
Review

Challenges for Data Quality in the Clinical Data Life Cycle: Systematic Review

Doyeon An¹, MPH; Minsik Lim¹, BA; Suehyun Lee², PhD

¹Department of IT Convergence, Graduate School, Gachon University, Seongnam, Republic of Korea

²Department of Computer Engineering, College of IT Convergence, Gachon University, Seongnam, Republic of Korea

Corresponding Author:

Suehyun Lee, PhD

Department of Computer Engineering

College of IT Convergence

Gachon University

Gyeonggi-do

Seongnam, 13120

Republic of Korea

Phone: 82 01090129364

Email: leesh@gachon.ac.kr

Abstract

Background: Electronic health record (EHR) data are anticipated to inform the development of health policy systems across countries and furnish valuable insights for the advancement of health and medical technology. As the current paradigm of clinical research is shifting toward data centrality, the utilization of health care data is increasingly emphasized.

Objective: We aimed to review the literature on clinical data quality management and define a process for ensuring the quality management of clinical data, especially in the secondary utilization of data.

Methods: A systematic review of PubMed articles from 2010 to October 2023 was conducted. A total of 82,346 articles were retrieved and screened based on the inclusion and exclusion criteria, narrowing the number of articles to 851 after title and abstract review. Articles focusing on clinical data quality management life cycles, assessment methods, and tools were selected.

Results: We reviewed 105 papers describing the clinical data quality management process. This process is based on a 4-stage life cycle: planning, construction, operation, and utilization. The most frequently used dimensions were completeness, plausibility, concordance, security, currency, and interoperability.

Conclusions: Given the importance of the secondary use of EHR data, standardized quality control methods and automation are necessary. This study proposes a process to standardize data quality management and develop a data quality assessment system.

(*J Med Internet Res* 2025;27:e60709) doi: [10.2196/60709](https://doi.org/10.2196/60709)

KEYWORDS

clinical research informatics; data quality; data accuracy; electronic health records; frameworks; quality of health care

Introduction

As data continue to accumulate, the question of how to use neglected data has received increasing attention. In particular, the need for quality control in the use of electronic health record (EHR) data has been emphasized. EHR data are expected to facilitate the development of national health policy systems and provide useful information for improving public health and medical technology [1]. As the current clinical research paradigm shifts to one of data centrality, the use of EHR data has increasingly been emphasized [2].

The quality of EHR data research depends on the quality of the generated data, which is a major research limitation. EHR data are essential in preclinical research, which is conducted to study the future of diseases and draft policies. Therefore, integrated data must be used seamlessly and incorporate different types of data. Currently, various methods for integrated data management are being developed [3-9], but quality control standards are set differently for each data type, and discussions in this regard are challenging because of the nature of EHR data [10-13].

Although research into EHR data quality management is actively underway, a gold standard for assessing data quality remains absent. Inconsistencies in data formats and terminology, a lack of standardization, security issues, and challenges in processing large-scale data persist as major obstacles to establishing standardized EHR data management practices [14,15]. Another critical challenge in EHR data management is achieving consistency across data sets from different hospitals and health care systems [16]. The variability in data collection methods and formats among institutions complicates the integration of data sets, undermining the reproducibility and reliability of research [17].

The consistent quality of EHR data is a critical factor in the performance of data analytics. Meeting data quality standards requires a management system that is appropriate for each stage of the data life cycle [18,19]. However, no standardized approach is available to assess the quality of EHR data [14]. For accurate and consistent research on EHR data, common data models (CDMs) such as the Observational Medical Outcomes Partnership CDM and Sentinel CDM are being built [20,21]. However, CDMs are evaluated individually depending on their type [22-24].

The quality of clinical data depends on the quality of the data on which they are built, and such dependence is another major research limitation. A data quality management process defines the basic principles of data management and enables accurate, consistent control of data quality [25]. High-quality data can be defined as such when they are not built piecemeal but are managed throughout the entire process of operation and use.

Textbox 1. Search terms.

```
'quality[ti]' AND ('data quality' OR 'data accuracy' OR 'Quality of Health Care' OR 'Quality Indicators' OR 'quality control') AND (EHR OR electronic medical record OR computerized medical record OR medical records systems, computerized [mh]) AND English[lang] NOT (review OR Clinical Trial OR Documents OR Books)
```

A total of 82,346 articles were retrieved from PubMed. To select articles suitable for our research purpose, we referred to previous studies and applied the inclusion and exclusion criteria listed in Textbox 2 [14,27-29]. The studies were evaluated based on their relevance to the assessment and management of data quality of EHR data. This was done by applying inclusion and exclusion criteria to the titles and abstracts of the studies. This process was conducted by an author with a degree in public health (DA) and cross-checked by another author specializing

This study aimed to understand the importance of clinical data quality management and the life cycle-based clinical data quality management process. Accordingly, the existing literature on EHRs and clinical data quality was reviewed, and the guidelines for the predefined clinical data quality management processes of planning, implementation, operation, and utilization [26] were subsequently considered.

Methods

Definition of the Clinical Data Life Cycle

In the context of systematic data quality management, we defined the life cycle of clinical data quality management [26] as the quality management activities for health care data that include a series of steps from data construction to operation and use [26].

Literature Review on Data Quality

We aimed to identify articles that extensively discussed the generation and quality of EHR data. In this study, an EHR refers to all electronically stored records of patient health information, encompassing both electronic medical records and personal health records. To conduct the literature review, we followed the methods of previous studies that closely reviewed previous EHR data [14,27-29]. A PubMed literature search was conducted by the first author in October 2023. The keywords for the search were text words and Medical Subject Headings such as “data quality,” “data accuracy,” “quality indicators,” “quality of health care,” “quality control,” and combinations of these terms (Textbox 1). The literature search was limited to articles published in English.

in health informatics (MS) to minimize bias. In cases of disagreement in study selection, final decisions were made through thorough discussion. A total of 851 articles were selected after the first review. In the second review, all articles were manually reviewed by the first author to ensure they met the criteria. Subsequently, all papers related to data quality were selected and classified based on the following 4 keywords: “data quality,” “EHR assessment,” “treatment quality,” and “hospital quality.”

Textbox 2. Inclusion and exclusion criteria.

Inclusion criteria

- Original research using data quality assessment methods
- Focus on data derived from electronic health records or related systems

Exclusion criteria

- Guidelines limited to one medical area (eg, cardiology) without generalization to other areas
- Review papers
 - Guidance aimed at governing bodies
 - Published before 2010
 - Papers not in the English language
 - No full text available
- Not a paper on data quality issues

To focus on data quality management for clinical data analysis, we reviewed the full text of each article containing 2 of the 4 keywords, that is, “data quality” and “EHR assessment.” In this process, we reviewed medical data quality and 13 relevant guidelines. Ultimately, 105 studies were included.

For each article, we described the category, definition of data quality, data quality management methods, and quality control procedures. The literature categories included the main perspectives, research methods, and research findings. For efficiency, we reviewed the articles by classifying them into the following 4 topics: “framework,” “quality measures,” “quality tool,” and “interview.” Framework papers included articles addressing general procedures for data quality, while papers on quality measures included those involving data evaluation. Articles on quality tools included those that developed data evaluation tools, while interview articles included those that evaluated data based on the opinions of experts in actual hospital settings.

We abstracted the general methods and procedures for data quality management based on data life cycle and evaluation methods in each paper. To establish standards for the data life cycle, we analyzed the literature related to data frameworks and identified ways to construct data quality management procedures. The data quality evaluation criteria, quality evaluation methods, data types, and vocabulary used in each article were also collected. The content of the articles was then repeatedly reviewed to define their quality control dimensions.

To organize the overall data quality assessment methodology, we reviewed the literature that mentioned the data life cycle; however, finding articles offering a clear definition was difficult.

Data quality must be consistently defined [30]. The literature shows how clinical data are constructed and evaluated according to different processes. Studies have been conducted to define methods for evaluating data; however, the series of processes through which data are generated and used has not been considered. We realized that consistent data quality management could be implemented by identifying and defining the data characteristics highlighted in the literature. Our study attempted to define a set of processes through which data are constructed, operated, and used through a literature review and to include all commonly occurring concepts. We then reviewed all articles to collect data on the use of the newly defined processes and dimensions.

