

Review

Digital Health Data Quality Issues: Systematic Review

Rehan Syed^{1*}, PhD; Rebekah Eden^{1*}, PhD; Tendai Makasi¹, PhD; Ignatius Chukwudi¹, PhD; Azumah Mamudu¹, MSc; Mostafa Kamalpour¹, PhD; Dakshi Kapugama Geeganage¹, PhD; Sareh Sadeghianasl¹, PhD; Sander J J Leemans², PhD; Kanika Goel¹, PhD; Robert Andrews¹, PhD; Moe Thandar Wynn¹, PhD; Arthur ter Hofstede¹, PhD; Trina Myers¹, PhD

¹School of Information Systems, Faculty of Science, Queensland University of Technology, Brisbane, Australia

²Rheinisch-Westfälische Technische Hochschule, Aachen University, Aachen, Germany

*these authors contributed equally

Corresponding Author:

Rehan Syed, PhD

School of Information Systems

Faculty of Science

Queensland University of Technology

2 George Street

Brisbane, 4000

Australia

Phone: 61 7 3138 9360

Email: r.syed@qut.edu.au

Abstract

Background: The promise of digital health is principally dependent on the ability to electronically capture data that can be analyzed to improve decision-making. However, the ability to effectively harness data has proven elusive, largely because of the quality of the data captured. Despite the importance of data quality (DQ), an agreed-upon DQ taxonomy evades literature. When consolidated frameworks are developed, the dimensions are often fragmented, without consideration of the interrelationships among the dimensions or their resultant impact.

Objective: The aim of this study was to develop a consolidated digital health DQ dimension and outcome (DQ-DO) framework to provide insights into 3 research questions: What are the dimensions of digital health DQ? How are the dimensions of digital health DQ related? and What are the impacts of digital health DQ?

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, a developmental systematic literature review was conducted of peer-reviewed literature focusing on digital health DQ in predominately hospital settings. A total of 227 relevant articles were retrieved and inductively analyzed to identify digital health DQ dimensions and outcomes. The inductive analysis was performed through open coding, constant comparison, and card sorting with subject matter experts to identify digital health DQ dimensions and digital health DQ outcomes. Subsequently, a computer-assisted analysis was performed and verified by DQ experts to identify the interrelationships among the DQ dimensions and relationships between DQ dimensions and outcomes. The analysis resulted in the development of the DQ-DO framework.

Results: The digital health DQ-DO framework consists of 6 dimensions of DQ, namely accessibility, accuracy, completeness, consistency, contextual validity, and currency; interrelationships among the dimensions of digital health DQ, with consistency being the most influential dimension impacting all other digital health DQ dimensions; 5 digital health DQ outcomes, namely clinical, clinician, research-related, business process, and organizational outcomes; and relationships between the digital health DQ dimensions and DQ outcomes, with the consistency and accessibility dimensions impacting all DQ outcomes.

Conclusions: The DQ-DO framework developed in this study demonstrates the complexity of digital health DQ and the necessity for reducing digital health DQ issues. The framework further provides health care executives with holistic insights into DQ issues and resultant outcomes, which can help them prioritize which DQ-related problems to tackle first.

(*J Med Internet Res* 2023;25:e42615) doi: [10.2196/42615](https://doi.org/10.2196/42615)

KEYWORDS

data quality; digital health; electronic health record; eHealth; systematic reviews

Introduction

Background

The health care landscape is changing globally owing to substantial investments in health information systems that seek to improve health care outcomes [1]. Despite the rapid adoption of health information systems [2] and the perception of digital health as a panacea [3] for improving health care quality, the outcomes have been mixed [4,5]. As Reisman [6] noted, despite substantial investment and effort and widespread application of digital health, many of the promised benefits have yet to be realized.

The promise of digital health is principally dependent on the ability to electronically capture data that can be analyzed to improve decision-making at the local, national [6], and global levels [7]. However, the ability to harness data effectively and meaningfully has proven difficult and elusive, largely because of the quality of the data captured. Darko-Yawson and Ellingsen [8] highlighted that digital health has resulted in more bad data rather than improving the quality of data. It is widely accepted that the data from digital health are plagued by accuracy and completeness concerns [9-12]. Poor data quality (DQ) can be detrimental to continuity of care [13], patient safety [14], clinician productivity [15], and research [16].

To assess DQ, scholars have developed numerous DQ taxonomies, which evaluate the extent to which the data contained within digital health systems adhere to multiple dimensions (ie, measurable components of DQ). Weiskopf and Weng [17] identified 5 dimensions of DQ spanning completeness, correctness, concordance, plausibility, and currency. Subsequently, Weiskopf et al [18] refined the typology to consist of only 3 dimensions: completeness, correctness, and currency. Similarly, Puttkammer et al [13] focused on completeness, accuracy, and timeliness, whereas Kahn et al [19] examined conformance, completeness, and plausibility. Others identified “fitness of use” [20] and the validity of data to a specific context [21] as key DQ dimensions. Overall, there are wide-ranging definitions of DQ, with an agreed-upon taxonomy evading the literature. In this paper, upon synthesizing the literature, we define DQ as the extent to which digital health data are accessible, accurate, complete, consistent, contextually valid, and current. When consolidated frameworks are developed, the dimensions are often treated in a fragmented manner, with few attempts to understand the relationships between the dimensions and the resultant outcomes. This is substantiated by Bettencourt-Silva et al [22], who indicated that DQ is not systematically or consistently assessed.

Research Aims and Questions

Failure of health organizations to leverage high-quality data will compromise the sustainability of an already strained health care system [23]. Therefore, we undertook a systematic literature review to answer the following research questions: (1) What are the dimensions of digital health DQ? (2) How are the dimensions of digital health DQ related? and (3) What are the impacts of digital health DQ? The aim of this research was to develop, from synthesizing the literature, a consolidated digital health DQ dimension and outcome (DQ-DO) framework, which

demonstrates the DQ dimensions and their interrelationships as well as their impact on core health care outcomes. The consolidated DQ-DO framework will be beneficial to both research and practice. For researchers, our review consolidates the digital health DQ literature and provides core areas for future research to rigorously evaluate and improve digital health DQ. For practice, this study provides health care executives and strategic decision makers with insights into both the criticality of digital health DQ by exemplifying the impacts and the complexity of digital health DQ by demonstrating the interrelationships between the dimensions. [Multimedia Appendix 1](#) [24] provides a list of common acronyms used in this study.

This paper is structured as follows: first, we provide details of the systematic literature review method; second, in line with the research questions, we present our 3 key findings—(1) DQ dimensions, (2) DQ interrelationships, and (3) DQ outcomes; and third, we compare the findings of our study with those of previous studies and discuss the implications of this work.

Methods

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and the guidelines proposed by Webster and Watson [25] for systematic literature reviews. Specifically, consistent with Templier and Paré [26], this systematic literature review was developmental in nature with the goal of developing a consolidated digital health DQ framework.

Literature Search and Selection

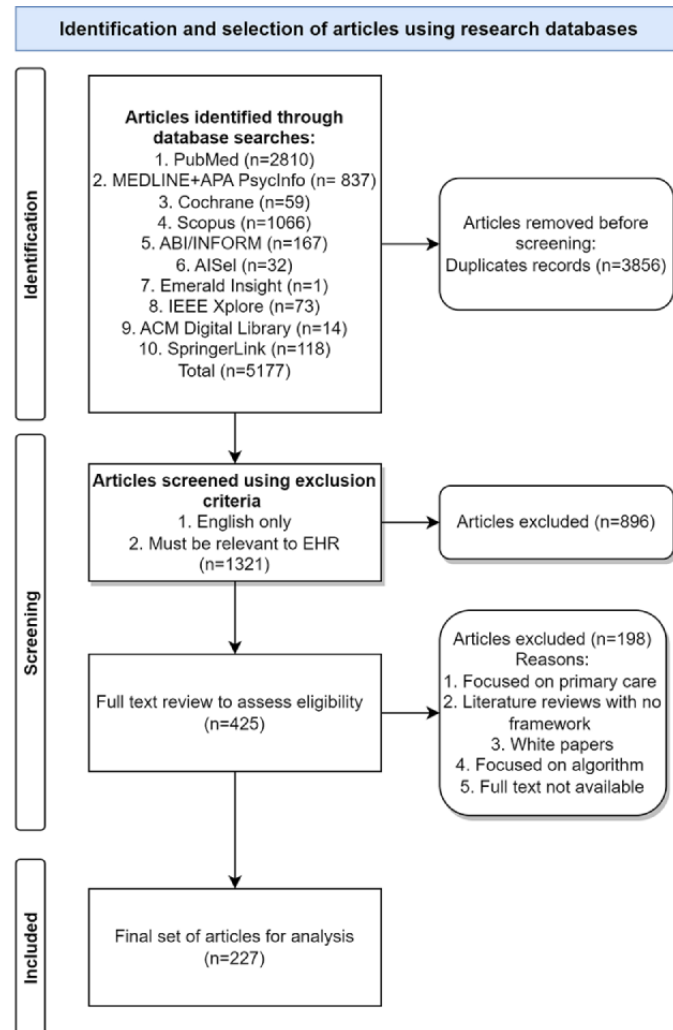
To ensure the completeness of the review [25] and consistent with interdisciplinary reviews, the literature search spanned multiple fields and databases (ie, PubMed, Public Health, Cochrane, SpringerLink, EBSCOhost [MEDLINE and PsycInfo], ABI/INFORM, AISel, Emerald Insight, IEEE Xplore digital library, Scopus, and ACM Digital Library). The search was conducted in October 2021 and was not constrained by the year of publication because the concept of DQ has a long-standing academic history. The search terms were reflective of our research topic and research questions. To ensure comprehensiveness, the search terms were broadened by searching their synonyms. For example, we used search terms such as “electronic health record,” “digital health record,” “e-health,” “electronic medical record,” “EHR,” “EMR,” “data quality,” “data reduction,” “data cleaning,” “data pre-processing,” “information quality,” “data cleansing,” “data preparation,” “intelligence quality,” “data wrangling,” and “data transformation.” Keywords and search queries were reviewed by the reference librarian and subject matter experts in digital health ([Multimedia Appendix 2](#)).

The papers returned from the search were narrowed down in a 4-step process ([Figure 1](#)). In the identification step, 5177 articles were identified through multiple database searches, and from these, 3856 (74.48%) duplicates were removed, resulting in 1321 (25.42%) articles. These 1321 articles were randomly divided into 6 batches, which were assigned to separate researchers, who applied the inclusion and exclusion criteria

(Textbox 1). As a result of abstract screening, 67.83% (896/1321) of articles were excluded, resulting in 425 (32.17%) articles. Following an approach to the abstract screening, the 425 articles were again randomly divided into 6 batches and assigned to 1 of the 6 researchers to read and assess the relevance of the article in line with the selection criteria. The assessment of each of the 425 articles was then verified by the research team, resulting in a final set of 227 (53.4%) relevant

articles. During this screening phase (ie, abstract and full text), daily meetings were held with the research team in which any uncertainties were raised and discussed until consensus was reached by the team as to whether the article should be included or excluded from the review. In line with Tempplier and Paré [26], as this systematic literature review was developmental in nature rather than an aggregative meta-analysis, quality appraisals were not performed on individual articles.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) inclusion process. EHR: electronic health record.



Textbox 1. Inclusion and exclusion criteria.

<p>Inclusion criteria</p> <ul style="list-style-type: none">• Specifically focuses on data quality in digital health• Empirical papers or review articles where conceptual frameworks were either developed or assessed• Considers digital health within hospital settings• Published in peer-reviewed outlets within any time frame• Published in English <p>Exclusion criteria</p> <ul style="list-style-type: none">• Development of algorithms for advanced analytics techniques (eg, machine learning and artificial intelligence) without application within hospital settings• Descriptive papers without a conceptual framework or an empirical analysis• Focused only on primary care (eg, general practice)• Pre-go-live considerations (eg, software development)• Theses and non-peer-reviewed publications (eg, white papers and editorials)

Literature Analysis

The relevant articles were imported to NVivo (version 12; QSR International), where the analysis was iteratively performed. To ensure reliability and consistency in coding, a coding rule book [27] was developed and progressively updated to guide the coding process. The analysis involved 6 steps (Figure 2).

In the first step of the analysis, the research team performed open coding [27], where relevant statements from each article were extracted using verbatim codes and grouped based on similarities [28]. The first round of coding resulted in 1298 open codes. Second, the open codes were segmented into 2 high-level themes: the first group contained 1044 (80.43%) open codes pertaining directly to DQ dimensions (eg, data accuracy), and the second group contained 254 (19.57%) open codes pertaining to DQ outcomes (eg, financial outcomes).

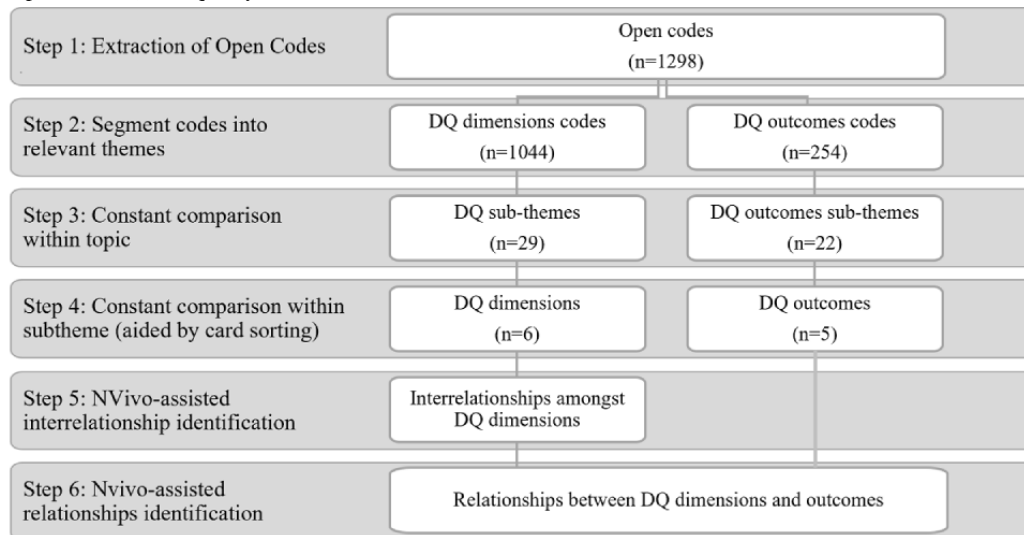
In the third step, through constant comparison [29], the 1044 raw DQ codes were combined into 29 DQ subthemes based on commonalities (eg, contextual DQ, fitness for use, granularity, relevancy, accessibility, and availability). In the fourth step, again by performing iterative and multiple rounds of constant comparison, the 254 open codes related to DQ outcomes were used to construct 22 initial DQ outcome subthemes (eg, patient safety, clinician-patient relationship, and continuity of care). The DQ outcome subthemes were further compared with each other, resulting in 5 DQ outcome dimensions (eg, clinical, business process, research-related, clinician, and organizational outcomes). For the DQ subthemes, a constant comparison was performed using the card-sorting method [30], where an expert panel of 8 DQ researchers split into 4 groups assessed the subthemes for commonalities and differences. The expert groups

presented their categorization to each other until a consensus was reached. This resulted in a consolidated set of 6 DQ dimensions (accuracy, consistency, completeness, contextual validity, accessibility, and currency). [Multimedia Appendix 3 \[9,12,13,15,16,18,19,21,31-65\]](#) provides an example of how the open codes were reflected in the subthemes and themes.

