

On some aspects of medical data quality

Constantin Gaidric, Galina Magariu, Tatiana Verlan

Abstract

This paper examines the specific problem of the quality of medical data of patients when they are hospitalized for treatment through the lens of the general problem of data quality. The need to apply international standards in data quality is discussed, taking into account the specifics of the medical field and the national standards and regulations. This necessity is considered with respect to the problem of medical institutions switching to IT systems with electronic medical records and electronic health records. The specifics of the procedures for filling in the files of hospitalized patients are highlighted.

Keywords: medical data, data quality, medical information systems, electronic medical record, electronic health record.

MSC 2020: 92C50, 68P27.

1 Introduction

The problem of data quality is becoming more and more important in all areas of human activity, even more so in healthcare. Even though virtually all of society understands the need to move to a digital economy, data is not treated as an essential and precious asset.

Data quality is not a simple scalar measure; it is defined by several dimensions, which reflect certain aspects with special meanings for different users, depending on the purposes for which the data is examined. Evaluation of the quality of data consists in establishing a value for each dimension by which it is appreciated how an aspect, a characteristic quality, is achieved, that allows understanding or decision-making according to the proposed objective.

In most medical institutions (ambulances, clinics, hospitals, laboratories) the data of patients' medical investigations are collected. Depending on the institution's management system, they are contained in the databases to which users from those institutions or the network of medical communities have secure access.

The quality of data in medical institutions has two aspects: the first one refers to the implications for the patient (without correct data we cannot expect the correctness of the diagnosis, and, therefore, the treatment process and, ultimately, the patient's health); the second one refers to the efficiency of the institution's management that ensures its functioning adequately to the tasks within the limits of the available sources. In this paper, the first aspect is examined. The impact of poor data quality can affect decision-making on treatment tactics and has a strategic influence on the treatment duration.

One of the most valuable assets in current business, as well as in planning for enterprises and institutions, is data. High-quality data is essential for each individual also. Data quality is naturally an evolving concept.

However, the issue of data quality is extremely important for correct decision-making and needs to be taken seriously. One example of this attitude is the UK, where this problem is considered at the government level. The Government Data Quality Framework was created, which provides "a more structured approach to understanding, documenting and improving the quality of its data" [12] (about data to which public bodies have access). "It presents a set of principles for effective data quality management, and provides practical advice to support their implementation." [12] In the development of this important framework as part of the National Data Strategy, the Data Management Association of the UK (DAMA UK) took part. All this shows the level of seriousness in approaching the problem of data quality in the UK.

The Data Management Association of the United Kingdom (DAMA UK) considers data quality dimensions to be "measurable features or characteristics of data" [12]. They are used to assess data quality and identify problems related to data quality, and therefore, the quality of important decisions. DAMA UK defines six basic dimensions of data quality (Completeness, Uniqueness, Consistency, Timeliness, Validity,

Accuracy), which can be and are completed according to user needs.

Understanding what good data means and how it can be measured and improved, if necessary, is quite a difficult problem for several reasons. There are a variety of definitions, and the number of dimensions taken into account differs from one domain to another depending on the context. Also, it depends on the vision of those who manipulate the data and for what purpose the data is applied [1].

Quality properties related to data usefulness are called “dimensions” in the literature on data quality.

A dimension is a measurable property of quality that represents some aspect of data (relevance, accuracy, consistency, etc.) and can be used to judge quality. Thus, some concrete data may be considered of high quality in certain respects according to one set of dimensions and less qualitative according to another set of dimensions. Probably, completeness, if this is not one of the most important dimensions, without which it is impossible to talk about the quality of some data, then this is the most frequently requested and encountered one. ISO 8000 [2] is the generally accepted standard for data quality in businesses and organizations.

In the last decades, in scientific publications and those of practical applications in management, increased attention is paid to the problem of the quality of data and information in databases. It is an undeniably important fact, but no less important is the quality of the data used in hospitals in the initial period, at the admission of the patient when a local database is formed from the data contained in the referral form from the family doctor and in the examination of the doctor in the admission department.

Data is fundamental to making correct, effective, evidence-based decisions. Perfect data quality is not always achievable, and therefore, the decision maker, in our case the physician, should understand what additional data would be needed to still ensure the achievement of the possible goal. So, a structured approach is needed to understand, document, and improve the quality of the data we intend to use.

