MAIN BARRIERS FOR QUALITY DATA COLLECTION IN EHR
A Review

Rui Mendes[1,2] and Pedro Pereira Rodrigues[1,3]

[1]*Faculty of Medicine of the University of Porto, Al. Prof. Hernâni Monteiro, 4200-319 Porto, Portugal*
[2]*Faculty of Sciences of the University of Porto, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal*
[3]*LIAAD - INESC Porto, L.A. & CINTESIS - Center for Research in Health*
Technologies and Information Systems, University of Porto, Porto, Portugal

Abstract: The volume of health data is rising and health information technologies which include electronic health records are a promising solution, on data management and collection, to achieve greater quality outcomes. However, they often cause errors instead of preventing them. To study the main barriers to high quality data collection from electronic health records, a qualitative review study was conducted using 5 different database engines having only considered data quality and documentation issues, opportunities and challenges for proper data collection, electronic health records data and corresponding databases quality. It were included 16 articles from which data availability, format, accuracy and data accessibility were the most focused problems to address. Still, solutions are available: early recognition of those problems, well structured and designed EHRs, standard coding use, periodic accuracy monitoring and feedback and broad use of such systems for the most daily tasks possible, among others. Altogether they can improve EHR data quality for everyday use.

1 INTRODUCTION

Technological progress, health services commitment to their use, improvements in clinician's skills and willingness to use health information technologies contribute to raise data quality to the point it can be used for research investigation. The support of the health information technology (HIT), with the use of a well structured and designed electronic health record (EHR), allows the possibility to process more accurately, effectively and with more efficiency the large amount of information being produced and managed every day and translate it to a better quality care. (de Lusignan and van Weel, 2006).As these systems are set, we should need a better physician and patient education and also a better clarification of what kind of information is necessary and wanted by both of them (Berner, 2005).

Quality improvement and error reduction are two of the justifications for health care information technologies. However, researchers evaluating the problematic implementation of clinical information systems often find situations where they're responsible for errors, instead of preventing them, affecting high quality data collection (Stead, 2007).

On this paper we aim to review the main barriers to high quality data collection from EHRs. Then we expect to have a better understanding why and what is possible to be made to achieve better outcomes.

2 SEARCH STRATEGY AND STUDY SELECTION

A qualitative review study was conducted on the found literature about the main barriers in high quality data collection from EHRs.

The database research was held on Google, Google Scholar, PubMed, Scopus, ISI Web of Knowledge and ScienceDirect using the following key-words: barriers, high quality data, data collection, EHR, Electronic Health Records. At PubMed the following MesH Terms were used - "Medical Records Systems, Computerized", "data collection" – plus the key-words – "quality" and "barriers".

Some queries were applied in order to refine the search. From a selection of 116 eligible articles (17 PubMed, 69 ScienceDirect, 2 ISI Web of

Knowledge, 21 SCOPUS and 7 Google Scholar) a title and abstract analysis was performed from which a total of 35 articles were selected regarding data quality collection and health information structures required to data quality. Workshop analysis, forum presentations, letters and papers which regarded implementation issues and situation analysis reports of a given institution, general perspectives on quality care improvement following the use of electronic health records and specific workflow analysis were excluded. After full-text review of 27 articles concerning data quality and documentation issues, opportunities and challenges for proper data collection, electronic health records data and corresponding databases quality, a selection of 16 articles was made.

3 MAIN BARRIERS TO HIGH QUALITY DATA COLLECTION

Some problems may arise along the path from collecting raw data into useful information, called information quality problems. Aiming to data quality we should define it accordingly with the definition given in total data quality management (TDQM) as data fit for purpose by its consumers (Strong et al., 1997a, de Lusignan et al., 2006; Cruz-Correia, 2009). On several occasions "information" normally relates to both data and information; but "data" usually refers to information in its early stages of processing, and "information" to the product at a later stage (Strong et al., 1997a).

The data flow process has several actors who influence the quality information obtained from such data at a later stage (Cruz-Correia, 2009). Having the right information when and where it's needed comes with certain demands like the increased need for its correct filtering, context-sensitive decision support, legal and ethical guidelines regarding obligations to obtain and use the information, achieve real patient-physician expectations regarding the use and usefulness of the information, and enhancing data accuracy. Health care is an information-based science. Many clinical practice acts involve gathering, synthesizing, and acting on information (Hersh, 2002). Since patient information has traditionally been incomplete and fragmented, lifelong EHR stands as a promising solution to achieve complete and accessible information (Berner, 2005).

3.1 Electronic Health Records

According with the International Organization for Standardization (ISO) definition for EHR, it is a repository of patient data in digital form, stored and exchanged securely, and accessible by multiple authorized users. It has retrospective, concurrent, and retrospective information and its primary purpose is to support continuing, efficient and quality integrated health care (Häyrinen, 2008).

