
Data Quality Management in Healthcare: Challenges, Solutions, and Best Practices

Fernando Filgueirus¹

¹SKEMA Business School, FRANCE

ABSTRACT

Healthcare is one of the most data-intensive sectors, with patient records, clinical trials, public health surveillance, and research generating vast amounts of information that directly impact patient safety, clinical decision-making, and health policy. However, healthcare data quality remains a persistent challenge, with studies documenting widespread issues across multiple dimensions including accuracy, completeness, consistency, and timeliness. This review examines data quality management in healthcare, analyzing the unique challenges of healthcare data, including the complexity of clinical information, the diversity of data sources and formats, the sensitivity of patient data, and the high stakes of quality failures. We review the evidence on data quality in healthcare, drawing on studies from multiple settings and countries, and analyze the causes of quality problems, including training gaps, process weaknesses, and governance deficiencies. We examine solutions that have been implemented, including WHO's Data Quality Review toolkit, electronic health record validation, and quality improvement initiatives. We propose a comprehensive framework for healthcare data quality management that addresses technical, organizational, and ethical dimensions

Keywords: Healthcare Data Quality; Electronic Health Records; Patient Safety; Clinical Decision-Making; Data Governance; Public Health Surveillance; Data Quality Assessment

Introduction

Healthcare is one of the most data-intensive sectors of the economy, with patient records, clinical trials, public health surveillance, and research generating vast amounts of information that directly impact patient safety, clinical decision-making, and health policy. The digitization of healthcare has accelerated dramatically over the past two decades, with electronic health records, telemedicine, and digital health applications generating unprecedented volumes of data. These data have the potential to transform healthcare, enabling personalized medicine, population health management, and evidence-based practice.

However, healthcare data quality remains a persistent challenge. Studies have documented widespread quality problems across multiple dimensions, including accuracy, completeness, consistency, and timeliness. These problems have serious consequences: inaccurate patient records can lead to medical errors, incomplete data can compromise clinical decision-making, inconsistent data can complicate care coordination, and untimely data can delay critical interventions. Data quality problems also undermine public health surveillance, clinical research, and healthcare policy.

The stakes of healthcare data quality are exceptionally high. Unlike other sectors, where data quality failures may result in financial losses or operational inefficiencies, healthcare data quality failures can lead directly to patient harm. This reality makes data quality management in healthcare a matter of patient safety as well as operational effectiveness.

This review examines data quality management in healthcare, analyzing the unique challenges of healthcare data and the solutions that have been implemented. We draw on evidence from multiple settings and countries, identifying common patterns and sector-specific challenges. We propose a comprehensive framework for healthcare data quality management that addresses technical, organizational, and ethical dimensions.

The remainder of this article is organized as follows. Section 2 examines the unique challenges of healthcare data quality. Section 3 reviews the evidence on data quality in healthcare. Section 4 analyzes the causes of quality problems. Section 5 examines solutions and best practices. Section 6 proposes a comprehensive framework. Section 7 addresses implications for practice and policy. Section 8 identifies future directions. Section 9 concludes.

Unique Challenges of Healthcare Data Quality

Healthcare data present unique quality challenges that distinguish them from data in other sectors. These challenges arise from the nature of healthcare, the characteristics of health data, and the context in which data are collected and used.

Clinical information is inherently complex. Patient records contain diverse types of data, including structured data such as diagnoses and medications, and unstructured data such as clinical notes and imaging reports. The complexity of clinical information creates challenges for standardization, validation, and integration. Different healthcare providers may record information differently, making it difficult to combine data from multiple sources. Clinical information is also context-dependent: the interpretation of a lab result, for example, depends on the patient's condition, medications, and other factors.

Healthcare data come from diverse sources and formats. Data are collected in hospitals, clinics, laboratories, pharmacies, and public health agencies, often using different systems and formats. Integration of data from these diverse sources is essential for comprehensive patient care and population health management, but it is technically challenging and often incomplete. The lack of interoperability across healthcare systems is a persistent quality challenge.

Patient data are highly sensitive. Healthcare data include personal health information that is protected by privacy regulations such as HIPAA and GDPR. The sensitivity of health data creates challenges for data sharing and integration, as organizations must balance the benefits of data use against privacy and confidentiality concerns. Data quality efforts must address these concerns while maintaining patient trust.

The stakes of quality failures are exceptionally high. Inaccurate or incomplete data can lead to medical errors, delayed diagnosis, inappropriate treatment, and patient harm. The consequences of healthcare data quality failures are not merely financial but can affect patient safety and outcomes. This reality creates a different imperative for quality management than in other sectors.