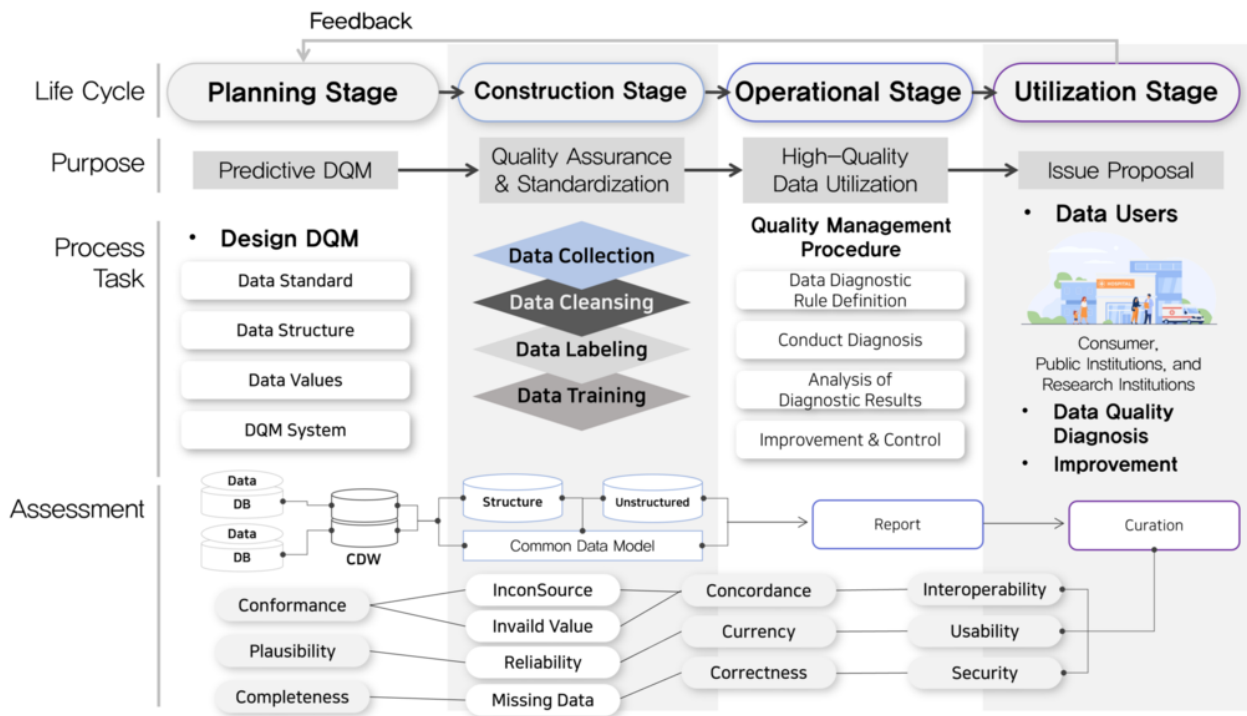
Results

Data Quality Assessment Framework Based on the Clinical Data Life Cycle

Data quality can be defined as “the level that can continuously meet the various activity purposes or satisfaction of users using data” [31]. Data quality management refers to a set of activities that ensure data quality. With the goal of developing and implementing high-quality data, data quality management encompasses all data-related management activities, from data creation to use [26].

Figure 1 illustrates the life cycle of clinical data and defines the data quality management methods according to the life cycle stage. We used the clinical data life cycle, which consists of the planning, construction, operation, and utilization stages [26]. In producing high-quality data, data must be managed according to the data life cycle and governance principles [26].

Figure 1. Life cycle of clinical data quality management (DQM). CDW: common data warehouse; DB: database.



We established the definitions for each clinical data life cycle stage by reviewing the literature (Table 1). The literature included in the review often described the data life cycle for

improving hospital EHR quality, quality measurement, and clinical decision support [32-36].

Table 1. Defining the life cycle of clinical data quality management.

Life cycle stage	Definition	References
Planning stage	Defining data standards based on the direction of data and creating a clear strategy for establishing quality management activities	[6,32,33,37]
Construction stage	Considering the characteristics among data sets, collecting data, and proceeding with overall data construction and management that reflect clinical attributes	[32,37-43]
Operation stage	Conducting data quality assessments on the constructed data and reviewing them from various angles and perspectives	[32,33,37,44,45]
Utilization stage	Sharing the outcomes of data quality validation, implementing data quality enhancement activities, and recalibrating the overall data quality	[32,33,37,46]

Planning Stage

In the planning stage of data quality management, key issues such as the data to be generated and their documentation and organization, storage and security, stewardship, and accessibility for reuse and sharing are considered [47]. Developing a data management plan should involve describing how data will be handled throughout the life of the project and after completion and establishing principles that are easy to implement [48].

Construction Stage

The construction stage involves quality control. It is also called the big data life cycle stage [25] (Figure 1). This data life cycle stage consists of 4 stages: data collection, data cleaning, data labeling, and data learning. At each stage of the life cycle, the tasks to be performed vary. For example, data quality control standards must be established and reflected in the data collection stage.

Operation Stage

Managing constructed data is the most active phase of data quality management. When building quality data, quality control must be implemented starting from the planning stage. However, not all data are built with quality control in mind from the planning stage. In data quality management, the operational stage involves activities to diagnose and improve the quality of the data loaded in data construction projects.

Utilization Stage

The main users of public medical data are public institutions and research institutes. Data quality management organizations must continuously implement improvements to provide high-quality data by adhering to the requirements of both data providers and consumers. Moreover, data must be continuously and accurately managed to provide high-quality medical services [9]. Accordingly, a support system must be institutionalized to continuously communicate with researchers on the use of

medical data, and a foundation such as medical data standards must be established to ensure the uninterrupted provision of high-quality data.

Proposed Data Framework Based on the Clinical Data Life Cycle

In our literature review, we found one commonality: All stages are interrelated and emphasize the need to manage data from a holistic, life cycle perspective [26]. The plan-do-study-act (PDSA) cycle, which was frequently mentioned in most of the articles we reviewed, is primarily used for short-term processes, such as data construction or operation [33,38,46]. Therefore, the PDSA cycle, which is mainly used in the data construction stage, could not be applied in our study. The clinical data life cycle proposed in this study is designed to manage data comprehensively from a governance perspective. It is structured in a mutually organic manner, allowing for the reapplication of improvements after EHR data planning, construction, and secondary use. A set of procedures, such as the data framework, provides an environment for researchers to understand data, identify quality issues, and address them effectively [49]. As

data significantly influence research outcomes, they must meaningfully be evaluated and managed throughout their life cycle [30]. Some studies did not consider data from a life cycle perspective [34,35,50-52]. Nevertheless, they considered the ecological use of data. They also considered the impact of data on hospital treatment processes [34,35]. Thus, data operations are organically linked, reflecting the interplay between different stages.

Dimensions of the Data Life Cycle and Clinical Data Quality Management

The set of reviewed papers comprised 44 papers on data framework, 32 papers on quality measures, 20 papers on quality tools, and 9 papers on interviews (Figures 2 and 3; Multimedia Appendix 1). Completeness was identified as the most commonly used indicator, particularly in 94 papers (Table 2 and Table 3). Research using data quality dimensions can be classified according to the stage of the clinical data life cycle, with the greatest amount of research occurring in the planning and implementation phase (Table 3).

Figure 2. Diagram of the literature review process for clinical data quality management.

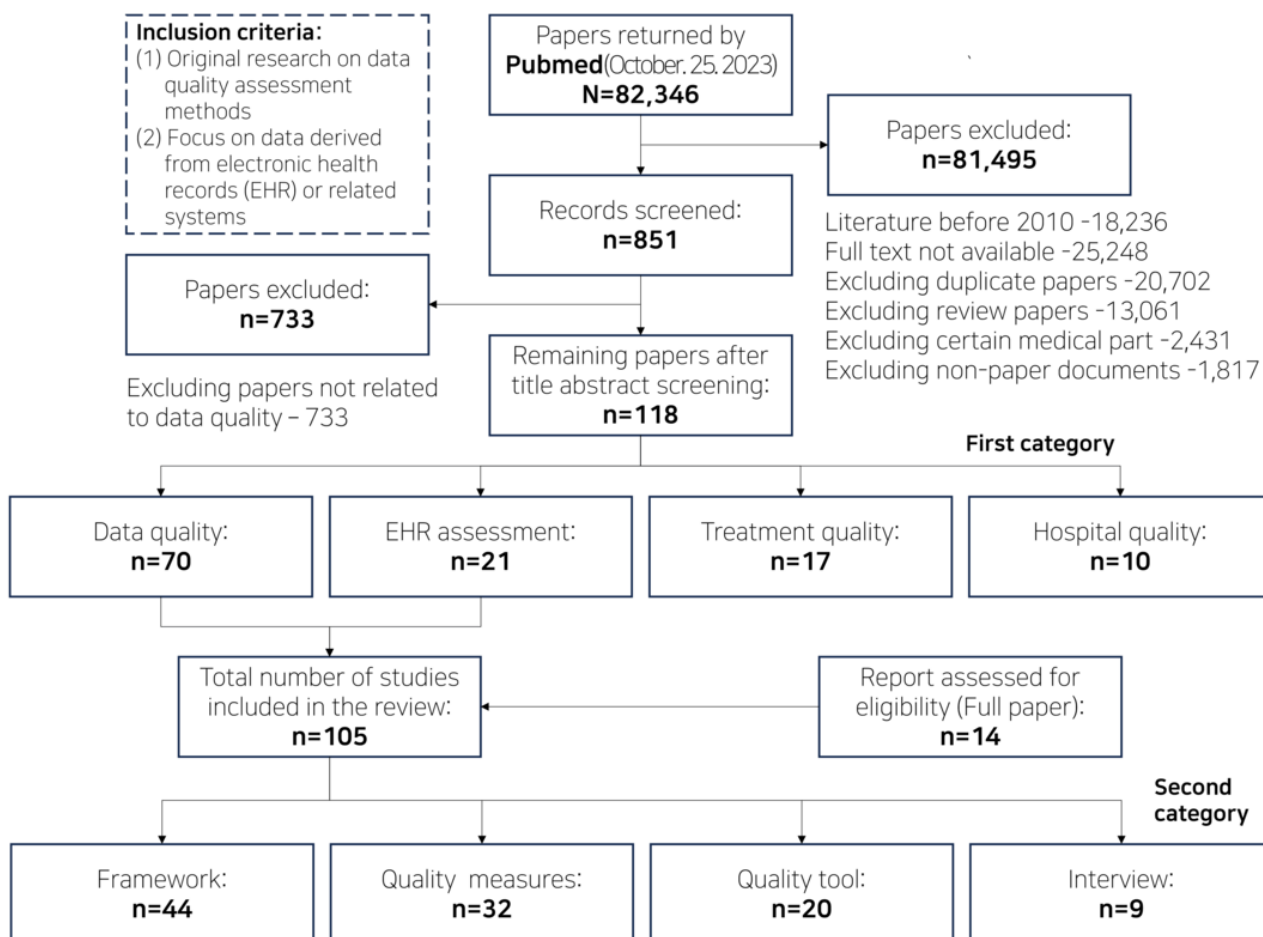


Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram.