According to many authors including [2], quality in use is generally considered the degree to which a product or a system can be used by users to achieve objectives with effectiveness, efficiency, lack of risk,

and satisfaction in specific contexts of use. The properties of quality in use are classified depending on the specifics of the field of activity through different sets of characteristics, among which the most common are: timeliness, precision, traceability, effectiveness, efficiency, and availability.

The standard ISO/IEC 25010:2011 “Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models” [2] defines:

- A model of quality in use and interaction of the product in a specific context, applicable to the human-computer system, including the mode of operation of the system.
- A product quality model composed of eight characteristics (which are subdivided into sub-characteristics) that relate to the static properties of software and dynamic properties of the information system.

The features in both models are relevant to both software products and any computer systems in the field. Terminology of characteristics (Accuracy, Completeness, Reliability, Relevance, Timeliness) and sub-characteristics is defined so that to be applied for measuring and evaluating quality (see Table 1). Thus, the set of quality characteristics can be selected for each context and compared if it corresponds to the applied standard.

We insistently promote the idea of using the ISO/IEC 25010:2011 standard even for traditional medical records (on paper), because their gradual transition to electronic support is inevitable. Information systems for electronic medical records keeping can ensure the high quality of medical data, their safety, accuracy, reliability, and easy exchange of information between patient and physician.

Nowadays, there is a world tendency of using patients’ medical data in different healthcare institutions in electronic format. According to data from The Commonwealth Fund [10], in developed countries, primary care physicians actively use this format in their practice: Australia – 97% of physicians; Canada – 86 %; England, Netherlands, and

Table 1. Characteristics of data quality in the ISO standard. Definitions adapted and modified according to [3]

Characteristic of data quality in the ISO standard	Definition
Correctness	the degree to which the data correctly represents the true value of the attribute
Completeness	data has values for all attributes expected in a specific usage context
Consistency	the degree to which the data is consistent and free of contradiction with other data
Credibility	the degree to which the data is considered true and credible by the user
Timeliness	the degree to which the data is age-appropriate in a specific context of use
Accessibility	data can be accessed in the context used
Conformity	the degree to which the data complies with the standards or regulations in force at the institution that maintains an information system
Confidentiality	the extent to which data is accessible and interpretable only by authorized users
Efficiency	the degree to which the data can be processed and provide the expected levels of performance
Precision	the degree to which the data is accurate
Traceability	the degree to which data access and changes are ensured
Understanding, comprehension	the degree to which data can be read and interpreted
Availability	the degree to which the data can be accessed
Portability	the degree to which the data allows installation, replacement or migration from one system to another maintaining the existing quality in a specific context of use
Recovery	the degree to which a certain level of operations and data quality is allowed to be maintained and preserved, even in the event of failure

Norway – 99 %; France and Germany – 88 %, Sweden – 98 %, New Zealand – 100 %, United States – 91 %.

Developing countries show a lower percentage of such use because of the problems in the level of health care and understanding the new technologies and their potential in resolving challenges in the domain [11].

We believe that in view of the transition over time of any system for patient records maintenance to the electronic version, it can be recommended to maintain and use these standards for patient records, a fact that would not require changes in the process of transition to electronic records.

Standard ISO/IEC 25012 [4] defines the data quality model that includes the same 15 characteristics and sub-characteristics from two points of view: inherent, which includes all the intrinsic characteristics, with the potential to explicitly and implicitly conform to the needs under the specified conditions; and dependent on the system, which refers to how data quality is obtained and maintained in an information system under specified conditions.

Decisions accompanying our daily lives should be based on high-quality data, so estimating data quality is a fundamental element in ensuring the relevance of decisions based on the data used. To increase confidence in data-driven decisions, it is necessary to measure and know the quality of the data used with appropriate tools [5].

A wide variety of commercial, open-source, and academic data quality tools have been developed based on scientific research. The range of functions offered by these tools varies widely, as the term “data quality” is context-dependent and not always used consistently.

Some tools exclusively offer data cleaning and enhancement functionalities that specifically address measurement capabilities, i.e., detection of quality issues and, most commonly, automated data modification (e.g., data cleaning). But this is not usually possible in information systems.