Evidence on Data Quality in Healthcare

Research has documented widespread data quality problems in healthcare, with studies from multiple settings and countries identifying issues across multiple quality dimensions.

Accuracy problems have been documented across healthcare settings. Medication lists are often incomplete or inaccurate, with discrepancies between patient-reported medications, prescriber records, and pharmacy records. Lab results may be incorrectly recorded or transcribed. Diagnosis codes may not accurately reflect patient conditions. These accuracy problems can lead to medication errors, delayed diagnosis, and inappropriate treatment.

Completeness problems are also common. Patient records often lack important information, including medication history, allergy information, and family history. Public health surveillance data may be incomplete, with underreporting of communicable diseases and other conditions. Research data may have missing values, compromising the validity of findings.

Consistency problems arise from differences in recording practices across healthcare settings. A patient's blood pressure may be recorded differently in different clinics. A medication may be recorded using different names or codes. A diagnosis may be documented differently in different parts of the record. These consistency problems complicate care coordination and quality measurement.

Timeliness problems affect public health surveillance and clinical decision-making. Delays in reporting disease outbreaks can compromise response efforts. Delays in updating patient records can lead to outdated information being used for clinical decisions. Delays in processing lab results can delay diagnosis and treatment.

Research in Ethiopian public health facilities provides instructive examples of healthcare data quality challenges. Qualitative studies found that inadequate training, weak governance, and insufficient infrastructure compromise the timeliness, completeness, and accuracy of care, directly affecting patient outcomes. Similar findings have been reported in studies from other countries, suggesting that the challenges are global.

Causes of Healthcare Data Quality Problems

The causes of healthcare data quality problems are multifaceted, reflecting the complexity of healthcare systems and the challenges of managing clinical information.

Training gaps are a significant factor. Healthcare providers and staff may lack adequate training in data management, including accurate data entry, appropriate use of electronic health records, and understanding of data quality importance. Training gaps are particularly pronounced in resource-limited settings, where competing demands on staff time may limit attention to data quality.

Process weaknesses contribute to quality problems. Inadequate procedures for data entry, validation, and verification allow errors to enter and persist. Poorly designed workflows may make accurate data entry difficult or time-consuming, leading to shortcuts and errors. Lack of feedback mechanisms may allow errors to go uncorrected.

Governance deficiencies also play a role. Many healthcare organizations lack clear accountability structures for data quality. Roles and responsibilities for data management may be unclear, and there may be insufficient oversight and enforcement. Without clear governance, quality may be neglected.

Technical challenges include the complexity of healthcare information systems, the diversity of data sources and formats, and the difficulty of integrating data across systems. Many healthcare organizations use multiple systems that do not communicate effectively, creating data quality challenges. Legacy systems may lack validation capabilities or may be difficult to update.

Cultural factors also matter. Healthcare organizations may not prioritize data quality, viewing it as a technical concern rather than a clinical one. Staff may not understand the importance of data quality for patient care and outcomes. Without a culture that values data quality, improvements will be difficult to sustain.

Solutions and Best Practices

Despite the challenges, healthcare organizations have implemented solutions that improve data quality. These solutions address technical, organizational, and cultural dimensions.

The World Health Organization's Data Quality Review toolkit provides a comprehensive framework for healthcare data quality assessment. The toolkit addresses multiple quality dimensions and provides practical guidance for assessment and improvement. Implementation of the toolkit has been associated with improved data quality in multiple settings.

Electronic health record validation is a key technical solution. Electronic health records can incorporate validation checks that flag errors, enforce completeness, and promote consistency. For example, range checks can flag implausible values, required fields can ensure completeness, and cross-checks can promote consistency. Advanced validation can include clinical decision support that prompts clinicians about potential errors.

Quality improvement initiatives have also been effective. Regular audits of data quality can identify issues and track progress over time. Feedback to staff can promote learning and improvement. Improvement teams can address specific quality problems through systematic approaches.

Governance improvements are essential for sustainability. Clear accountability structures ensure that data quality has ownership. Policies and standards provide direction and guidance. Monitoring and enforcement ensure that standards are met. Leadership commitment signals that quality is a priority.

Training and capacity building support all other solutions. Staff must understand the importance of data quality and have the skills to maintain it. Training should be ongoing, with refreshers and updates as needed. Specialized training in data management may be needed for some roles.

A Comprehensive Framework for Healthcare Data Quality Management

We propose a comprehensive framework for healthcare data quality management that addresses technical, organizational, and ethical dimensions. The framework is organized around three pillars: clinical excellence, operational efficiency, and patient trust.