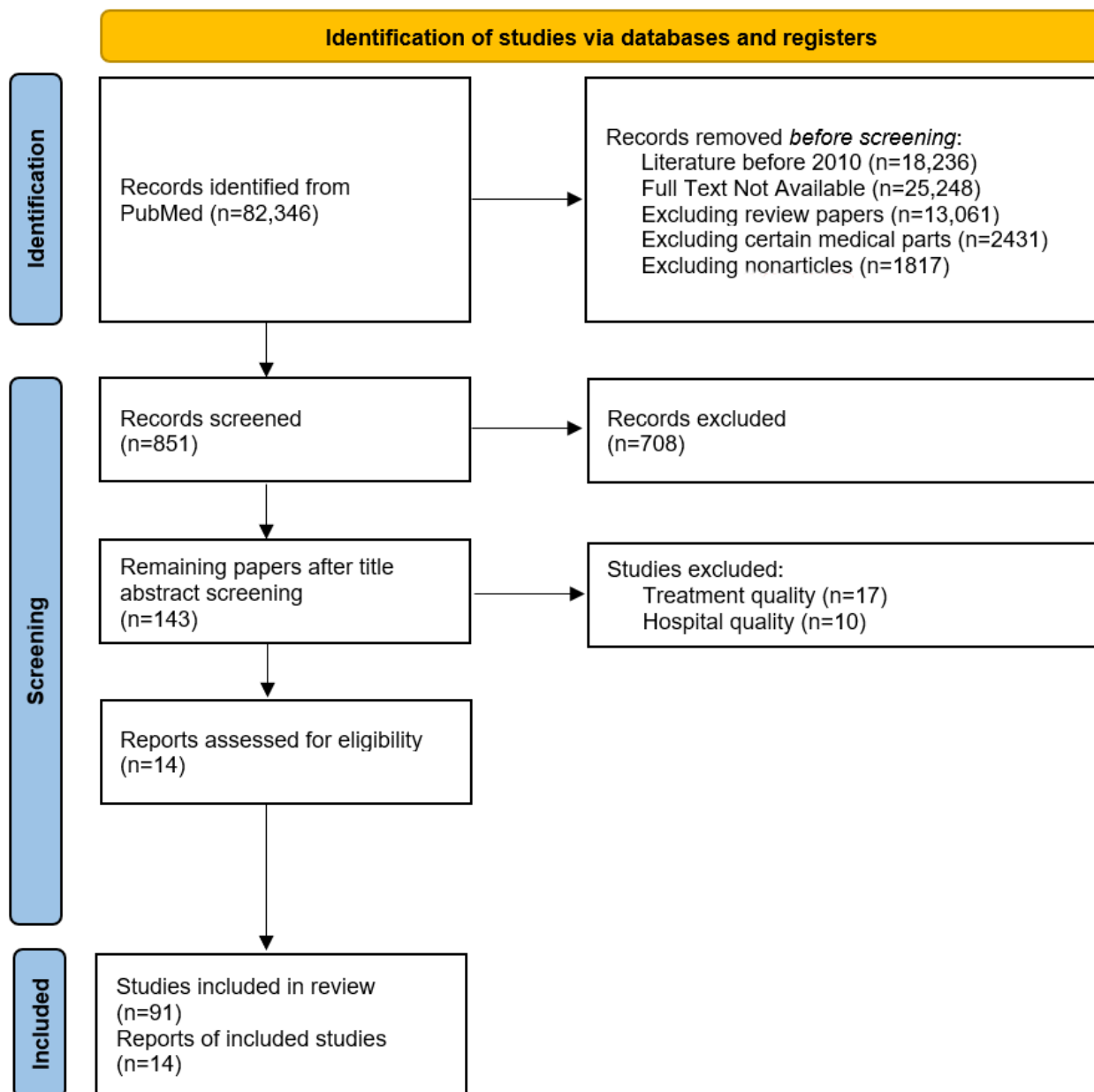


Table 2. Definitions of the life cycle of clinical data quality management and dimensions of data quality.

Dimension	Definition	Synonyms
Completeness	Assessing the extent to which data have been fully constructed in accordance with their characteristics and intended design	Completeness, correctness, conformance, incompleteness, consistency
Plausibility	Degree of reliability in data values and the significance of the associated information	Accuracy, consistency, relevance
Concordance	The extent to which data can be stored in accordance with their characteristics based on standards	Structure, standardization
Security	The extent to which data are trustworthy and accessible only to authorized users	Security, availability, confidentiality, representation, confidentiality, trustworthiness
Currency	The extent to which data can be provided promptly when needed	Currency, timeliness, currentness
Interoperability	The degree to which data operation is flexible, providing a sufficient and useful level of information that satisfies users	Availability, manageability, variability

Table 3. Life cycle of clinical data quality management and dimensions of data quality.

Dimension	Planning stage (n=69)		Construction stage (n=99)		Operation stage (n=95)		Utilization stage (n=72)	
	Mentions, n (%) ^a	Articles	Mentions, n (%) ^a	Articles	Mentions, n (%) ^a	Articles	Mentions, n (%) ^a	Articles
Completeness (n=107)	22 (20.6)	[6,7,18,19,25,32,33,43,45,53-65]	34 (31.8)	[7,9,18,19,25,32,39,40,43,45,53-77]	30 (28)	[7,9,15,19,25,32,34,43,49,50,55-57,59,63,65,70,72,75,76,78-85]	21 (19.6)	[7,16,18,19,22,25,32,43,49,50,56,63,65-67,75,78,84-87]
Plausibility (n=72)	19 (26.4)	[6,7,11,17-19,25,32,33,43,45,51,54,56,61,63-65,88]	25 (34.7)	[7,9,11,17,19,22,25,43,45,51,54-56,61,63-66,68-70,75,76,88,89]	26 (36.1)	[7,9,11,15,17,19,25,43,45,46,49,51,56,63,65,70,75,76,79-83,88,90,91]	19 (26.4)	[7,11,16-19,25,32,43,46,49,56,65,66,75,86,88,90]
Concordance (n=81)	18 (22.2)	[6,7,17-19,25,32,33,43,45,51,56,57,59,62,63,65]	22 (27.2)	[7,9,17,19,25,43-45,51,55-57,59,62,63,65,67,70,75,76]	23 (28.4)	[7,9,17,19,25,32,43,44,49,51,56,57,59,63,65,70,75,76,79,80,85,90]	18 (22.2)	[7,16-19,25,32,43,49,56,63,65,67,75,85,86,90]
Security (n=33)	8 (24.2)	[17,19,25,32,45,51,58,60,63]	9 (27.3)	[17,19,25,45,51,58,60,63,89,91]	7 (21.2)	[17,19,25,51,63,79,90,91]	9 (27.3)	[16,17,19,25,32,63,86,87,90]
Currency (n=42)	9 (21.4)	[33,43,45,52,54,57,62,63,92]	14 (33.3)	[11,15,17,43,52,55,57,62,63,67,71,72,92,93]	10 (23.8)	[11,15,17,43,55,57,63,72,79,85,93]	9 (21.4)	[11,16,17,32,43,57,63,67,85,86]
Interoperability (n=35)	7 (20)	[17,33,35,36,62,63,94]	8 (22.9)	[17,35,36,46,55,63,74,79,80,94]	10 (28.6)	[17,35,36,46,55,63,74,79,80,94]	10 (28.6)	[17,32,35,36,46,63,75,86,90,94]

^aDistribution of each dimension across the stages of the clinical data life cycle (planning, construction, operation, and utilization), calculated as a proportion of each dimension's total.

Completeness

Completeness was mainly used in the construction or operation stage and was used as an indicator for EHR evaluation [66,67], data quality system development [7,53,78], data recognition [17], and comparative evaluation [50]. The related terms used in the articles included correctness, conformance, incompleteness, and consistency.

Plausibility

Plausibility was the second most frequently used indicator, with 72 references mentioning it. It was often used in data evaluation during the operation phase of the data life cycle. It was mainly mentioned in the literature on data tool development [54,55], framework presentation [45,68], data measurement [69], and data quality assessment [7,66,89].

Concordance

Similar to completeness and plausibility, concordance was frequently mentioned in the construction and utilization stages. Concordance can be considered an indicator that determines

whether the characteristics of different data are best expressed and stored based on standards. Concordance was mentioned in the studies that developed, experimented with, and evaluated quality management tools [9,51,54,56,57,70,90]. The related terms mentioned in the articles included structure and standardization.

Security

As EHR data are sensitive, great attention must be paid to ethical issues and data leakage. Therefore, the security of EHR data is crucial. In contrast to the aforementioned 3 indicators, which reflect the completeness of data, security was most frequently mentioned in the construction and utilization stages. The related terms mentioned in the articles included availability, confidentiality, representation, and trustworthiness.