After identifying the DQ dimensions and outcomes, the next stage of coding progressed to identifying the interrelationships (step 5) among the DQ dimensions and the relationships (step 6) between the DQ dimensions and DQ outcomes. To this end, the matrix coding query function using relevant Boolean operators (AND and NEAR) in NVivo was performed. The outcomes of the matrix queries were reviewed and verified by an expert researcher in the health domain.

Throughout the analysis, steps for providing credibility to our findings were performed. First, before commencing the analysis, the research team members who extracted the verbatim codes initially independently reviewed 3 common articles and then convened to review any variations in coding. In addition, they reconvened multiple times a week to discuss their coding and update the codebook to ensure that a consistent approach was followed. Coder corroboration was performed throughout the analysis, with 2 experienced researchers independently verifying all verbatim codes until a consensus was reached [27]. Subsequent coder corroboration was performed by 2 experienced researchers to ensure that the open codes were accurately mapped to the themes and dimensions. This served to provide internal reliability. Steps for improving external reliability were also performed [66]. Specifically, the card-sorting method provided an expert appraisal. In addition, the findings were presented to and confirmed by 3 digital health care professionals.

Figure 2. Analysis process. DQ: data quality.



Results

Overview

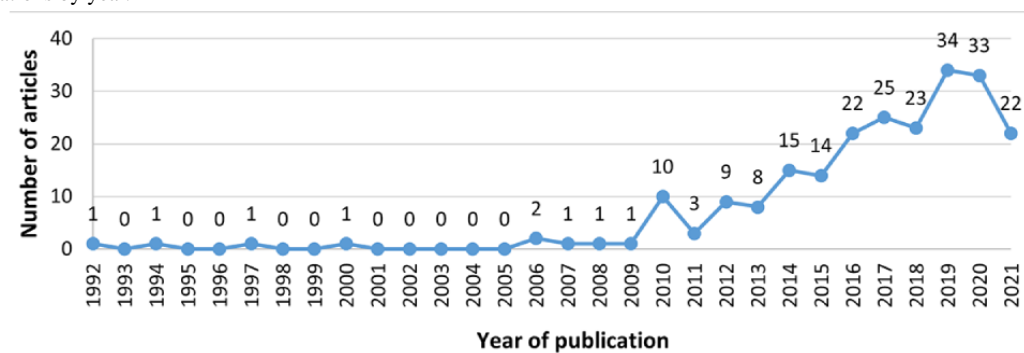
The vast majority of relevant articles were published in journal outlets (169/227, 74.4%), followed by conference proceedings (42/227, 18.5%) and book sections (16/227, 7%). The 169 journal articles were published in 107 journals, with 12% (n=13) of the journals publishing >1 study (these journals are *BMC Medical Informatics and Decision Making*, *eGEMS*, *International Journal of Medical Informatics*, *Applied Clinical Informatics*, *Journal of Medical Internet Research*, *Journal of the American Medical Information Association*, *PLOS One*, *BMC Emergency Medicine*, *Computer Methods and Programs in Biomedicine*, *International Journal of Population Data Science*, *JCO Clinical Cancer Informatics*, *Perspectives in Health Information Management*, *Studies in Health Technology*

and *Informatics*, *Australian Health Review*, *MBC Health Services Research*, *BMJ Open*, *Decision Support Systems*, *Health Informatics Journal*, *International Journal of Information Management*, *JAMIA Open*, *JMIR Medical Informatics*, *Journal of Biomedical Informatics*, *Journal of Medical Systems*, *Malawi Medical Journal*, *Medical Care*, *Online Journal of Public Health Informatics*, and *Telemedicine and e-Health*). A complete breakdown of the number of articles published in each outlet is provided in [Multimedia Appendix 4](#).

Overall, as illustrated in [Figure 3](#), the interest in digital health DQ has been increasing over time, with sporadic interest before 2006.

In the subsequent sections, we provide an overview of the DQ definitions, DQ dimensions, their interrelationships, and DQ outcomes to develop a consolidated digital health DQ framework.

Figure 3. Publications by year.



DQ Definitions

Multiple definitions of DQ were discussed in the literature ([Multimedia Appendix 5](#) [17,18,20-22,31,54,67-77]). There was no consensus on a single definition of DQ; however, an analysis of the definitions revealed two perspectives, which we labeled as the (1) context-agnostic perspective and (2) context-aware perspective. The context-agnostic perspective defines DQ based on a set of dimensions, regardless of the context within which the data are used. For instance, as Abiy et al [67] noted “documentation and contents of data within an

electronic medical record (EMR) must be accurate, complete, concise, consistent and universally understood by users of the data, and must support the legal business record of the organization by maintaining the required parameters such as consistency, completeness and accuracy.” By contrast, the context-aware perspective evaluates the dimensions of DQ with recognition of the context within which the data are used. For instance, as the International Organization for Standardization and Liu et al [78] noted, DQ is “the degree to which data satisfy the requirements defined by the product-owner organization”

and can be reflected through its dimensions such as completeness and accuracy.

DQ Dimensions

Overview

In total, 30 subthemes were identified and grouped into 6 DQ dimensions: accuracy, consistency, completeness, contextual

validity, accessibility, and currency (Table 1; Multimedia Appendix 6 [8-12,14-16,18-22,31-62,67,69,71,72,76,79-168]). Consistency (164/227, 72.2%), completeness (137/227, 60.4%), and accuracy (123/227, 54.2%) were the main DQ dimensions. Comparatively, less attention was paid to accessibility (28/227, 12.3%), currency (18/227, 7.9%), and contextual validity (26/227, 11.5%).

Table 1. Description of the data quality (DQ) dimensions.

Dimension	Description	Subthemes
Accuracy	“The degree to which data reveal the truth about the event being described” [31]	Validity, correctness, integrity, conformance, plausibility, veracity, and accurate diagnostic data
Consistency	“Absence of differences between data items representing the same objects based on specific information requirements. Consistent data contain the same data values when compared between different databases” [31]	Inconsistent data capturing, standardization, concordance, uniqueness, data variability, temporal variability, system differences, semantic consistency, structuredness, and representational consistency
Completeness	“The absence of data at a single moment over time or when measured at multiple moments over time” [79]	Missing data, level of completeness, representativeness, fragmentation, and breadth of documentation
Contextual validity	Assessment of DQ is “dependent on the task at hand” [18]	Contextual DQ, fitness for use, granularity, and relevancy
Accessibility	How “feasible it is for users to extract the data of interest” [18]	Accessible DQ and availability
Currency	“The degree to which data represent reality from the required point in time” [32]	Timeliness

DQ Dimension: Accessibility

The accessibility dimension (28/227, 12.3%) is composed of both the accessibility (15/28, 54%) and availability (13/28, 46%) subthemes, reflecting the feasibility for users to extract data of interest [18]. Scholars regularly view the *accessibility* subtheme favorably, with the increased adoption of electronic health record (EHR) systems overcoming physical and chronological boundaries associated with paper records by allowing access to information from multiple locations at any time [33,80]. Top et al [33] noted that EHR made it possible for nurses to access patient data, resulting in improved decision-making. Furthermore, Rosenlund et al [81] noted that EHRs benefit health care professionals by providing increased opportunities for searching and using information. The *availability* subtheme is an extension of the accessibility subtheme and examines whether data exist and whether the existing data are in a format that is readily usable [34]. For instance, Dentler et al [34] noted that pathology reports, although accessible, are recorded in a nonstructured, free-text format, making it challenging to readily use the data. Although structuredness may make data more available, Yoo et al [82] highlighted that structured data entry in the form of drop-down lists and check boxes tends to reduce the narrative description of patients' medical conditions. Although not explicitly investigating accessibility, Makeleni and Cilliers [31] also noted the challenges associated with structured data entry.

DQ Dimension: Accuracy

The accuracy dimension (123/227, 54.2%) is composed of 7 subthemes, namely correctness (42/123, 34.1%), validity (23/123, 18.7%), integrity (19/123, 15.4%), plausibility (17/123, 13.8%), accurate diagnostic data (13/123, 10.6%), conformance (7/123, 5.7%), and veracity (2/123, 1.6%). Accuracy refers to

the extent to which data reveal the truth about the event being described [31] and conform to their actual value [83].

Studies often referred to accuracy as the “*correctness*” of data, which is the degree to which data correctly communicate the parameter being represented [32]. By contrast, other studies focused on *plausibility*, which is the extent to which data points are believable [35]. Although accuracy concerns were present for all forms of digital health data, some studies focused specifically on *inaccuracies in diagnostic data* and stated that “the accurate and precise assignment of structured [diagnostic] data within EHRs is crucial” [84] and is “key to supporting secondary clinical data” [36].

To assess accuracy, the literature regularly asserts that data must be *validated* against metadata constraints, system assumptions, and local knowledge [19] and *conform* to structural and syntactical rules. According to Kahn et al [19] and Sirgo et al [85], conformance focuses on the compliance of data with internal or external formatting and relational or computational definitions. Accurate, verified, and validated data as well as data conforming to standards contribute to the *integrity* of the data. Integrity requires that the data stored in health information systems are accurate and consistent, where the “improper use of [health information systems] can jeopardise the integrity of a patient's information” [31]. An emerging subtheme of accuracy is the veracity of data, which represents uncertainty in the data owing to inconsistency, ambiguity, latency, deception, and model approximations [21]. It is particularly important in the context of the secondary use of big data, where “data veracity issues can arise from attempts to preserve privacy,...and is a function of how many sources contributed to the data” [86].

DQ Dimension: Completeness

The completeness dimension (114/227, 50.2%) is composed of 5 subthemes: missing data (66/114, 57.9%), level of completeness (25/114, 21.9%), representativeness (13/114, 11.4%), fragmentation (8/114, 7%), and breadth of documentation (2/114, 1.8%). A well-accepted definition of data completeness considers 4 perspectives: documentation (the presence of observations regarding a patient in data), breadth (the presence of all desired forms of data), density (the presence of a desired frequency of data values over time), and prediction (the presence of sufficient data to predict an outcome) [169]. Our analysis revealed that these 4 perspectives, although accepted, are rarely systematically examined in the extant literature; rather, papers tended to discuss completeness or the lack thereof as a whole.

Missing data is a prominent subtheme and represents a common problem in EHR data. For instance, Gloyd et al [87] argued that incomplete, missing, and implausible data “was by far the most common challenge encountered.” Scholars regularly identified that data fragmentation contributed to incompleteness, with a patient’s medical record deemed incomplete owing to data being required from multiple systems and EHRs [18,37,88-93]. “Data were also considered hidden within portals, outside systems, or multiple EHRs, frustrating efforts to assemble a complete clinical picture of the patient” [89]. Positive perspectives pertaining to data completeness focus on the level of completeness, with studies reporting relatively high completeness rates in health data sets [34,38,80,94,95,170]. For data to be considered complete, it needs to be captured at sufficient breadth and depth over time [12,18].

Some studies have proposed techniques for improving completeness, including developing fit-for-purpose user interfaces [68,96,97], standardizing documentation practices, [98,99], automating documentation [100], and performing quality control [99].

In some instances, the *level of completeness* and *extent of missing data* differed depending on the health status of the patient [15,16,18,20,39-43,86,90,101,170,171], which we classified into the subtheme of *representativeness*. It has been found that there is “a statistically significant relationship between EHR completeness and patient health status” [42], with more data recorded for patients who are sick than for patients with less-acute conditions. This strongly aligns with the subtheme of contextual validity.

DQ Dimension: Consistency

The consistency dimension (157/227, 69.2%) is composed of 10 subthemes: inconsistent data capturing (33/157, 21%), standardization (28/157, 17.8%), concordance (22/157, 14%), uniqueness (14/157, 8.9%), data variability (14/157, 8.9%), temporal variability (13/157, 8.3%), system differences (12/157, 7.6%), semantic consistency (10/157, 6.4%), structuredness (7/157, 4.5%), and representational consistency (4/157, 2.5%).

Inconsistent data capturing is a prevalent subtheme caused by the manual nature of data entry in health care settings [86], especially when data entry involve multiple times, teams, and goals [102]. Inconsistent data capturing results in *data variability*

and *temporal variability*. Data variability refers to inconsistencies in the data captured within and between health information systems, whereas temporal variability reflects inconsistencies that occur over time and may be because of changes in policies or medical guidelines [20,44-46,87,103-105]. *Semantic inconsistency* (ie, data with logical contradictions) and *representational inconsistency* (ie, data variations owing to multiple formats) can also result from inconsistent data capturing [47].

Standardization in terms of terminology, diagnostic codes, and workflows [99] are proffered to minimize inconsistency in data entry, yet in practice, there is a “lack of standardized data and terminology” [9] and “even with a set standard in place not all staff accept and follow the routine” [99]. The lack of standardization is further manifested because of health information system differences across settings [106]. As a result of the differences between systems, *concordance*—the extent of “agreement between elements in the EHR, or between the EHR and another data source”—is hampered [107].

Furthermore, inconsistent data entry can be caused by redundancy within the system because of structured versus unstructured data [108], which we label as the subtheme “*structuredness*,” and duplication across systems [39,48,104,109,172,173], which we label as the subtheme “*uniqueness*.” Although structured data entry “facilitates information retrieval” [33] and is “in a format that enables reliable extraction” [18], the presence of unstructured fields leads to data duplication efforts, hampering uniqueness, as data are recorded in multiple places with varying degrees of granularity and level of detail.

DQ Dimension: Contextual Validity

The contextual validity dimension (26/227, 11.5%) is composed of 4 subthemes: fitness for use (11/26, 42%), contextual DQ (9/26, 35%), granularity (4/26, 15%), and relevancy (2/26, 8%). Contextual validity requires a deep understanding of the context that gives rise to data [86], including technical, organizational, behavioral, and environmental factors [174].

Contextual DQ is often described as “fitness of use” [20], for which understanding the context in which data are collected is deemed important [18,90]. Another factor that contributes to data being fit for use is the *granularity* of data. Adequate *granularity* of time stamps [49], patient information [16], and data present in EHR (eg, diagnostic code [16]) was considered important to make data fit for use. Finally, for data to be fit for use, they must be *relevant*. As indicated by Schneeweiss and Glynn [41], for data to be meaningful, health care databases need to contain relevant information of sufficient quality, which can help answer specific questions. The literature clearly demonstrates the need to take context into consideration when analyzing data and the need to adapt technologies to the health care context so that appropriate data are collected for reliable analysis to be performed.