According to [8], tools that detect and report data quality problems are needed and a large number (667) are found, of which 50.82% are domain-specific, dedicated to certain types of data. Among the most common services they offer are data profiling, data quality measure-

ment, and continuous data quality monitoring, but these services are applicable only in specific areas.

2 Quality of data included in medical records

In healthcare, in recent years, many countries have transitioned or are in the process of transitioning from paper to digital records by hospitals, doctors' offices, clinics, and health care facilities. In the specialized literature, when talking about medical information systems, sources of data, and information about the patient, as a rule, the terms *electronic medical record* and *electronic health record* are used.

Electronic medical record (EMR) is a digital version of a patient's paper medical record that contains the patient's limited medical history completed mainly by the family doctor and the specialist doctors to whom the patient was referred for diagnosis and treatment. Patients' EMRs are typically owned and completed by primary institutions of specialized medical care, regardless of size. EMRs are not transmitted outside the institution unless the patient is admitted for inpatient treatment.

The electronic health record (EHR) contains patient information from all medical institutions that, over time, have been involved in patient care.

The EHR can include medical history, vital signs, progress notes, diagnoses, medications, immunization data, allergies, laboratory data, and imaging reports. It may also contain other relevant information such as insurance information and demographics. When talking about health care reform, it is necessary to emphasize the meaningful use of EHR. EHRs are designed as interoperable systems, which allow data from different systems to be accessed and used. Each medical institution can have access to the complete medical history of the patient, even if he was treated by other institutions. So the patient's medical information reaches every specialist, laboratory, or center, to which he calls. The EHR should be a comprehensive source of medical information on the general health status of patients, designed to be accessible by any authorized institution for diagnosis and medical care.

In Moldova, there are some popular EHR implementations and

many small EMR medical information systems installed in both private and public hospitals and diagnostic centers [6].

An example of an EHR is the DICOM Network. This is a distributed medical image preprocessing and archiving system.

The DICOM network was launched in Moldova in 2012 with the aim of providing access to collected imaging data for medical personnel with access rights and also access for patients when they need their personal data. Today, the system is implemented in many hospitals in Moldova, collects and processes more than 5 TB of data per month collected from hospitals equipped with different types of medical equipment [7].

Most of these systems contain both the data from the medical records of the patients registered at these medical institutions and also different types of medical image collections. The patient's personal data is the most sensitive and important information that should correspond to the main dimensions of data quality (completeness, correctness, timeliness, accessibility, conformity, etc.) in order to be used with confidence by each medical institution to which the patient has referred, for external consultations, when medical analyzes or images should be transferred to another medical institution, or when the patient wants to familiarize himself with the analyses and images of his latest investigations. A mandatory condition is securing access to any data and images from the EHR.

In the medical institutions of the Republic of Moldova, personal data is protected by the regulation NCPPD (National Center for the Protection of Personal Data of the Republic of Moldova), which is based on national legislation.

In most hospital institutions, internal standards for maintaining data from patient records are established. For example, in the Timofei Moșneaga Republican Clinical Hospital (Chisinau) in order to standardize data from different departments and to establish a standard method of completing and maintaining the inpatient medical record, the *Standard Operating Procedure* was approved (<https://scr.md/page/ro-proceduri-operaionale-269>). One can talk about pragmatic quality which refers to how the data enable the medical staff to achieve their goals and how easy to understand, how clear, how usable they are.

This procedure aims to ensure the quality of patient data and, first of all, the correctness and completeness of the medical care process at all stages in order to ensure the quality of the treatment given to the patient during his hospitalization and monitoring.

The most sensitive information present is medical data about patients. Denise Silber [7] finds that more than 86% of eHealth data errors are administrative or process errors: wrong recording of patient information, wrong test orders, and wrong drug prescriptions. The conclusions are based on a study conducted by 42 doctors over a 20-week period in 2000, which highlighted errors that occur outside the hospital setting.

Data quality is an important concern in any application scenario being both frequent and potentially costly. It is estimated that 2% of customer records become obsolete in a month due to data “degradation”. Medical centers are plagued by unresolved data quality issues. For medical centers, the most pressing need is to estimate data completeness.

Patients’ health can be adversely affected by a lack of documentation on treatments and medications. Other data quality considerations are not to be neglected. These domain-specific problems stem from deficiencies in data management processes or technical restrictions that may also occur in other domains. Data quality, in general, depends on context, that is, on notions of “good” or “poor” data that cannot be separated from the context in which they were produced and used (medical analysis data can be considered good as long as they fall within the limits for each type of analysis, for example, temperature 35-43°C, blood sugar – in the range of 70-120 mg per 100 ml).