Currency

Currency was mentioned most often during the data construction stage. In particular, the availability of data must be determined during data construction. Having readily available data is critical

for the research process. The terms representing currency included timeliness.

Interoperability

The most cited limitation of EHR data is the difficulty with linking data between hospitals. By combining and sharing data already in use, more resources can be utilized. The indicator representing this relation is interoperability. The literature review in this study revealed a strong emphasis on interoperability, but it was not mentioned in articles defining other data quality indicators.

Discussion

Principal Findings

This study reviewed the existing literature, focusing on the importance of quality management from the EHR data life cycle perspective. Accordingly, an EHR data life cycle framework was defined, and 6 quality indicators were identified.

Data quality ensures the validity of research findings and provides information to demonstrate the appropriateness of EHR data use [49]. In this study, we identified the requirements for each stage of the data life cycle, including cycle-specific objectives, tasks, and evaluation metrics, to determine the validity of data. Data quality is a fundamental element for determining whether data have been constructed for their intended purpose [95]. Quality management must be applied at every stage of data processing to ensure that all data are reliable and appropriately handled [96].

The metrics identified in this study were frequently mentioned in the literature. We mapped the categories proposed in this study for currency and interoperability, which differ from the indicators proposed in previous studies. An accurate definition of these dimensions is essential for data quality. The definition of completeness alone can vary the completeness ratio of data depending on the type of data or the purpose for which quality is defined [28,86]. Dimensions have been developed to clearly define and automatically measure data [45]. Currency and interoperability metrics are not entirely new. They were mentioned repeatedly in various studies [33,43,45,52,54,57,62,63,92]. Currency refers to information about current data [63] and is primarily used for temporal information when representing the lifetime of data [16]. Temporal factors exert a significant effect on research results. In addition, currency should be considered when visualizing data quality results [42].

This study proposes a total of 6 data quality dimensions based on a comprehensive literature review. These indicators are not universally applicable across all data sets; additional dimensions may be warranted depending on specific conditions (Multimedia Appendix 2). For instance, bias can emerge based on data construction or the research environment. Addressing bias is crucial and has been emphasized in numerous studies on data quality [14,16,83]. In this regard, assessing task relevance is vital to verify that the constructed data meet their intended objectives and are effective for their purpose [45]. Furthermore, if data are integrated from multiple sources rather than generated from a single system, it is critical to evaluate consistency across

data sets using the variability dimension [57]. In clinical settings, the validity and reliability of data are fundamental to the development of safe and accurate predictive models [57]. It is also necessary to assess usability to confirm that researchers in clinical environments can use data both effectively and efficiently [7,42] (Multimedia Appendix 3). Before using and measuring any data quality dimension, the purpose and research objectives of the data must be thoroughly understood, and the indicators must be selected accordingly. Systematic data quality assessments are essential at each phase of the data life cycle to ensure comprehensive data utilization. Each dimension can play a vital role in ensuring data accuracy, reliability, and efficiency, thereby enhancing the reproducibility and validity of the research. Developing a well-defined data quality plan minimizes unnecessary processes and costs and directly enhances data transparency and trustworthiness.

The majority of discussions on the quality of EHR data have centered on 3 key areas: conformance, plausibility, and completeness [6,42,49,70,86]. However, the actual quality of data can vary significantly depending on the measurement methods and management strategies used, due to factors such as the type and volume of data, data construction environment, characteristics of the disease, and type of system in which the data are generated [75,94]. A substantial body of research has proposed and developed a multitude of indicators. Through a comprehensive review of the literature, we identified that dimensions such as accuracy, consistency, completeness, and currency are closely interrelated according to data characteristics. Additionally, these indicators may vary in relevance depending on the data life cycle stage. Many studies, however, have overlooked these aspects. Recognizing the interdependence between dimensions while accounting for the unique characteristics of the data is crucial to establishing high-quality data.

When ensuring effective data quality management, simplified data guidelines that can be easily applied must be considered. Data quality management frameworks and guidelines are being developed in a data-specific manner [12,18,19,25,65]. From the data life cycle perspective, data quality management must be coordinated from a governance perspective throughout the entire life cycle. Several different types of data exist. To actively manage the quality of different data, more diverse data quality management methodologies must be developed [97]. Meanwhile, ensuring that data are usable and consistent requires clearly targeted and planned quality control procedures [48]. Regarding ensuring the scalability of data connections, quality control for integrated data using standardized procedures should be implemented from the planning stage [98].

In our study, we emphasized the importance of interoperability in the use of EHR data. The use of EHR helps researchers conduct their studies involving large amounts of data at a low cost [99] and facilitates the analysis of health information from thousands of individuals. Ideally, EHRs should be accurate and complete because they contain all health records [100]; however, EHR data face numerous quality issues [4,101]. In addition, challenges arise from the use of different EHR systems across hospitals and the heterogeneity of data, resulting in limited interoperability. Limited interoperability and inconsistent data

exchange across settings are significant barriers to quality improvement [102]. The interoperability of EHRs with medical data is becoming increasingly valuable because of its potential to exponentially increase the availability of data or directly impact the activation of research. EHR systems can efficiently support data structuring and quality measurement results and have a great impact on patients and their time [102]. Interoperability among EHR systems refers to the linking of data, which improves data usability. Therefore, regulating the data structure or transfer standards between systems is essential to improve data quality and interoperability.

Considerable effort has been made to improve the quality of EHRs. These efforts include the development of automated data quality assessment systems [9,42,103], organization of quality indicator events, and development of metrics. Data must be sufficiently flexible to be used for multiple purposes. Moreover, data must be managed according to user needs, and diagnoses must be made based on the users' purpose. When producing high-quality data, the data must be thoroughly examined from a data life cycle perspective, starting from data construction, to ensure that data standards are well established and applied, data are consistently secured, and errors are minimized [104].

Establishing criteria for data quality is critical because the data sources for research questions represent a major determinant of research outcomes. Several factors necessitate the establishment of data quality standards. First, the types of data required vary according to the research topic, and data types and structure are significantly diverse. In addition, medical practices and health care systems vary widely worldwide, and their differences can affect the relevance of data to research questions [12]. Data must be managed continuously and accurately to provide high-quality medical services [9]. Consequently, the perspectives for measuring the level of data quality must be defined, and the criteria for what should be measured must be established [25].

Investing in EHR data quality management improves clinical outcomes [34]. As hospital resources are limited, data preprocessing and quality assessment must be automated to avoid wasting resources. Many hospital researchers have focused on automating data quality assessment [3,6,8,9,59,77,105]. However, automation across all data sets lacks a unified

standard, and different tools have been developed for different data types and languages. Given the diverse criteria and forms of EHR data, such approaches are not pragmatic [14]. Accurately defining the domains and task ontologies for measuring data quality in the automation process is critical [45,59]. Various methodologies and quality criteria have been identified [29]. Nevertheless, flexible tools that consider interoperability must be developed, and existing methodologies must be used to create a unified automation tool [14].

Limitations

Our literature review has several limitations that need to be considered. First, the literature selection was conducted solely by the first author, which may introduce subjectivity to the process and result in classifications that other reviewers might not agree with. Although cross-review efforts were made, the lack of a multireviewer approach may limit the generalizability of the findings. Second, in this study, we conducted the literature search using only one database. Due to the use of a single source, there may be a risk of missing other relevant studies. However, prior to conducting our study, we performed the same search in other databases and observed similar results to those obtained from PubMed, the database ultimately used in this research. Third, the quality dimensions identified in this review, derived solely from existing literature, have not been validated by clinical experts. The absence of expert validation may limit the practical applicability of these dimensions in clinical settings, indicating a need for further expert review.

Conclusion

As the value of EHR data increases, the demand for high-quality data also rises. Standardized quality management and automation of data quality assessment are necessary to produce high-quality data and improve their usability. This study focuses on the secondary use of EHR data, reviews the existing literature, and redefines quality management indicators from a data life cycle perspective. As data quality assessment methods based on the data life cycle perspective have not yet been developed, future work should focus on developing data quality assessment systems with an emphasis on standardized frameworks and tools that consider the specific characteristics of the data.

Acknowledgments

We attest that there was no use of GenAI technology in the generation of text, figures, or other informational content of this manuscript.

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2022-KH125153) and This work was supported by the Gachon University research fund of 2023 (GCU-202400550001).

Data Availability

The data supporting this article are available upon request from the corresponding author.

Conflicts of Interest

None disclosed.

Multimedia Appendix 1

Data Search List.

[[ZIP File \(Zip Archive\), 139 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

Additional Quality Dimension.

[[DOCX File , 29 KB-Multimedia Appendix 2](#)]

Multimedia Appendix 3

Term of Data Quality Management.

[[XLSX File \(Microsoft Excel File\), 13 KB-Multimedia Appendix 3](#)]

Multimedia Appendix 4

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.