Information manufacturing process encompasses three main roles: information producers, responsible for generating and providing information; information custodians, who provide and manage computing resources for storage, maintenance, and securing information and information consumers, who access and utilize information for their tasks (Strong et al., 1997a).

There are four major aspects known to information quality and fifteen dimensions underlying them (Strong et al., 1997b). These four major characteristics related to high-quality data are: intrinsic data quality, data quality context, data quality representation and data quality accessibility. A quality data problem is defined when any difficulty is encountered along one or more quality dimensions that turn data completely or largely unfit for use (Strong et al., 1997b).

From the selected literature data availability (on 15 articles), data format (on 15 articles), data accessibility (on 14 articles), data accuracy (on 12 articles) constitute the main barriers in contrast data validation, revenue cycle management, auditing (only on 6) and data cleansing (on 4) were less focused.

3.1.1 Data Sources/Availability

There are difficulties associated in storing over time large amounts (not necessarily better) of varied information which often has conflicting or ambiguous concepts across different computer systems: lack of pieces of information, different values or representations (formats or codes), aggregated and non-aggregated impaired data movement across the industry due to lack of mapping and connecting different and inconsistent sources of data, ineffective data collection mechanisms for some required fields; errors at data entry from users and no data entry validation mechanisms at that point; lack on the use of international terminologies and resulting poor semantic interoperability.(Strong et al., 1997a; Weiner, 2007; Häyrinen, 2008; Vaughan, 2009).

Common data dictionaries and data warehouses are a current solution to distributed system problems (Strong et al., 1997a). The alternative is constant maintenance of data and systems to address changing data requirements (Strong et al., 1997b). To improve data aggregation should be considered a standardized infrastructure and moving to a single comprehensive controlled vocabulary for structured data, making data transfer between different services easier (Hersh, 2002; de Lusignan et al., 2006). Dedicated technology and human resources are necessary to monitor, catch, and correct errors at the point of transfer (AHIMA, 2008). In the event of systems failure, business continuity planning, policies, and procedures for healthcare documentation are fundamental assets for data and documentation quality (AHIMA, 2008).

3.1.2 Data Format

There are four methods for data capture in EHRs: entering data directly, including templates or screens completed by the user; scanning handwritten documents; transcribing text reports created by using dictation or speech recognition; interfacing or feeding data from other information systems, such as laboratory systems, radiology systems, blood pressure monitors, or electrocardiographs. Each one of these methods has strengths and weaknesses that may have an impact on data quality (AHIMA, 2008).

Direct data entry produces discrete, structured data that can easily be analyzed and reported. However, such data may be less accurate and negatively impact the quality of documentation (McDonald, 1997; AHIMA, 2008). On the other hand much information is stored as unstructured, narrative data. Such data are difficult to use reliably in queries for several reasons, including among others misspellings, synonyms, homonyms and negation (Weiner, 2007). "Coded" data are needed to better represent a clinical concept, since there are many forms to represent it, giving the necessary attention to coding systems dynamics – new codes are added all the time without old ones being removed (Strong et al., 1997a; de Lusignan and van Weel, 2006; Häyrinen, 2008). At present, there isn't a single standard system for recording structured data, a standard approach to coding and classification (de Lusignan and van Weel, 2006).

3.1.3 Data Accuracy

An accurate electronic health record can eliminate rework by capturing data once at the source and presenting it for reuse as needed later on, but is rarely achieved in practice (de Lusignan and van Weel, 2006; Stead, 2007). Accuracy of system documentation is normally calculated using two measures: the proportion of documented observations in the system that are correct (true) – correctness; and the proportion of observations that are documented – completeness (Berner, 2005; Stead, 2007). Common causes for data inaccuracy include placing a question in the wrong person's workflow; not allowing for clinically relevant answers; reflecting what the physician ordered but not what the patient really did; among other gaps in information about care by providers who are not using the system (Stead, 2007). Also establishing the order of events and the time lapse between each one is also problematic, especially when are used several unsynchronized mechanisms to tell the time (Cruz-Correia, 2009). Another problem comes when we don't know where and who entered such data. Bayesian inference, the development of terminology and minimal data set standards and also structured data entry may improve data completeness (Strong et al., 1997a; Berner, 2005; Weiner, 2007; Häyrinen, 2008).

3.1.4 Data Accessibility

Data accessibility (filtered by ethical issues like data ownership, security, confidentiality and privacy) is surely an obstacle to research investigation by third parties, as this issue is still unclear, without access to them analysts can't do research and managers can't make decisions, like the unclear details about the research methods employed by researchers, not allowing studies replication (Strong et al., 1997a; de Lusignan and van Weel, 2006; Kaplan and Harris-Salamone, 2009). Structured notes allows easier information retrieval; as when an information system is used, and semantic tagging of information is used (Häyrinen, 2008). Policies and procedures development should also consider, data capture and access control methods, determine when a record is complete, auditing, evaluation and maintenance of code sets, attend to which components refer to the legal health record and privacy and security regarding integrity issues as well (AHIMA, 2008). These permissions are also barriers to accessibility and affect the overall reputation and value of this data (Strong et al., 1997b).