3 Some procedures for filling in the records of inpatients

The doctor from the inpatient department, after examining the patient and data from the accompanying documents, enters the *inpatient diagnosis* in the medical record of the inpatient.

Upon inpatient admission, the nurse at the registry completes the

inpatient's medical record, in which the *referral diagnosis* is entered according to the referral extract from the outpatient's medical record and the *inpatient diagnosis* (or *Admission Diagnosis* – diagnosis given to patient on admission to hospital).

The attending physician prescribes the exhaustive examination of the patient according to the admission diagnosis, and after the first 72 hours of admission specifies the diagnosis.

The medical record of the inpatient is completed by the staff from different subdivisions of the institution, who fix the information related to the diagnostic process, the dynamic evolution of the pathological process, and the applied treatment. The data in the record are intended to exhaustively reflect the assistance provided, the correctness and completeness of the data about the doctors' prescriptions, and the care provided by the nurses.

Given that the correct diagnosis and care of the patient are directly influenced by the quality of the data in the patient record, we will try to evaluate at what moments and on whom the quality of the inscriptions in the record depends.

Currently, not all medical institutions maintain patient records in electronic format and there are no databases of medical institutions connected in a functional national network.

The trends of managers and healthcare workers in the digital transformation of services and the use of artificial intelligence can only be achieved if they are based on good-quality data throughout the entire process: *preliminary diagnosis – hospitalization – evidence-based diagnosis – treatment – recommendations for the discharged patient and the family physician – storing data and information in the EHR* for subsequent decision-making and service delivery.

However, it is imperative to take into account the ISO/IEC 25010 standard that replaced ISO/IEC 9126-1:2001 and was developed to maintain the quality and evaluation process of the IT product. The standard includes a quality model in use composed of five characteristics (effectiveness, efficiency, satisfaction, freedom from risk, and context coverage) and a quality model defined by the ISO/IEC 25012 standard which includes 15 characteristics valid for any field. We will operate with only those specific to inpatient records for specifying,

measuring, and evaluating quality.

Although the scope of the product quality model is for IT systems, many of the characteristics are also relevant to wider systems and services.

The first step in completing the inpatient record is to enter the *referral diagnosis* and *admission diagnosis*.

That is why the data used to argue the referral diagnosis will be checked in terms of validity (if the defined parameters are according to the format of the given hospital and fall within the time interval in which the analyses are considered current), completeness (if they are sufficient to give credibility to the referral diagnosis), accuracy/ up-to-date (if they are not older than 3 days, so they can be used), compliance (the standards or regulations in force regarding the specific context of the case are respected), correctness (the data correctly represent the value of the attributes what confirms the diagnosis) and credibility (the degree to which the staff trusts the included data and reflects the doctor's needs in establishing the diagnosis).

For admission diagnosis, compliance is checked. The data included in the patient record are those that allow the attending physician to initiate the treatment process and prescribe his examination according to the protocol.

The analyses data and the attending physician's own observations lead to specifying the patient's diagnosis based on which the treatment continues.

The attending physician must be sure that the data he accesses and based on which he will propose the tactics of the patient's treatment are sufficient and satisfactory to obtain a result in a reasonable time.

Since the attending physician acts according to the protocol, and all investigations including analyses are carried out following the internal regulations of the hospital, the problem of trust in the quality of the data used to specify the diagnosis and the treatment of the patient becomes less acute. When the patient is discharged, it should be determined whether the data used by the doctor should be archived. In such a case, they must be: 1) accessible to the medical institutions to which the patient later turns and 2) secured. It would also be useful to store the data together with the information about its quality. How-

ever, in order to ensure traceability and interoperability of the system and compliance with all the compartments of the ISO 8000 standard, it is necessary to monitor the elements that characterize such dimensions as accuracy, promptness, conformity, and completeness.

4 Conclusions

This paper is the extended and revised version of the conference paper [13] presented at WIIS 2023. In this article, the problem of the quality of medical data was examined with the aim of facilitating the transition process to integrated health data systems that would provide the possibility for the attending physician to dispose of all the necessary medical investigations of the patient regardless of the location and the medical unit (laboratory, clinic) at which they were performed.