[[PDF File \(Adobe PDF File\), 82 KB-Multimedia Appendix 4](#)]

References

1. Kwon T, Jeong Y, Lee D. Standardization and quality evaluation of health and medical big data. Osong: Korea Health Industry Development Institute. Nov 25, 2019. URL: <https://www.khiss.go.kr/board/view/pgNm=1&rowGr=10&col=314&tblId=1750&multi=MENU005&schIpe=0&schIex=&trdStye=&catgryId=&contnt=&sdSaChr=&sdFrChr=&conty=> [accessed 2023-10-25]
2. Choi MS, Lee SH. Current status and issues of data management plan in Korea. The Journal of the Korea Contents Association. 2020;20:220-229. [doi: [10.5392/JKCA.2020.20.06.220](https://doi.org/10.5392/JKCA.2020.20.06.220)]
3. Tute E, Scheffner I, Marscholke M. A method for interoperable knowledge-based data quality assessment. BMC Med Inform Decis Mak. Mar 09, 2021;21(1):93. [FREE Full text] [doi: [10.1186/s12911-021-01458-1](https://doi.org/10.1186/s12911-021-01458-1)] [Medline: [33750371](https://pubmed.ncbi.nlm.nih.gov/33750371/)]
4. Weiskopf NG, Khan FJ, Woodcock D, Dorr DA, Cigarroa JE, Cohen AM. A mixed methods task analysis of the implementation and validation of EHR-based clinical quality measures. AMIA Annu Symp Proc. 2016;2016:1229-1237. [FREE Full text] [Medline: [28269920](https://pubmed.ncbi.nlm.nih.gov/28269920/)]
5. Devine EB, Van Eaton E, Zadworny ME, Symons R, Devlin A, Yanez D, et al. Automating electronic clinical data capture for quality improvement and research: the CERTAIN validation project of real world evidence. EGEMS (Wash DC). May 22, 2018;6(1):8. [FREE Full text] [doi: [10.5334/egems.211](https://doi.org/10.5334/egems.211)] [Medline: [29881766](https://pubmed.ncbi.nlm.nih.gov/29881766/)]
6. Khare R, Utidjian L, Ruth B, Kahn M, Burrows E, Marsolo K, et al. A longitudinal analysis of data quality in a large pediatric data research network. J Am Med Inform Assoc. Nov 01, 2017;24(6):1072-1079. [FREE Full text] [doi: [10.1093/jamia/ocx033](https://doi.org/10.1093/jamia/ocx033)] [Medline: [28398525](https://pubmed.ncbi.nlm.nih.gov/28398525/)]
7. Kapsner LA, Mang JM, Mate S, Seuchter SA, Vengadeswaran A, Bathelt F, et al. Linking a consortium-wide data quality assessment tool with the MIRACUM metadata repository. Appl Clin Inform. Aug 25, 2021;12(4):826-835. [FREE Full text] [doi: [10.1055/s-0041-1733847](https://doi.org/10.1055/s-0041-1733847)] [Medline: [34433217](https://pubmed.ncbi.nlm.nih.gov/34433217/)]
8. Chiang J, Lin J, Yang C. Automated evaluation of electronic discharge notes to assess quality of care for cardiovascular diseases using Medical Language Extraction and Encoding System (MedLEE). J Am Med Inform Assoc. May 01, 2010;17(3):245-252. [FREE Full text] [doi: [10.1136/jamia.2009.000182](https://doi.org/10.1136/jamia.2009.000182)] [Medline: [20442141](https://pubmed.ncbi.nlm.nih.gov/20442141/)]
9. Mang JM, Seuchter SA, Gulden C, Schild S, Kraska D, Prokosch H, et al. DQAgui: a graphical user interface for the MIRACUM data quality assessment tool. BMC Med Inform Decis Mak. Aug 11, 2022;22(1):213. [FREE Full text] [doi: [10.1186/s12911-022-01961-z](https://doi.org/10.1186/s12911-022-01961-z)] [Medline: [35953813](https://pubmed.ncbi.nlm.nih.gov/35953813/)]
10. Mohamed Y, Song X, McMahon TM, Sahil S, Zozus M, Wang Z, et al. Tailoring rule-based data quality assessment to the Patient-Centered Outcomes Research Network (PCORnet) common data model (CDM). AMIA Annu Symp Proc. 2022;2022:775-784. [FREE Full text] [Medline: [37128433](https://pubmed.ncbi.nlm.nih.gov/37128433/)]
11. Davoudi S, Dooling J, Glondys B, Jones T, Kadlec L, Overgaard S, et al. Data quality management model (Updated). J AHIMA. Oct 2015;86(10):62-65. [Medline: [26665585](https://pubmed.ncbi.nlm.nih.gov/26665585/)]
12. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products. Food and Drug Administration. Jul 2024. URL: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory> [accessed 2025-03-06]
13. Madnick SE, Wang RY, Lee YW, Zhu H. Overview and framework for data and information quality research. J. Data and Information Quality. Jun 2009;1(1):1-22. [doi: [10.1145/1515693.1516680](https://doi.org/10.1145/1515693.1516680)]

14. Lewis AE, Weiskopf N, Abrams ZB, Foraker R, Lai AM, Payne PRO, et al. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc*. Sep 25, 2023;30(10):1730-1740. [FREE Full text] [doi: [10.1093/jamia/ocad120](https://doi.org/10.1093/jamia/ocad120)] [Medline: [37390812](https://pubmed.ncbi.nlm.nih.gov/37390812/)]
15. Johnson S, Speedie S, Simon G, Kumar V, Westra B. Application of an ontology for characterizing data quality for a secondary use of EHR data. *Appl Clin Inform*. Dec 16, 2017;07(01):69-88. [doi: [10.4338/aci-2015-08-ra-0107](https://doi.org/10.4338/aci-2015-08-ra-0107)]
16. 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*. Jul 2012;50 Suppl:S21-S29. [FREE Full text] [doi: [10.1097/MLR.0b013e318257dd67](https://doi.org/10.1097/MLR.0b013e318257dd67)] [Medline: [22692254](https://pubmed.ncbi.nlm.nih.gov/22692254/)]
17. Callahan T, Barnard J, Helmkamp L, Maertens J, Kahn M. Reporting data quality assessment results: identifying individual and organizational barriers and solutions. *EGEMS (Wash DC)*. Sep 04, 2017;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/)]
18. Public data preventive quality management guide. Sejong, South Korea. Ministry of the Interior and Safety; Mar 18, 2021.
19. Public Data Quality Management Manual ver 2.0. Ministry of Interior And Safety. Jan 15, 2018. URL: https://www.data.go.kr/en/bbs/rcr/selectRecsroom.do?pageIndex=1&originId=PDS_0000000000000516 [accessed 2025-03-24]
20. Data Quality Review and Characterization Programs. Sentinel Initiative. Jun 06, 2024. URL: <https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model/data-quality-review-and-characterization-programs> [accessed 2025-03-06]
21. Makadia R, Ryan PB. Transforming the premier perspective hospital database into the Observational Medical Outcomes Partnership (OMOP) common data model. *EGEMS (Wash DC)*. 2014;2(1):1110. [FREE Full text] [doi: [10.13063/2327-9214.1110](https://doi.org/10.13063/2327-9214.1110)] [Medline: [25848597](https://pubmed.ncbi.nlm.nih.gov/25848597/)]
22. Huser V, DeFalco FJ, Schuemie M, Ryan PB, Shang N, Velez M, et al. Multisite evaluation of a data quality tool for patient-level clinical data sets. *EGEMS (Wash DC)*. 2016;4(1):1239. [FREE Full text] [doi: [10.13063/2327-9214.1239](https://doi.org/10.13063/2327-9214.1239)] [Medline: [28154833](https://pubmed.ncbi.nlm.nih.gov/28154833/)]
23. Estiri H, Stephens KA, Klann JG, Murphy SN. Exploring completeness in clinical data research networks with DQe-c. *J Am Med Inform Assoc*. Jan 01, 2018;25(1):17-24. [FREE Full text] [doi: [10.1093/jamia/ocx109](https://doi.org/10.1093/jamia/ocx109)] [Medline: [29069394](https://pubmed.ncbi.nlm.nih.gov/29069394/)]
24. Bian J, Lyu T, Loiacono A, Viramontes TM, Lipori G, Guo Y, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. *J Am Med Inform Assoc*. Dec 09, 2020;27(12):1999-2010. [FREE Full text] [doi: [10.1093/jamia/ocaa245](https://doi.org/10.1093/jamia/ocaa245)] [Medline: [33166397](https://pubmed.ncbi.nlm.nih.gov/33166397/)]
25. Lee Y, Son K, Yoo S, Kim E, Co. Big Data Platform and Center Data Quality Management Guide. In: platform DoB, editor. Daegu. National Information Society Agency; Jun 2, 2022:7-143.
26. Lee S, Roh G, Kim J, Ho Lee Y, Woo H, Lee S. Effective data quality management for electronic medical record data using SMART DATA. *Int J Med Inform*. Dec 2023;180:105262. [FREE Full text] [doi: [10.1016/j.ijmedinf.2023.105262](https://doi.org/10.1016/j.ijmedinf.2023.105262)] [Medline: [37871445](https://pubmed.ncbi.nlm.nih.gov/37871445/)]
27. Arts DGT, De Keizer NF, Scheffer G. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Assoc*. 2002;9(6):600-611. [FREE Full text] [doi: [10.1197/jamia.m1087](https://doi.org/10.1197/jamia.m1087)] [Medline: [12386111](https://pubmed.ncbi.nlm.nih.gov/12386111/)]
28. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. Jan 01, 2013;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/)]
29. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med*. Jan 10, 2022;5(1):2. [FREE Full text] [doi: [10.1038/s41746-021-00549-7](https://doi.org/10.1038/s41746-021-00549-7)] [Medline: [35013569](https://pubmed.ncbi.nlm.nih.gov/35013569/)]
30. Liaw S, Guo JGN, Ansari S, Jonnagaddala J, Godinho MA, Borelli AJ, et al. Quality assessment of real-world data repositories across the data life cycle: a literature review. *J Am Med Inform Assoc*. Jul 14, 2021;28(7):1591-1599. [FREE Full text] [doi: [10.1093/jamia/ocaa340](https://doi.org/10.1093/jamia/ocaa340)] [Medline: [33496785](https://pubmed.ncbi.nlm.nih.gov/33496785/)]
31. English L. Information quality management: The next frontier. ProQuest. 2001. URL: <https://www.proquest.com/openview/5ee454e132571ff609fe90509f73abfa/1?cbi=39817&pq-origsite=gscholar> [accessed 2023-10-25]
32. Winter A, Takabayashi K, Jahn F, Kimura E, Engelbrecht R, Haux R, et al. Quality requirements for electronic health record systems. *Methods Inf Med*. Jan 31, 2018;56(S 01):e92-e104. [doi: [10.3414/me17-05-0002](https://doi.org/10.3414/me17-05-0002)]
33. Jedwab RM, Franco M, Owen D, Ingram A, Redley B, Dobroff N. Improving the quality of electronic medical record documentation: development of a compliance and quality program. *Appl Clin Inform*. Aug 07, 2022;13(4):836-844. [FREE Full text] [doi: [10.1055/s-0042-1756369](https://doi.org/10.1055/s-0042-1756369)] [Medline: [36070801](https://pubmed.ncbi.nlm.nih.gov/36070801/)]
34. Damberg CL, Shortell SM, Raube K, Gillies RR, Rittenhouse D, McCurdy RK, et al. Relationship between quality improvement processes and clinical performance. *Am J Manag Care*. Aug 2010;16(8):601-606. [FREE Full text] [Medline: [20712393](https://pubmed.ncbi.nlm.nih.gov/20712393/)]
35. Ramirez A, Sulieman L, Schlueter D, Halvorson A, Qian J, Ratsimbazafy F, et al. All of Us Research Program. The research program: data quality, utility, and diversity. *Patterns (N Y)*. Aug 12, 2022;3(8):100570. [FREE Full text] [doi: [10.1016/j.patter.2022.100570](https://doi.org/10.1016/j.patter.2022.100570)] [Medline: [36033590](https://pubmed.ncbi.nlm.nih.gov/36033590/)]