DQ Dimension: Currency

The currency dimension (18/227, 7.9%) is composed of a single subtheme: *timeliness*. Currency, or timeliness, is defined by Afshar et al [32] and Makeleni and Cilliers [31] as the degree

to which data represent reality from the required point in time. From an EHR perspective, data should be up to date, available, and reflect the profile of the patient at the time when the data are accessed [32,50]. Lee et al [35] extended this to include the recording of an event at the time when it occurs such that a value is deemed current if it is representative of the clinically relevant time of the event. Frequently mentioned causes for lack of currency of data include (1) recording of events (long) after the event actually occurred [91,99,110,111], (2) incomplete recording of patient characteristics over time [16], (3) system or interface design not matching workflow and impeding timely recording of data [99], (4) mixed-mode recording—paper and electronic [99], and (5) lack of time stamp metadata, meaning that the temporal sequence of events is not reflected in the recorded data [16].

Interrelationships Among the DQ Dimensions

As illustrated in Figure 4 and Multimedia Appendix 7 [16,34,40,42,78,80,90,91,109], interrelationships were found among the digital health DQ dimensions.

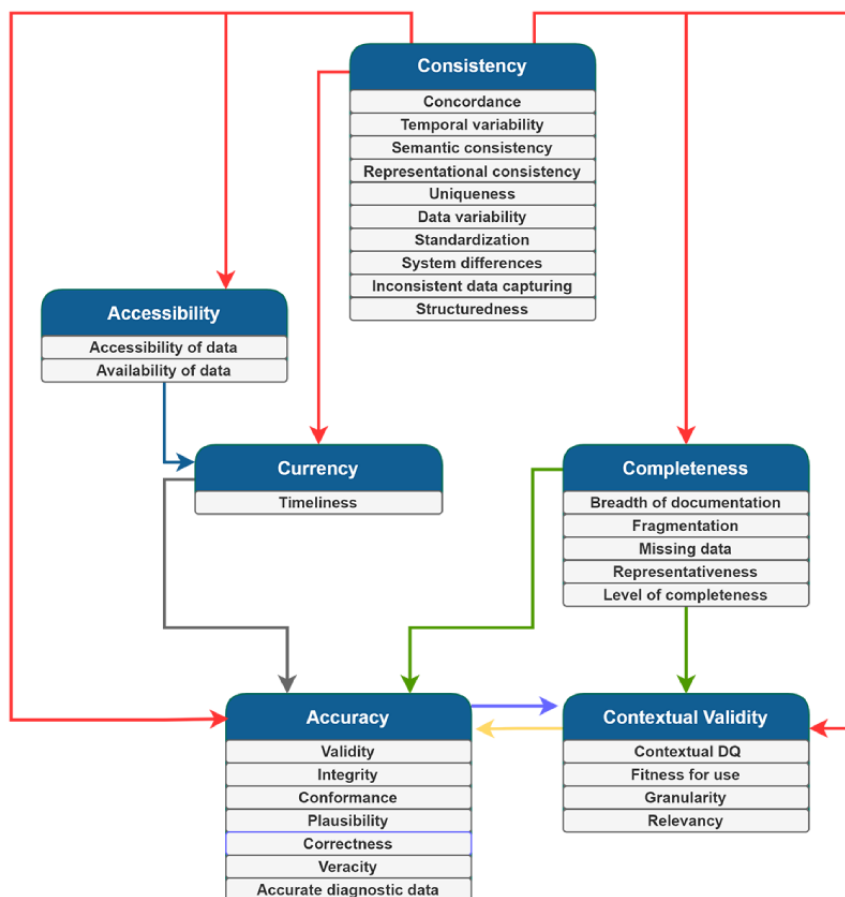
Consistency influenced all the DQ dimensions. Commonly, these relationships were expressed in terms of the presence of structured and consistent data entry, which prompts complete and accurate data to be entered into the health information system and provides more readily accessible and current data for health care professionals when treating patients. As Roukema et al [80] noted, “structured data entry applications can prompt for completeness, provide greater accuracy and better ordering for searching and retrieval, and permit validity checks for DQ monitoring, research, and especially decision support.” When data are entered inconsistently, it impedes the accuracy of the medical record and the contextual validity for secondary uses of the data [40].

Accessibility of data was found to influence the currency dimension of DQ. When data are not readily accessible, they seldom satisfy the timeliness of information for health care or research purposes [34]. Currency also influenced the accuracy of data. In a study investigating where DQ issues in EHR arise, it was found that “false negatives and false positives in the problem list sometimes arose when the problem list...[was] out-of-date, either because a resolved problem was not removed or because an active problem was not added” [90].

Furthermore, completeness influences the accuracy of data; as Makeleni and Cilliers [31] noted, “data should be complete to ensure it is accurate.” The presence of inaccurate data was regularly linked to information fragmentation [88], incomplete data entry [109], and omissions [35]. Completeness also influenced contextual validity, as it is necessary to have all the data available to complete specific tasks [78]. When it comes to the secondary use of EHR data, evaluation of “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” [42].

Accuracy and contextual validity exhibited a bidirectional relationship with each other. The literature suggests that accuracy influences contextual validity; however, data cannot simply be extracted from structured form fields, and free-text fields will also need to be consulted. For instance, Kim and Kim [112] identified “it is sometimes thought that structured data are more completely optimized for clinical research. However, this is not always the case, particularly given that extracted EMR data can still be unstable and contain serious errors.” By contrast, other studies suggest that when only a segment of information regarding a specific clinical event (ie, contextual validity) is captured, inaccuracy can ensue [16].

Figure 4. Interrelationships between the data quality (DQ) dimensions.



Outcomes of Digital Health DQ

The analysis of the literature identified five types of digital health DQ outcomes: (1) clinical, (2) business process, (3) clinician, (4) research-related, and (5) organizational outcomes (Multimedia Appendix 8 [15,16,20,31,33,39,40,42,51,52,55,57, 58,61,63,64,84,90,105,113,166,175-178]). Using NVivo’s built-in cross-tab query coupled with subject matter expert analysis, it was identified that different DQ dimensions were related to DQ outcomes in different ways (Table 2). Currency was the only dimension that did not have a direct effect on DQ outcomes. However, as shown in Figure 5, it is plausible that currency affects DQ outcomes by impacting other DQ dimensions. In the subsequent paragraphs, we discuss each DQ dimension and its respective outcomes.

We identified that the accessibility DQ dimension influenced clinical, clinician, business process, research-related, and organizational outcomes. In terms of *clinical outcomes*, Roukema et al [80] indicated that EHRs substantially enhance the quality of patient care by improving the accessibility and legibility of health care data. The increased accessibility of medical records during the delivery of patient care is further proffered to benefit *clinicians* by reducing the data entry burden [33]. By contrast, inconsistency in the availability of data across health settings increases clinician workload; as Wiebe et al [15] noted, “given the predominantly electronic form of communication between hospitals and general practitioners in Alberta, the inconsistency in availability of documentation in one single location can delay processes for practitioners

searching for important health information.” When data are accessible and available, they can improve *business processes* (eg, quality assurance) and *research-related* (eg, outcome-oriented research) *outcomes* and can support *organizational outcomes* with improved billing and financial management [179].

The literature demonstrates that data accuracy influences *clinical outcomes* [14,39,51] and *research-related outcomes* [14,113]; as Wang et al [14] described, “errors in healthcare data are numerous and impact secondary data use and potentially patient care and safety.” Downey et al [39] observed the negative impact on quality of care (ie, *clinical outcomes*) resulting from incorrect data and stated, “manual data entry remains a primary mechanism for acquiring data in EHRs, and if the data is incorrect then the impact to patients and patient care could be significant” [39]. Poor data accuracy also diminishes the quality of *research outcomes*. Precise data are beneficial in producing high-quality research outcomes. As Gibby [113] explained, “computerized clinical information systems have considerable advantages over paper recording of data, which should increase the likelihood of their use in outcomes research. Manual records are often inaccurate, biased, incomplete, and illegible.” Closely related to accuracy, contextual validity is an important DQ dimension that considers the fitness for *research*; as stated by Weiskopf et al [42], “[w]hen repurposed for secondary use, however, the concept of ‘fitness for use’ can be applied.”

The consistency DQ dimension was related to all DQ outcomes. It was commonly reported that inconsistency in data negatively

impacts the *reusability* of EHR data for research purposes, hindering *research-related outcomes* and negatively impacting *business processes* and *organizational outcomes*. For example, Kim et al [114] acknowledged that inconsistent data labeling in EHR systems may hinder accurate research results, noting that “a system may use local terminology that allows unmanaged synonyms and abbreviations...If local data are not mapped to terminologies, performing multicentre research would require extensive labour.” Alternatively, von Lucadou et al [16] indicated the impact of inconsistency on *clinical outcomes*, reporting that the existence of inconsistencies in captured data “could explain the varying number of diagnoses throughout the encounter history of some subjects,” whereas Diaz-Garelli et al [84] demonstrated the negative impact that inconsistency has on *clinicians* in terms of increased workload.

interprofessional communication), *research-related outcomes* (eg, research facilitation), and *organizational outcomes* (eg, key performance indicators related to readmissions) [15]. For example, while reviewing the charts of 3011 nonobstetric inpatients, Wiebe et al [15] found that missing discharge summary within an EHR “can present several issues for healthcare processes, including hindered communication between hospitals and general practitioners, heightened risk of readmissions, and poor usability of coded health data,” among other widespread implications. Furthermore, Liu et al [69] reported that “having incomplete data on patients’ records has posed the greatest threat to patient care.” Owing to the heterogeneous nature (with multiple data points) of EHR data, Richesson et al [20] emphasized that access to large, complete data will allow clinical investigators “to detect smaller clinical effects, identify and study rare disorders, and produce robust, generalisable results.”

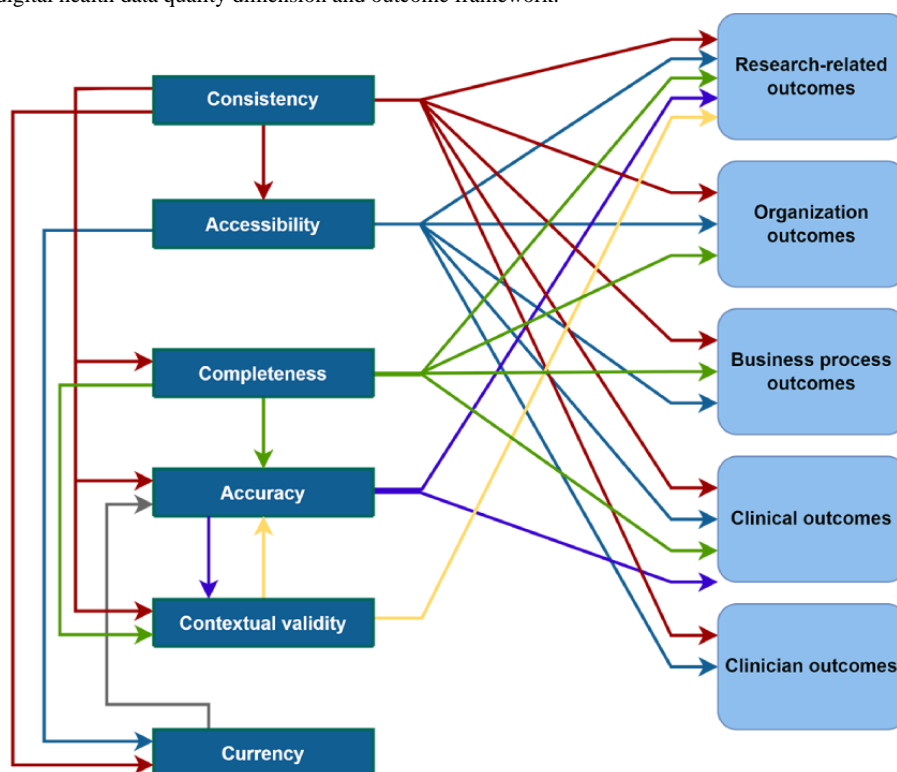
Incomplete EMR data were found to impact *clinical outcomes* (eg, reduced quality of care), *business process outcomes* (eg,

Table 2. The relationships between *data quality (DQ)* dimensions and data outcomes.

DQ dimension	Outcomes ^a				
	Research	Organizational	Business process	Clinical	Clinician
Accessibility	✓	✓	✓	✓	✓
Accuracy	✓			✓	
Completeness	✓	✓	✓	✓	
Consistency	✓	✓	✓	✓	✓
Contextual validity	✓				
Currency					

^aThe checkmark symbol indicates that the relationship between the DQ dimension and the outcome is reported in the literature. Blank cells indicate that there is no evidence to support the relationship.

Figure 5. Consolidated digital health data quality dimension and outcome framework.



Discussion

Overview

The following sections describe the three main findings of this research: the (1) dimensions of DQ, (2) interrelationships among the dimensions of DQ, and (3) outcomes of DQ. As described in the *Summary of Key Findings* section, these 3 findings led to the development of the DQ-DO framework. Subsequently, we compared the DQ-DO framework with related works. This leads to implications for future research. The *Discussion* section concludes with a reflection on the limitations of this study.

Summary of Key Findings

In summary, we unearthed 3 core findings. First, we identified 6 dimensions of DQ within the digital health domain: consistency, accessibility, completeness, accuracy, contextual validity, and currency. These dimensions were synthesized from 30 subthemes described in the literature. We found that consistency, completeness, and accuracy are the predominant dimensions of DQ. Comparatively, limited attention has been paid to the dimensions of accessibility, currency, and contextual validity. Second, we identified the interrelationships among these 6 dimensions of digital health DQ (Table 2). The literature indicates that the consistency dimension can influence all other DQ dimensions. The accessibility of data was found to influence the currency of data. Completeness impacts accuracy and contextual validity, with these dimensions serving as dependent variables and exhibiting a bidirectional relationship with each other. Third, we identified 5 types of data outcomes (Table 2; Multimedia Appendix 8): research-related, organizational, business process, clinical, and clinician outcomes. Consistency was found to be a very influential dimension, impacting all types of DQ outcomes. By contrast, contextual validity was shown to be particularly important for data reuse (eg, performance measurement and outcome-oriented research). Although currency does not directly impact any outcomes, it impacts the accuracy of data, which impacts clinical and research-related outcomes. Therefore, if currency issues are not resolved, accuracy issues would still prevail. Consistency, accessibility, and completeness were shown to be important considerations for achieving the goal of improving organizational outcomes. Through consolidating our 3 core findings, we developed a consolidated DQ-DO framework (Figure 5).

Comparison With Literature

Our findings extend those of previous studies on digital health DQ in 3 ways. First, through our rigorous approach, we identified a comprehensive set of DQ dimensions, which both confirmed and extended the existing literature. For instance, Weiskopf and Weng [17] identified 5 DQ dimensions, namely completeness, correctness, concordance, plausibility, and currency, all of which are present within our DQ framework, although in some instances, we use slightly different terms (referring to correctness as accuracy and concordance as consistency). Extending the framework of Weiskopf and Weng [17], we view plausibility as a subtheme of accuracy and disentangle accessibility from completeness, and we also stress the importance of contextual validity per Richesson et al [20].

Others have commonly had a narrower perspective of DQ, focusing on completeness, correctness, and currency [18] or on completeness, timeliness, and accuracy [13]. In other domains of digital health, such as physician rating systems, Wang and Strong's [180] DQ dimensions of intrinsic, contextual, representational, and accessibility have been adopted. Such approaches to assessing DQ are appropriate, although they remove a level of granularity that is necessary to understand relationships and outcomes. This is particularly necessary given the salience of consistency in our data set and the important role it plays in generating outcomes.

