

# Data Quality Methods and Applications in Health Care System: A Systematic Literature Review

Paolo Pietro Biancone<sup>1</sup>, Silvana Secinaro<sup>1</sup>, Valerio Brescia<sup>1</sup> & Davide Calandra<sup>1</sup>

<sup>1</sup> School of Management and Economics, University of Turin, Turin, Italy

Correspondence: Davide Calandra, School of Management and Economics, University of Turin, Turin, Italy.  
E-mail: [davide.calandra@unito.it](mailto:davide.calandra@unito.it)

Received: January 3, 2019

Accepted: February 12, 2019

Online Published: March 8, 2019

doi:10.5539/ijbm.v14n4p35

URL: <https://doi.org/10.5539/ijbm.v14n4p35>

## Abstract

The use of data cataloging tools allows keeping different records of both qualitative and quantitative information. However, the large amount of data is not always synonymous with quality, in the medical field this argument becomes even more critical if it is considered the consequences that the lack of a systematic and rigorous process can have for patients. The analysis was conducted through a systematic review includes several general cases and practical methodologies of data quality analysis in the health context. The search for results was made using the keywords "data quality" and "health." The study considers publications made from 2014 to 2018, topic related to Business, Management and Accounting, exclusively in the case of the Tutto platform peer-review journals were chosen, English language of publication. Efficient use of information requires databases that can collect and order health information. However, this is the first step, data quality attempts to go further through the creation of qualitative or statistical control processes and indicators able to ascertain the lack of data or identify potential anomalies. The conducted analysis sets the stage for future quality implementation in the clinical pathway and patient management.

**Keywords:** data quality, health data management, healthcare, applications

## 1. Introduction

The availability of data and their quantity is today one of the issues most addressed by different sectors. Health is no exception, in recent years there has been more significant interest in the correct management of qualitative and quantitative data, considered vital to improving service and health outcomes (Frost & Sullivan, 2012).

Before starting the review, we believe it is necessary to highlight some definitions. According to the distinction made by Davenport (1998) for "data", we mean a discrete and objective fact on events, by "information" we mean in the same way a data but transformed by the processes of adding value, contextualization, categorization, calculation, correction, and condensation.

The term data quality means the ability of data and information to respond optimally to the intended purpose; in particular, we often refer to a process characterized above all by a precise knowledge of the elements, and secondly by their management and analysis (Davenport, 1998). The main phase of data quality assessment, however, is the verification of all data management phases, the ultimate goal is to identify any deficiencies and increase their quality by reducing the costs of non-quality (Batini & Scannapieco, 2006; Wills, 2014; Biancone, Secinaro & Brescia; 2018).

The reference legislation derives from the International Organization for Standardization 8000, the documents contained in it intend to provide first of all an overview of the concept of data quality and subsequently to delineate the entirety of the characteristics. The indications, for the reasons explained above, are generic and applicable to any context (ISO, 2015).

The multidisciplinary nature of the subject requires, however, to verify the specific best practices for each business. In the sanitary case, the World Health Organization has outlined the main characteristics and indicators (World Health Organisation, 2017a, 2017b, 2017c) through a dual approach that includes both the theoretical part and the practical application.

The basic framework is therefore characterized by two types of sources: the first ISO 8000 generic and applicable to different contexts and the second directly applicable to the healthcare world (Figure 1).

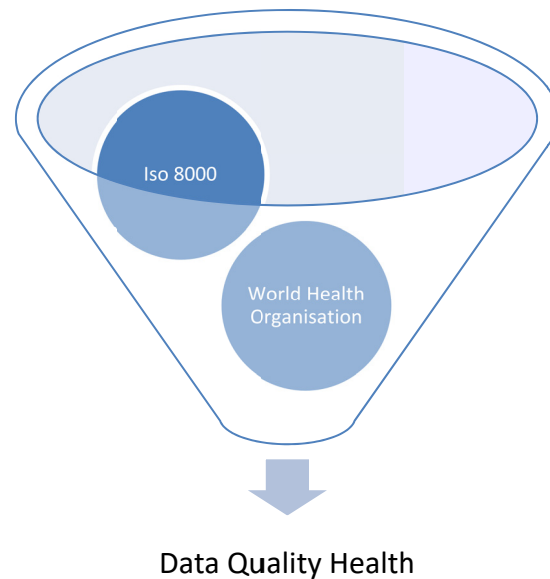


Figure 1. Theoretical background for data quality health

In this case, the metrics provided by the World Health Organisation (2017a) identify which data quality indicators:

1. The completeness and timeliness of the data;
2. The internal coherence of data;
3. External coherence of data;
4. Comparisons of data on the entire population.