36. Lin Y, Staes CJ, Shields DE, Kandula V, Welch BM, Kawamoto K. Design, development, and initial evaluation of a terminology for clinical decision support and electronic clinical quality measurement. *AMIA Annu Symp Proc.* 2015;2015:843-851. [[FREE Full text](#)] [Medline: [26958220](#)]
37. Chelico JD, Wilcox AB, Vawdrey DK, Kuperman GJ. Designing a clinical data warehouse architecture to support quality improvement initiatives. *AMIA Annu Symp Proc.* 2016;2016:381-390. [[FREE Full text](#)] [Medline: [28269833](#)]
38. Knight AW, Szucs C, Dhillon M, Lembke T, Mitchell C. The eCollaborative: using a quality improvement collaborative to implement the National eHealth Record System in Australian primary care practices. *Int J Qual Health Care.* Aug 12, 2014;26(4):411-417. [[FREE Full text](#)] [doi: [10.1093/intqhc/mzu059](#)] [Medline: [24925685](#)]
39. Randall SM, Ferrante AM, Boyd JH, Semmens JB. The effect of data cleaning on record linkage quality. *BMC Med Inform Decis Mak.* Jun 05, 2013;13:64. [[FREE Full text](#)] [doi: [10.1186/1472-6947-13-64](#)] [Medline: [23739011](#)]
40. Schepens MHJ, Trompert AC, van Hooff ML, van der Velde E, Kallewaard M, Verberk-Jonkers IJAM, et al. Using existing clinical information models for Dutch quality registries to reuse data and follow COUMT paradigm. *Appl Clin Inform.* Mar 2023;14(2):326-336. [[FREE Full text](#)] [doi: [10.1055/s-0043-1767681](#)] [Medline: [37137338](#)]
41. Kukhareva PV, Kawamoto K, Shields DE, Barfuss DT, Halley AM, Tippetts TJ, et al. Clinical Decision Support-based Quality Measurement (CDS-QM) framework: prototype implementation, evaluation, and future directions. *AMIA Annu Symp Proc.* 2014;2014:825-834. [[FREE Full text](#)] [Medline: [25954389](#)]
42. Engel N, Wang H, Jiang X, Lau CY, Patterson J, Acharya N, et al. EHR data quality assessment tools and issue reporting workflows for the 'All of Us' research program clinical data research network. *AMIA Jt Summits Transl Sci Proc.* 2022;2022:186-195. [[FREE Full text](#)] [Medline: [35854725](#)]
43. Diaz-Garelli J, Bernstam EV, Lee M, Hwang KO, Rahbar MH, Johnson TR. DataGauge: a practical process for systematically designing and implementing quality assessments of repurposed clinical data. *EGEMS (Wash DC).* Jul 25, 2019;7(1):32. [[FREE Full text](#)] [doi: [10.5334/egems.286](#)] [Medline: [31367649](#)]
44. Fife CE, Walker D, Thomson B. Electronic health records, registries, and quality measures: What? Why? How? *Adv Wound Care (New Rochelle).* Dec 2013;2(10):598-604. [[FREE Full text](#)] [doi: [10.1089/wound.2013.0476](#)] [Medline: [24761335](#)]
45. 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;2015:1937-1946. [[FREE Full text](#)] [Medline: [26958293](#)]
46. Carroll A, Johnson D. Know it when you see it: identifying and using special cause variation for quality improvement. *Hosp Pediatr.* Nov 2020;10(11):e8-e10. [[FREE Full text](#)] [doi: [10.1542/hpeds.2020-002303](#)] [Medline: [33051243](#)]
47. Holles JH, Schmidt L. Graduate Research Data Management Course Content: Teaching the Data Management Plan (DMP). 2018. Presented at: ASEE Annual Conference & Exposition; June 23-27, 2018; Salt Lake City, UT. [doi: [10.18260/1-2--30562](#)]
48. Michener WK. Ten simple rules for creating a good data management plan. *PLoS Comput Biol.* Oct 22, 2015;11(10):e1004525. [[FREE Full text](#)] [doi: [10.1371/journal.pcbi.1004525](#)] [Medline: [26492633](#)]
49. Huang Y, Voorham J, Haaijer-Ruskamp FM. Using primary care electronic health record data for comparative effectiveness research: experience of data quality assessment and preprocessing in The Netherlands. *J Comp Eff Res.* Jul 2016;5(4):345-354. [[FREE Full text](#)] [doi: [10.2217/cer-2015-0022](#)] [Medline: [27346480](#)]
50. Bell EJ, Takhar SS, Beloff JR, Schuur JD, Landman AB. Information technology improves emergency department patient discharge instructions completeness and performance on a national quality measure: a quasi-experimental study. *Appl Clin Inform.* 2013;4(4):499-514. [[FREE Full text](#)] [doi: [10.4338/ACI-2013-07-RA-0046](#)] [Medline: [24454578](#)]
51. Rahbar MH, Gonzales NR, Ardjomand-Hessabi M, Tahanan A, Sline MR, Peng H, et al. The University of Texas Houston Stroke Registry (UTHSR): implementation of enhanced data quality assurance procedures improves data quality. *BMC Neurol.* Jun 15, 2013;13(1):61. [[FREE Full text](#)] [doi: [10.1186/1471-2377-13-61](#)] [Medline: [23767957](#)]
52. Kvale M, Hesselton S, Hoffmann T, Cao Y, Chan D, Connell S, et al. Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics.* Aug 2015;200(4):1051-1060. [[FREE Full text](#)] [doi: [10.1534/genetics.115.178905](#)] [Medline: [26092718](#)]
53. Van Batavia JP, Weiss DA, Long CJ, Madison J, McCarthy G, Plachter N, et al. Using structured data entry systems in the electronic medical record to collect clinical data for quality and research: Can we efficiently serve multiple needs for complex patients with spina bifida? *PRM.* Dec 13, 2018;11(4):303-309. [doi: [10.3233/prm-170525](#)]
54. Johnson S, Speedie S, Simon G, Kumar V, Westra B. Quantifying the effect of data quality on the validity of an eMeasure. *Appl Clin Inform.* Dec 14, 2017;08(04):1012-1021. [doi: [10.4338/aci-2017-03-ra-0042](#)]
55. 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.* Aug 23, 2020;11(4):622-634. [[FREE Full text](#)] [doi: [10.1055/s-0040-1715567](#)] [Medline: [32968999](#)]
56. Razzaghi H, Greenberg J, Bailey L. Developing a systematic approach to assessing data quality in secondary use of clinical data based on intended use. *Learn Health Syst.* Jan 2022;6(1):e10264. [[FREE Full text](#)] [doi: [10.1002/lrh2.10264](#)] [Medline: [35036548](#)]
57. Fu S, Wen A, Schaeferle GM, Wilson PM, Demuth G, Ruan X, et al. Assessment of data quality variability across two EHR systems through a case study of post-surgical complications. *AMIA Jt Summits Transl Sci Proc.* 2022;2022:196-205. [[FREE Full text](#)] [Medline: [35854735](#)]