Second, unlike previous studies on DQ dimensions, we also demonstrate how these dimensions are all related to each other. By analyzing the interrelationships between these DQ dimensions, we can determine how a particular dimension influences another and in which direction this relationship is unfolding. This is an important implication for digital health practitioners, as although several studies have examined how to validate [38] and resolve DQ issues [16], resolving issues with a specific DQ dimension requires awareness of the interrelated DQ dimensions. For instance, to improve accuracy, one also needs to consider improving consistency and completeness.

Third, although previous studies describe how DQ can impact a particular outcome (eg, the studies by Weiskopf et al [18], Johnson et al [52], and Dantanarayana and Sahama [115]), they largely focus broadly on DQ, a specific dimension of DQ, or a specific outcome. For instance, Sung et al [181] noted that poor-quality data were a prominent barrier hindering the adoption of digital health systems. By contrast, Kohane et al [182] focused on research-related outcomes in terms of publication potential and identified that incompleteness and inconsistency can serve as core impediments. To summarize, the DQ-DO framework (Figure 5) developed through this review provides not only the dimensions and the outcomes but also the interrelationships between these dimensions and how they influence outcomes.

Implications for Future Work

Implication 1: Equal Consideration Across DQ Dimensions

This study highlights the importance of each of the 6 DQ dimensions: consistency, accessibility, completeness, accuracy, contextual validity, and currency. These dimensions have received varying amounts of attention in the literature. Although we observe that some DQ dimensions such as accessibility, contextual validity, and currency are discussed less frequently than others, it does not mean that these dimensions are not important for assessment. This is evident in Figure 5, which shows that all DQ dimensions except for currency directly influence DQ outcomes. Although we did not identify a direct relationship between the currency of data and the 6 types of data outcomes, it is likely that the currency of data influences the accuracy of data, which subsequently influences the research-related and clinical outcomes. Future research, including consultation with a range of stakeholders, needs to further delve into understanding the underresearched DQ

dimensions. For instance, both the currency and accessibility of data are less frequently discussed dimensions in the literature; however, with the advances in digital health technologies, both have become highly relevant for real-time clinical decisions [21,53].

Implication 2: Empirical Investigations of the Impact of the DQ Dimensions

The DQ-DO framework identified in this study has been developed through a rigorous systematic literature review process, which synthesized the literature related to digital health DQ. To extend this study, we advocate for empirical mixed methods case studies to validate the framework, including an examination of the interrelationships between DQ dimensions and DQ outcomes, based on real-life data and consultation with a variety of stakeholders. Existing approaches can be used to identify the presence of issues related to DQ dimensions within digital health system logs [38,183]. The DQ outcomes could be assessed by extracting prerecorded key performance indicators from case hospitals and be triangulated with interview data to capture patients', clinicians', and hospitals' perspectives of the impacts of DQ [184]. This could then be incorporated into a longitudinal study in which data collection is performed before and after a DQ improvement intervention, which would provide efficacy to the digital health DQ intervention.

Implication 3: Understanding the Root Causes of DQ Challenges

Although this study provides a first step toward a more comprehensive understanding of DQ dimensions for digital health data and their influences on outcomes, it does not explore the potential causes of such DQ challenges. Without understanding the reasons behind these DQ issues, the true potential of evidence-based health care decision-making remains unfulfilled. Future research should examine the root causes of DQ challenges in health care data to prevent such challenges from occurring in the first place. A framework that may prove useful in illuminating the root causes of DQ issues is the Odigos framework, which indicates that DQ issues emanate from the social world (ie, macro and situational structures, roles, and norms), material world (eg, quality of the EHR system and technological infrastructure), and personal world (eg, characteristics and behaviors of health care professionals) [183]. These insights could then be incorporated into a data governance roadmap for digital hospitals.

Implication 4: Systematic Assessment and Remedy of DQ Issues

Although prevention remains better than the cure (refer to implication 3), not all DQ errors can be prevented or mitigated. It is common for many health care organizations to dedicate resources to data cleaning to obtain high-quality data in a timely manner, and this will remain necessary (although hopefully to a lesser degree). Some studies (eg, Weiskopf et al [18]) advocate evidence-based guidelines and frameworks for a detailed

assessment of the quality of digital health data. However, few studies have focused on a systematic and automated method of assessing and remedying common DQ issues. Future research should also focus on evidence-based guidelines, best practices, and automated means to assess and remedy digital health data.

Limitations

This review is scoped to studying digital health data generated within a hospital setting and not those generated within other health care settings. This is necessary because of the vast differences between acute health care settings and primary care. Future research should seek to investigate the digital health data of primary care settings to identify the DQ dimensions and outcomes relevant to these settings. In addition, this literature review has been scoped to peer-reviewed outlets, with "grey" literature excluded, which could have led to publication bias. Although this scoping may have resulted in the exclusion of some relevant articles, it was necessary to ensure the quality behind the development of the digital health DQ framework. An additional limitation that may be raised by our method is that because of the sheer number of articles returned by our search, we did not perform double coding (where 2 independent researchers analyze the same article). To mitigate this limitation, steps were taken to minimize bias by conducting coder corroboration sessions and group validation, as mentioned in the *Methods* section, with the objective of improving internal and external reliability [66]. To further improve internal reliability, 2 experienced researchers verified the entirety of the analysis in NVivo and to improve external reliability, card-sorting assessments were performed with DQ experts, and the findings were presented and confirmed by 3 digital health care professionals. Furthermore, empirical validation of the framework is required, both in terms of real-life data and inputs from a range of experts.

Conclusions

The multidisciplinary systematic literature review conducted in this study resulted in the development of a consolidated digital health DQ framework comprising 6 DQ dimensions, the interrelationships among these dimensions, 6 DQ outcomes, and the relationships between these dimensions and outcomes. We identified four core implications to motivate future research: specifically, researchers should (1) pay equal consideration to all dimensions of DQ, as the dimensions can both directly and indirectly influence DQ outcomes; (2) seek to empirically assess the DQ-DO framework using a mixed methods case study design; (3) identify the root causes of the digital health DQ issues; and (4) develop interventions for mitigating DQ issues or preventing them from arising. The DQ-DO framework provides health care executives (eg, chief information officers and chief clinical informatics officers) with insights into DQ issues and which digital health-related outcomes they have an impact on, and this can help them prioritize tackling DQ-related problems.

Acknowledgments

The authors acknowledge the support provided by the Centre of Data Science, Queensland University of Technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of key terms.

[\[DOCX File , 17 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Verification of search strategy.

[\[DOCX File , 27 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Data coding structures.

[\[DOCX File , 202 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Publication outlets.

[\[DOCX File , 18 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Data quality definitions.

[\[DOCX File , 16 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

Evidence of the subtheme for each data quality dimension.

[\[DOCX File , 19 KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

Evidence for the interrelationships among the dimensions of data quality.

[\[DOCX File , 17 KB-Multimedia Appendix 7\]](#)

Multimedia Appendix 8

Evidence for the outcomes of data quality.

[\[DOCX File , 19 KB-Multimedia Appendix 8\]](#)

References

1. Duncan R, Eden R, Woods L, Wong I, Sullivan C. Synthesizing dimensions of digital maturity in hospitals: systematic review. *J Med Internet Res* 2022 Mar 30;24(3):e32994 [FREE Full text] [doi: [10.2196/32994](https://doi.org/10.2196/32994)] [Medline: [35353050](https://pubmed.ncbi.nlm.nih.gov/35353050/)]
2. Eden R, Burton-Jones A, Scott I, Staib A, Sullivan C. Effects of eHealth on hospital practice: synthesis of the current literature. *Aust Health Rev* 2018 Sep;42(5):568-578. [doi: [10.1071/AH17255](https://doi.org/10.1071/AH17255)] [Medline: [29986809](https://pubmed.ncbi.nlm.nih.gov/29986809/)]
3. Rahimi K. Digital health and the elusive quest for cost savings. *Lancet Digit Health* 2019 Jul;1(3):e108-e109 [FREE Full text] [doi: [10.1016/S2589-7500\(19\)30056-1](https://doi.org/10.1016/S2589-7500(19)30056-1)] [Medline: [33323258](https://pubmed.ncbi.nlm.nih.gov/33323258/)]
4. Eden R, Burton-Jones A, Scott I, Staib A, Sullivan C. The impacts of eHealth upon hospital practice: synthesis of the current literature. Deeble Institute for Health Policy Research. 2017 Oct 17. URL: https://ahha.asn.au/system/files/docs/publications/impacts_of_ehealth_2017.pdf [accessed 2023-03-01]
5. Zheng K, Abraham J, Novak LL, Reynolds TL, Gettinger A. A survey of the literature on unintended consequences associated with health information technology: 2014-2015. *Yearb Med Inform* 2016 Nov 10(1):13-29 [FREE Full text] [doi: [10.15265/IY-2016-036](https://doi.org/10.15265/IY-2016-036)] [Medline: [27830227](https://pubmed.ncbi.nlm.nih.gov/27830227/)]
6. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. *P T* 2017 Sep;42(9):572-575 [FREE Full text] [Medline: [28890644](https://pubmed.ncbi.nlm.nih.gov/28890644/)]
7. Global Strategy of Digital Health 2020-2025. World Health Organization. 2021. URL: <https://www.who.int/docs/default-source/documents/g4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf> [accessed 2021-10-15]

8. Darko-Yawson S, Ellingsen G. Assessing and improving EHRs data quality through a socio-technical approach. *Procedia Comput Sci* 2016;98:243-250. [doi: [10.1016/j.procs.2016.09.039](https://doi.org/10.1016/j.procs.2016.09.039)]
9. Afzal M, Hussain M, Ali Khan W, Ali T, Lee S, Huh EN, et al. Comprehensible knowledge model creation for cancer treatment decision making. *Comput Biol Med* 2017 Mar 01;82:119-129. [doi: [10.1016/j.combiomed.2017.01.010](https://doi.org/10.1016/j.combiomed.2017.01.010)] [Medline: [28187294](https://pubmed.ncbi.nlm.nih.gov/28187294/)]
10. Assale M, Dui LG, Cina A, Seveso A, Cabitza F. The revival of the notes field: leveraging the unstructured content in electronic health records. *Front Med (Lausanne)* 2019 Apr 17;6:66 [FREE Full text] [doi: [10.3389/fmed.2019.00066](https://doi.org/10.3389/fmed.2019.00066)] [Medline: [31058150](https://pubmed.ncbi.nlm.nih.gov/31058150/)]
11. Savitz ST, Savitz LA, Fleming NS, Shah ND, Go AS. How much can we trust electronic health record data? *Healthc (Amst)* 2020 Sep;8(3):100444. [doi: [10.1016/j.hjdsi.2020.100444](https://doi.org/10.1016/j.hjdsi.2020.100444)] [Medline: [32919583](https://pubmed.ncbi.nlm.nih.gov/32919583/)]
12. Beauchemin M, Weng C, Sung L, Pichon A, Abbott M, Hershman DL, et al. Data quality of chemotherapy-induced nausea and vomiting documentation. *Appl Clin Inform* 2021 Mar;12(2):320-328 [FREE Full text] [doi: [10.1055/s-0041-1728698](https://doi.org/10.1055/s-0041-1728698)] [Medline: [33882585](https://pubmed.ncbi.nlm.nih.gov/33882585/)]
13. Puttkammer N, Baseman JG, Devine EB, Valles JS, Hyppolite N, Garilus F, et al. An assessment of data quality in a multi-site electronic medical record system in Haiti. *Int J Med Inform* 2016 Feb;86:104-116. [doi: [10.1016/j.ijmedinf.2015.11.003](https://doi.org/10.1016/j.ijmedinf.2015.11.003)] [Medline: [26620698](https://pubmed.ncbi.nlm.nih.gov/26620698/)]
14. Wang Z, Penning M, Zozus M. Analysis of anesthesia screens for rule-based data quality assessment opportunities. *Stud Health Technol Inform* 2019;257:473-478 [FREE Full text] [Medline: [30741242](https://pubmed.ncbi.nlm.nih.gov/30741242/)]
15. Wiebe N, Xu Y, Shaheen AA, Eastwood C, Boussat B, Quan H. Indicators of missing electronic medical record (EMR) discharge summaries: a retrospective study on Canadian data. *Int J Popul Data Sci* 2020 Dec 11;5(1):1352 [FREE Full text] [doi: [10.23889/ijpds.v5i3.1352](https://doi.org/10.23889/ijpds.v5i3.1352)] [Medline: [34007880](https://pubmed.ncbi.nlm.nih.gov/34007880/)]
16. von Lucadou M, Ganslandt T, Prokosch HU, Toddenroth D. Feasibility analysis of conducting observational studies with the electronic health record. *BMC Med Inform Decis Mak* 2019 Oct 28;19(1):202 [FREE Full text] [doi: [10.1186/s12911-019-0939-0](https://doi.org/10.1186/s12911-019-0939-0)] [Medline: [31660955](https://pubmed.ncbi.nlm.nih.gov/31660955/)]
17. 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 Jan 01;20(1):144-151 [FREE Full text] [doi: [10.1136/amiajnl-2011-000681](https://doi.org/10.1136/amiajnl-2011-000681)] [Medline: [22733976](https://pubmed.ncbi.nlm.nih.gov/22733976/)]
18. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A data quality assessment guideline for electronic health record data reuse. *EGEMS (Wash DC)* 2017 Sep 04;5(1):14 [FREE Full text] [doi: [10.5334/egems.218](https://doi.org/10.5334/egems.218)] [Medline: [29881734](https://pubmed.ncbi.nlm.nih.gov/29881734/)]
19. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)* 2016 Sep 11;4(1):1244 [FREE Full text] [doi: [10.13063/2327-9214.1244](https://doi.org/10.13063/2327-9214.1244)] [Medline: [27713905](https://pubmed.ncbi.nlm.nih.gov/27713905/)]
20. Richesson RL, Hammond WE, Nahm M, Wixted D, Simon GE, Robinson JG, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J Am Med Inform Assoc* 2013 Dec;20(e2):e226-e231 [FREE Full text] [doi: [10.1136/amiajnl-2013-001926](https://doi.org/10.1136/amiajnl-2013-001926)] [Medline: [23956018](https://pubmed.ncbi.nlm.nih.gov/23956018/)]
21. Reimer AP, Madigan EA. Veracity in big data: how good is good enough. *Health Informatics J* 2019 Dec;25(4):1290-1298 [FREE Full text] [doi: [10.1177/1460458217744369](https://doi.org/10.1177/1460458217744369)] [Medline: [29388495](https://pubmed.ncbi.nlm.nih.gov/29388495/)]
22. Bettencourt-Silva JH, Clark J, Cooper CS, Mills R, Rayward-Smith VJ, de la Iglesia B. Building data-driven pathways from routinely collected hospital data: a case study on prostate cancer. *JMIR Med Inform* 2015 Jul 10;3(3):e26 [FREE Full text] [doi: [10.2196/medinform.4221](https://doi.org/10.2196/medinform.4221)] [Medline: [26162314](https://pubmed.ncbi.nlm.nih.gov/26162314/)]
23. Britt H, Pollock AJ, Wong C. Can Medicare sustain the health of our ageing population. *The Conversation*. 2015 Nov 3. URL: <https://theconversation.com/can-medicare-sustain-the-health-of-our-ageing-population-49579> [accessed 2023-03-01]
24. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med* 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: [10.1371/journal.pmed.1000387](https://doi.org/10.1371/journal.pmed.1000387)] [Medline: [21267058](https://pubmed.ncbi.nlm.nih.gov/21267058/)]
25. Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. *MIS Q* 2002 Jun;26(2):xiii-xxiii. [doi: [10.2307/4132319](https://doi.org/10.2307/4132319)]
26. Templier M, Paré G. A framework for guiding and evaluating literature reviews. *Commun Assoc Inf Syst* 2015;37(1):6. [doi: [10.17705/1cais.03706](https://doi.org/10.17705/1cais.03706)]
27. Saldaña J. *The Coding Manual For Qualitative Researchers*. 3rd edition. Thousand Oaks, CA, USA: Sage Publications; 2015.
28. Dubois A, Gadde LE. “Systematic combining”—a decade later. *J Bus Res* 2014 Jun;67(6):1277-1284. [doi: [10.1016/j.jbusres.2013.03.036](https://doi.org/10.1016/j.jbusres.2013.03.036)]
29. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research: notes on the Gioia methodology. *Organ Res Methods* 2013 Jan;16(1):15-31. [doi: [10.1177/1094428112452151](https://doi.org/10.1177/1094428112452151)]
30. Best P, Badham J, McConnell T, Hunter RF. Participatory theme elicitation: open card sorting for user led qualitative data analysis. *Int J Soc Res Methodol* 2022;25(2):213-231. [doi: [10.1080/13645579.2021.1876616](https://doi.org/10.1080/13645579.2021.1876616)]
31. Makeleni N, Cilliers L. Critical success factors to improve data quality of electronic medical records in public healthcare institutions. *S Afr J Inf Manag* 2021 Mar 03;23(1):a1230. [doi: [10.4102/sajim.v23i1.1230](https://doi.org/10.4102/sajim.v23i1.1230)]