For completeness it means the measure of the entities envisaged and supposed to be within a database or a collection of data, the scope of application may be local, district, regional or provincial (World Health Organisation, 2017a).

Timeliness refers to the presentation of the entities envisaged by the deadline (World Health Organisation, 2017a).

Internal coherence is understood as the information capacity of the same data at different times or concerning the documents of origin and databases (World Health Organisation, 2017a).

External coherence reflects the comparison between two sources of data measuring the same health indicator (World Health Organisation, 2017a).

In order to make comparisons, it may be necessary to determine the adequacy of the data relative to the study population compared to a larger population, in this sense the obvious comparisons represent a benchmark and comparison measure (World Health Organisation, 2017a).

The measures presented remain functional to the set-up of a health data quality model and will be verified in the results included in this review.

## 2. Method

The systematic review was conducted in October, and November 2018 and two databases were considered: Scopus and Tutto (It is the database used by the University of Turin and brings together all the available collections. The search criteria used will be described for greater clarity and expository transparency). The search for results was made using the keywords "data quality" AND "health."

For the selection of the results the following inclusion criteria were used:

- publications made from 2014 to 2018;
- topic related to Business, Management and Accounting;
- exclusively in the case of the Tutto platform peer-review journals were chosen;

- English language of publication.

Results that are not directly related to the theoretical or practical meaning of the data quality process in the health sector have been excluded.

Tutto is a platform able to put together several collections, and the research team has isolated the results deriving from Scopus as already present in the systematic review.

To this end the results belonging to the collections of Taylor & Francis (Bai et al., 2018), BioMed Central (Ledikwe et al., 2014), Springer (Mitsunaga et al., 2015), Journal of Health Management & Informatics (Firouraghi et al., 2018), BioMed Central (Skyttberg et al., 2016), Elsevier (Verma, 2014), Oxford University Press (Watson et al., 2017), American College of Healthcare Executives (Wills, 2014) and National Institutes of Health (NIH) Health Care Systems Research Collaboratory (Zozus M. N., et al., 2014).

### *2.1 Data Extraction*

The research team summarized the results by extracting the information of interest. For each article, the information concerning the author, the year, the type of study, the main characteristics, the measures used, the applications, the data collection method and the reference country (Table 2) were selected.

With the same criterion, it was possible to analyse the compatibility of the selected results with the metrics provided by the World Health Organisation (2017a) and mentioned in our introduction (Table 3).

### *2.2 Excluded Results*

The search criteria returned 840 results.

The reading of the title and abstracts allowed to validate for the systematic review 16 sources equal to 1.90%; the rejection rate was 98.10% for a total of 824 items.

In the reading of the results, a precise classification by subject was assigned (Table 1, Figure 2).

During the analysis, a strong noisiness of the results was noted. In particular, 36.41% of the cases analyses patient satisfaction, the quality of the treatments and services provided by the health system, as well as reflections on the quality of the healthcare brand. Moreover, 19.17% of results are involved in analysing new IT systems to increase the quality of healthcare and the use of algorithms for managing medical records and medical needs; in 8.50% of cases, however, the articles that were presented by the database were not relevant to any of the keywords chosen for the systematic review.

Afterward, 7.89% of the excluded results refer to the quality of jobs and health leadership, 7.65% is inherent to scientific studies on the environment and on the impact concerning the quality of life of magnetic waves, of PM 2.5 and 10.