58. Joukes E, de Keizer NF, de Bruijne MC, Abu-Hanna A, Cornet R. Impact of electronic versus paper-based recording before EHR implementation on health care professionals' perceptions of EHR use, data quality, and data reuse. *Appl Clin Inform.* Mar 2019;10(2):199-209. [FREE Full text] [doi: [10.1055/s-0039-1681054](https://doi.org/10.1055/s-0039-1681054)] [Medline: [30895574](https://pubmed.ncbi.nlm.nih.gov/30895574/)]
59. Wu Q, Ganz C, Li L. Data quality control for electronic pathology reporting. *J Registry Manag.* 2022;49(3):95-96. [FREE Full text] [Medline: [37260925](https://pubmed.ncbi.nlm.nih.gov/37260925/)]
60. Devine EB, Capurro D, van Eaton E, Alfonso-Cristancho R, Devlin A, Yanez ND, et al. Preparing electronic clinical data for quality improvement and comparative effectiveness research: the SCOAP CERTAIN automation and validation project. *EGEMS (Wash DC).* 2013;1(1):1025. [FREE Full text] [doi: [10.13063/2327-9214.1025](https://doi.org/10.13063/2327-9214.1025)] [Medline: [25848565](https://pubmed.ncbi.nlm.nih.gov/25848565/)]
61. Huser V, Williams ND, Mayer CS. Linking provider specialty and outpatient diagnoses in Medicare claims data: data quality implications. *Appl Clin Inform.* Aug 2021;12(4):729-736. [FREE Full text] [doi: [10.1055/s-0041-1732404](https://doi.org/10.1055/s-0041-1732404)] [Medline: [34348410](https://pubmed.ncbi.nlm.nih.gov/34348410/)]
62. 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.* Jun 04, 2016;16:61. [FREE Full text] [doi: [10.1186/s12911-016-0305-4](https://doi.org/10.1186/s12911-016-0305-4)] [Medline: [27260476](https://pubmed.ncbi.nlm.nih.gov/27260476/)]
63. Fadahunsi KP, Wark PA, Mastellos N, Neves AL, Gallagher J, Majeed A, et al. Assessment of clinical information quality in digital health technologies: international eDelphi study. *J Med Internet Res.* Dec 06, 2022;24(12):e41889. [FREE Full text] [doi: [10.2196/41889](https://doi.org/10.2196/41889)] [Medline: [36472901](https://pubmed.ncbi.nlm.nih.gov/36472901/)]
64. Garies S, Cummings M, Forst B, McBrien K, Soos B, Taylor M, et al. Achieving quality primary care data: a description of the Canadian Primary Care Sentinel Surveillance Network data capture, extraction, and processing in Alberta. *Int J Popul Data Sci.* Jul 29, 2019;4(2):1132. [FREE Full text] [doi: [10.23889/ijpds.v4i2.1132](https://doi.org/10.23889/ijpds.v4i2.1132)] [Medline: [34095540](https://pubmed.ncbi.nlm.nih.gov/34095540/)]
65. Agency N. Data Quality Management Guidelines v2.0 for Artificial Intelligence Learning - Quality Management Guide. In: Division ID. editor. Daegu. National Information Society Agency; Mar 14, 2022.
66. Garies S, McBrien K, Quan H, Manca D, Drummond N, Williamson T. A data quality assessment to inform hypertension surveillance using primary care electronic medical record data from Alberta, Canada. *BMC Public Health.* Feb 02, 2021;21(1):264. [FREE Full text] [doi: [10.1186/s12889-021-10295-w](https://doi.org/10.1186/s12889-021-10295-w)] [Medline: [33530975](https://pubmed.ncbi.nlm.nih.gov/33530975/)]
67. 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.* Jul 2021;4(3):oob057. [FREE Full text] [doi: [10.1093/jamiaopen/oob057](https://doi.org/10.1093/jamiaopen/oob057)] [Medline: [34350392](https://pubmed.ncbi.nlm.nih.gov/34350392/)]
68. Winnenburg R, Bodenreider O. Metrics for assessing the quality of value sets in clinical quality measures. *AMIA Annu Symp Proc.* 2013;2013:1497-1505. [FREE Full text] [Medline: [24551422](https://pubmed.ncbi.nlm.nih.gov/24551422/)]
69. Gadde MA, Wang Z, Zozus M, Talburt JB, Greer ML. Rules based data quality assessment on claims database. *Stud Health Technol Inform.* Jun 26, 2020;272:350-353. [FREE Full text] [doi: [10.3233/SHTI200567](https://doi.org/10.3233/SHTI200567)] [Medline: [32604674](https://pubmed.ncbi.nlm.nih.gov/32604674/)]
70. Rogers JR, Callahan TJ, Kang T, Bauck A, Khare R, Brown JS, et al. A data element-function conceptual model for data quality checks. *EGEMS (Wash DC).* Apr 23, 2019;7(1):17. [FREE Full text] [doi: [10.5334/egems.289](https://doi.org/10.5334/egems.289)] [Medline: [31065558](https://pubmed.ncbi.nlm.nih.gov/31065558/)]
71. Li M, Cai H, Zhi Y, Fu Z, Duan H, Lu X. A configurable method for clinical quality measurement through electronic health records based on openEHR and CQL. *BMC Med Inform Decis Mak.* Feb 10, 2022;22(1):37. [FREE Full text] [doi: [10.1186/s12911-022-01763-3](https://doi.org/10.1186/s12911-022-01763-3)] [Medline: [35144618](https://pubmed.ncbi.nlm.nih.gov/35144618/)]
72. Keegan N, Vasselmann S, Barnett E, Nweji B, Carbone E, Blum A, et al. Clinical annotations for prostate cancer research: defining data elements, creating a reproducible analytical pipeline, and assessing data quality. *Prostate.* Aug 2022;82(11):1107-1116. [FREE Full text] [doi: [10.1002/pros.24363](https://doi.org/10.1002/pros.24363)] [Medline: [35538298](https://pubmed.ncbi.nlm.nih.gov/35538298/)]
73. Dixon BE, Siegel JA, Oemig TV, Grannis SJ. Electronic health information quality challenges and interventions to improve public health surveillance data and practice. *Public Health Rep.* 2013;128(6):546-553. [FREE Full text] [doi: [10.1177/003335491312800614](https://doi.org/10.1177/003335491312800614)] [Medline: [24179266](https://pubmed.ncbi.nlm.nih.gov/24179266/)]
74. Cusick MM, Sholle ET, Davila MA, Kabariti J, Cole CL, Champion TR. A method to improve availability and quality of patient race data in an electronic health record system. *Appl Clin Inform.* Oct 2020;11(5):785-791. [FREE Full text] [doi: [10.1055/s-0040-1718756](https://doi.org/10.1055/s-0040-1718756)] [Medline: [33241548](https://pubmed.ncbi.nlm.nih.gov/33241548/)]
75. Arvidsson E, Dijkstra R, Klemenc-Ketiš Z. Measuring quality in primary healthcare - opportunities and weaknesses. *Zdr Varst.* Sep 2019;58(3):101-103. [FREE Full text] [doi: [10.2478/sjph-2019-0013](https://doi.org/10.2478/sjph-2019-0013)] [Medline: [31275436](https://pubmed.ncbi.nlm.nih.gov/31275436/)]
76. Lewis J, Stephens J, Musick B, Brown S, Malateste K, Ha Dao Ostinelli C, et al. on the behalf of IeDEA. The IeDEA harmonist data toolkit: a data quality and data sharing solution for a global HIV research consortium. *J Biomed Inform.* Jul 2022;131:104110. [FREE Full text] [doi: [10.1016/j.jbi.2022.104110](https://doi.org/10.1016/j.jbi.2022.104110)] [Medline: [35680074](https://pubmed.ncbi.nlm.nih.gov/35680074/)]
77. Paulsen A, Overgaard S, Lauritsen JM. Quality of data entry using single entry, double entry and automated forms processing--an example based on a study of patient-reported outcomes. *PLoS One.* Apr 6, 2012;7(4):e35087. [FREE Full text] [doi: [10.1371/journal.pone.0035087](https://doi.org/10.1371/journal.pone.0035087)] [Medline: [22493733](https://pubmed.ncbi.nlm.nih.gov/22493733/)]
78. 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;2019:630-638. [FREE Full text] [Medline: [31259018](https://pubmed.ncbi.nlm.nih.gov/31259018/)]

