Challenges of Data Quality in Clinical Data Life Cycle: A Systematic Review

Doyeon An, Minsik Lim, Suehyun Lee

Submitted to: Journal of Medical Internet Research
on: May 19, 2024

Table of Contents

Challenges of Data Quality in Clinical Data Life Cycle: A Systematic Review

Doyeon An[1] MPH; Minsik Lim[1]; Suehyun Lee[1] PhD

[1]Department of Computer Engineering College of IT Convergence Gachon University Seongnam KR

Corresponding Author:
Suehyun Lee PhD
Department of Computer Engineering
College of IT Convergence
Gachon University
AI·Engineering Building, 317A
1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do
Seongnam
KR

Abstract

Background: It is anticipated that electronic health record (EHR) data will 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-centricity, the utilization of healthcare data is becoming 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 to assess the quality of electronic health record (EHR) and clinical data. Articles that defined quality management procedures based on the life cycle of clinical data quality management and discussed quality management assessment methods and tools were selected. The articles were categorized into four themes.

Results: We reviewed 105 papers describing the clinical data quality management process. This process is based on a four-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.

(JMIR Preprints 19/05/2024:60709)
DOI: https://doi.org/10.2196/preprints.60709

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Review

Challenges of Data Quality in Clinical Data Life Cycle: A Systematic Review

Abstract

Background: It is anticipated that electronic health record (EHR) data will 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-centricity, the utilization of healthcare data is becoming 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 to assess the quality of electronic health record (EHR) and clinical data. Articles that defined quality management procedures based on the life cycle of clinical data quality management and discussed quality management assessment methods and tools were selected. The articles were categorized into four themes.
Results: We reviewed 105 papers describing the clinical data quality management process. This process is based on a four-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.

Keywords: clinical research informatics; data quality; electronic health records; frameworks

Introduction

As data continue to accumulate, the question of how to utilize 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 a national health policy system and provide useful information for improving public health and medical technology[1]. As the current clinical research paradigm shifts to data centricity, the use of EHR data has increasingly been emphasized[2].
The quality of EHR data research depends on the quality of the generated data and 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 with different types of data. Currently, various methods for integrated data management are being developed[3-9], but quality control standards are also set differently for each data, and discussions in this regard are challenging because of the nature of EHR data[10-13].
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[14, 15]. However, no standardized approach is available to assess the quality of EHR data[16]. 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[17, 18]. However, CDMs are evaluated individually depending on their type[19-21].

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 the accurate and consistent control of data quality[22]. 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.

The current study aims 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[23] were subsequently considered.

Methods

Definition of clinical data life cycle

In the context of systematic data quality management, we defined the life cycle of clinical data quality management[23] as the quality management activities for healthcare data that include a series of steps from data construction to operation and use[23].

Literature review on data quality

We aimed to identify articles that extensively discuss the generation and quality of EHR data. To conduct the literature review (Figure 1), we followed the methods of previous studies that closely reviewed previous EHR data[16, 24-26]. 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.

'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 Table 1[16, 24-26]. 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 that they met the criteria. Subsequently, all papers related to data quality were selected and classified based on four keywords: “data quality,” “EHR assessment,” “treatment quality,” and “hospital quality.”

Table 1. Inclusion and Exclusion Criteria

Inclusion criteria	
Original research using data quality assessment methods	
Focus on data derived from EHRs[a] or related systems	
Exclusion criteria	
Guidelines limited to one medical area (e.g., cardiology) without generalization to other areas	
Review papers	Guidance aimed at governing bodies
	Published before 2010
	Non-English language paper
	Does not provide full text
Not a paper on data quality issues	Does not address data quality issues

[a]EHR: electronic health

record

To focus on data quality management for clinical data analysis, we reviewed the full text of each article containing two of the four 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 four 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 mentions data life cycle; however, finding articles offering a clear definition was difficult. Data quality must be consistently defined[27]. 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 utilized 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.

