

SYSTEMATIC REVIEW

Open Access



# Common data quality elements for health information systems: a systematic review

Hossein Ghalavand<sup>1\*</sup>, Saied Shirshahi<sup>2</sup>, Alireza Rahimi<sup>2</sup>, Zarrin Zarrinabadi<sup>1</sup> and Fatemeh Amani<sup>3</sup>

## Abstract

**Background** Data quality in health information systems has a complex structure and consists of several dimensions. This research conducted for identify Common data quality elements for health information systems.

**Methods** A literature review was conducted and search strategies run in Web of Knowledge, Science Direct, Emerald, PubMed, Scopus and Google Scholar search engine as an additional source for tracing references. We found 760 papers, excluded 314 duplicates, 339 on abstract review and 167 on full-text review; leaving 58 papers for critical appraisal.

**Results** Current review shown that 14 criteria are categorized as the main dimensions for data quality for health information system include: Accuracy, Consistency, Security, Timeliness, Completeness, Reliability, Accessibility, Objectivity, Relevancy, Understandability, Navigation, Reputation, Efficiency and Value- added. Accuracy, Completeness, and Timeliness, were the three most-used dimensions in literature.

**Conclusions** At present, there is a lack of uniformity and potential applicability in the dimensions employed to evaluate the data quality of health information system. Typically, different approaches (qualitative, quantitative and mixed methods) were utilized to evaluate data quality for health information system in the publications that were reviewed. Consequently, due to the inconsistency in defining dimensions and assessing methods, it became imperative to categorize the dimensions of data quality into a limited set of primary dimensions.

**Keywords** Data quality, Health Information System, Systematic review

## Background

Appropriate planning in the health sector relies on the existence of accurate data and the quality of the data must be continuously controlled. The World Health Organization has tried to ensure the quality of health data by providing a toolkit. This toolkit supports countries to assess and improve the quality of health data [1, 2].

The existence of accurate, complete, and timely data plays an important role in health care management [3–5]. Data quality is often only considered a component of the effectiveness of health information systems, and hiding the value of data quality in other parts of the health field can lead to incorrect decision-making [6–9]. Previous studies have confirmed that data quality

\*Correspondence:  
Hossein Ghalavand  
Hosseinghalavand@gmail.com

<sup>1</sup>Department of Medical library and Information Science, Abadan University of Medical Sciences, Abadan, Iran

<sup>2</sup>Department of Medical library and Information Science, School of Health Management and Information Sciences, Isfahan University of Medical Sciences, Isfahan, Iran

<sup>3</sup>Student Research Committee, Abadan University of Medical Sciences, Abadan, Iran



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

is a multidimensional concept. Data quality assessment requires familiarity with different subjective and objective criteria and both subjective perceptions of people and objective measurements of information must be addressed [10, 11]. Qualitative evaluations of subjective data reflect the needs and experiences of stakeholders, and objective evaluations reflect the needs of managers and stakeholders [12].

Adverse effects on the quality of care, increasing costs, creating liability risks, and reducing the benefits of investing in health information systems can be identified as the negative effects of poor-quality data [13–16]. Defects in data quality can lead to incorrect diagnosis and intervention in health care [4, 13, 17, 18]. The quality of healthcare depends on the existence of quality data, which ultimately leads to a significant impact on customer satisfaction [13, 19].

Data quality in health information systems has a complex structure and consists of several dimensions and some critical factors performance such as environmental and organizational, technical and behavioral affected on data quality in health information system [20–22]. As we mentioned later, previous studies have sporadically reported some data quality elements in health information systems. There is no comprehensive agreement on its dimensions and there is no unique accepted definition of data quality among researchers for health information systems. However, there is still a lack of a review compiling and synthesizing all elements introduced in the literature. In this study, a more comprehensive understanding of the elements for quality of data in health information systems has been done using a systematic review method. The findings of this study can provide opportunities for health policy maker to become familiar with various data quality elements in health information. This systematic review specifically answered the following research questions:

- 1- What are the common data quality elements for health information systems?
- 2- What are the roles of common data quality elements to improve the performance of health information systems?

## Methods

In this review, we used a systematic approach to retrieve the relevant research studies. Our reporting strategy follows the PRISMA guidelines [23].

### Eligibility criteria

In this study the inclusion criteria were: (1) Data quality components were showcased within a health information system; (2) published from the year 2003 to 2024; (3) empirical studies that answered the research questions or tested the hypothesis and conducted on specific health system. The exclusion criteria were: (1) Research that did not outline data quality dimensions in health management systems; (2) Content presented in a format other than a scientific article such as Conference papers, book sections, and ...; (4) Methodologies deemed to be deficient in terms of quality; (5) Publication language not in English; and (7) The full text was unavailable.

### Information sources

The literature search was conducted between September and October 2023, using the following five electronic scientific databases: Web of Knowledge, Science Direct, Emerald, PubMed, Scopus and Google Scholar search engine as an additional source for tracing references.

### Search strategy

This study used a systematized review approach to identify common data quality elements for health information systems. The following keywords were used in the search strategy: Data quality, Health, clinic, Hospital, Medical, Information system. The keywords chosen were searched using various combinations and in the fields of title, abstract, subject, and keyword. We considered the search features in each database and used the Boolean operators (AND, OR) to combine and search selected keywords. An example of the search strategy was given in Table 1.