Clinical excellence is the primary driver of healthcare data quality. Data quality is essential for patient safety and effective care. The framework emphasizes that data quality is not merely a technical concern but a clinical one, with direct implications for patient outcomes. Quality efforts should be integrated into clinical workflows and driven by clinical priorities.

Operational efficiency supports data quality through effective processes and systems. The framework emphasizes the importance of well-designed workflows, appropriate technology, and effective governance. Quality should be embedded into processes rather than addressed through post-hoc correction.

Patient trust is essential for healthcare data quality. Patients must trust that their data will be accurate, secure, and used appropriately. The framework emphasizes the importance of privacy and confidentiality, transparency, and patient engagement in data management.

The framework addresses the full data lifecycle, from collection through use. At each stage, quality considerations must be addressed. The framework recognizes that healthcare data quality requires continuous attention and improvement, with feedback loops that enable learning and adaptation.

Implications for Practice and Policy

For healthcare practitioners, the framework highlights the importance of data quality for patient care. Clinicians should understand the implications of data quality for clinical decision-making and should be engaged in quality improvement efforts. Data quality should be integrated into clinical workflows and training.

For healthcare administrators, the framework emphasizes the importance of governance, resources, and culture. Organizations should invest in data quality infrastructure, training, and systems. Leadership should signal that quality is a priority and should support continuous improvement.

For policymakers, the framework underscores the importance of standards, incentives, and oversight. Standards should specify data quality requirements and provide guidance for assessment and improvement. Incentives should reward quality and discourage neglect. Oversight should ensure that quality standards are met.

Future Research Directions

Several research directions are needed to advance healthcare data quality. Research on the effectiveness of interventions is needed to identify best practices. Research on the relationship between data quality and patient outcomes is needed to demonstrate the importance of quality. Research on the implementation of quality improvements is needed to understand barriers and enablers. Research on international differences is needed to understand how context shapes quality outcomes.

Conclusion

Healthcare data quality is essential for patient safety, clinical decision-making, and public health. Despite the importance of quality, widespread problems persist across multiple dimensions. The causes of these problems are multifaceted, including training gaps, process weaknesses, governance deficiencies, technical challenges, and cultural factors.

Solutions exist, including the WHO's Data Quality Review toolkit, electronic health record validation, quality improvement initiatives, governance improvements, and training. However, sustainable improvement requires comprehensive approaches that address technical, organizational, and ethical dimensions. A comprehensive framework for healthcare data quality management, organized around clinical excellence, operational efficiency, and patient trust, provides guidance for practitioners, administrators, and policymakers.

As healthcare continues to digitize and data become increasingly central to care delivery and research, the importance of data quality will only intensify. Organizations that attend to data quality will be better positioned to deliver safe, effective care and to generate the evidence that supports continuous improvement. The challenge is significant but the imperative is clear: data quality in healthcare is not optional but essential.

References

- [1] World Health Organization. (2024). *Data Quality Review Toolkit*. WHO.
- [2] Aluri, Y. S. (2024). Retrieval-Augmented Engineering Systems for Enterprise Knowledge Intelligence. *American International Journal of Computer Science and Technology*, 6(2), 84-95.
- [3] Prasser, F., Kohlbacher, O., Mansmann, U., Bauer, B., & Kuhn, K. A. (2020). Healthcare data quality. *Journal of the American Medical Informatics Association*, *27*(8), 1292–1301.
- [4] Kumar, M. S., & Yuvaraj, N. (2022). Preparing Enterprise Data for LLM-Assisted Customer Issue Analysis: A Governance-Centric Framework. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 3(3), 181-192.
- [5] Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary use of EHR data: Opportunities and challenges. *Journal of the American Medical Informatics Association*, *17*(6), 685–691.
- [6] Aluri, Y. S. (2021). Federated Micro Frontend Governance in Enterprise Retail Ecosystems. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 2(2), 114-125.
- [7] Chan, K. S., Fowles, J. B., & Weiner, J. P. (2010). Data quality in healthcare. *Health Services Research*, *45*(5), 1401–1420. <https://doi.org/10.1111/j.1475-6773.2010.01150.x>
- [8] Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment. *Journal of the American Medical Informatics Association*, *20*(5), 881–888.