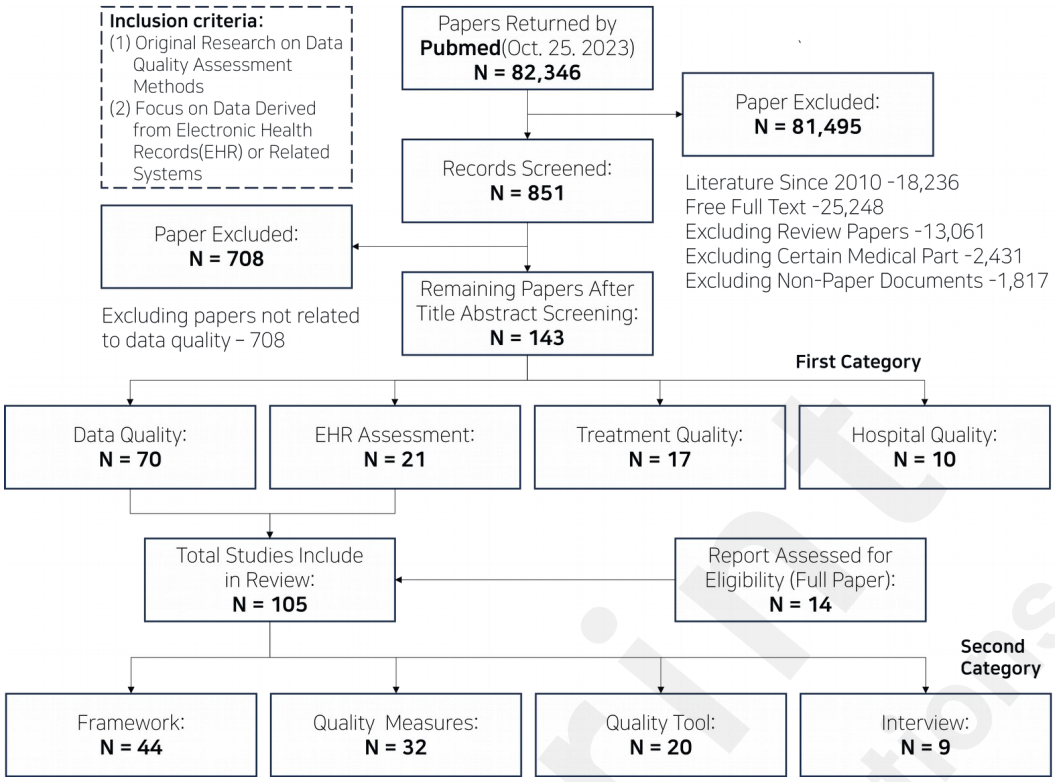


Figure 1. Diagram of the literature review process for clinical data quality management

Results

Data quality assessment framework based on 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”[28]. 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[23].

Figure 2 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[23]. In producing high-quality data, data must be managed according to the data life cycle and governance principles[23].

We established the definitions for each clinical data life cycle stage by reviewing the literature (Table 2). The literature included in the review often describes the data life cycle for improving hospital EHR quality, quality measurement, and clinical decision support[29-33].

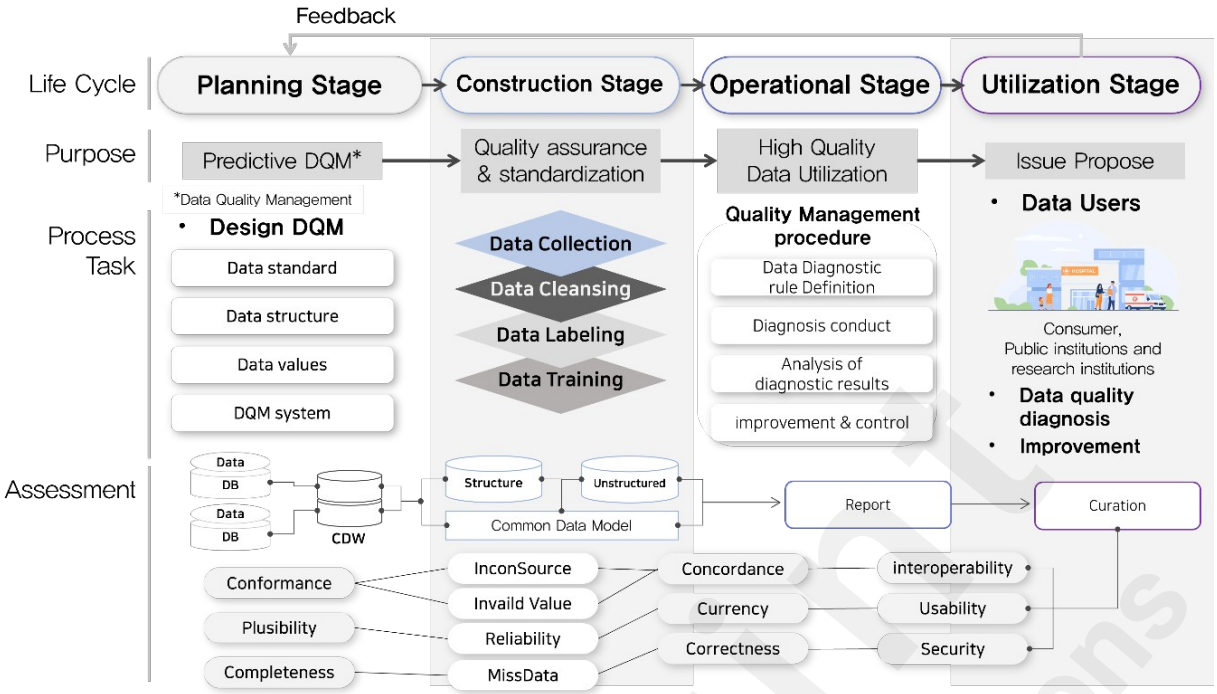


Figure 2. Life cycle of clinical data quality management

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[34]. 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[35].

Construction stage

The construction stage involves quality control. It is also called the big data life cycle stage[22] (Figure 2). This data life cycle stage consists of four 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. In 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 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[23]. 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[30, 36, 37]. 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 the current 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[38]. As data significantly influence research outcomes, they must meaningfully be evaluated and managed throughout their life cycle[27]. Some studies did not consider data from a life cycle perspective[31, 32, 39-41]. Nevertheless, they considered the ecological use of data. They also considered the impact of data on hospital treatment processes[31, 32]. Thus, data operations are organically linked, reflecting the interplay between different stages.

Table 2. Defining the Life Cycle of Clinical Data Quality Management

Life Cycle	Definition	
Planning stage	Defining data standards based on the direction of data and creating a clear strategy for establishing quality management activities	[6, 29, 30, 42]
Construction Stage	Considering the characteristics among datasets, collecting data, and proceeding with overall data construction and management that reflects clinical attributes	[29, 36, 42-47]
Operation stage	Conducting data quality assessments on the constructed data and reviewing them from various angles and perspectives	[29, 30, 42, 48, 49]
Utilization Stage	Sharing the outcomes of data quality validation, implementing data quality enhancement activities, and recalibrating the overall data quality	[29, 30, 37, 42]

Dimensions of 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 (Figure 1). Completeness was identified as the most commonly used indicator, particularly in 94 papers (Table 3).