### Study selection

All the results were imported into EndNote reference management software. The duplicate and non-journal papers were removed. Next, the title and abstract of the remaining articles were screened to detect subject relevance with the research objectives. The selected articles were analyzed based on the inclusion and exclusion

**Table 1** Example of search strategies

Web of Knowledge	TS= (((data quality) AND (health* OR clinic* OR hospital OR medical) AND (information system*)) AND LANGUAGE: (English) AND Indexes = SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI
Science Direct	Title, abstract, keywords: (data quality) AND (health* OR clinic* OR hospital OR medical) AND (information system*)
Emerald	(content-type: article) AND (data quality AND (health* OR clinic* OR hospital OR medical) AND (information system*))
PubMed	(((data quality[Title/Abstract]) AND (health*[Title/Abstract] OR clinic*[Title/Abstract] OR hospital[Title/Abstract] OR medical[Title/Abstract])) AND (information system*[Title/Abstract]))
Scopus	TITLE-ABS-KEY (data quality) AND (health* OR clinic* OR hospital OR medical) AND (information system*))

criteria. Finally, the reference lists of all identified articles were searched for additional studies. Two researchers undertook the screening of titles and abstracts obtained through the searches. A sample of just over 20% of articles was double screened in order to assess the level of agreement between the researchers. Disagreements were resolved through discussion or consultation with a third researcher.

#### Data collection process

Data extraction was completed independently by two assessors. The data were extracted from including four sections: bibliographic information, methodology, and the data quality elements investigated, and key findings. Each study was treated as a single unit of analysis and the relevant information in each study was extracted using a designated data extraction form.

#### Data items

Information was extracted from each included study (including first author, title, publication date, type of study, methodology, processes of knowledge management that were studied and selected results). We emphasize the results of selected papers that have reported elements for assessment data quality in health information systems.

#### Risk of bias in individual studies

In this study, we used the Joanna Briggs Institute (JBI) checklist [24] for quality assessment. The authors assessed the included studies with a further random examination by two independent reviewers. The results of the quality assessment were compared any disagreements between the reviewers were addressed through discussion or by involving a third reviewer.

#### Synthesis of results

In this review, by adopting similar identifies elements as broader themes, the results of the included studies were analyzed and categorized. Finally, the homogeneous data quality elements in health information systems were synthesized and described.

#### Risk of bias within studies

The JBI checklist was applied to all 58 studies; none were excluded based on quality assessment and all studies were rated as unclear or high risk of bias. In 16% of studies, we cannot find "statement locating the researcher culturally or theoretically" and in 37%, "influence of the researcher on the research" is not addressed.

## Results

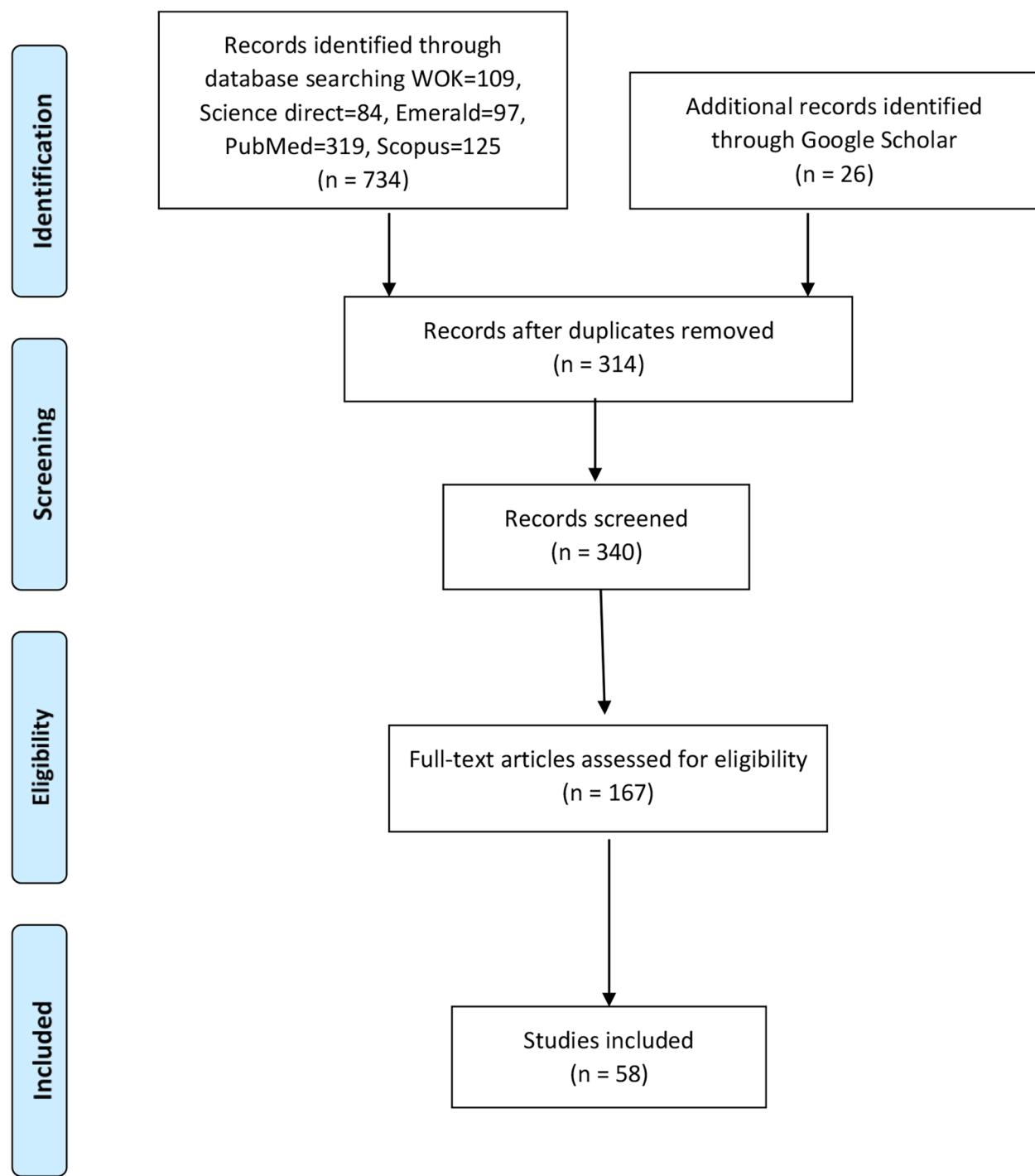
The search for systematic reviews identified 734 references published between 2003 and 2024. Title and abstract review selected 167 references for full text review. In the analysis, it was found that 68 papers did not address research questions or test hypotheses, 32 papers lacked discussion on data quality dimensions in health management systems, and nine documents presented content in a format other than a scientific article.