32. Afshar AS, Li Y, Chen Z, Chen Y, Lee JH, Irani D, et al. An exploratory data quality analysis of time series physiologic signals using a large-scale intensive care unit database. *JAMIA Open* 2021 Jul;4(3):ooab057 [FREE Full text] [doi: [10.1093/jamiaopen/ooab057](https://doi.org/10.1093/jamiaopen/ooab057)] [Medline: [34350392](https://pubmed.ncbi.nlm.nih.gov/34350392/)]
33. Top M, Yilmaz A, Gider Ö. Electronic medical records (EMR) and nurses in Turkish hospitals. *Syst Pract Action Res* 2013;26(3):281-297. [doi: [10.1007/s11213-012-9251-y](https://doi.org/10.1007/s11213-012-9251-y)]
34. Dentler K, Cornet R, ten Teije A, Tanis P, Klinkenbijl J, Tytgat K, et al. Influence of data quality on computed Dutch hospital quality indicators: a case study in colorectal cancer surgery. *BMC Med Inform Decis Mak* 2014 Apr 11;14:32 [FREE Full text] [doi: [10.1186/1472-6947-14-32](https://doi.org/10.1186/1472-6947-14-32)] [Medline: [24721489](https://pubmed.ncbi.nlm.nih.gov/24721489/)]
35. Lee K, Weiskopf N, Pathak J. A framework for data quality assessment in clinical research datasets. *AMIA Annu Symp Proc* 2018 Apr 16;2017:1080-1089 [FREE Full text] [Medline: [29854176](https://pubmed.ncbi.nlm.nih.gov/29854176/)]
36. Diaz-Garelli F, Strowd R, Ahmed T, Lycan Jr TW, Daley S, Wells BJ, et al. What oncologists want: identifying challenges and preferences on diagnosis data entry to reduce EHR-induced burden and improve clinical data quality. *JCO Clin Cancer Inform* 2021 May;5:527-540 [FREE Full text] [doi: [10.1200/CCI.20.00174](https://doi.org/10.1200/CCI.20.00174)] [Medline: [33989015](https://pubmed.ncbi.nlm.nih.gov/33989015/)]
37. Rule A, Rick S, Chiu M, Rios P, Ashfaq S, Calvitti A, et al. Validating free-text order entry for a note-centric EHR. *AMIA Annu Symp Proc* 2015 Nov 5;2015:1103-1110 [FREE Full text] [Medline: [26958249](https://pubmed.ncbi.nlm.nih.gov/26958249/)]
38. van Hoeven LR, de Bruijne MC, Kemper PF, Koopman MM, Rondeel JM, Leyte A, et al. Validation of multisource electronic health record data: an application to blood transfusion data. *BMC Med Inform Decis Mak* 2017 Jul 14;17(1):107 [FREE Full text] [doi: [10.1186/s12911-017-0504-7](https://doi.org/10.1186/s12911-017-0504-7)] [Medline: [28709453](https://pubmed.ncbi.nlm.nih.gov/28709453/)]
39. Downey S, Indulska M, Sadiq S. Perceptions and challenges of EHR clinical data quality. In: *Proceedings of the 30th Australasian Conference on Information Systems*. 2019 Presented at: ACIS '19; December 9-11, 2019; Perth, Australia p. 233-243.
40. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. *Summit Transl Bioinform* 2010 Mar 01;2010:1-5 [FREE Full text] [Medline: [21347133](https://pubmed.ncbi.nlm.nih.gov/21347133/)]
41. Schneeweiss S, Glynn RJ. Real-world data analytics fit for regulatory decision-making. *Am J Law Med* 2018 May;44(2-3):197-217. [doi: [10.1177/0098858818789429](https://doi.org/10.1177/0098858818789429)] [Medline: [30106649](https://pubmed.ncbi.nlm.nih.gov/30106649/)]
42. Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. *AMIA Annu Symp Proc* 2013 Nov 16;2013:1472-1477 [FREE Full text] [Medline: [24551421](https://pubmed.ncbi.nlm.nih.gov/24551421/)]
43. Lingren T, Sadhasivam S, Zhang X, Marsolo K. Electronic medical records as a replacement for prospective research data collection in postoperative pain and opioid response studies. *Int J Med Inform* 2018 Mar;111:45-50 [FREE Full text] [doi: [10.1016/j.ijmedinf.2017.12.014](https://doi.org/10.1016/j.ijmedinf.2017.12.014)] [Medline: [29425633](https://pubmed.ncbi.nlm.nih.gov/29425633/)]
44. Baker K, Dunwoodie E, Jones RG, Newsham A, Johnson O, Price CP, et al. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. *Int J Med Inform* 2017 Jul;103:32-41 [FREE Full text] [doi: [10.1016/j.ijmedinf.2017.03.011](https://doi.org/10.1016/j.ijmedinf.2017.03.011)] [Medline: [28550999](https://pubmed.ncbi.nlm.nih.gov/28550999/)]
45. Callahan A, Shah NH, Chen JH. Research and reporting considerations for observational studies using electronic health record data. *Ann Intern Med* 2020 Jun 02;172(11 Suppl):S79-S84 [FREE Full text] [doi: [10.7326/M19-0873](https://doi.org/10.7326/M19-0873)] [Medline: [32479175](https://pubmed.ncbi.nlm.nih.gov/32479175/)]
46. Deng X, Lin WH, E-Shyong T, Eric KY, Salloway MK, Seng TC. From descriptive to diagnostic analytics for assessing data quality: an application to temporal data elements in electronic health records. In: *Proceedings of the 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics*. 2016 Presented at: BHI '16; February 24-27, 2016; Las Vegas, NV, USA p. 236-239. [doi: [10.1109/bhi.2016.7455878](https://doi.org/10.1109/bhi.2016.7455878)]
47. Jetley G, Zhang H. Electronic health records in IS research: quality issues, essential thresholds and remedial actions. *Decis Support Syst* 2019 Nov;126:113137. [doi: [10.1016/j.dss.2019.113137](https://doi.org/10.1016/j.dss.2019.113137)]
48. El Fadly A, Lucas N, Rance B, Verplancke P, Lastic PY, Daniel C. The REUSE project: EHR as single datasource for biomedical research. *Stud Health Technol Inform* 2010;160(Pt 2):1324-1328. [Medline: [20841899](https://pubmed.ncbi.nlm.nih.gov/20841899/)]
49. Suriadi S, Mans RS, Wynn MT, Partington A, Karnon J. Measuring patient flow variations: a cross-organisational process mining approach. In: *Proceedings of the 2nd Asia Pacific Conference on Business Process Management*. 2014 Presented at: AP-BPM '14; July 3-4, 2014; Brisbane, Australia p. 43-58. [doi: [10.1007/978-3-319-08222-6_4](https://doi.org/10.1007/978-3-319-08222-6_4)]
50. Chiasera A, Toai TJ, Bogoni LP, Armellini G, Jara JJ. Federated EHR: how to improve data quality maintaining privacy. In: *Proceedings of the 2011 IST-Africa Conference*. 2011 Presented at: IST-Africa '11; May 11-13, 2011; Gaborone, Botswana p. 1-8.
51. Koepke R, Petit AB, Ayele RA, Eickhoff JC, Schauer SL, Verdon MJ, et al. Completeness and accuracy of the Wisconsin immunization registry: an evaluation coinciding with the beginning of meaningful use. *J Public Health Manag Pract* 2015;21(3):273-281. [doi: [10.1097/PHH.0000000000000216](https://doi.org/10.1097/PHH.0000000000000216)] [Medline: [25590511](https://pubmed.ncbi.nlm.nih.gov/25590511/)]
52. Johnson SG, Pruinelli L, Hoff A, Kumar V, Simon GJ, Steinbach M, et al. A framework for visualizing data quality for predictive models and clinical quality measures. *AMIA Jt Summits Transl Sci Proc* 2019 May 6;2019:630-638 [FREE Full text] [Medline: [31259018](https://pubmed.ncbi.nlm.nih.gov/31259018/)]
53. Top M, Yilmaz A, Karabulut E, Otieno OG, Saylam M, Bakır S, et al. Validation of a nurses' views on electronic medical record systems (EMR) questionnaire in Turkish health system. *J Med Syst* 2015 Jun;39(6):67. [doi: [10.1007/s10916-015-0250-2](https://doi.org/10.1007/s10916-015-0250-2)] [Medline: [25957164](https://pubmed.ncbi.nlm.nih.gov/25957164/)]

54. Wang Z, Talburt JR, Wu N, Dagtas S, Zozus MN. A rule-based data quality assessment system for electronic health record data. *Appl Clin Inform* 2020 Aug;11(4):622-634 [FREE Full text] [doi: [10.1055/s-0040-1715567](https://doi.org/10.1055/s-0040-1715567)] [Medline: [32968999](https://pubmed.ncbi.nlm.nih.gov/32968999/)]
55. Toftdahl AK, Pape-Haugaard L, Palsson T, Villumsen M. Collect once - use many times: the research potential of low back pain patients' municipal electronic healthcare records. *Stud Health Technol Inform* 2018;247:211-215. [Medline: [29677953](https://pubmed.ncbi.nlm.nih.gov/29677953/)]
56. Diaz-Garelli F, Strowd R, Lawson VL, Mayorga ME, Wells BJ, Lycan Jr TW, et al. Workflow differences affect data accuracy in oncologic EHRs: a first step toward detangling the diagnosis data babel. *JCO Clin Cancer Inform* 2020 Jun;4:529-538 [FREE Full text] [doi: [10.1200/CCI.19.00114](https://doi.org/10.1200/CCI.19.00114)] [Medline: [32543899](https://pubmed.ncbi.nlm.nih.gov/32543899/)]
57. Estiri H, Klann JG, Murphy SN. A clustering approach for detecting implausible observation values in electronic health records data. *BMC Med Inform Decis Mak* 2019 Jul 23;19(1):142 [FREE Full text] [doi: [10.1186/s12911-019-0852-6](https://doi.org/10.1186/s12911-019-0852-6)] [Medline: [31337390](https://pubmed.ncbi.nlm.nih.gov/31337390/)]
58. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A data quality ontology for the secondary use of EHR data. *AMIA Annu Symp Proc* 2015 Nov 5;2015:1937-1946 [FREE Full text] [Medline: [26958293](https://pubmed.ncbi.nlm.nih.gov/26958293/)]
59. Fu S, Leung LY, Raulli AO, Kallmes DF, Kinsman KA, Nelson KB, et al. Assessment of the impact of EHR heterogeneity for clinical research through a case study of silent brain infarction. *BMC Med Inform Decis Mak* 2020 Mar 30;20(1):60 [FREE Full text] [doi: [10.1186/s12911-020-1072-9](https://doi.org/10.1186/s12911-020-1072-9)] [Medline: [32228556](https://pubmed.ncbi.nlm.nih.gov/32228556/)]
60. Fort D, Wilcox AB, Weng C. Could patient self-reported health data complement EHR for phenotyping? *AMIA Annu Symp Proc* 2014 Nov 14;2014:1738-1747 [FREE Full text] [Medline: [25954446](https://pubmed.ncbi.nlm.nih.gov/25954446/)]
61. Yamamoto K, Matsumoto S, Yanagihara K, Teramukai S, Fukushima M. A data-capture system for post-marketing surveillance of drugs that integrates with hospital electronic health records. *Open Access J Clin Trials* 2011 Apr 29;2011(3):21-26. [doi: [10.2147/oajct.s19579](https://doi.org/10.2147/oajct.s19579)]
62. Hawley G, Jackson C, Hepworth J, Wilkinson SA. Sharing of clinical data in a maternity setting: how do paper hand-held records and electronic health records compare for completeness? *BMC Health Serv Res* 2014 Dec 21;14:650 [FREE Full text] [doi: [10.1186/s12913-014-0650-x](https://doi.org/10.1186/s12913-014-0650-x)] [Medline: [25528664](https://pubmed.ncbi.nlm.nih.gov/25528664/)]
63. Black AS, Sahama T. Chronicling the patient journey: co-creating value with digital health ecosystems. In: *Proceedings of the Australasian Computer Science Week Multiconference*. 2016 Presented at: ACSW '16; February 1-5, 2016; Canberra, Australia p. 60. [doi: [10.1145/2843043.2843381](https://doi.org/10.1145/2843043.2843381)]
64. Almeida JR, Fajarda O, Pereira A, Oliveira JL. Strategies to access patient clinical data from distributed databases. In: *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies*. 2019 Presented at: BIOSTEC '19; February 22-24, 2019; Prague, Czech Republic p. 466-473. [doi: [10.5220/0007576104660473](https://doi.org/10.5220/0007576104660473)]
65. Chang IC, Li YC, Wu TY, Yen DC. Electronic medical record quality and its impact on user satisfaction — Healthcare providers' point of view. *Gov Inf Q* 2012 Apr;29(2):235-242. [doi: [10.1016/j.giq.2011.07.006](https://doi.org/10.1016/j.giq.2011.07.006)]
66. Yin RK. *Qualitative Research from Start to Finish*. 2nd edition. New York, NY, USA: Guilford Press; 2015.
67. Abiy R, Gashu K, Asemaw T, Mitiku M, Fekadie B, Abebaw Z, et al. A comparison of electronic medical record data to paper records in antiretroviral therapy clinic in Ethiopia: what is affecting the quality of the data? *Online J Public Health Inform* 2018 Sep 21;10(2):e212 [FREE Full text] [doi: [10.5210/ojphi.v10i2.8309](https://doi.org/10.5210/ojphi.v10i2.8309)] [Medline: [30349630](https://pubmed.ncbi.nlm.nih.gov/30349630/)]
68. Perimal-Lewis L, Teubner D, Hakendorf P, Horwood C. Application of process mining to assess the data quality of routinely collected time-based performance data sourced from electronic health records by validating process conformance. *Health Informatics J* 2016 Dec;22(4):1017-1029 [FREE Full text] [doi: [10.1177/1460458215604348](https://doi.org/10.1177/1460458215604348)] [Medline: [26460104](https://pubmed.ncbi.nlm.nih.gov/26460104/)]
69. Liu C, Zowghi D, Talaei-Khoei A. Empirical evaluation of the influence of EMR alignment to care processes on data completeness. In: *Proceedings of the 53rd Hawaii International Conference on System Sciences*. 2020 Presented at: HICSS '20; January 7-10, 2020; Maui, HI, USA p. 3519-3528. [doi: [10.24251/hicss.2020.432](https://doi.org/10.24251/hicss.2020.432)]
70. Liaw ST, Taggart J, Yu H. EHR-based disease registries to support integrated care in a health neighbourhood: an ontology-based methodology. *Stud Health Technol Inform* 2014;205:171-175. [Medline: [25160168](https://pubmed.ncbi.nlm.nih.gov/25160168/)]
71. Muthee V, Bochner AF, Osterman A, Liku N, Akhwale W, Kwach J, et al. The impact of routine data quality assessments on electronic medical record data quality in Kenya. *PLoS One* 2018 Apr 18;13(4):e0195362 [FREE Full text] [doi: [10.1371/journal.pone.0195362](https://doi.org/10.1371/journal.pone.0195362)] [Medline: [29668691](https://pubmed.ncbi.nlm.nih.gov/29668691/)]
72. Ryan JJ, Doster B, Daily S, Lewis C. Data quality assurance via perioperative EMR reconciliation. In: *Proceedings of the 24th Americas Conference on Information Systems*. 2018 Presented at: AMCIS '18; August 16-18, 2018; New Orleans, LA, USA.
73. Landis-Lewis Z, Manjomo R, Gadabu OJ, Kam M, Simwaka BN, Zickmund SL, et al. Barriers to using eHealth data for clinical performance feedback in Malawi: a case study. *Int J Med Inform* 2015 Oct;84(10):868-875 [FREE Full text] [doi: [10.1016/j.ijmedinf.2015.07.003](https://doi.org/10.1016/j.ijmedinf.2015.07.003)] [Medline: [26238704](https://pubmed.ncbi.nlm.nih.gov/26238704/)]
74. Chunyan D. Patient privacy protection in China in the age of electronic health records. *Hong Kong Law J* 2013;43(1):245-278.
75. Penning ML, Blach C, Walden A, Wang P, Donovan KM, Garza MY, et al. Near real time EHR data utilization in a clinical study. *Stud Health Technol Inform* 2020 Jun 16;270:337-341 [FREE Full text] [doi: [10.3233/SHTI200178](https://doi.org/10.3233/SHTI200178)] [Medline: [32570402](https://pubmed.ncbi.nlm.nih.gov/32570402/)]
76. Rajamani S, Kayser A, Emerson E, Solarz S. Evaluation of data exchange process for interoperability and impact on electronic laboratory reporting quality to a state public health agency. *Online J Public Health Inform* 2018 Sep 21;10(2):e204 [FREE Full text] [doi: [10.5210/ojphi.v10i2.9317](https://doi.org/10.5210/ojphi.v10i2.9317)] [Medline: [30349622](https://pubmed.ncbi.nlm.nih.gov/30349622/)]