Completeness

Completeness was mainly used in the construction or operation stage and was employed as an indicator for EHR evaluation[50, 51] data quality system development[7, 52, 53] data recognition[54], and comparative evaluation[39]. 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[55, 56], framework presentation[49, 57], data measurement[58], and data quality assessment[7, 50, 59].

Concordance

Similar to completeness and plausibility, concordance was frequently mentioned in the construction and utilization stages. Concordance can be considered as 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 on, and evaluated quality management tools[9, 40, 55, 60-63]. The related terms mentioned in the articles paper included structure and standardization.

Security

As EHR data are sensitive, great attention must be paid to ethical issues or data leakage. Therefore, the security of EHR data is crucial. In contrast to the above three 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 are the difficulty in 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 the current study revealed a strong emphasis on interoperability, but it was not mentioned in articles defining other data quality indicators.

Table 3. Life Cycle of Clinical Data Quality Management and Dimensions of Data Quality

Dimension	Completeness	Plausibility	Concordance	Security	Currency	Interoperability	Total
Definition	Assessing the extent to which data have been fully constructed in accordance with their characteristics and intended design.	Degree of reliability in data values and the significance of the associated information.	The extent to which data can be stored in accordance with their characteristics based on standards.	The extent to which data are trustworthy and accessible only to authorized users.	The extent to which data can be provided promptly when needed.	The degree to which data operation is flexible, providing a sufficient and useful level of information that satisfies users.	
	Completeness, Correctness, Conformance, Incompleteness, Consistency	Accuracy, Consistency, Relevance	Structure, Standardization	Security, Availability, Confidentiality, Representation, Confidentiality, Trustworthiness	Currency, Timeliness, Currentness	Availability, Manageability, Variability	

linical Data Life Cycle							
lanning tage	22[6, 7, 14, 15, 22, 29, 30, 47, 49, 52, 55, 56, 61, 63-71]	19[6, 7, 11, 14, 15, 22, 29, 30, 40, 47, 49, 54, 55, 61, 67, 69-72]	18[6, 7, 14, 15, 22, 29, 30, 40, 47, 49, 54, 61, 63, 65, 68, 69, 71]	8[15, 22, 29, 40, 49, 54, 64, 66, 69]	9[30, 41, 47, 49, 55, 63, 68, 69, 73]	7[30, 32, 33, 54, 68, 69, 74]	69
onstruction tage	34[7, 9, 14, 15, 22, 29, 43, 44, 47, 49-52, 55-58, 61-71, 75-81]	25[7, 9, 11, 15, 19, 22, 40, 47, 49, 50, 54-59, 61, 62, 67, 69-72, 79, 80]	22[7, 9, 15, 22, 40, 47-49, 51, 54, 56, 61-63, 65, 68, 69, 71, 79, 80]	9[15, 22, 40, 49, 54, 59, 64, 66, 69, 82]	14[11, 41, 47, 51, 54, 56, 63, 68, 69, 73, 75, 76, 83, 84]	8[32, 33, 37, 54, 56, 69, 74, 78, 85, 86]	99
peration tage	30[7, 9, 15, 22, 29, 31, 38, 39, 47, 53, 56, 61-63, 65, 69, 71, 76, 79, 80, 83, 85-91]	26[7, 9, 11, 15, 22, 37, 38, 40, 47, 49, 54, 60-62, 69, 71, 72, 79, 80, 82, 83, 85-89]	23[7, 9, 15, 22, 29, 38, 40, 47, 48, 54, 60-63, 65, 69, 71, 79, 80, 85, 86, 91]	7[15, 22, 40, 54, 60, 69, 82, 85]	10[11, 47, 54, 56, 63, 69, 76, 83-85, 91]	10[32, 33, 37, 54, 56, 69, 74, 78, 85, 86]	95
tilization tage	21[7, 14, 15, 19, 22, 29, 38, 39, 47, 50, 51, 53, 61, 69, 71, 79, 90-94]	19[7, 11, 14, 15, 22, 29, 37, 38, 47, 50, 54, 60, 61, 71, 72, 79, 92, 93]	18[7, 14, 15, 22, 29, 38, 47, 51, 54, 60, 61, 69, 71, 79, 91-93]	9[15, 22, 29, 54, 60, 69, 92-94]	9[11, 29, 47, 51, 54, 63, 69, 91-93]	10[29, 32, 33, 37, 54, 60, 69, 74, 79, 92]	72
otal	107	72	81	33	42	35	

Discussion

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 six quality indicators were identified.

Data quality ensures the validity of research findings and provides information to demonstrate the appropriateness of EHR data use[38]. In the current 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 in 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 the current 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[25, 92]. Data metrics have been developed to clearly define and automatically measure data[49]. Currency and interoperability metrics are not entirely new. They have been mentioned repeatedly in various studies [30, 41, 47, 49, 55, 63, 68, 69, 73]. Currency refers to information about current data[69] and is primarily used for

temporal information when representing the lifetime of data[93]. Temporal factors exert a significant effect on research results. In addition, currency should be considered when visualizing data quality results[46].

In 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, 14, 15, 22, 71]. 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[35]. As for ensuring the scalability of data connections, quality control for integrated data should be implemented from the planning stage using standardized procedures[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 on 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 increase the availability of data exponentially or directly impact the activation of research. EHR systems can efficiently support data structuring and quality measurement results and exert 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 for improving 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, 46, 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 users' purpose. In 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 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 healthcare 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[22].