Out of the 58 selected paper for final review, 42 were released between 2013 and 2024 [1, 4, 5, 7–11, 14–18, 21, 22, 25–53]. Thirteen papers looked at information quality [7, 11, 14, 27–29, 31, 37, 52, 54–56], five at content quality [7, 15, 21, 43, 50], and thirty-six at data quality [4, 5, 10, 14, 17, 20, 21, 27–29, 31–33, 36, 37, 42–44, 47, 49–53, 55, 57–60]. None of the publications, however, made a distinction between "data" and "information," or between "data quality" and "information quality." As a result, "information quality" and "data quality" were used synonymously [21]. The search results and the study selection process are presented in Fig. 1.

Evaluating the quality of the data was the primary goal of the reviewed studies [4, 5, 10, 13–15, 17–21, 27–33, 35–39, 41–45, 47–53, 55–57, 59–66]. Two paper focused on information quality in health systems [11, 52]. Methods for evaluating the quality of data were presented in eight publications [10, 20, 21, 35, 38, 41, 51, 52], 19 publications tended to conduct on the health information [5, 8, 10, 11, 16, 17, 20–22, 26, 31, 37, 42, 47, 49–51, 55, 57, 60, 66] and eight paper focus on health or medical records as an information system in health context [13, 19, 25, 38, 44, 45, 64, 67].

To describe data quality, the studies employed a total of 57 dimensions. The first data quality attribute for health information system that was most often used was accuracy [4, 5, 15, 17, 19, 28, 29, 32–34, 37, 41, 43, 45, 46, 49, 51, 53, 59], second is completeness [4, 5, 20, 28–30, 41, 44–46, 48, 49, 51–53, 56], and third most-frequently criterion is timeliness [5, 28, 41, 44, 45, 51]. Table 2 displays the common dimensions of data quality in health information systems that derived from existing literature.

Data accuracy measures the extent to which information accurately represents the objects or events. The accuracy of the information that is gathered, utilized, and stored is assessed through data accuracy. It is imperative for records to serve as a dependable source of information and to facilitate the generation of valuable insights through analysis. Maintaining high data accuracy guarantees that records and datasets meet the standards for reliability and trustworthiness, allowing for their use in decision-making and various applications [4, 5, 17, 28, 29, 32, 34]. Correctness, precision, free of error, validity, believability and integrity are common terms that use for describe data accuracy [21]. Data believability relates to



**Fig. 1** Flow diagram of study selection process

whether the data is regarded as being true, real, and credible. Data believability is based on user's perceptions [1, 36, 40].

Data consistency is the state in which all copies or instances of data are identical across various information systems. This uniformity is crucial in maintaining the

accuracy, currency, and coherence of data across different platforms and applications. It is essential for instilling trust in users accessing the data. Implementing data validation rules, employing data standardization techniques, and utilizing data synchronization processes are some strategies to uphold data consistency. By ensuring data

**Table 2** The common elements of data quality in health information systems

Dimension	Definitions	Sample of studies
Accuracy	data are correct, reliable and free of error	[4, 5, 17, 28, 29]
Consistency	data is presented in the same format and compatible with previous data	[45, 48, 51, 65]
Security	access to data is restricted appropriately to maintain its security	[5, 48, 56]
Timeliness	the data is sufficiently up-to-date for the task at hand	[28, 41, 44, 45, 51]
Completeness	data is not missing and is of sufficient for the task at hand	[4, 28–30, 44]
Reliability	data is correct and reliable	[49, 53, 57, 59, 65]
Accessibility	data is available, or easily and quickly retrievable	[4, 26, 29, 50, 57]
Objectivity	data is unbiased, unprejudiced and impartial	[36, 41, 44–46]
Relevancy	data is applicable and helpful for the task at hand	[29, 45, 48, 51]
Understandability	data are clear without ambiguity and easily comprehended	[34, 37, 46]
Navigation	data are easily found and linked to other data	[4, 50, 65]
Reputation	data is highly regarded in terms of source or content	[42, 60, 61]
Efficiency	data are able to quickly meet the needs for the task at hand	[7, 28, 29, 48]
Value- added	data is beneficial, provides advantages from its use	[5, 45]

consistency, organizations can provide users with reliable information for making informed decisions, streamline operations, minimize errors, and enhance efficiency [9, 45, 48, 51, 52, 65].