77. Hart R, Kuo MH. Better data quality for better healthcare research results - a case study. *Stud Health Technol Inform* 2017;234:161-166. [Medline: [28186034](#)]
78. Liu C, Zowghi D, Talaei-Khoei A, Jin Z. Empirical study of data completeness in electronic health records in China. *Pac Asia J Assoc Inf Syst* 2020 Jun 30;12(2):104-130. [doi: [10.17705/1thci.12204](#)]
79. Estiri H, Stephens KA, Klann JG, Murphy SN. Exploring completeness in clinical data research networks with DQe-c. *J Am Med Inform Assoc* 2018 Jan 01;25(1):17-24 [FREE Full text] [doi: [10.1093/jamia/ocx109](#)] [Medline: [29069394](#)]
80. 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 Jan;117(1):15-21. [doi: [10.1542/peds.2004-2741](#)] [Medline: [16396855](#)]
81. Rosenlund M, Kivekäs E, Mikkonen S, Arvonen S, Reponen J, Saranto K. Health professionals' perceptions of information quality in the health village portal. *Stud Health Technol Inform* 2019 Jul 04;262:300-303. [doi: [10.3233/SHTI190078](#)] [Medline: [31349327](#)]
82. Yoo CW, Goo J, Huang D, Behara R. Explaining task support satisfaction on electronic patient care report (ePCR) in emergency medical services (EMS): an elaboration likelihood model lens. In: *Proceedings of the 38th International Conference on Information Systems*. 2017 Presented at: ICIS '17; December 10-13, 2017; Seoul, South Korea. [doi: [10.1016/j.im.2020.103336](#)]
83. Almutiry O, Wills G, Alwabel A, Crowder R, Walters R. Toward a framework for data quality in cloud-based health information system. In: *Proceedings of the 2013 International Conference on Information Society*. 2013 Presented at: i-Society '13; June 24-26, 2013; Toronto, Canada p. 153-157.
84. Diaz-Garelli JF, Strowd R, Ahmed T, Wells BJ, Merrill R, Laurini J, et al. A tale of three subspecialties: diagnosis recording patterns are internally consistent but Specialty-Dependent. *JAMIA Open* 2019 Oct;2(3):369-377 [FREE Full text] [doi: [10.1093/jamiaopen/ooz020](#)] [Medline: [31984369](#)]
85. Sirgo G, Esteban F, Gómez J, Moreno G, Rodríguez A, Blanch L, et al. Validation of the ICU-DaMa tool for automatically extracting variables for minimum dataset and quality indicators: the importance of data quality assessment. *Int J Med Inform* 2018 Apr;112:166-172. [doi: [10.1016/j.ijmedinf.2018.02.007](#)] [Medline: [29500016](#)]
86. Sukumar SR, Natarajan R, Ferrell RK. Quality of big data in health care. *Int J Health Care Qual Assur* 2015;28(6):621-634. [doi: [10.1108/IJHCQA-07-2014-0080](#)] [Medline: [26156435](#)]
87. Gloyd S, Wagenaar BH, Woelk GB, Kalibala S. Opportunities and challenges in conducting secondary analysis of HIV programmes using data from routine health information systems and personal health information. *J Int AIDS Soc* 2016 Jul 20;19(5 Suppl 4):20847 [FREE Full text] [doi: [10.7448/IAS.19.5.20847](#)] [Medline: [27443274](#)]
88. Fox F, Aggarwal VR, Whelton H, Johnson O. A data quality framework for process mining of electronic health record data. In: *Proceedings of the 2018 IEEE International Conference on Healthcare Informatics*. 2018 Presented at: ICHI '18; June 4-7, 2018; New York, NY, USA p. 12-21. [doi: [10.1109/ichi.2018.00009](#)]
89. McCormack JL, Ash JS. Clinician perspectives on the quality of patient data used for clinical decision support: a qualitative study. *AMIA Annu Symp Proc* 2012;2012:1302-1309 [FREE Full text] [Medline: [23304409](#)]
90. Weiskopf NG, Cohen AM, Hannan J, Jarmon T, Dorr DA. Towards augmenting structured EHR data: a comparison of manual chart review and patient self-report. *AMIA Annu Symp Proc* 2020 Mar 4;2019:903-912 [FREE Full text] [Medline: [32308887](#)]
91. Wilk M, Marsh DW, De Freitas S, Prowle J. Predicting length of stay in hospital using electronic records available at the time of admission. *Stud Health Technol Inform* 2020 Jun 16;270:377-381. [doi: [10.3233/SHTI200186](#)] [Medline: [32570410](#)]
92. Kindermann A, Tute E, Benda S, Löpprich M, Richter-Pechanski P, Dieterich C. Preliminary analysis of structured reporting in the HiGHmed use case cardiology: challenges and measures. *Stud Health Technol Inform* 2021 May 24;278:187-194. [doi: [10.3233/SHTI210068](#)] [Medline: [34042893](#)]
93. Lanzola G, Parimbelli E, Micieli G, Cavallini A, Quaglini S. Data quality and completeness in a web stroke registry as the basis for data and process mining. *J Healthc Eng* 2014;5(2):163-184 [FREE Full text] [doi: [10.1260/2040-2295.5.2.163](#)] [Medline: [24918182](#)]
94. Alwhaibi M, Balkhi B, Alshammari TM, AlQahtani N, Mahmoud MA, Almetwazi M, et al. Measuring the quality and completeness of medication-related information derived from hospital electronic health records database. *Saudi Pharm J* 2019 May;27(4):502-506 [FREE Full text] [doi: [10.1016/j.jsps.2019.01.013](#)] [Medline: [31061618](#)]
95. Corey KM, Helmkamp J, Simons M, Curtis L, Marsolo K, Balu S, et al. Assessing quality of surgical real-world data from an automated electronic health record pipeline. *J Am Coll Surg* 2020 Mar;230(3):295-305.e12. [doi: [10.1016/j.jamcollsurg.2019.12.005](#)] [Medline: [31945461](#)]
96. Hosseini N, Mostafavi SM, Zendehelel K, Eslami S. Factors affecting clinicians' adherence to principles of diagnosis documentation: a concept mapping approach for improved decision-making. *Health Inf Manag* 2022 Sep;51(3):149-158. [doi: [10.1177/1833358321991362](#)] [Medline: [33845621](#)]
97. Bae CJ, Griffith S, Fan Y, Dunphy C, Thompson N, Urchek J, et al. The challenges of data quality evaluation in a joint data warehouse. *EGEMS (Wash DC)* 2015 May 22;3(1):1125 [FREE Full text] [doi: [10.13063/2327-9214.1125](#)] [Medline: [26290882](#)]

98. Skyttberg N, Chen R, Blomqvist H, Koch S. Exploring vital sign data quality in electronic health records with focus on emergency care warning scores. *Appl Clin Inform* 2017 Aug 30;8(3):880-892 [[FREE Full text](#)] [doi: [10.4338/ACI-2017-05-RA-0075](https://doi.org/10.4338/ACI-2017-05-RA-0075)] [Medline: [28853764](#)]
99. Skyttberg N, Vicente J, Chen R, Blomqvist H, Koch S. How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments. *BMC Med Inform Decis Mak* 2016 Jun 04;16:61 [[FREE Full text](#)] [doi: [10.1186/s12911-016-0305-4](https://doi.org/10.1186/s12911-016-0305-4)] [Medline: [27260476](#)]
100. Skyttberg N, Chen R, Koch S. Man vs machine in emergency medicine - a study on the effects of manual and automatic vital sign documentation on data quality and perceived workload, using observational paired sample data and questionnaires. *BMC Emerg Med* 2018 Dec 13;18(1):54 [[FREE Full text](#)] [doi: [10.1186/s12873-018-0205-2](https://doi.org/10.1186/s12873-018-0205-2)] [Medline: [30545312](#)]
101. Polubriaginof F, Salmasian H, Albert DA, Vawdrey DK. Challenges with collecting smoking status in electronic health records. *AMIA Annu Symp Proc* 2018 Apr 16;2017:1392-1400 [[FREE Full text](#)] [Medline: [29854208](#)]
102. Garg N, Kuperman G, Onyile A, Lowry T, Genes N, DiMaggio C, et al. Validating health information exchange (HIE) data for quality measurement across four hospitals. *AMIA Annu Symp Proc* 2014 Nov 14;2014:573-579 [[FREE Full text](#)] [Medline: [25954362](#)]
103. Duan R, Cao M, Wu Y, Huang J, Denny JC, Xu H, et al. An empirical study for impacts of measurement errors on EHR based association studies. *AMIA Annu Symp Proc* 2017 Feb 10;2016:1764-1773 [[FREE Full text](#)] [Medline: [28269935](#)]
104. Gumede-Moyo S, Todd J, Bond V, Mee P, Filteau S. A qualitative inquiry into implementing an electronic health record system (SmartCare) for prevention of mother-to-child transmission data in Zambia: a retrospective study. *BMJ Open* 2019 Sep 06;9(9):e030428 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2019-030428](https://doi.org/10.1136/bmjopen-2019-030428)] [Medline: [31494616](#)]
105. Sáez C, Gutiérrez-Sacristán A, Kohane I, García-Gómez JM, Avillach P. EHRtemporalVariability: delineating temporal data-set shifts in electronic health records. *Gigascience* 2020 Aug 01;9(8):giaa079 [[FREE Full text](#)] [doi: [10.1093/gigascience/giaa079](https://doi.org/10.1093/gigascience/giaa079)] [Medline: [32729900](#)]
106. Deakne Davies SJ, Grundmeier RW, Campos DA, Hayes KL, Bell J, Alessandrini EA, Pediatric Emergency Care Applied Research Network. The pediatric emergency care applied research network registry: a multicenter electronic health record registry of pediatric emergency care. *Appl Clin Inform* 2018 Apr;9(2):366-376 [[FREE Full text](#)] [doi: [10.1055/s-0038-1651496](https://doi.org/10.1055/s-0038-1651496)] [Medline: [29791930](#)]
107. Reimer AP, Milinovich A, Madigan EA. Data quality assessment framework to assess electronic medical record data for use in research. *Int J Med Inform* 2016 Jun;90:40-47 [[FREE Full text](#)] [doi: [10.1016/j.ijmedinf.2016.03.006](https://doi.org/10.1016/j.ijmedinf.2016.03.006)] [Medline: [27103196](#)]
108. Chen ES, Carter EW, Sarkar IN, Winden TJ, Melton GB. Examining the use, contents, and quality of free-text tobacco use documentation in the electronic health record. *AMIA Annu Symp Proc* 2014 Nov 14;2014:366-374 [[FREE Full text](#)] [Medline: [25954340](#)]
109. Hamamura FD, Withy K, Hughes K. Identifying barriers in the use of electronic health records in Hawai'i. *Hawaii J Med Public Health* 2017 Mar;76(3 Suppl 1):28-35 [[FREE Full text](#)] [Medline: [28435756](#)]
110. Boyle DI, Cunningham SG. Resolving fundamental quality issues in linked datasets for clinical care. *Health Informatics J* 2002 Jun;8(2):73-77. [doi: [10.1177/146045820200800205](https://doi.org/10.1177/146045820200800205)]
111. Tilahun B, Fritz F. Modeling antecedents of electronic medical record system implementation success in low-resource setting hospitals. *BMC Med Inform Decis Mak* 2015 Aug 01;15:61 [[FREE Full text](#)] [doi: [10.1186/s12911-015-0192-0](https://doi.org/10.1186/s12911-015-0192-0)] [Medline: [26231051](#)]
112. Kim HS, Kim JH. Proceed with caution when using real world data and real world evidence. *J Korean Med Sci* 2019 Jan 28;34(4):e28 [[FREE Full text](#)] [doi: [10.3346/jkms.2019.34.e28](https://doi.org/10.3346/jkms.2019.34.e28)] [Medline: [30686950](#)]
113. Gibby GL. Anesthesia information-management systems: their role in risk-versus cost assessment and outcomes research. *J Cardiothorac Vasc Anesth* 1997 Apr;11(2 Suppl 1):2-25. [doi: [10.1016/s1053-0770\(97\)80002-x](https://doi.org/10.1016/s1053-0770(97)80002-x)] [Medline: [9106006](#)]
114. Kim M, Shin SY, Kang M, Yi BK, Chang DK. Developing a standardization algorithm for categorical laboratory tests for clinical big data research: retrospective study. *JMIR Med Inform* 2019 Aug 29;7(3):e14083 [[FREE Full text](#)] [doi: [10.2196/14083](https://doi.org/10.2196/14083)] [Medline: [31469075](#)]
115. Dantanarayana G, Sahama T. Metrics for eHealth services improvement. In: *Proceedings of the IEEE 18th International Conference on e-Health Networking, Applications and Services*. 2016 Presented at: Healthcom '16; September 14-16, 2016; Munich, Germany p. 1-6. [doi: [10.1109/healthcom.2016.7749425](https://doi.org/10.1109/healthcom.2016.7749425)]
116. García-de-León-Chocano R, Sáez C, Muñoz-Soler V, Oliver-Roig A, García-de-León-González R, García-Gómez JM. Robust estimation of infant feeding indicators by data quality assessment of longitudinal electronic health records from birth up to 18 months of life. *Comput Methods Programs Biomed* 2021 Aug;207:106147 [[FREE Full text](#)] [doi: [10.1016/j.cmpb.2021.106147](https://doi.org/10.1016/j.cmpb.2021.106147)] [Medline: [34020376](#)]
117. Liaw ST, Chen HY, Maneze D, Taggart J, Dennis S, Vagholkar S, et al. The quality of routinely collected data: using the "principal diagnosis" in emergency department databases as an example. *Electron J Health Inform* 2012;7(1):1-8.
118. Baernholdt M, Dunton N, Hughes RG, Stone PW, White KM. Quality measures: a stakeholder analysis. *J Nurs Care Qual* 2018;33(2):149-156. [doi: [10.1097/NCQ.0000000000000292](https://doi.org/10.1097/NCQ.0000000000000292)] [Medline: [28915223](#)]