Investing in EHR data quality management improves clinical outcomes[31]. 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, 65, 81, 105]. However, automation across all datasets 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[16]. Accurately defining the domains and task ontologies for measuring data quality in the automation process is critical[49, 65]. Various methodologies and quality criteria have been identified[26]. Nevertheless, flexible tools that consider interoperability must be developed, and existing methodologies must be employed to create a unified

automation tool[16].

Limitations

Our literature review has several limitations that need to be considered. First, the literature selection was performed by the first author alone. This approach may introduce subjectivity into the selection process, and other reviewers may not agree with the chosen classifications. Second, the literature search was conducted at a single location; hence, other relevant studies might have been missed. Third, additional reviews by other experts were not conducted. The indicators selected in this study result from a literature review and thus need to be verified by clinical experts.

Conclusion

As the value of EHR data increase, the demand for high-quality data also rise. 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.

Acknowledgements

SL and DA designed this study. DA conducted the literature review and analysis, drafted and revised the manuscript, and prepared the tables and figures. ML interpreted the literature and drafted the manuscript. All authors reviewed and revised the manuscript and approved the final version for submission.

Conflicts of Interest

No potential conflicts of interest are disclosed.

Abbreviations

EHR: Electronic Health Record
PDSA: Plan–Do–Study–Act

Funding

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (No. 2021R1I1A3044287141298207687010 4).

Multimedia Appendix 1

Multimedia Appendix 1. Data search list

References

1. Kwon T, Jeong Y, Lee D. Standardization and quality evaluation of health and medical big data. Osong: Korea Health Industry Development Institute, 2019.
2. Myung S, Lee S-h. Current Status and Issues of Data Management Plan in Korea. Journal of The Korea Contents Association. 2020;20(6):220-9. doi: https://doi.org/10.5392/JKCA.2020.20.06.220.
3. Tute E, Scheffner I, Marschollek M. A method for interoperable knowledge-based data quality assessment. BMC Medical Informatics and Decision Making. 2021 2021/03/09;21(1):93. doi: 10.1186/s12911-021-01458-1.

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;2016:1229-37. PMID: 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). 2018 May 22;6(1):8. PMID: 29881766. doi: 10.5334/egems.211.
6. Khare R, Utidjian L, Ruth BJ, Kahn MG, Burrows E, Marsolo K, et al. A longitudinal analysis of data quality in a large pediatric data research network. Journal of the American Medical Informatics Association. 2017;24(6):1072-9. doi: 10.1093/jamia/ocx033.
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. 2021 Aug;12(4):826-35. PMID: 34433217. doi: 10.1055/s-0041-1733847.
8. Chiang J-H, Lin J-W, Yang C-W. Automated evaluation of electronic discharge notes to assess quality of care for cardiovascular diseases using Medical Language Extraction and Encoding System (MedLEE). Journal of the American Medical Informatics Association. 2010;17(3):245-52. doi: 10.1136/jamia.2009.000182.
9. Mang JM, Seuchter SA, Gulden C, Schild S, Kraska D, Prokosch H-U, et al. DQAgui: a graphical user interface for the MIRACUM data quality assessment tool. BMC Medical Informatics and Decision Making. 2022 2022/08/11;22(1):213. doi: 10.1186/s12911-022-01961-z.
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;2022:775-84. PMID: 37128433.
11. Davoudi S, Dooling J, Glondys B, Jones T, Kadlec L, Overgaard S, et al. Data Quality Management Model (Updated). Journal of AHIMA. 2015 10/01;86:62-5.
12. Administration FaD. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products Draft Guidance for Industry. In: Services USDoHaH, editor.: Food and Drug Administration; 2021.
13. Madnick SE, Wang RY, Lee YW, Zhu H. Overview and Framework for Data and Information Quality Research. J Data and Information Quality. 2009;1(1):Article 2. doi: 10.1145/1515693.1516680.
14. Safety MotIa, Agency NIS. Public data preventive quality management guide. In: Policy DoPD, editor. Sejong: Ministry of the Interior and Safety; 2021.
15. Agency NIS. Public Data Quality Management Manual ver 2.0. Daegu: National Information Society Agency; 2018.
16. 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. Journal of the American Medical Informatics Association. 2023;30(10):1730-40. doi: 10.1093/jamia/ocad120.
17. Center SO. Data Quality Review and Characterization Programs. Search Sentinel Common Data Model. 3.3.1 ed. Sentinel Operations Center: U.S. Food and Drug Administration; 2023.
18. 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. PMID: 25848597. doi: 10.13063/2327-9214.1110.
19. 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. PMID: 28154833. doi: 10.13063/2327-9214.1239.