Data security is the practice of protecting information from corruption, theft, or unauthorized access throughout its life cycle. This involves safeguarding hardware, software, storage devices, and user devices, as well as implementing access controls, administrative controls, and organizational policies. By utilizing tools and technologies that enhance visibility of data usage, such as data masking, encryption, and redaction, organizations can ensure the security of their data. Moreover, data security assists organizations in streamlining auditing procedures and complying with data protection regulations, ultimately reducing the risk of cyber-attacks, human error, and insider threats [5, 48, 56]. Secure access, safe, confidentiality and privacy are common terms that use for describe data security [21].

Data timeliness denotes the currency and availability of data at the required time for its intended use. This is critical for enabling health organizations to make swift and accurate decisions based on the most up-to-date information. The timeliness of data has an impact on data quality as it determines the reliability and usefulness of information systems. Moreover, timely data can lead to cost savings as organizations can utilize real-time data to effectively manage inventories, optimize delivery routes, and coordinate with suppliers, thus reducing the risk of stock outs, minimizing delivery delays, and ensuring smooth operations [5, 25, 28, 41, 44, 45, 51].

Completeness of data refers to the extent to which information includes all necessary elements and observations for a specific purpose. This factor enhances the integrity and reliability of analyses, preventing gaps in understanding and supporting more robust decision-making processes. In a complete dataset, all variables

relevant to the presentation of information should be present and fully populated with valid data values. Any missing, incorrect, or incomplete entries in the dataset can compromise the quality of analyses, interpretations, and decisions based on that data [4, 5, 9, 28–30, 41, 44, 45, 52]. Coverage, comprehensiveness, appropriate amount, adequate, appropriate amount of data and integrity are common terms that use for describe data completeness [21]. The amount of data indicates the extent of data sets obtained for analysis and processing. In present-day information systems, these sets of data are frequently observed to be escalating in size, reaching capacities such as terabytes and petabytes [4, 29, 50, 57].

Data reliability pertains to the uniformity of data across various records, programs, or platforms, as well as the credibility of the data source. Reliable data remains consistently accurate, while unreliable data may not always be valid, making it challenging to ascertain its accuracy. Consequently, organizations cannot depend on unreliable data for decision-making. Data reliability, also referred to as data observability, represents the trustworthiness of data and the insights derived from it for enabling sound decision-making. Reliability is characterized by two other fundamental elements of data quality include accuracy and consistency [9, 49, 53, 57, 59, 65].

Data accessibility refers to the ease with which users can locate, retrieve, comprehend, and utilize data within an organization's information systems. This is crucial in the modern digital landscape, where data is valuable for decision-making, strategic planning, and operational efficiency. Ensuring data accessibility involves creating an environment where data is available, understandable, and usable by individuals with varying levels of technical expertise. This approach is closely tied to data democratization, which aims to break down silos and make data available across different levels and departments of an organization. A well-implemented data accessibility

strategy ensures that data is not locked away in isolated information systems but is integrated and accessible, contributing to a more informed and agile organizational structure. The ultimate goal is to empower users to leverage data in their daily tasks and decision-making processes, thus fostering a data-driven culture [4, 26, 29, 33, 50, 57].

Data Objectivity refers to the extent to which data is free from personal biases, emotions, and subjective interpretations. Objective data is verifiable, reliable, and accurate, meaning that it can be verified independently by multiple parties. In other words, objective data is based on facts rather than opinions or judgments. In the context of information systems, data objectivity is crucial because it enables organizations to make informed decisions based on accurate and reliable information. Objective data helps to reduce errors, inconsistencies, and uncertainties, ensuring that business processes are efficient, effective, and compliant with regulatory requirements. Data objectivity in information systems is often hindered by biases in data collection, data quality issues, information overload, and lack of standardization. Biases may arise from human error, sampling errors, or deliberate data manipulation during the collection process. Inaccuracies, inconsistencies, and incompleteness resulting from poor data quality can compromise the objectivity of the information. The overwhelming amount of data available can make it challenging to differentiate between objective and subjective information. Inconsistencies in data representation and interpretation may occur due to the use of different systems or formats [36, 41, 44–46].

Data relevancy is an aspect of data quality that determines whether the data used or generated are relevant to add to the new target system and how usable it is for users [9, 29, 45, 48, 51]. Ease of operation, Usability, applicable, utility, Usefulness, Perceived usefulness and importance are common terms that use for describe data relevancy [21]. The concept of data usability revolves around a user's ability to obtain meaningful information from various systems. When data is stored in text files that demand prolonged and intricate processing before it can be analyzed, its usability is limited. Conversely, data that is conveniently displayed on a performance dashboard for immediate interpretation is classified as highly usable [4, 25, 29, 45, 48, 50]. The concept of data usefulness denotes the level at which data, post-analysis, aligns with the intended purpose within a given context for its user or consumer. In most cases, data usefulness is attained when all criteria related to data quality, such as dependability, thoroughness, uniformity, and others, are fulfilled [43, 50, 52].

