental PBRN Practices. Applied clinical informatics, 11(2), 305–314.

Van House, N. A. (2003). Digital libraries and collaborative knowledge construction. In A. P. Bishop, N. A. Van House & B. P. Buttenfield (Eds.), Digital library use: Social practice in design and evaluation (pp. 271-295). Cambridge, Mass.: MIT Press.

Wagner, M. M., & Hogan, W. R. (1996). The accuracy of medication data in an outpatient electronic medical record. Journal of the American Medical Informatics Association : JAMIA, 3(3), 234–244. https://doi.org/10.1136/jamia.1996.96310637

Weiskopf, N.G. (2015). Enabling the Reuse of Electronic Health Record Data through Data Quality Assessment and Transparency, [Doctoral dissertation, Columbia University]. Columbia Academic Commons. https://academiccommons.columbia.edu/doi/10.7916/D8RF5SS2

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the American Medical Informatics Association : JAMIA, 20(1), 144–151.