20. Estiri H, Stephens KA, Klann JG, Murphy SN. Exploring completeness in clinical data research networks with DQe-c. Journal of the American Medical Informatics Association. 2017;25(1):17-24. doi: 10.1093/jamia/ocx109.
21. 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. Journal of the American Medical Informatics Association. 2020;27(12):1999-2010. doi: 10.1093/jamia/ocaa245.
22. Lee Y, Son K-m, Yoo S, Kim E, Co. C. Big Data Platform and Center Data Quality Management Guide. In: platform DoB, editor. Daegu: National Information Society Agency; 2022. p. 7-143.
23. Lee S, Roh G-H, Kim J-Y, Ho Lee Y, Woo H, Lee S. Effective data quality management for electronic medical record data using SMART DATA. International Journal of Medical Informatics. 2023 2023/12/01/;180:105262. doi: https://doi.org/10.1016/j.ijmedinf.2023.105262.
24. Arts DGT, de Keizer NF, Scheffer G-J. Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. Journal of the American Medical Informatics Association. 2002;9(6):600-11. doi: 10.1197/jamia.M1087.
25. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the American Medical Informatics Association. 2013;20(1):144-51. doi: 10.1136/amiajnl-2011-000681.
26. 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 Digital Medicine. 2022 2022/01/10;5(1):2. doi: 10.1038/s41746-021-00549-7.
27. Liaw S-T, Guo JGN, Ansari S, Jonnagaddala J, Godinho MA, Borelli AJ, Jr, et al. Quality assessment of real-world data repositories across the data life cycle: A literature review. Journal of the American Medical Informatics Association. 2021;28(7):1591-9. doi: 10.1093/jamia/ocaa340.
28. English LP, editor. Information quality management: The next frontier. ASQ World Conference on Quality and Improvement Proceedings; 2001: American Society for Quality.
29. Winter A, Takabayashi K, Jahn F, Kimura E, Engelbrecht R, Haux R, et al. Quality Requirements for Electronic Health Record Systems. A Japanese-German Information Management Perspective. 2017 2018/01/31;56(S 01):e92-e104. doi: 10.3414/ME17-05-0002.
30. 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. 2022 2022/09/07;13(04):836-44. doi: 10.1055/s-0042-1756369.
31. 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. 2010 Aug;16(8):601-6. PMID: 20712393.
32. Ramirez AH, Sulieman L, Schlueter DJ, Halvorson A, Qian J, Ratsimbazafy F, et al. The All of Us Research Program: Data quality, utility, and diversity. Patterns. 2022 2022/08/12/;3(8):100570. doi: https://doi.org/10.1016/j.patter.2022.100570.
33. 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;2015:843-51. PMID: 26958220.
34. Holles JH, Schmidt L, editors. Graduate research data management course content: teaching the Data Management Plan (DMP). 2018 ASEE Annual Conference & Exposition;

2018.
35. Michener WK. Ten simple rules for creating a good data management plan. PLoS computational biology. 2015;11(10):e1004525.
36. 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. 2014 Aug;26(4):411-7. PMID: 24925685. doi: 10.1093/intqhc/mzu059.
37. Carroll AR, Johnson DP. Know It When You See It: Identifying and Using Special Cause Variation for Quality Improvement. Hospital Pediatrics. 2020;10(11):e8-e10. doi: 10.1542/hpeds.2020-002303.
38. 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. Journal of Comparative Effectiveness Research. 2016;5(4):345-54. doi: doi:10.2217/cer-2015-0022.
39. 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. PMID: 24454578. doi: 10.4338/aci-2013-07-ra-0046.
40. 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 Neurology. 2013 2013/06/15;13(1):61. doi: 10.1186/1471-2377-13-61.
41. Kvale MN, Hesselson S, Hoffmann TJ, 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. 2015;200(4):1051-60. doi: 10.1534/genetics.115.178905.
42. Chelico JD, Wilcox AB, Vawdrey DK, Kuperman GJ. Designing a Clinical Data Warehouse Architecture to Support Quality Improvement Initiatives. AMIA Annu Symp Proc. 2016;2016:381-90. PMID: 28269833.
43. Randall SM, Ferrante AM, Boyd JH, Semmens JB. The effect of data cleaning on record linkage quality. BMC Medical Informatics and Decision Making. 2013 2013/06/05;13(1):64. doi: 10.1186/1472-6947-13-64.
44. Schepens MHJ, Trompert AC, van Hooff ML, van der Velde E, Kallewaard M, Verberk-Jonkers I, et al. Using Existing Clinical Information Models for Dutch Quality Registries to Reuse Data and Follow COUMT Paradigm. Appl Clin Inform. 2023 Mar;14(2):326-36. PMID: 37137338. doi: 10.1055/s-0043-1767681.
45. 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;2014:825-34. PMID: 25954389.
46. 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;2022:186-95. PMID: 35854725.
47. Diaz-Garelli JF, 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). 2019 Jul 25;7(1):32. PMID: 31367649. doi: 10.5334/egems.286.
48. Fife CE, Walker D, Thomson B. Electronic Health Records, Registries, and Quality