Data Understandability refer to the level at which data exhibits qualities that facilitate understanding and analysis by users, and are presented in relevant languages,

symbols, and measurements within a defined context of utilization [22, 34, 37, 46]. Interpretability, ease of understanding, granularity and transparency are common terms that use for describe data understandability [21].

Data navigation refers to the process of searching, locating, and extracting relevant data from a vast pool of information to support decision-making, problem-solving, or analysis. It involves the utilization of different techniques and tools to navigate through extensive data, identify patterns, trends, and correlations, and present the information in a meaningful and actionable way. The success of data navigation is contingent upon several dimensions, including technical, domain knowledge, systems, methodological, and human dimensions. The technical dimension involves mastering programming languages like SQL and Python, utilizing data visualization software such as Tableau and Power BI, and implementing data mining techniques like machine learning algorithms. Domain knowledge dimension stresses the importance of expertise in specific fields. Information system dimension highlights the role of databases, data warehouses, cloud storage platforms, and other technologies in facilitating data navigation by storing, managing, and providing access to data. Methodological dimension focuses on statistical analysis, data mining techniques, and data visualization methods as key approaches to navigating data. Lastly, human dimension recognizes the significance of communication skills, collaboration, and critical thinking in the process of data navigation [4, 50, 65, 68].

Data reputation is the evaluation of the trustworthiness, reliability, and credibility of data in an information system. It signifies the extent to which stakeholders, such as users, decision-makers, and other systems, perceive the data as accurate, reliable, and complete. Within an information system, data reputation plays a crucial role in decision-making, trust, system performance, and data sharing [42, 60, 61].

The concept of data efficiency revolves around an organization's effectiveness in maximizing the value obtained from its data, while simultaneously minimizing the resources essential for processing, storing, and up keeping that data. Put simply, data efficiency focuses on streamlining the collection, storage, analysis, and utilization of data to meet objectives. When considering an information system, data efficiency can be examined from various angles, such as efficiency in data acquisition, storage, processing, analysis, visualization, security, retention, and archiving [7, 28, 29, 48].

Data value-added pertains to the process of refining raw data into more useful, meaningful, and valuable information that can support decision-making, drive business outcomes, and create a competitive advantage. This process involves extracting insights, patterns,

or trends from large datasets and presenting them in a manner that is easy to understand and act upon. By prioritizing these dimensions of data value-added within an information system, organizations can ensure that their data is transformed into valuable insights that support informed decision-making and drive business outcomes [5, 22, 25, 45].

## Discussion

In a few papers, the concept of “fitness for use” was applied to data quality [6, 55, 69]. Two viewpoints can be used to characterize data quality: (1) the inherent quality of the data elements and set, and (2) how the set satisfies the needs of the user. The definition provided by the International Standards Organization best captures the accepted meaning of data quality, which is “the totality of features and characteristics of an entity that bears on its ability to satisfy stated and implied needs” [4, 15, 28, 33, 53].

Current review study identified 14 common dimensions for data quality in health information system. In related research data quality dimensions classified on four dimensions include: intrinsic (accuracy, objectivity, reputation), contextual timeliness, completeness, and relevancy), representational (representational format, understandability, consistency), and accessibility (accessibility, security) categories [53, 60, 69–71]. There exists a certain level of intersection between the aspects of data quality recognized in this review and those research in prior classifications of data quality.

Previous literature has often discussed intrinsic data quality in terms of the absence of defects, as indicated by various dimensions such as accuracy, perfection, freshness, and uniformity [72], and “completeness, unambiguity, meaningless and correctness” [54, 73, 74]. The Canadian Institute for Health Information put forth a set of 69 quality criteria, organized into 24 quality characteristics, and further classified into 6 quality dimensions: accuracy, timeliness, comparability, usability, relevance, and privacy & security [58, 71]. Research on data quality has primarily concentrated on recognizing general quality traits like accuracy, currency, completeness, correctness, consistency, and timeliness as fundamental aspects of data quality applicable across different fields. Nevertheless, existing reviews reveal a lack of consensus regarding the conceptual framework and definition of data quality [70, 73]. However, our previous review shows there is a lack of consensus conceptual framework and definition for data quality [1, 71].

In this study, the three most-frequently used dimensions of data quality were accuracy, completeness and timeliness, respectively. This arrangement is somewhat different from previous literature in which the three most-frequently used dimensions were arranged in the

order of completeness, accuracy, and timeliness, respectively [43, 51, 53]. Furthermore, the absence of a precise definition of the data quality dimensions led to complexities in evaluating them. The definitions of dimensions and their associated metrics were occasionally based on intuition, past experiences, or the underlying goals. These results indicate that data quality is a multi-faceted phenomenon. Likewise, other scholars argue that data quality is a multi-dimensional notion [5, 28, 38, 52, 61].

## Conclusions

The Health Information Systems heavily rely on data, as they perform essential functions like generation, compilation, analysis, synthesis, communication, and data application to support decision-making. The literature frequently evaluates the dimensions of data quality, but there is currently a lack of consistency and potential generalizability in using these dimensions and methods to assess data quality in Health Information Systems. In this review of the literature, the data quality for health information system were examined and identified 14 common dimension include: Accuracy, Consistency, Security, Timeliness, Completeness, Reliability, Accessibility, Objectivity, Relevancy, Understandability, Navigation, Reputation, Efficiency and Value-added.