119. Curtis MD, Griffith SD, Tucker M, Taylor MD, Capra WB, Carrigan G, et al. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv Res* 2018 Dec;53(6):4460-4476 [[FREE Full text](#)] [doi: [10.1111/1475-6773.12872](https://doi.org/10.1111/1475-6773.12872)] [Medline: [29756355](https://pubmed.ncbi.nlm.nih.gov/29756355/)]
120. Gupta S, Liu L, Patterson OV, Earles A, Bustamante R, Gawron AJ, et al. A framework for leveraging "Big Data" to advance epidemiology and improve quality: design of the VA colonoscopy collaborative. *EGEMS (Wash DC)* 2018 Apr 13;6(1):4 [[FREE Full text](#)] [doi: [10.5334/egems.198](https://doi.org/10.5334/egems.198)] [Medline: [29881762](https://pubmed.ncbi.nlm.nih.gov/29881762/)]
121. Knake LA, Ahuja M, McDonald EL, Ryckman KK, Weathers N, Burstain T, et al. Quality of EHR data extractions for studies of preterm birth in a tertiary care center: guidelines for obtaining reliable data. *BMC Pediatr* 2016 Apr 29;16:59 [[FREE Full text](#)] [doi: [10.1186/s12887-016-0592-z](https://doi.org/10.1186/s12887-016-0592-z)] [Medline: [27130217](https://pubmed.ncbi.nlm.nih.gov/27130217/)]
122. Rees S, Akbari A, Collins H, Lee SC, Marchant A, Rees A, et al. Developing a standardised approach to the aggregation of inpatient episodes into person-based spells in all specialties and psychiatric specialties. *BMC Med Inform Decis Mak* 2019 Nov 29;19(1):246 [[FREE Full text](#)] [doi: [10.1186/s12911-019-0953-2](https://doi.org/10.1186/s12911-019-0953-2)] [Medline: [31783849](https://pubmed.ncbi.nlm.nih.gov/31783849/)]
123. Salomon RM, Blackford JU, Rosenbloom ST, Seidel S, Clayton EW, Dilts DM, et al. Openness of patients' reporting with use of electronic records: psychiatric clinicians' views. *J Am Med Inform Assoc* 2010;17(1):54-60 [[FREE Full text](#)] [doi: [10.1197/jamia.M3341](https://doi.org/10.1197/jamia.M3341)] [Medline: [20064802](https://pubmed.ncbi.nlm.nih.gov/20064802/)]
124. Tantoso E, Wong WC, Tay WH, Lee J, Sinha S, Eisenhaber B, et al. Hypocrisy around medical patient data: issues of access for biomedical research, data quality, usefulness for the purpose and omics data as game changer. *Asian Bioeth Rev* 2019 Jun;11(2):189-207 [[FREE Full text](#)] [doi: [10.1007/s41649-019-00085-3](https://doi.org/10.1007/s41649-019-00085-3)] [Medline: [33717311](https://pubmed.ncbi.nlm.nih.gov/33717311/)]
125. Zagher J, Rodrigues-Jr JF, Goeriot L, Amer-Yahia S. Real-world patient trajectory prediction from clinical notes using artificial neural networks and UMLS-based extraction of concepts. *J Healthc Inform Res* 2021 Dec;5(4):474-496 [[FREE Full text](#)] [doi: [10.1007/s41666-021-00100-z](https://doi.org/10.1007/s41666-021-00100-z)] [Medline: [35419508](https://pubmed.ncbi.nlm.nih.gov/35419508/)]
126. Cohen B, Vawdrey DK, Liu J, Caplan D, Furuya EY, Mis FW, et al. Challenges associated with using large data sets for quality assessment and research in clinical settings. *Policy Polit Nurs Pract* 2015 Aug;16(3-4):117-124 [[FREE Full text](#)] [doi: [10.1177/1527154415603358](https://doi.org/10.1177/1527154415603358)] [Medline: [26351216](https://pubmed.ncbi.nlm.nih.gov/26351216/)]
127. Smith N, Coleman KJ, Lawrence JM, Quinn VP, Getahun D, Reynolds K, et al. Body weight and height data in electronic medical records of children. *Int J Pediatr Obes* 2010 May 03;5(3):237-242. [doi: [10.3109/17477160903268308](https://doi.org/10.3109/17477160903268308)] [Medline: [19961272](https://pubmed.ncbi.nlm.nih.gov/19961272/)]
128. Welch G, von Recklinghausen F, Taenzer A, Savitz L, Weiss L. Data cleaning in the evaluation of a multi-site intervention project. *EGEMS (Wash DC)* 2017 Dec 15;5(3):4 [[FREE Full text](#)] [doi: [10.5334/egems.196](https://doi.org/10.5334/egems.196)] [Medline: [29881755](https://pubmed.ncbi.nlm.nih.gov/29881755/)]
129. Wennberg S, Karlsten LA, Stalfors J, Bratt M, Bugten V. Providing quality data in health care - almost perfect inter-rater agreement in the Norwegian tonsil surgery register. *BMC Med Res Methodol* 2019 Jan 07;19(1):6 [[FREE Full text](#)] [doi: [10.1186/s12874-018-0651-2](https://doi.org/10.1186/s12874-018-0651-2)] [Medline: [30616535](https://pubmed.ncbi.nlm.nih.gov/30616535/)]
130. Aalsma MC, Schwartz K, Haight KA, Jarjoura GR, Dir AL. Applying an electronic health records data quality framework across service sectors: a case study of juvenile justice system data. *EGEMS (Wash DC)* 2019 Jul 11;7(1):26 [[FREE Full text](#)] [doi: [10.5334/egems.258](https://doi.org/10.5334/egems.258)] [Medline: [31328133](https://pubmed.ncbi.nlm.nih.gov/31328133/)]
131. Alves DS, Maranhão PA, Pereira AM, Bacelar-Silva GM, Silva-Costa T, Beale TW, et al. Can openEHR represent the clinical concepts of an obstetric-specific EHR - obscure software? *Stud Health Technol Inform* 2019 Aug 21;264:773-777. [doi: [10.3233/SHTI190328](https://doi.org/10.3233/SHTI190328)] [Medline: [31438029](https://pubmed.ncbi.nlm.nih.gov/31438029/)]
132. Khetarpal S. Clinical research using an information system: the multicenter perioperative outcomes group. *Anesthesiol Clin* 2011 Sep;29(3):377-388. [doi: [10.1016/j.anclin.2011.06.002](https://doi.org/10.1016/j.anclin.2011.06.002)] [Medline: [21871400](https://pubmed.ncbi.nlm.nih.gov/21871400/)]
133. O'Shea MP, Kennedy C, Relihan E, Harkin K, Hennessy M, Barry M. Assessment of an electronic patient record system on discharge prescribing errors in a Tertiary University Hospital. *BMC Med Inform Decis Mak* 2021 Jun 21;21(1):195 [[FREE Full text](#)] [doi: [10.1186/s12911-021-01551-5](https://doi.org/10.1186/s12911-021-01551-5)] [Medline: [34154570](https://pubmed.ncbi.nlm.nih.gov/34154570/)]
134. Ward MJ, Froehle CM, Hart KW, Lindsell CJ. Operational data integrity during electronic health record implementation in the ED. *Am J Emerg Med* 2013 Jul;31(7):1029-1033. [doi: [10.1016/j.ajem.2013.03.027](https://doi.org/10.1016/j.ajem.2013.03.027)] [Medline: [23707000](https://pubmed.ncbi.nlm.nih.gov/23707000/)]
135. Kamdje-Wabo G, Gradinger T, Löbe M, Lodahl R, Seuchter SA, Sax U, et al. Towards structured data quality assessment in the German medical informatics initiative: initial approach in the MII demonstrator study. *Stud Health Technol Inform* 2019 Aug 21;264:1508-1509. [doi: [10.3233/SHTI190508](https://doi.org/10.3233/SHTI190508)] [Medline: [31438205](https://pubmed.ncbi.nlm.nih.gov/31438205/)]
136. Diaz-Garelli JF, Strowd R, Wells BJ, Ahmed T, Merrill R, Topaloglu U. Lost in translation: diagnosis records show more inaccuracies after biopsy in oncology care EHRs. *AMIA Jt Summits Transl Sci Proc* 2019 May 6;2019:325-334 [[FREE Full text](#)] [Medline: [31258985](https://pubmed.ncbi.nlm.nih.gov/31258985/)]
137. DeShazo JP, Hoffman MA. A comparison of a multistate inpatient EHR database to the HCUP Nationwide Inpatient Sample. *BMC Health Serv Res* 2015 Sep 15;15:384 [[FREE Full text](#)] [doi: [10.1186/s12913-015-1025-7](https://doi.org/10.1186/s12913-015-1025-7)] [Medline: [26373538](https://pubmed.ncbi.nlm.nih.gov/26373538/)]
138. Funkner AA, Egorov MP, Fokin SA, Orlov GM, Kovalchuk SV. Citywide quality of health information system through text mining of electronic health records. *Appl Netw Sci* 2021 Jul 17;6(1):53. [doi: [10.1007/s41109-021-00395-2](https://doi.org/10.1007/s41109-021-00395-2)]
139. Potter LE, Purdie C, Nielsen S. The view from the trenches: satisfaction with eHealth systems by a group of health professionals. In: *Proceedings of the 23rd Australasian Conference on Information Systems*. 2012 Presented at: ACIS '12; December 3-5, 2012; Geelong, Australia p. 41.