Measures: What? Why? How? Adv Wound Care (New Rochelle). 2013 Dec;2(10):598-604. PMID: 24761335. doi: 10.1089/wound.2013.0476.
49. 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;2015:1937-46. PMID: 26958293.
50. 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. 2021 2021/02/02;21(1):264. doi: 10.1186/s12889-021-10295-w.
51. 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;4(3). doi: 10.1093/jamiaopen/ooab057.
52. 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? Journal of Pediatric Rehabilitation Medicine. 2018;11:303-9. doi: 10.3233/PRM-170525.
53. 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;2019:630-8. PMID: 31259018.
54. 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). 2017 Sep 4;5(1):16. PMID: 29881736. doi: 10.5334/egems.214.
55. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. Quantifying the Effect of Data Quality on the Validity of an eMeasure. Appl Clin Inform. 2017 Oct;8(4):1012-21. PMID: 29241241. doi: 10.4338/aci-2017-03-ra-0042.
56. 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 2020/09/23;11(04):622-34. doi: 10.1055/s-0040-1715567.
57. Winnenburg R, Bodenreider O. Metrics for assessing the quality of value sets in clinical quality measures. AMIA Annu Symp Proc. 2013;2013:1497-505. PMID: 24551422.
58. Gadde MA, Wang Z, Zozus M, Talburt JB, Greer ML. Rules Based Data Quality Assessment on Claims Database. Stud Health Technol Inform. 2020 Jun 26;272:350-3. PMID: 32604674. doi: 10.3233/shti200567.
59. Barkhuysen P, de Grauw W, Akkermans R, Donkers J, Schers H, Biermans M, et al. Is the quality of data in an electronic medical record sufficient for assessing the quality of primary care? Journal of the American Medical Informatics Association. 2013;21(4):692-8. doi: 10.1136/amiajnl-2012-001479.
60. 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. 2022 Jun 6;290:173-7. PMID: 35672994. doi: 10.3233/shti220055.
61. Razzaghi H, Greenberg J, Bailey LC. Developing a systematic approach to assessing data quality in secondary use of clinical data based on intended use. Learning Health Systems. 2022;6(1):e10264. doi: https://doi.org/10.1002/lrh2.10264.
62. 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). 2019 Apr 23;7(1):17. PMID: 31065558. doi: 10.5334/egems.289.
63. 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. PMID: 35854735.
64. 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. 2019 Mar;10(2):199-209. PMID: 30895574. doi: 10.1055/s-0039-1681054.
65. Wu Q, Ganz C, Li L. Data Quality Control for Electronic Pathology Reporting. J Registry Manag. 2022 Fall;49(3):95-6. PMID: 37260925.
66. 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. PMID: 25848565. doi: 10.13063/2327-9214.1025.
67. Huser V, Williams ND, Mayer CS. Linking Provider Specialty and Outpatient Diagnoses in Medicare Claims Data: Data Quality Implications. Appl Clin Inform. 2021 Aug;12(4):729-36. PMID: 34348410. doi: 10.1055/s-0041-1732404.
68. 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 Medical Informatics and Decision Making. 2016 2016/06/04;16(1):61. doi: 10.1186/s12911-016-0305-4.
69. 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. 2022;24(12):e41889. PMID: 36472901. doi: 10.2196/41889.
70. 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. 2019 Jul 29;4(2):1132. PMID: 34095540. doi: 10.23889/ijpds.v4i2.1132.
71. ICT MoSa, Agency NIS. Data Quality Management Guidelines v2.0 for Artificial Intelligence Learning - Quality Management Guide. In: Division ID, editor. Daegu: National Information Society Agency; 2022.
72. Upadhyay S, Hu HF. 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. PMID: 35273449. doi: 10.1177/11786329211070722.
73. Greiver M, Drummond N, Birtwhistle R, Queenan J, Lambert-Lanning A, Jackson D. Using EMRs to fuel quality improvement. Can Fam Physician. 2015 Jan;61(1):92, e68-9. PMID: 25609529.
74. 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. 2021 Dec 28;29(1):80-8. PMID: 34648005. doi: 10.1093/jamia/ocab233.
75. 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 Medical Informatics and Decision Making. 2022 2022/02/10;22(1):37. doi: 10.1186/s12911-022-01763-3.
76. Keegan NM, Vasselman SE, Barnett ES, Nweji B, Carbone EA, Blum A, et al. Clinical annotations for prostate cancer research: Defining data elements, creating a reproducible analytical pipeline, and assessing data quality. The Prostate. 2022;82(11):1107-16. doi: https://doi.org/10.1002/pros.24363.
77. Dixon BE, Siegel JA, Oemig TV, Grannis SJ. Electronic Health Information Quality Challenges and Interventions to Improve Public Health Surveillance Data and Practice.