The quality of data in health information systems is indispensable for healthcare institutions to make well-informed decisions and provide patients with optimal care. Accurate and timely data assists healthcare organizations and professionals in identifying patterns, predicting outcomes, and enhancing patient results. Conversely, inadequate data quality in healthcare or other data-related issues can lead to inaccurate diagnoses, inappropriate treatments, and harm to patients. To ensure data quality in healthcare, organizations must prioritize investments in data governance, data management, and data analysis tools, while also maintaining a continuous process of monitoring and improving data quality in health information systems.

It is essential to have high-quality data in order to ensure the safe and dependable delivery of healthcare services. Health facility data plays a crucial role in monitoring performance. While various organizations may prioritize different aspects of data quality, it is important to acknowledge that no health data, regardless of its source, can be deemed flawless. All data are susceptible to various limitations related to data quality, including missing values, bias, measurement error, and human errors in data entry and computation. These limitations are associated with technical, behavioral, and organizational factors [75].

This study has limitations. Firstly, the number of articles with complete data was relatively small. Secondly, assessing the quality of some studies were difficult because the

quality assessment criteria were not clearly identified. We have proposed four fundamental implications to inspire future research. Firstly, it is crucial for researchers to give equal attention to all dimensions of data quality, as these dimensions can have both direct and indirect effects on data quality outcomes. Secondly, researchers should aim to evaluate the existing data quality models and frameworks through a combination of mixed methods and case study designs. Thirdly, it is important to identify the underlying causes of data quality issues in health information systems. Lastly, efforts should be made to develop interventions that can effectively address and prevent data quality issues from occurring.

#### Abbreviations

JBI Joanna Briggs Institute

#### Acknowledgements

Not applicable.

#### Author contributions

Hossein Ghalavand and Saeid Shirshahi Conceived the study, prepared the analysis plan, conducted the analysis, and prepared the draft manuscript. Alireza Rahimi, Zarrin Zarrinabadi and Fatemeh Amani Conceived the study, prepared the analysis plan, performed the literature search, screening for study inclusion/exclusion, and risk of bias assessment, conducted the analysis, and prepared the draft manuscript. All authors contributed to the final version of the manuscript.

#### Funding

This study was supported by Abadan University of medical sciences, Research code: 1557.

#### Data availability

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

This research study was approved by Ethics Committee in Biomedical Research at Abadan University of Medical Sciences (Ethical code: IR.ABADANUMS.REC.1401.122).

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 17 April 2024 / Accepted: 21 August 2024

Published online: 02 September 2024

#### References

- Liaw S-T, et al. Quality assessment of real-world data repositories across the data life cycle: a literature review. *J Am Med Inform Assoc.* 2021;28(7):1591–9.
- WHO. *Data Quality Assurance (DQA)*. Health Service Data 2022 [cited 2022 2022]; <https://www.who.int/data/data-collection-tools/health-service-data/data-quality-assurance-dqa#:~:text=WHO%20has%20produced%20the%20Data,annual%20data%20quality%20desk%20review>
- FMoH E. *Health sector transformation plan*. 2015, Addis Ababa, Ethiopia.
- Rumisha SF, et al. Data quality of the routine health management information system at the primary healthcare facility and district levels in Tanzania. *BMC Med Inf Decis Mak.* 2020;20(1):340.
- Chekol A, et al. Data quality and associated factors of routine health information system among health centers of West Gojjam Zone, northwest Ethiopia, 2021. *Front Health Serv.* 2023;3:1059611.
- Pipino LL, Lee YW, Wang RY. Data quality assessment. *Commun ACM.* 2002;45(4):211–8.