Acknowledgments. The project Ref. Nr. 20.80009.5007.22 “Intelligent Information systems for solving ill structured problems, knowledge and Big Data processing” has supported part of the research for this paper.

References

- [1] C. Batini, M. Palmonari, and G. Viscusi, “Opening the closed world: A survey of information quality research in the wild,” in *The Philosophy of Information Quality*, Springer International Publishing, 2014, pp. 43–73.
- [2] ISO/IEC 25010:11. *Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models*, 2011. [Online]. Available: <https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en>.
- [3] Tuomo Uotila and Helina” Melkas, “Complex Knowledge Conversion Processes and Information Quality in Regional Innovation Networks,” *Knowledge and Process Management*, vol. 15, no. 4, pp. 224–234, 2008, www.interscience.wiley.com. DOI: 10.1002/kpm.317.

- [4] ISO/IEC 25012 [ISO/IEC, “Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model,” ISO/IEC, Tech.Rep. ISO/IEC 25012, 2008.
- [5] Bernd Heinrich, Diana Hristova, Mathias Klier, Alexander Schiller, and Michael Szubartowicz, “Requirements for Data Quality Metrics,” *Journal of Data and Information Quality*, vol. 9, no. 2, pp. 12:1–12:32, January 2018.
- [6] Alexandr Golubev, Petru Bogatencov, Grigore Secrieru, Ecaterina Matenco, “Incident Handling and Personal Data Protection in Medical Images systems,” in *Intelligent information systems for solving weakly-structured problems, processing knowledge and big data*, Chisinau, Editors: S. Cojocar, C. Gaindric, I. Tițchiev, and T. Verlan. Chișinău: Vladimir Andrunachievi Institute of Mathematics and Computer Science, 2022, pp. 223–228. ISBN 978-9975-68-415-6. 004+519.7(082).
- [7] Peter Bogatencov, Nicolai Iliuha, Grigore Secrieru, and Alexandr Golubev, “DICOM Network for Medical Imagistic Investigations Storage, Access and Processing,” *Networking in Education and Research Proceedings of the 11th RoEduNet IEEE International Conference*, (Sinaia, Romania), January 17-19, 2013, pp. 38–42.
- [8] Lisa Ehrlinger and Wolfram Wöß, “A Survey of Data Quality Measurement and Monitoring Tools,” in *Front. Big Data*, vol. 5, Sec. *Data Mining and Management*, 31 March 2022, <https://doi.org/10.3389/fdata.2022.850611>.
- [9] Guide to Health and the Best Doctors. [Online]. Available: <http://www.hbroussais.fr/HEGP>. (in French)
- [10] The Commonwealth Fund. “Primary Care Physicians’ Use of Electronic Medical Records, Selected Health & System Statistics,” June 05, 2020. [Online]. Available: <https://www.commonwealthfund.org/international-health-policy-center/system-stats/primary-care-physicians-emrs>.

- [11] Mosharop Hossian, “Electronic Medical Record and Electronic Health Record have potential to improve quality of care,” *ACADEMIA Letters*, July 2021. [Online]. Available: <https://doi.org/10.20935/AL1565>.
- [12] Government Data Quality Hub, “Guidance. The Government Data Quality Framework,” gov.uk, 3 December 2020. [Online]. Available: <https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework>.
- [13] Constantin Gaidric, Galina Magariu, and Tatiana Verlan, “On some aspects of medical data quality,” in *Proceedings of Workshop on Intelligent Information Systems: WIIS2023*, (Chisinau, October 19-21, 2023), 2023, pp. 128–140. ISBN 978-9975-68-492-7.

Constantin Gaidric¹, Galina Magariu²,
Tatiana Verlan³

Received August 24, 2023
Accepted November 29, 2023

^{1,2,3}Vladimir Andrunachievici Institute of Mathematics
and Computer Science, SUM

5, Academiei street, Chisinau, Republic of Moldova, MD 2028

¹ ORCID: <https://orcid.org/0009-0003-2893-9626>

E-mail: constantin.gaidric@math.md

² ORCID: <https://orcid.org/0009-0001-4860-8715>

E-mail: galina.magariu@math.md

³ ORCID: <https://orcid.org/0009-0006-4519-1105>

E-mail: tatiana.verlan@math.md