79. Dentler K, Numans ME, ten Teije A, Cornet R, de Keizer NF. Formalization and computation of quality measures based on electronic medical records. *J Am Med Inform Assoc*. Mar 01, 2014;21(2):285-291. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2013-001921](https://doi.org/10.1136/amiajnl-2013-001921)] [Medline: [24192317](#)]
80. Kahn M, Ranade D. The impact of electronic medical records data sources on an adverse drug event quality measure. *J Am Med Inform Assoc*. 2010;17(2):185-191. [[FREE Full text](#)] [doi: [10.1136/jamia.2009.002451](https://doi.org/10.1136/jamia.2009.002451)] [Medline: [20190062](#)]
81. 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;16(9):e0257677. [[FREE Full text](#)] [doi: [10.1371/journal.pone.0257677](https://doi.org/10.1371/journal.pone.0257677)] [Medline: [34570811](#)]
82. Thuraisingam S, Chondros P, Dowsey MM, Spelman T, Garies S, Choong PF, et al. Assessing the suitability of general practice electronic health records for clinical prediction model development: a data quality assessment. *BMC Med Inform Decis Mak*. Oct 30, 2021;21(1):297. [[FREE Full text](#)] [doi: [10.1186/s12911-021-01669-6](https://doi.org/10.1186/s12911-021-01669-6)] [Medline: [34717599](#)]
83. Wang H, Belitskaya-Levy I, Wu F, Lee JS, Shih M, Tsao PS, et al. A statistical quality assessment method for longitudinal observations in electronic health record data with an application to the VA million veteran program. *BMC Med Inform Decis Mak*. Oct 20, 2021;21(1):1. [doi: [10.1186/S12911-021-01643-2](https://doi.org/10.1186/S12911-021-01643-2)]
84. Wang J, Cha J, Sebek K, McCullough C, Parsons A, Singer J, et al. Factors related to clinical quality improvement for small practices using an EHR. *Health Serv Res*. Dec 2014;49(6):1729-1746. [[FREE Full text](#)] [doi: [10.1111/1475-6773.12243](https://doi.org/10.1111/1475-6773.12243)] [Medline: [25287906](#)]
85. Kiogou SD, Chi C, Zhang R, Ma S, Adam TJ. Clinical data cohort quality improvement: the case of the medication data in the University of Minnesota's clinical data repository. *AMIA Jt Summits Transl Sci Proc*. 2022;2022:293-302. [[FREE Full text](#)] [Medline: [35854717](#)]
86. Alwhaibi M, Balkhi B, Alshammari T, AlQahtani N, Mahmoud M, Almetwazi M, et al. Measuring the quality and completeness of medication-related information derived from hospital electronic health records database. *Saudi Pharm J*. May 2019;27(4):502-506. [[FREE Full text](#)] [doi: [10.1016/j.jsps.2019.01.013](https://doi.org/10.1016/j.jsps.2019.01.013)] [Medline: [31061618](#)]
87. Tak YW, Han JH, Park YJ, Kim D, Oh JS, Lee Y. Examining final-administered medication as a measure of data quality: a comparative analysis of death data with the Central Cancer Registry in Republic of Korea. *Cancers (Basel)*. Jun 27, 2023;15(13):3371. [[FREE Full text](#)] [doi: [10.3390/cancers15133371](https://doi.org/10.3390/cancers15133371)] [Medline: [37444480](#)]
88. Upadhyay S, Hu H. A qualitative analysis of the impact of electronic health records (EHR) on healthcare quality and safety: clinicians' lived experiences. *Health Serv Insights*. 2022;15:11786329211070722. [[FREE Full text](#)] [doi: [10.1177/11786329211070722](https://doi.org/10.1177/11786329211070722)] [Medline: [35273449](#)]
89. Barkhuysen P, de Grauw W, Akkermans R, Donkers J, Schers H, Biermans M. Is the quality of data in an electronic medical record sufficient for assessing the quality of primary care? *J Am Med Inform Assoc*. 2014;21(4):692-698. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2012-001479](https://doi.org/10.1136/amiajnl-2012-001479)] [Medline: [24145818](#)]
90. Fu S, Wen A, Pagali S, Zong N, St Sauver J, Sohn S, et al. The implication of latent information quality to the reproducibility of secondary use of electronic health records. *Stud Health Technol Inform*. Jun 06, 2022;290:173-177. [[FREE Full text](#)] [doi: [10.3233/SHTI220055](https://doi.org/10.3233/SHTI220055)] [Medline: [35672994](#)]
91. Dy S, Lorenz K, O'Neill S, Asch S, Walling A, Tisnado D, et al. Cancer Quality-ASSIST supportive oncology quality indicator set: feasibility, reliability, and validity testing. *Cancer*. Jul 01, 2010;116(13):3267-3275. [[FREE Full text](#)] [doi: [10.1002/cncr.25109](https://doi.org/10.1002/cncr.25109)] [Medline: [20564637](#)]
92. Greiver M, Drummond N, Birtwhistle R, Queenan J, Lambert-Lanning A, Jackson D. Using EMRs to fuel quality improvement. *Can Fam Physician*. Jan 2015;61(1):92, e68-92, e69. [[FREE Full text](#)] [Medline: [25609529](#)]
93. Capurro D, Yetisgen M, van Eaton E, Black R, Tarczy-Hornoch P. Availability of structured and unstructured clinical data for comparative effectiveness research and quality improvement: a multisite assessment. *EGEMS (Wash DC)*. 2014;2(1):1079. [[FREE Full text](#)] [doi: [10.13063/2327-9214.1079](https://doi.org/10.13063/2327-9214.1079)] [Medline: [25848594](#)]
94. Romero L, Carneiro PB, Riley C, Clark H, Uy R, Park M, et al. Building capacity of community health centers to overcome data challenges with the development of an agile COVID-19 public health registry: a multistate quality improvement effort. *J Am Med Inform Assoc*. Dec 28, 2021;29(1):80-88. [[FREE Full text](#)] [doi: [10.1093/jamia/ocab233](https://doi.org/10.1093/jamia/ocab233)] [Medline: [34648005](#)]
95. Merino J, Caballero I, Rivas B, Serrano M, Piattini M. A data quality in use model for big data. *Future Generation Computer Systems*. Oct 2016;63:123-130. [[FREE Full text](#)] [doi: [10.1016/j.future.2015.11.024](https://doi.org/10.1016/j.future.2015.11.024)]
96. E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1). Food and Drug Administration. Mar 2018. URL: <https://www.fda.gov/media/93884/download> [accessed 2025-03-06]
97. Colditz RR, Conrad C, Wehrmann T, Schmidt M, Dech S. TiSeG: a flexible software tool for time-series generation of MODIS data utilizing the quality assessment science data set. *IEEE Trans. Geosci. Remote Sensing*. Oct 2008;46(10):3296-3308. [doi: [10.1109/tgrs.2008.921412](https://doi.org/10.1109/tgrs.2008.921412)]
98. Final NIH Policy for Data Management and Sharing. National Institutes of Health. Oct 29, 2020. URL: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html> [accessed 2025-03-06]
99. Häyriinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform*. May 2008;77(5):291-304. [[FREE Full text](#)] [doi: [10.1016/j.ijmedinf.2007.09.001](https://doi.org/10.1016/j.ijmedinf.2007.09.001)] [Medline: [17951106](#)]

100. Miran S, Nelson S, Redd D, Zeng-Treitler Q. Using multivariate long short-term memory neural network to detect aberrant signals in health data for quality assurance. *Int J Med Inform.* Mar 2021;147:104368. [FREE Full text] [doi: [10.1016/j.ijmedinf.2020.104368](https://doi.org/10.1016/j.ijmedinf.2020.104368)] [Medline: [33401168](https://pubmed.ncbi.nlm.nih.gov/33401168/)]
101. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. *Summit Transl Bioinform.* Mar 01, 2010;2010:1-5. [FREE Full text] [Medline: [21347133](https://pubmed.ncbi.nlm.nih.gov/21347133/)]
102. Torda P, Tinoco A. Achieving the promise of electronic health record-enabled quality measurement: a measure developer's perspective. *EGEMS (Wash DC).* 2013;1(2):1031. [FREE Full text] [doi: [10.13063/2327-9214.1031](https://doi.org/10.13063/2327-9214.1031)] [Medline: [25848574](https://pubmed.ncbi.nlm.nih.gov/25848574/)]
103. co. Bigdata Quality Evaluation Tool Development. In: *SW Computing Industry Source Technology Development Project 3rd Year Final Report.* Korea. Ministry of Science and ICT; Feb 14, 2020:1-283.
104. Gong X, Shroff N. Incentivizing Truthful Data Quality for Quality-Aware Mobile Data Crowdsourcing. 2018. Presented at: Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing; June 26-29, 2018; Los Angeles, CA. URL: <https://scienceon.kisti.re.kr/srch/selectPORSrchReport.do?cn=TRKO202100007277#>; [doi: [10.1145/3209582.3209599](https://doi.org/10.1145/3209582.3209599)]
105. Amoah AO, Amirfar S, Silfen SL, Singer J, Wang JJ. Applied use of composite quality measures for EHR-enabled practices. *EGEMS (Wash DC).* 2015;3(1):1118. [FREE Full text] [doi: [10.13063/2327-9214.1118](https://doi.org/10.13063/2327-9214.1118)] [Medline: [26290881](https://pubmed.ncbi.nlm.nih.gov/26290881/)]

Abbreviations

- CDM:** common data model
- DQM:** data quality management
- EHR:** electronic health record
- PDSA:** plan-do-study-act

Edited by X Ma; submitted 19.05.24; peer-reviewed by I Adeleke, D Yoon, J Lee, J Declerck; comments to author 27.09.24; revised version received 01.11.24; accepted 26.01.25; published 23.04.25

Please cite as:

An D, Lim M, Lee S

Challenges for Data Quality in the Clinical Data Life Cycle: Systematic Review

J Med Internet Res 2025;27:e60709

URL: <https://www.jmir.org/2025/1/e60709>

doi: [10.2196/60709](https://doi.org/10.2196/60709)

PMID: [40266662](https://pubmed.ncbi.nlm.nih.gov/40266662/)

©Doyeon An, Minsik Lim, Suehyun Lee. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 23.04.2025. 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 (ISSN 1438-8871), 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.