- Ouedraogo M, et al. A quality assessment of Health Management Information System (HMIS) data for maternal and child health in Jimma Zone, Ethiopia. *PLoS ONE.* 2019;14(3):e0213600.
- Lemma S, et al. Improving quality and use of routine health information system data in low- and middle-income countries: a scoping review. *PLoS ONE.* 2020;15(10):e0239683.
- Bammidi TR, et al. The crucial role of Data Quality in Automated decision-making systems. *Int J Manage Educ Sustainable Dev.* 2024;7(7):22.
- Adane A, et al. Exploring data quality and use of the routine health information system in Ethiopia: a mixed-methods study. *BMJ open.* 2021;11(12):e050356.
- Mohammed SA, Yusof MM. Towards an evaluation framework for information quality management (IQM) practices for health information systems—evaluation criteria for effective IQM practices. *J Eval Clin Pract.* 2013;19(2):379–87.
- Long J, Seko C. *A New Method for Database Data Quality Evaluation at the Canadian Institute for Health Information (CIHI)*. in *ICIQ*. 2002. Citeseer.
- Adeleke IT, et al. Data quality assessment in healthcare: a 365-day chart review of inpatients' health records at a Nigerian tertiary hospital. *J Am Med Inform Assoc.* 2012;19(6):1039–42.
- Singh M, et al. Health management information system data quality under NRHM in District Sonipat, Haryana. *Int J Health Sci Res (IJHSR).* 2016;6(9):11–4.
- Harrison K, Rahimi N. Carolina Danovaro-Holliday, *factors limiting data quality in the expanded programme on immunization in low and middle-income countries: a scoping review*. *Vaccine.* 2020;38(30):4652–63.
- Shama AT, et al. Assessment of quality of routine health information system data and associated factors among departments in public health facilities of Harari region, Ethiopia. *BMC Med Inf Decis Mak.* 2021;21(1):1–12.
- Bosch-Capblanch X, et al. Does an innovative paper-based health information system (PHISICC) improve data quality and use in primary healthcare? Protocol of a multicountry, cluster randomised controlled trial in sub-Saharan African rural settings. *BMJ Open.* 2021;11(7):e051823.
- Ehsani-Moghaddam B, Martin K, Queenan JA. Data quality in healthcare: a report of practical experience with the Canadian Primary Care Sentinel Surveillance Network data. *Health Inform Manage J.* 2021;50(1–2):88–92.
- Brown PJB, Warmington V. Data quality probes—exploiting and improving the quality of electronic patient record data and patient care. *Int J Med Informatics.* 2002;68(1):91–8.
- Lima CR, et al. [Review of data quality dimensions and applied methods in the evaluation of health information systems]. *Cad Saude Publica.* 2009;25(10):2095–109.
- Alipour J, Ahmadi M. Dimensions and assessment methods of data quality in health information systems. *Acta Med Mediterranea.* 2017;33(2):313–20.
- Tolera A, et al. Barriers to healthcare data quality and recommendations in public health facilities in dire Dawa city administration, eastern Ethiopia: a qualitative study. *Front Digit Health.* 2024. 6.
- Vrabel M. M. *Preferred reporting items for systematic reviews and meta-analyses*. In *Oncology nursing forum*. Oncology Nursing Society; 2015.
- JBI QARI Critical appraisal checklist for interpretive & critical research. The Joanna Briggs Institute, Adelaide 2018; <http://joannabriggs.org/research/critical-appraisal-tools.html>
- Fraser HSF, et al. Factors Influencing Data Quality in Electronic Health Record Systems in 50 Health Facilities in Rwanda and the role of clinical Alerts: cross-sectional observational study. *JMIR Public Health Surveill.* 2024;10:e49127.
- Madandola OO, et al. The relationship between electronic health records user interface features and data quality of patient clinical information: an integrative review. *J Am Med Inform Assoc.* 2023;31(1):240–55.
- Getachew N, Erkalo B, Garedew MG. Data quality and associated factors in the health management information system at health centers in Shashoghi district, Hadiya Zone, southern Ethiopia, 2021. Volume 22. *BMC Medical Informatics and Decision Making.* 2022. pp. 1–9. 1.
- Solomon M, et al. Data quality assessment and associated factors in the health management information system among health centers of Southern Ethiopia. *PLoS ONE.* 2021;16(10):e0255949.
- Moukéné A, et al. Health management information system (HMIS) data quality and associated factors in Massaguet district, Chad. *BMC Med Inf Decis Mak.* 2021;21(1):326.

