


A Review of the State of the Art of Data Quality in Healthcare


Caihua Liu, Guilin University of Electronic Technology, China

 <https://orcid.org/0000-0002-9641-4712>

Amir Talaei-Khoei, University of Nevada, Reno, USA

Veda C. Storey, Georgia State University, USA

Guochao Peng, Sun Yat-sen University, China*

 <https://orcid.org/0000-0003-1206-5509>

ABSTRACT

Effective implementation of strategic data-driven health analysis initiatives is heavily dependent on the quality of the electronic medical records that serve as the foundation from which to improve clinical decisions and, in turn, the quality of care. Although there is a large body of research on the quality of healthcare data, a systematical understanding of the methods used to address the issues of data quality is missing. This study analyzes research articles in health information systems/healthcare informatics on data quality to derive a set of dimensions for understanding data quality. Issues related to each dimension are identified and methods used to address them summarized. The issues and methods can inform healthcare professionals of how to improve data practices.

KEYWORDS

Data Quality, Healthcare, Literature Review

INTRODUCTION

Organizations and individuals increasingly rely on information systems (IS) for data-driven decision-making (ZareRavasan & Krčál, 2021), making it critical to ensure that the data being used is of sufficient quality. Poor quality data can cause the loss of revenue and even put lives at risk, especially in healthcare. For example, entry errors in a provider system resulted in an inappropriate treatment, with a patient developing seizures and requiring intubation (ECRI Institute, 2015). Duplicated drug orders entered into two separate prescribing systems used by a hospital resulted in nurses administering an excessive amount of insulin to a patient, resulting in death (Rowland, 2014). Furthermore, data not collected systematically in electronic medical records (EMR) or other operational systems can be of poorer quality, limiting its reuse (Kahn et al., 2016).

The poor delivery of evidence-based practice (James, 2013) and the high rate of adverse events (Landrigan et al., 2010) in hospital admissions and general practices motivate service providers to

DOI: 10.4018/JGIM.316236

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

optimize clinical handovers between healthcare settings. This requires extensive efforts to improve the internal quality of data and advance the quality of information exchange programs. A substantial body of literature has identified the issues that may arise when undertaking these efforts, resulting in poor quality of medical or healthcare data (Michel-Verkerke, 2012; Vilic et al., 2016). Research on data quality for healthcare has developed several methods to resolve these issues (Prasser et al., 2018; Zięba, 2014). However, a systematic understanding of the methods is not available in the healthcare IS/informatics literature. It would be useful to link methods to data quality issues for two main reasons.

First, using EMR has created the potential to improve the quality of care and address cost-effectiveness. However, the scope of new risks (e.g., the quality of the data used in healthcare and research settings) for patients is still not fully understood (Rowland, 2014). Data quality issues are more prominent than in earlier healthcare technology initiatives but have not been taken seriously by practice. The widespread use of EMR technology is inevitable, so addressing data quality issues needs to be a priority.

Second, previous literature reviews of data quality for healthcare (Arts et al., 2002; Chen et al., 2014; Johnson et al., 2015; Kahn et al., 2016; Liaw et al., 2013; Thiru et al., 2003; Weiskopf & Weng, 2013) focus on describing and assessing data quality but do not provide a holistic picture that matches the methods to specific issues associated with different dimensions of data quality. As a result, there is a lack of studies to depict an overall understanding of methods used to address data quality issues in healthcare. Such a study could help healthcare and IT professionals ascertain what work has been done to address data quality issues in this field and identify possible gaps for further exploration.

In addition, a challenge arises in that the current data quality literature uses different terms to describe data quality dimensions, thus limiting our understanding of different or similar data quality dimensions discussed (Kahn et al., 2016; Liaw et al., 2013; Wang & Strong, 1996; Weiskopf & Weng, 2013). Furthermore, the lack of consistent use of terms that describe data quality dimensions makes it difficult to explain different or similar issues for a specific data quality dimension and establish relationships between methods and issues associated with different dimensions. Thus, standardizing the terms used in the literature for data quality dimensions, identifying the issues associated with the different dimensions, and understanding the methods used to address these issues could improve our understanding of this phenomenon, as well as the strengths and limitations of current methods, and future research needs.

This paper aims to identify definitions proposed for data quality in healthcare, and related data quality issues and methods for resolving them. Focusing on the healthcare IS/informatics field, this research identifies statistical and computational methods used to address data quality. We propose four research questions (RQs) to guide this study:

RQ1: What are the dimensions of data quality in healthcare?

RQ2: What are the relevant issues related to these dimensions of data quality?

RQ3: What are the methods used to address these issues?

RQ4: What are the strengths and limitations of these methods?

We reviewed the current research to identify relevant articles published from the start of 2012 to June 2022, from highly ranked journals within healthcare IS/informatics journals following the guidelines of Wolfswinkel et al. (2013). The review includes relevant articles cited by these papers and articles that cite them.

Our study differs from prior literature reviews from three perspectives. First, we adopt a manual search from the top healthcare IS/informatics journals based on a set of inclusion and exclusion criteria. The consistency of the literature in the highly prestigious outlets could offer a better chance to structure and discuss statistical and computational methods to improve healthcare quality enabled by IT. Second, we refer to a well-established taxonomy development approach (Nickerson et al., 2013) in IS to categorize 21 data quality issues under seven unique dimensions of data quality and six

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-review-of-the-state-of-the-art-of-data-quality-in-healthcare/316236

Related Content

Lines of Advance in Global Information Technology Management: American/West European Approach

Ana R. del Aguila, Sebastián Bruqueand Antonio Padilla (2003). *Advanced Topics in Global Information Management, Volume 2* (pp. 87-111).

www.irma-international.org/chapter/lines-advance-global-information-technology/4513

An Exploratory Cross-National Study of Information Sharing and Human Resource Information Systems

Bongsug (Kevin) Chae, J. Bruce Prince, Jeffrey Katzand Rüdiger Kabst (2013). *Global Diffusion and Adoption of Technologies for Knowledge and Information Sharing* (pp. 280-308).

www.irma-international.org/chapter/exploratory-cross-national-study-information/72192

Offshoring: The Transition From Economic Drivers Toward Strategic Global Partnership and 24-Hour Knowledge Factory

Amar Gupta, Satwik Seshasai, Sourav Mukherjiand Auroop Ganguly (2008). *Global Information Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 3670-3677).

www.irma-international.org/chapter/offshoring-transition-economic-drivers-toward/19204

An Empirical Study of Factors Affecting RFID's Adoption in Taiwan

Dong-Her Shih, Yuh-Wen Chiu, She-I Changand David C. Yen (2008). *Journal of Global Information Management* (pp. 58-80).

www.irma-international.org/article/empirical-study-factors-affecting-rfid/3669

A Study on Effective Measurement of Search Results from Search Engines

Jin Zhang, Xin Cai, Taowen Le, Wei Feiand Feicheng Ma (2019). *Journal of Global Information Management* (pp. 196-221).

www.irma-international.org/article/a-study-on-effective-measurement-of-search-results-from-search-engines/215029