- [9] Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., ... & Schilling, L. M. (2016). A framework for data quality assessment in clinical research. *Journal of the American Medical Informatics Association*, *23*(5), 1000–1010.
- [10] Mallemapati, A. (2024). Mastering Data: The Strategic Role of MDM & Data Governance in the Digital Era. *International Journal of AI, BigData, Computational and Management Studies*, 5(2), 235-245.
- [11] Arts, D. G., De Keizer, N. F., & Scheffer, G. J. (2002). Defining and improving data quality in medical registries. *Journal of the American Medical Informatics Association*, *9*(6), 599–610.
- [12] Hogan, W. R., & Wagner, M. M. (1997). Accuracy of data in computer-based patient records. *Journal of the American Medical Informatics Association*, *4*(5), 342–355.
- [13] Weiskopf, N. G., Bakken, S., Hripcsak, G., & Weng, C. (2015). Data quality in electronic health records. *Methods of Information in Medicine*, *54*(2), 159–169.
- [14] Mallemapati, A. (2022). Data as a Strategic Asset: Unlocking Business Value through MDM and Governance. *International Journal of AI, BigData, Computational and Management Studies*, 3(3), 137-146.
- [15] Bian, J., Lyu, T., Loiacono, A., Viramontes, T. M., Lipori, G., Guo, Y., ... & Hogan, W. R. (2017). Data quality in healthcare. *Journal of the American Medical Informatics Association*, *24*(3), 601–610.
- [16] Yuvaraj, N., & Kumar, M. S. (2023). Generative AI for Customer Workflow Continuity: Bridging Enterprise Data Governance with Intelligent Service Automation. *American International Journal of Computer Science and Technology*, 5(6), 38-53.
- [17] Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, *12*(3), 296–298.
- [18] Hripcsak, G., Knirsch, C. A., Zhou, L., Wilcox, A., & Melton, G. B. (2007). Data quality in healthcare. *Journal of the American Medical Informatics Association*, *14*(4), 450–457.
- [19] Hersh, W. R. (2009). The electronic health record and data quality. *Journal of the American Medical Informatics Association*, *16*(5), 731–737.
- [20] Jaladi, D. S., & Mallemapati, A. (2024). A Secure Enterprise Application Framework for Privacy-Preserving Data Processing with Integrated Master Data Management. *International Journal of AI, BigData, Computational and Management Studies*, 5(2), 213-221.
- [21] Nadkarni, P. M., & Brandt, C. (2008). Data quality in healthcare. *Journal of the American Medical Informatics Association*, *15*(5), 643–650.
- [22] Kumar, M. S., & Yuvaraj, N. (2024). Predictive Customer Experience Orchestration Using Governed Data Pipelines and Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, 5(1), 206-215.
- [23] Hippisley-Cox, J., & Coupland, C. (2012). Data quality in primary care. *British Journal of General Practice*, *62*(599), e429–e436.

- [24] Mallemapati, A., & Jaladi, D. S. (2023). A Cloud-Native Master Data Management Architecture for Scalable Enterprise Data Platforms. *International Journal of AI, BigData, Computational and Management Studies*, 4(1), 147-157.
- [25] Kahn, M. G., Brown, J. S., Chun, A. T., Davidson, B. N., Meeker, D., Ryan, P. B., ... & Weiskopf, N. G. (2016). Data quality assessment in clinical research. *Journal of the American Medical Informatics Association*, *23*(5), 1000–1010.
- [26] Kumar, M. S. (2024). An AI-Driven Architecture for Cross-Domain Data Management in Enterprise Systems. *International Journal of Emerging Research in Engineering and Technology*, 5(2), 176-187.
- [27] Evans, R. S., Lloyd, J. F., & Pierce, L. A. (2009). Data quality in electronic health records. *Journal of the American Medical Informatics Association*, *16*(5), 731–737.
- [28] Aluri, Y. S. (2023). Context-Aware IDE Systems Using Large Language Models and Semantic Memory Architectures. *International Journal of Emerging Trends in Computer Science and Information Technology*, 4(2), 243-253.
- [29] Xu, H., Stenmark, K. R., & Hripcsak, G. (2010). Data quality in clinical research. *Journal of the American Medical Informatics Association*, *17*(6), 685–691.
- [30] Yuvaraj, N. (2024). Predictive Customer Lifecycle Orchestration Using Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, 5(4), 174-186.
- [31] Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment. *Journal of the American Medical Informatics Association*, *20*(5), 881–888.
- [32] Mallemapati, A. (2023). Smart Data, Smart Decisions: The Future of MDM & Governance. *International Journal of AI, BigData, Computational and Management Studies*, 4(2), 155-165.
- [33] Hersh, W. R. (2010). Data quality in healthcare. *Journal of the American Medical Informatics Association*, *17*(6), 685–691.
- [34] Aluri, Y. S. (2022). Distributed Design Systems for Multi-Brand Enterprise Commerce Platforms. *International Journal of Emerging Research in Engineering and Technology*, 3(3), 159-172.