140. Spuhl J, Doing-Harris K, Nelson S, Estrada N, Del Fiol G, Weir C. Concordance of electronic health record (EHR) data describing delirium at a VA hospital. *AMIA Annu Symp Proc* 2014 Nov 14;2014:1066-1071 [[FREE Full text](#)] [Medline: [25954416](#)]
141. Zakim D, Brandberg H, El Amrani S, Hultgren A, Stathakarou N, Nifakos S, et al. Computerized history-taking improves data quality for clinical decision-making-comparison of EHR and computer-acquired history data in patients with chest pain. *PLoS One* 2021 Sep 27;16(9):e0257677 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0257677](#)] [Medline: [34570811](#)]
142. Groenhof TK, Koers LR, Blasse E, de Groot M, Grobbee DE, Bots ML, UPOD, UCC-CVRM Study Groups. Data mining information from electronic health records produced high yield and accuracy for current smoking status. *J Clin Epidemiol* 2020 Feb;118:100-106 [[FREE Full text](#)] [doi: [10.1016/j.jclinepi.2019.11.006](#)] [Medline: [31730918](#)]
143. Sachdeva S, Batra S, Bhalla S. Evolving large scale healthcare applications using open standards. *Health Policy Technol* 2017 Dec;6(4):410-425. [doi: [10.1016/j.hlpt.2017.10.001](#)]
144. Li Y, Sperrin M, Martin GP, Ashcroft DM, van Staa TP. Examining the impact of data quality and completeness of electronic health records on predictions of patients' risks of cardiovascular disease. *Int J Med Inform* 2020 Jan;133:104033. [doi: [10.1016/j.ijmedinf.2019.104033](#)] [Medline: [31785526](#)]
145. Rockenschaub P, Nguyen V, Aldridge RW, Acosta D, García-Gómez JM, Sáez C. Data-driven discovery of changes in clinical code usage over time: a case-study on changes in cardiovascular disease recording in two English electronic health records databases (2001-2015). *BMJ Open* 2020 Feb 13;10(2):e034396 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2019-034396](#)] [Medline: [32060159](#)]
146. Shahid Ansari M, Jain D, Harikumar H, Rana S, Gupta S, Budhiraja S, et al. Identification of predictors and model for predicting prolonged length of stay in dengue patients. *Health Care Manag Sci* 2021 Dec;24(4):786-798 [[FREE Full text](#)] [doi: [10.1007/s10729-021-09571-3](#)] [Medline: [34389924](#)]
147. Flint A, Chaudhry NA, Riverso M, Pham A, Moser PP, Mramba LK, et al. Effective communication of cross-sectional imaging findings in Crohn's disease: comparing conventional EMR reporting to a published scoring system. *Abdom Radiol (NY)* 2018 Jul;43(7):1798-1806. [doi: [10.1007/s00261-017-1368-0](#)] [Medline: [29075825](#)]
148. Zozus MN, Young LW, Simon AE, Garza M, Lawrence L, Ounpraseuth ST, et al. Training as an intervention to decrease medical record abstraction errors multicenter studies. *Stud Health Technol Inform* 2019;257:526-539 [[FREE Full text](#)] [Medline: [30741251](#)]
149. 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. *Med Care* 2012 Jul;50 Suppl:S21-S29 [[FREE Full text](#)] [doi: [10.1097/MLR.0b013e318257dd67](#)] [Medline: [22692254](#)]
150. Reeves RM, FitzHenry F, Brown SH, Kotter K, Gobbel GT, Montella D, et al. Who said it? Establishing professional attribution among authors of Veterans' Electronic Health Records. *AMIA Annu Symp Proc* 2012;2012:753-762 [[FREE Full text](#)] [Medline: [23304349](#)]
151. Soil TA, Katsma CP, Stegwee RA, Albers EF, Freriks A, Ligt E. Value, participation and quality of electronic health records in the Netherlands. In: *Proceedings of the 43rd Hawaii International Conference on System Sciences*. 2010 Presented at: HICSS '10; January 5-8, 2010; Honolulu, HI, USA p. 1-10. [doi: [10.1109/hicss.2010.433](#)]
152. Erlirianto LM, Ali AH, Herdiyanti A. The implementation of the human, organization, and technology-fit (HOT-Fit) framework to evaluate the electronic medical record (EMR) system in a hospital. In: *Proceedings of the 3rd Information Systems International Conference*. 2015 Presented at: ISICO '15; November 2-4, 2015; Surabaya, Indonesia p. 580-587. [doi: [10.1016/j.procs.2015.12.166](#)]
153. Carr LL, Zelarny P, Meadows S, Kern JA, Long MB, Kern E. Development of a cancer care summary through the electronic health record. *J Oncol Pract* 2016 Feb;12(2):e231-e240. [doi: [10.1200/JOP.2015.006890](#)] [Medline: [26759492](#)]
154. Hawley S, Yu J, Bogetic N, Potapova N, Wakefield C, Thompson M, et al. Digitization of measurement-based care pathways in mental health through REDCap and electronic health record integration: development and usability study. *J Med Internet Res* 2021 May 20;23(5):e25656 [[FREE Full text](#)] [doi: [10.2196/25656](#)] [Medline: [34014169](#)]
155. Miettinen M, Korhonen M. Information quality in healthcare: coherence of data compared between organization's electronic patient records. In: *Proceedings of the 21st IEEE International Symposium on Computer-Based Medical Systems*. 2008 Presented at: CBMS '08; June 17-19, 2008; Jyväskylä, Finland p. 488-493. [doi: [10.1109/cbms.2008.64](#)]
156. Liaw ST, Chen HY, Maneze D, Taggart J, Dennis S, Vagholkar S, et al. Health reform: is routinely collected electronic information fit for purpose? *Emerg Med Australas* 2012 Feb;24(1):57-63. [doi: [10.1111/j.1742-6723.2011.01486.x](#)] [Medline: [22313561](#)]
157. Liu C, Zowghi D, Talaei-Khoei A. An empirical study of the antecedents of data completeness in electronic medical records. *Int J Inf Manage* 2020 Feb;50:155-170. [doi: [10.1016/j.ijinfomgt.2019.05.001](#)]
158. Piri S. Missing care: a framework to address the issue of frequent missing values;the case of a clinical decision support system for Parkinson's disease. *Decis Support Syst* 2020 Sep;136:113339. [doi: [10.1016/j.dss.2020.113339](#)]
159. Qiu H, Yu HY, Wang LY, Yao Q, Wu SN, Yin C, et al. Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. *Sci Rep* 2017 Nov 27;7(1):16417 [[FREE Full text](#)] [doi: [10.1038/s41598-017-16665-y](#)] [Medline: [29180800](#)]

160. Razavi AR, Gill H, Åhlfeldt H, Shahsavari N. A data pre-processing method to increase efficiency and accuracy in data mining. In: Proceedings of the 10th Conference on Artificial Intelligence in Medicine. 2005 Presented at: AIME '05; July 23-27, 2005; Aberdeen, UK p. 434-443. [doi: [10.1007/11527770_59](https://doi.org/10.1007/11527770_59)]
161. Salati M, Pompili C, Refai M, Xiumè F, Sabbatini A, Brunelli A. Real-time database drawn from an electronic health record for a thoracic surgery unit: high-quality clinical data saving time and human resources. *Eur J Cardiothorac Surg* 2014 Jun;45(6):1017-1019. [doi: [10.1093/ejcts/ezt577](https://doi.org/10.1093/ejcts/ezt577)] [Medline: [24394554](https://pubmed.ncbi.nlm.nih.gov/24394554/)]
162. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. *Stat Methods Med Res* 2018 Nov;27(11):3271-3285. [doi: [10.1177/0962280217696115](https://doi.org/10.1177/0962280217696115)] [Medline: [29298612](https://pubmed.ncbi.nlm.nih.gov/29298612/)]
163. Ni K, Chu H, Zeng L, Li N, Zhao Y. Barriers and facilitators to data quality of electronic health records used for clinical research in China: a qualitative study. *BMJ Open* 2019 Jul 02;9(7):e029314 [FREE Full text] [doi: [10.1136/bmjopen-2019-029314](https://doi.org/10.1136/bmjopen-2019-029314)] [Medline: [31270120](https://pubmed.ncbi.nlm.nih.gov/31270120/)]
164. Polubriaginof FC, Ryan P, Salmasian H, Shapiro AW, Perotte A, Safford MM, et al. Challenges with quality of race and ethnicity data in observational databases. *J Am Med Inform Assoc* 2019 Aug 01;26(8-9):730-736 [FREE Full text] [doi: [10.1093/jamia/ocz113](https://doi.org/10.1093/jamia/ocz113)] [Medline: [31365089](https://pubmed.ncbi.nlm.nih.gov/31365089/)]
165. Griffiths R, Schlüter DK, Akbari A, Cosgriff R, Tucker D, Taylor-Robinson D. Identifying children with cystic fibrosis in population-scale routinely collected data in Wales: a retrospective review. *Int J Popul Data Sci* 2020 Aug 11;5(1):1346 [FREE Full text] [doi: [10.23889/ijpds.v5i1.1346](https://doi.org/10.23889/ijpds.v5i1.1346)] [Medline: [33644411](https://pubmed.ncbi.nlm.nih.gov/33644411/)]
166. Callahan TJ, Barnard JG, Helmkamp LJ, Maertens JA, Kahn MG. Reporting data quality assessment results: identifying individual and organizational barriers and solutions. *EGEMS (Wash DC)* 2017 Sep 04;5(1):16 [FREE Full text] [doi: [10.5334/egems.214](https://doi.org/10.5334/egems.214)] [Medline: [29881736](https://pubmed.ncbi.nlm.nih.gov/29881736/)]
167. Hausvik GI, Thapa D, Munkvold BE. Information quality life cycle in secondary use of EHR data. *Int J Inf Manag* 2021 Feb;56:102227. [doi: [10.1016/j.ijinfomgt.2020.102227](https://doi.org/10.1016/j.ijinfomgt.2020.102227)]
168. Colquhoun DA, Shanks AM, Kapeles SR, Shah N, Saager L, Vaughn MT, et al. Considerations for integration of perioperative electronic health records across institutions for research and quality improvement: the approach taken by the multicenter perioperative outcomes group. *Anesth Analg* 2020 May;130(5):1133-1146 [FREE Full text] [doi: [10.1213/ANE.0000000000004489](https://doi.org/10.1213/ANE.0000000000004489)] [Medline: [32287121](https://pubmed.ncbi.nlm.nih.gov/32287121/)]
169. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform* 2013 Oct;46(5):830-836 [FREE Full text] [doi: [10.1016/j.jbi.2013.06.010](https://doi.org/10.1016/j.jbi.2013.06.010)] [Medline: [23820016](https://pubmed.ncbi.nlm.nih.gov/23820016/)]
170. Rutzner S, Ganslandt T, Fietkau R, Prokosch HU, Lubgan D. Noncurated data lead to misinterpretation of treatment outcomes in patients with prostate cancer after salvage or palliative radiotherapy. *JCO Clin Cancer Inform* 2019 Oct;3:1-11 [FREE Full text] [doi: [10.1200/CCL.19.00052](https://doi.org/10.1200/CCL.19.00052)] [Medline: [31599645](https://pubmed.ncbi.nlm.nih.gov/31599645/)]
171. Li S, Zhang L, Liu S, Hubbard R, Li H. Surveillance of noncommunicable disease epidemic through the integrated noncommunicable disease collaborative management system: feasibility pilot study conducted in the city of Ningbo, China. *J Med Internet Res* 2020 Jul 23;22(7):e17340 [FREE Full text] [doi: [10.2196/17340](https://doi.org/10.2196/17340)] [Medline: [32706706](https://pubmed.ncbi.nlm.nih.gov/32706706/)]
172. García-de-León-Chocano R, Muñoz-Soler V, Sáez C, García-de-León-González R, García-Gómez JM. Construction of quality-assured infant feeding process of care data repositories: construction of the perinatal repository (Part 2). *Comput Biol Med* 2016 Apr 01;71:214-222. [doi: [10.1016/j.compbimed.2016.01.007](https://doi.org/10.1016/j.compbimed.2016.01.007)] [Medline: [26950399](https://pubmed.ncbi.nlm.nih.gov/26950399/)]
173. Just BH, Marc D, Munns M, Sandefer R. Why patient matching is a challenge: research on master patient index (MPI) data discrepancies in key identifying fields. *Perspect Health Inf Manag* 2016 Apr 1;13(Spring):1e [FREE Full text] [Medline: [27134610](https://pubmed.ncbi.nlm.nih.gov/27134610/)]
174. Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. *Comput Methods Programs Biomed* 2019 Nov;181:104804. [doi: [10.1016/j.cmpb.2018.10.016](https://doi.org/10.1016/j.cmpb.2018.10.016)] [Medline: [30497872](https://pubmed.ncbi.nlm.nih.gov/30497872/)]
175. Mkalira Msiska KE, Kunitawa A, Kumwenda B. Factors affecting the utilisation of electronic medical records system in Malawian central hospitals. *Malawi Med J* 2017 Sep;29(3):247-253 [FREE Full text] [doi: [10.4314/mmj.v29i3.4](https://doi.org/10.4314/mmj.v29i3.4)] [Medline: [29872515](https://pubmed.ncbi.nlm.nih.gov/29872515/)]
176. Laird-Maddox M, Mitchell SB, Hoffman M. Integrating research data capture into the electronic health record workflow: real-world experience to advance innovation. *Perspect Health Inf Manag* 2014 Oct 1;11(Fall):1e [FREE Full text] [Medline: [25593571](https://pubmed.ncbi.nlm.nih.gov/25593571/)]
177. Mathieu A, Sauthier M, Jouvét P, Emeriaud G, Brossier D. Validation process of a high-resolution database in a paediatric intensive care unit-describing the perpetual patient's validation. *J Eval Clin Pract* 2021 Apr;27(2):316-324. [doi: [10.1111/jep.13411](https://doi.org/10.1111/jep.13411)] [Medline: [32372537](https://pubmed.ncbi.nlm.nih.gov/32372537/)]
178. Kim HS, Kim H, Jeong YJ, Kim TM, Yang SJ, Baik SJ, et al. Development of clinical data mart of HMG-CoA reductase inhibitor for varied clinical research. *Endocrinol Metab (Seoul)* 2017 Mar;32(1):90-98 [FREE Full text] [doi: [10.3803/EnM.2017.32.1.90](https://doi.org/10.3803/EnM.2017.32.1.90)] [Medline: [28256114](https://pubmed.ncbi.nlm.nih.gov/28256114/)]
179. Yoo CW, Huang CD, Goo J. Task support of electronic patient care report (ePCR) systems in emergency medical services: an elaboration likelihood model lens. *Inf Manag* 2020 Sep;57(6):103336. [doi: [10.1016/j.im.2020.103336](https://doi.org/10.1016/j.im.2020.103336)]

180. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. *J Manag Inf Syst* 1996;12(4):5-33. [doi: [10.1080/07421222.1996.11518099](https://doi.org/10.1080/07421222.1996.11518099)]
181. Sung M, He J, Zhou Q, Chen Y, Ji JS, Chen H, et al. Using an integrated framework to investigate the facilitators and barriers of health information technology implementation in noncommunicable disease management: systematic review. *J Med Internet Res* 2022 Jul 20;24(7):e37338 [FREE Full text] [doi: [10.2196/37338](https://doi.org/10.2196/37338)] [Medline: [35857364](https://pubmed.ncbi.nlm.nih.gov/35857364/)]
182. Kohane IS, Aronow BJ, Avillach P, Beaulieu-Jones BK, Bellazzi R, Bradford RL, Consortium For Clinical Characterization Of COVID-19 By EHR (4CE), et al. What every reader should know about studies using electronic health record data but may be afraid to ask. *J Med Internet Res* 2021 Mar 02;23(3):e22219 [FREE Full text] [doi: [10.2196/22219](https://doi.org/10.2196/22219)] [Medline: [33600347](https://pubmed.ncbi.nlm.nih.gov/33600347/)]
183. Andrews R, Emamjome F, ter Hofstede AH, Reijers HA. Root-cause analysis of process-data quality problems. *J Bus Anal* 2022;5(1):51-75. [doi: [10.1080/2573234x.2021.1947751](https://doi.org/10.1080/2573234x.2021.1947751)]
184. Poppe E, Pika A, Wynn MT, Eden R, Andrews R, ter Hofstede AH. Extracting best-practice using mixed-methods. *Bus Inf Syst Eng* 2021 Apr 20;63(6):637-651. [doi: [10.1007/s12599-021-00698-9](https://doi.org/10.1007/s12599-021-00698-9)]

Abbreviations

DQ: data quality

DQ-DO: data quality dimension and outcome

EHR: electronic health record

EMR: electronic medical record

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Leung; submitted 12.09.22; peer-reviewed by Z Wang, M Alshammari, Y Chu, C Sullivan; comments to author 18.11.22; revised version received 07.12.22; accepted 31.12.22; published 31.03.23

Please cite as:

Syed R, Eden R, Makasi T, Chukwudi I, Mamudu A, Kamalpour M, Kapugama Geeganage D, Sadeghianasl S, Leemans SJJ, Goel K, Andrews R, Wynn MT, ter Hofstede A, Myers T

Digital Health Data Quality Issues: Systematic Review

J Med Internet Res 2023;25:e42615

URL: <https://www.jmir.org/2023/1/e42615>

doi: [10.2196/42615](https://doi.org/10.2196/42615)

PMID: [37000497](https://pubmed.ncbi.nlm.nih.gov/37000497/)

©Rehan Syed, Rebekah Eden, Tendai Makasi, Ignatius Chukwudi, Azumah Mamudu, Mostafa Kamalpour, Dakshi Kapugama Geeganage, Sareh Sadeghianasl, Sander J J Leemans, Kanika Goel, Robert Andrews, Moe Thandar Wynn, Arthur ter Hofstede, Trina Myers. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org/>), 31.03.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.