4 MAIN FINDINGS AND RECOMMENDATIONS

From the point of view of this work intrinsic data quality, data quality context, data quality representation and data quality accessibility were identified as major data quality characteristics. Data availability, data format, data accuracy and data accessibility arise as major problems identified, relating to high-quality data collection on EHRs. There are solutions to solve such problems like early recognition of development of those problems and direct physician entry or physician entry control. Also, structured encounter forms and well structured and designed EHRs that include anticipatory prompts and that allow data linkage and aggregation to data consumers are part of the solutions available. A broad use of such systems for the most daily tasks possible without compromising the goal of compliant documentation and standard coding use are also to consider. Other relevant issues are periodic accuracy monitoring and feedback, better research methods explanation, evidence-based guidelines, automated data capture from patient information systems and others. If attended they can help reducing data quality problems in order to improve EHRs suitability for general everyday use.

REFERENCES

Ahima. (2008). Quality Data and Documentation for EHRs in Physician Practice. *Journal of AHIMA,* 79**,** 43-48. Retrieved from AHIMA Body of Knowledge.

Berner, E. S., Moss, J. (2005). Informatics Challenges for the Impending Patient Information Explosion. *J Am Med Inform Assoc,* 12**,** 614-617. doi:10.1197/jamia.M1873.

Cruz-Correia, R., Rodrigues, P. P., Freitas, A., Almeida, F. C., Chen, R., Costa-Pereira, A. (2009). Data Quality and Integration Issues in Electronic Health Records. *In:* HALL, H. V. C. A. (ed.) *Information Discovery on Electronic Health Records.* CRC Data Mining and Knowledge Discovery Series.

de Lusignan, S., Hague, N., Van Vlymen, J. & Kumarapeli, P. (2006). Routinely-collected general practice data are complex, but with systematic processing can be used for quality improvement and research. *Informatics in Primary Care,* 14**,** 59-66. Retrieved from: http://www.ingentaconnect.com/content/rmp/ipc/2006/00000014/00000001/art00008.

de Lusignan, S. & Van Weel, C. (2006). The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam. Pract.,* 23, 253-263. doi:10.1093/fampra/cmi106.

Häyrinen, K., Saranto, K., Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform.,* 77, 291-304. doi:10.1016/j.ijmedinf.2007.09.001.

Hersh, W. R. (2002). Medical Informatics: Improving Health Care Through Information. *JAMA,* 288**,** 1955-1958. doi:10.1001/jama.288.16.1955.

Kaplan, B. & Harris-Salamone, K. D. (2009). Health IT Success and Failure: Recommendations from Literature and an AMIA Workshop. *Journal of the American Medical Informatics Association,* 16**,** 291-299. doi:10.1197/jamia.M2997.

Mcdonald, C. J. (1997). The Barriers to Electronic Medical Record Systems and How to Overcome Them. *J Am Med Inform Assoc,* 4**,** 213-221. Retrieved from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC61236/

Orfanidis, L., Bamidis, P. D. & Eaglestone, B. (2004). Data Quality Issues in Electronic Health Records: An Adaptation Framework for the Greek Health System. *Health Informatics Journal,* 10**,** 23-36. doi:10.1177/146045804040665.

Pawlson, L. G. (2007). Health Information Technology: Does It Facilitate Or Hinder Rapid Learning? *Health Aff,* 26**,** w178-180. doi:10.1377/hlthaff.26.2.w178.

Stead, W. W. (2007). Rethinking Electronic Health Records to Better Achieve Quality and Safety Goals. *Annual Review of Medicine,* 58**,** 35-47. doi:10.1146/annurev.med.58.061705.144942.

Strong, D. M., Lee, Y. W. & Wang, R. Y. (1997a). 10 Potholes in the Road to Information Quality. *Computer,* 30**,** 38-46. doi:10.1109/2.607057.

Strong, D. M., Lee, Y. W. & Wang, R. Y. (1997b). Data quality in context. *Commun. ACM,* 40**,** 103-110. doi:http://doi.acm.org/10.1145/253769.253804.

Vaughan, C. (2009). *Three Barriers to Effectively Using Information Stored in EHRs.* HealthLeaders Media. Retrieved from: http://articles.icmcc.org/tag/secondary-data-use/

Weiner, M. G., Lyman, J. A., Murphy, S., Weineer, M. (2007). Electronic health records: high-quality electronic data for higher-quality clinical research. *British Computer Society.* Retrieved from IngentaConnect.