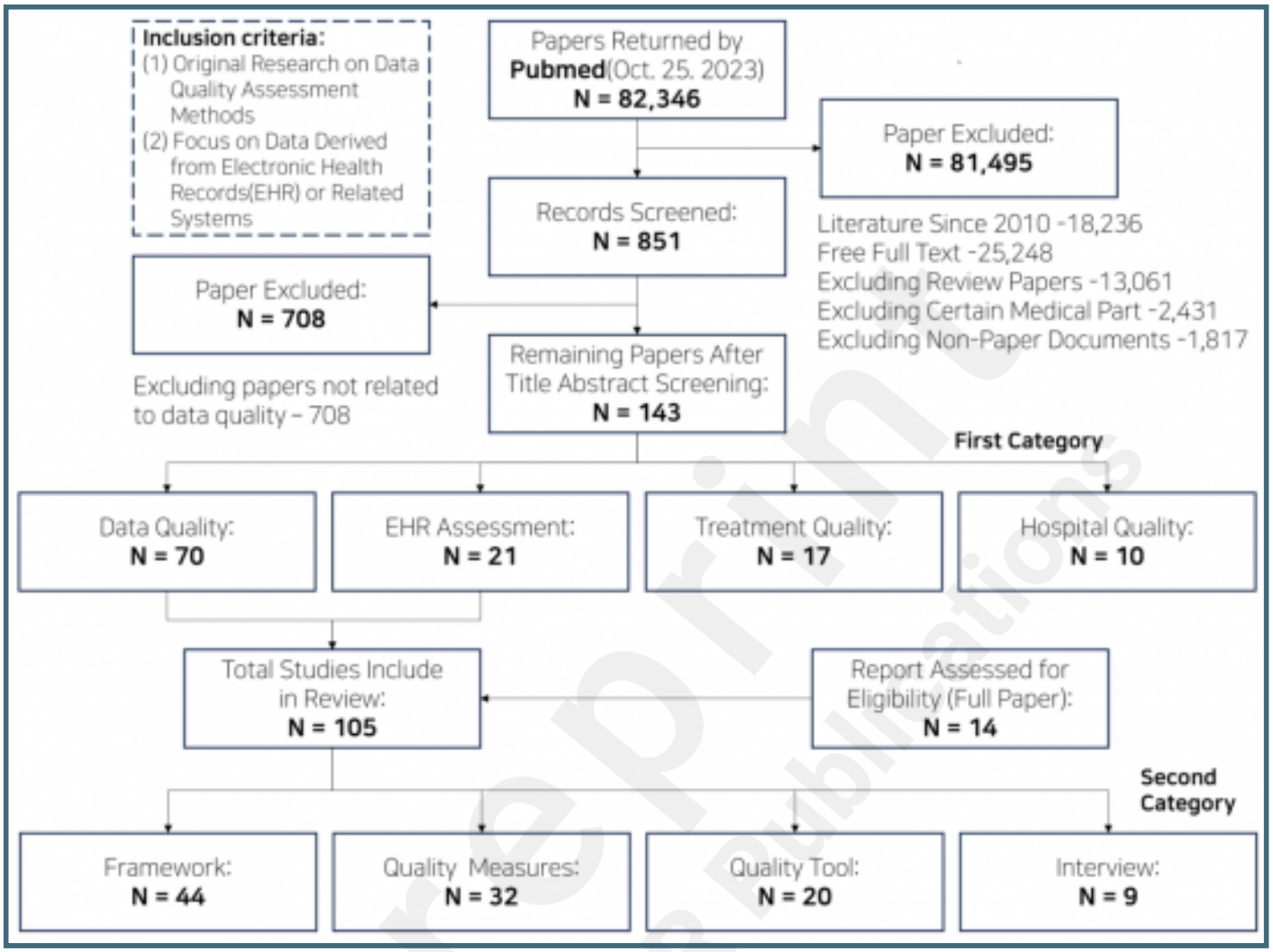
Public Health Reports. 2013;128(6):546-53. PMID: 24179266. doi: 10.1177/003335491312800614.
78. Cusick MM, Sholle ET, Davila MA, Kabariti J, Cole CL, Campion TR, Jr. A Method to Improve Availability and Quality of Patient Race Data in an Electronic Health Record System. Appl Clin Inform. 2020 Oct;11(5):785-91. PMID: 33241548. doi: 10.1055/s-0040-1718756.
79. Arvidsson E, Dijkstra R, Klemenc-Ketiš Z. Measuring Quality in Primary Healthcare - Opportunities and Weaknesses. Zdr Varst. 2019 Sep;58(3):101-3. PMID: 31275436. doi: 10.2478/sjph-2019-0013.
80. Lewis JT, Stephens J, Musick B, Brown S, Malateste K, Ha Dao Ostinelli C, et al. The IeDEA harmonist data toolkit: A data quality and data sharing solution for a global HIV research consortium. Journal of Biomedical Informatics. 2022 2022/07/01/;131:104110. doi: https://doi.org/10.1016/j.jbi.2022.104110.
81. 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. 2012;7(4):e35087. doi: 10.1371/journal.pone.0035087.
82. Dy SM, Lorenz KA, O'Neill SM, Asch SM, Walling AM, Tisnado D, et al. Cancer Quality-ASSIST supportive oncology quality indicator set. Cancer. 2010;116(13):3267-75. doi: https://doi.org/10.1002/cncr.25109.
83. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. Application of An Ontology for Characterizing Data Quality For a Secondary Use of EHR Data. Appl Clin Inform. 2016;7(1):69-88. PMID: 27081408. doi: 10.4338/aci-2015-08-ra-0107.
84. 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. PMID: 25848594. doi: 10.13063/2327-9214.1079.
85. Dentler K, Numans ME, ten Teije A, Cornet R, de Keizer NF. Formalization and computation of quality measures based on electronic medical records. Journal of the American Medical Informatics Association. 2013;21(2):285-91. doi: 10.1136/amiajnl-2013-001921.
86. Kahn MG, Ranade D. The impact of electronic medical records data sources on an adverse drug event quality measure. Journal of the American Medical Informatics Association. 2010;17(2):185-91. doi: 10.1136/jamia.2009.002451.
87. 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. PMID: 34570811. doi: 10.1371/journal.pone.0257677.
88. 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 Medical Informatics and Decision Making. 2021 2021/10/30;21(1):297. doi: 10.1186/s12911-021-01669-6.
89. Wang H, Belitskaya-Levy I, Wu F, Lee JS, Shih M-C, 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 Medical Informatics and Decision Making. 2021 2021/10/20;21(1):289. doi: 10.1186/s12911-021-01643-2.
90. Wang JJ, Cha J, Sebek KM, McCullough CM, Parsons AS, Singer J, et al. Factors Related to Clinical Quality Improvement for Small Practices Using an EHR. Health Services Research. 2014;49(6):1729-46. doi: https://doi.org/10.1111/1475-6773.12243.
91. Kiogou SD, Chi CL, 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. PMID: 35854717.
92. 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 Pharmaceutical Journal. 2019 2019/05/01/;27(4):502-6. doi: https://doi.org/10.1016/j.jsps.2019.01.013.
93. 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(0):S21-9. PMID: 22692254. doi: 10.1097/MLR.0b013e318257dd67.
94. Tak YW, Han JH, Park YJ, Kim D-H, 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. 2023;15(13):3371. PMID: doi:10.3390/cancers15133371.
95. Merino J, Caballero I, Rivas B, Serrano M, Piattini M. A Data Quality in Use model for Big Data. Future Generation Computer Systems. 2016 2016/10/01/;63:123-30. doi: https://doi.org/10.1016/j.future.2015.11.024.
96. Services USDoHaH. E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1) Guidance for Industry. OMB Control: Food and Drug Administration, 2018.
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 Transactions on Geoscience and Remote Sensing. 2008;46(10):3296-308. doi: 10.1109/TGRS.2008.921412.
98. Health NIo. Final NIH policy for data management and sharing. In: Director NGFOoT, editor. Bethesda, MD: National Institutes of Health; 2020.
99. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. International Journal of Medical Informatics. 2008 2008/05/01/;77(5):291-304. doi: https://doi.org/10.1016/j.ijmedinf.2007.09.001.
100. Miran SM, Nelson SJ, Redd D, Zeng-Treitler Q. Using multivariate long short-term memory neural network to detect aberrant signals in health data for quality assurance. International Journal of Medical Informatics. 2021 2021/03/01/;147:104368. doi: https://doi.org/10.1016/j.ijmedinf.2020.104368.
101. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. Summit Transl Bioinform. 2010 Mar 1;2010:1-5. PMID: 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. PMID: 25848574. doi: 10.13063/2327-9214.1031.
103. co. w. Bigdata Quality Evaluation Tool Development. In: IITP IfIcTPe, editor. Korea: Ministry of Science and ICT; 2020. p. 1-283.
104. Gong X, Shroff N. Incentivizing Truthful Data Quality for Quality-Aware Mobile Data Crowdsourcing. Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing; Los Angeles, CA, USA: Association for Computing Machinery; 2018. p. 161–70.
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. PMID: 26290881. doi: 10.13063/2327-9214.1118.