30. do Einloft N. Data quality and arbovirus infection associated factors in pregnant and non-pregnant women of childbearing age in Brazil: a surveillance database analysis. *One Health.* 2021;12:100244.
31. Ayele W et al. Data quality and its correlation with routine health information system structure and input at public health centers in Addis Ababa, Ethiopia. *Ethiop J Health Dev.* 2021; 35(1).
32. Mulissa Z, et al. Effect of data quality improvement intervention on health management information system data accuracy: an interrupted time series analysis. *PLoS ONE.* 2020;15(8):e0237703.
33. Yourkavitch J, Prosnitz D, Herrera S. Data quality assessments stimulate improvements to health management information systems: evidence from five African countries. *J Glob Health.* 2019;9(1):010806.
34. Endriyas M, et al. Understanding performance data: health management information system data accuracy in Southern Nations nationalities and people's Region, Ethiopia. *BMC Health Serv Res.* 2019;19(1):1–6.
35. Biancone P, et al. Data quality methods and applications in health care system: a systematic literature review. *Int J Bus Manage.* 2019;14(4):35–47.
36. Liu Y, et al. [Designing and implementation of the data quality control in the information system of air pollution and health impact monitoring]. *Wei Sheng Yan Jiu.* 2018;47(2):277–80.
37. Kumar M, et al. Research gaps in routine health information system design barriers to data quality and use in low- and middle-income countries: a literature review. *Int J Health Plann Manage.* 2018;33(1):e1–9.
38. Feder SL. Data quality in electronic health records research: quality domains and assessment methods. *West J Nurs Res.* 2018;40(5):753–66.
39. Watson NL, et al. Data management and data quality in PERCH, a large international case-control study of severe childhood pneumonia. *Clin Infect Dis.* 2017;64(suppl3):S238–44.
40. Wagenaar BH, et al. Data-driven quality improvement in low-and middle-income country health systems: lessons from seven years of implementation experience across Mozambique, Rwanda, and Zambia. *BMC Health Serv Res.* 2017;17:65–75.
41. Puttkammer N, et al. Identifying priorities for data quality improvement within Haiti's iSanté EMR system: comparing two methods. *Health Policy Technol.* 2017;6(1):93–104.
42. Finnegan K, et al. Barriers and facilitators of Data Quality and Use in Malawi's Health Information System. *Annals Global Health.* 2017;83(1):36–7.
43. Chen H, et al. Data Quality of the Chinese National AIDS Information System: a critical review. *Stud Health Technol Inf.* 2017;245:1352.
44. Woinarowicz M, Howell M. The impact of electronic health record (EHR) interoperability on immunization information system (IIS) data quality. *Online J Public Health Inf.* 2016;8(2):e184.
45. Puttkammer N, et al. An assessment of data quality in a multi-site electronic medical record system in Haiti. *Int J Med Informatics.* 2016;86:104–16.
46. Nicol E, Dudley L, Bradshaw D. Assessing the quality of routine data for the prevention of mother-to-child transmission of HIV: an analytical observational study in two health districts with high HIV prevalence in South Africa. *Int J Med Informatics.* 2016;95:60–70.
47. Wagenaar BH, et al. Effects of a health information system data quality intervention on concordance in Mozambique: time-series analyses from 2009–2012. *Popul Health Metr.* 2015;13:9.
48. Taggart J, Liaw S-T, Yu H. Structured data quality reports to improve EHR data quality. *Int J Med Informatics.* 2015;84(12):1094–8.
49. Glèlè Ahanhanzo Y, et al. Data quality assessment in the routine health information system: an application of the Lot Quality Assurance Sampling in Benin. *Health Policy Plan.* 2015;30(7):837–43.
50. Glèlè Ahanhanzo Y, et al. Factors associated with data quality in the routine health information system of Benin. *Arch Public Health.* 2014;72(1):25.
51. Chen H, et al. A review of data quality assessment methods for public health information systems. *Int J Environ Res Public Health.* 2014;11(5):5170–207.
52. Hahn D, Wanjala P, Marx M. Where is information quality lost at clinical level? A mixed-method study on information systems and data quality in three urban Kenyan ANC clinics. *Glob Health Action.* 2013;6:21424.
53. Chen H, Yu P, Wang N. Do we have the reliable data? An exploration of data quality for AIDS information system in China. *Stud Health Technol Inf.* 2013;192:1042.
54. Choquet R, et al. *The Information Quality Triangle: a methodology to assess clinical information quality*, in MEDINFO 2010. IOS; 2010. pp. 699–703.
55. Mettler T, Rohner P, Baacke L. *Improving data quality of health information systems: a holistic design-oriented approach.* 2008.
56. Sørensen HT, et al. Identification of cases of meningococcal disease: data quality in two Danish population-based information systems during a 14-year period. *Int J Risk Saf Med.* 1995;7(3):179–89.
57. Gimbel S, et al. An assessment of routine primary care health information system data quality in Sofala Province, Mozambique. *Popul Health Metr.* 2011;9:12.
58. Kerr K, Norris T, Stockdale R. *Data quality information and decision making: a healthcare case study.* ACIS 2007 proceedings, 2007: p. 98.
59. Ben Said M, et al. A multi-source information System via the internet for end-stage renal disease: Scalability and Data Quality. *Stud Health Technol Inf.* 2005;116:994–9.
60. Fletcher DM. Achieving data quality. How data from a pediatric health information system earns the trust of its users. *J Ahima.* 2004;75(10):22–6.
61. Bean KP. Data quality in hospital strategic information systems: a summary of survey findings. *Top Health Inf Manage.* 1994;15(2):13–25.
62. Kelly A, Becker W. Nutrition information systems and data quality requirements. *WHO Reg Publ Eur Ser.* 1991;34:15–24.
63. Leitheiser RL. *Data quality in health care data warehouse environments.* in *Proceedings of the 34th annual Hawaii international conference on system sciences.* 2001. IEEE.
64. Ndira S, Rosenberger K, Wetter T. Assessment of data quality of and staff satisfaction with an electronic health record system in a developing country (Uganda). *Methods Inf Med.* 2008;47(06):489–98.
65. Silva AA, et al. [Evaluation of data quality from the information system on live births in 1997–1998]. *Rev Saude Publica.* 2001;35(6):508–14.
66. Woelk GB, Moyo IM, Ray CS. A health information system revised. Part II: improving data quality and utilization. *Cent Afr J Med.* 1987;33(7):170–3.
67. Abbasi R, Khajouei R, Sadeqi M, Jabali. Timeliness and accuracy of information sharing from hospital information systems to electronic health record in Iran. *J Health Adm.* 2019;22(2):28–40.
68. Elavsky F, Nadolskis L, Moritz D. Data navigator: an accessibility-centered data navigation toolkit. *IEEE Trans Vis Comput Graph.* 2023;20(1):16–25.
69. Wang RY. A product perspective on total data quality management. *Commun ACM.* 1998;41(2):58–65.
70. Liaw S-T et al. *Data quality and fitness for purpose of routinely collected data—a general practice case study from an electronic practice-based research network (ePBRN).* in *AMIA Annual Symposium Proceedings.* 2011. American Medical Informatics Association.
71. Rahimi A, et al. Ontological specification of quality of chronic disease data in EHRs to support decision analytics: a realist review. *Decis Analytics.* 2014;1:1–31.
72. Redman TC. *Measuring data accuracy: A framework and review.* *Information quality.* 2014: pp. 21–36.
73. Orme AM, Yao H, Etzkorn LH. Indicating ontology data quality, stability, and completeness throughout ontology evolution. *J Softw Maintenance Evolution: Res Pract.* 2007;19(1):49–75.
74. Yao H, Orme AM, Etzkorn L. Cohesion metrics for ontology design and application. *J Comput Sci.* 2005;1(1):107–13.
75. Endriyas M, et al. Understanding performance data: health management information system data accuracy in Southern Nations nationalities and people's Region, Ethiopia. *BMC Health Serv Res.* 2019;19(175):1–6.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.