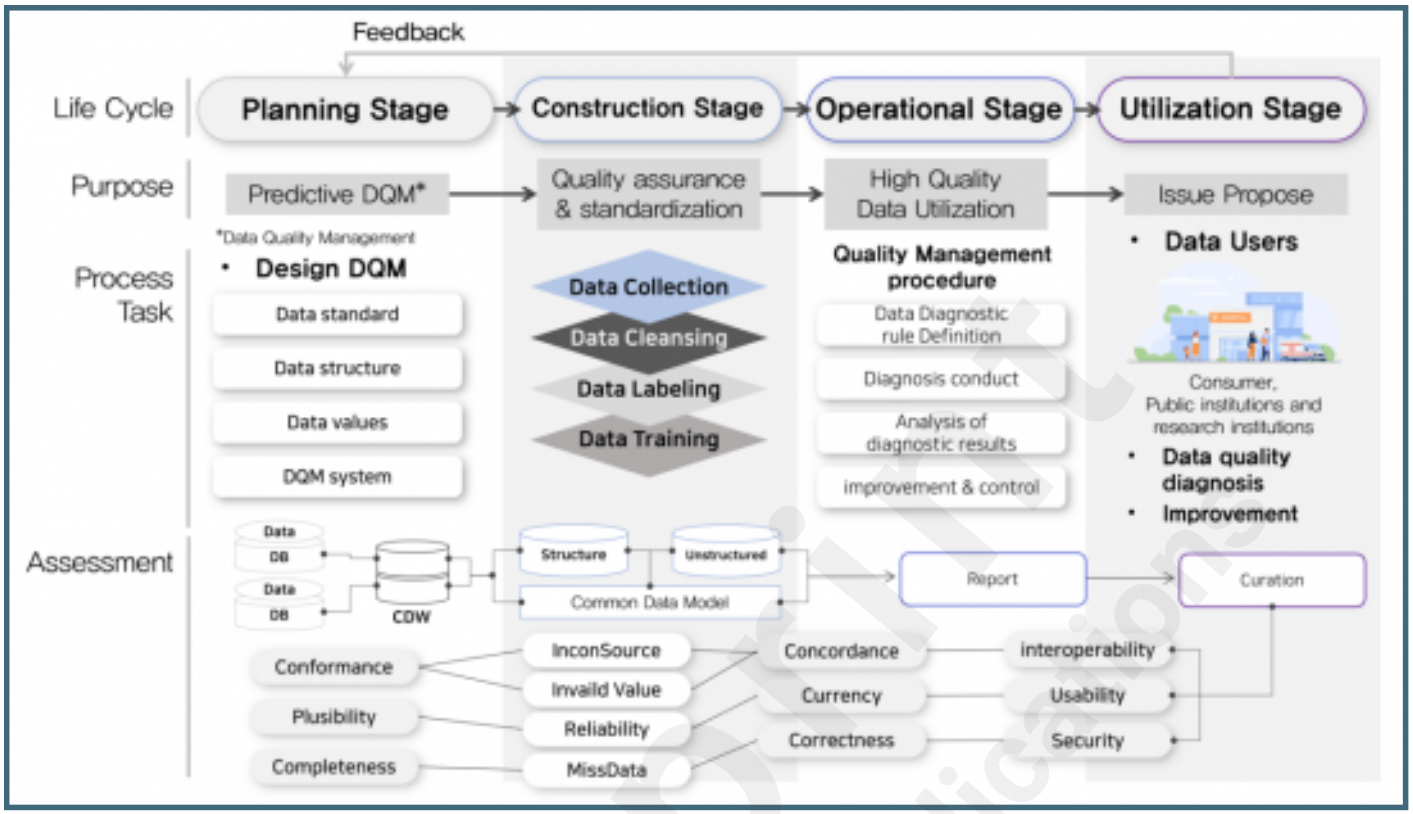
Supplementary Files

Figures

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



Life cycle of clinical data quality management.



Multimedia Appendixes

Data Search List.
URL: http://asset.jmir.pub/assets/1b97139b651453a71f577932a3cf3d89.zip