Furthermore, 6.07% examine the quality of food in the health context, 5.70% deal with analysing the quality of public services related to health issues. 2.67% take into consideration tourism, and health mobility and its correlation with the quality of care, 2.43% of the excluded analyse Big Data, 1.33% take into consideration the quality of the care service offered by the insurance companies, for example within the "Long Term Care" policies.

Finally, 1.21% analyse the impact of privacy on the quality of medical care, 0.61% refers to the supply chain process within the quality of the drug procurement process, and 0.36% of there are also cases that examine the quality of waste recovery.

Table 1. Results excluded by “Scopus” and “Tutto”

Results	Percentage
Health care quality	300 (36,41%)
Health care technology	158 (19,17%)
Other	70 (8,50%)
Work-relationship	65 (7,89%)
Environment	63 (7,65%)
Food	50 (6,07%)
Public service	47 (5,70%)
Health tourism	22 (2,67%)
Big data	20 (2,43%)
Insurance	11 (1,33%)
Privacy	10 (1,21%)
Supply chain	5 (0,61%)
Energy	3 (0,36%)

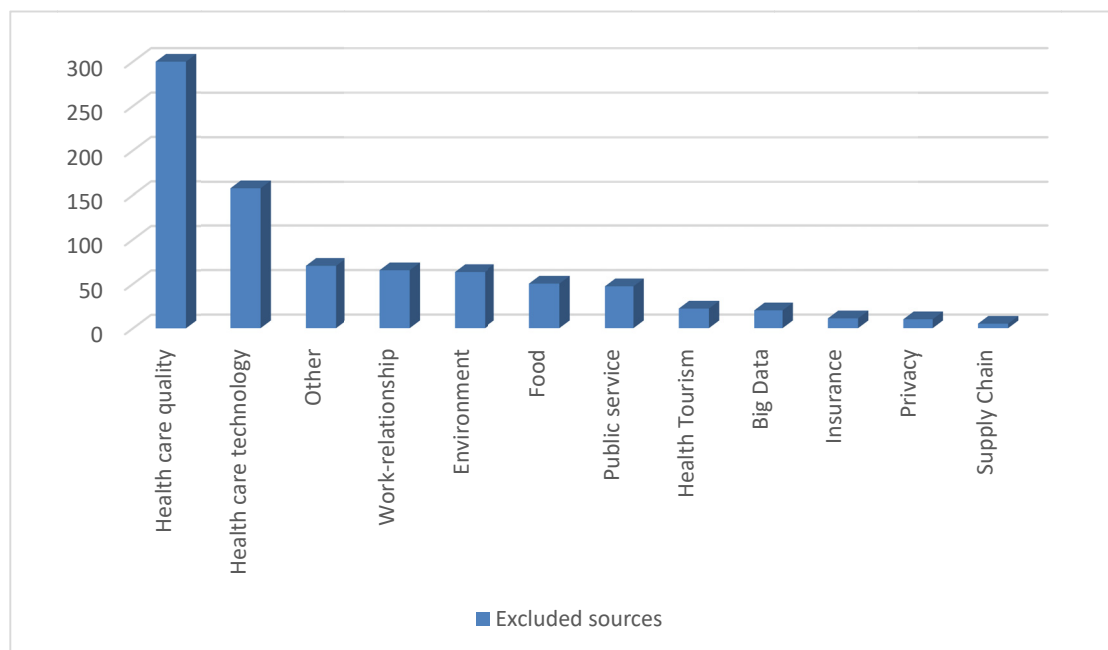


Figure 2. Excluded results

At the end of the complete reading of the results seven sources were rejected, in particular, after careful analysis, the research team showed that:

- A result deals with the data quality process within the UK social housing;
- One result referred to Big Data;
- One result gave an opinion exclusively on the legal impact of data in the health sector;
- Two results analysed the issue of data quality about the company, in particular in the SME's;
- Two results were related to reviews on data quality issues not directly related to health issues.

### 3. Results

Our research obtained 840 results. After the exclusion of foreign sources, 16 articles were accepted for the full reading. Seven sources were rejected, and therefore, nine results became part of the systematic review (Figure 3).

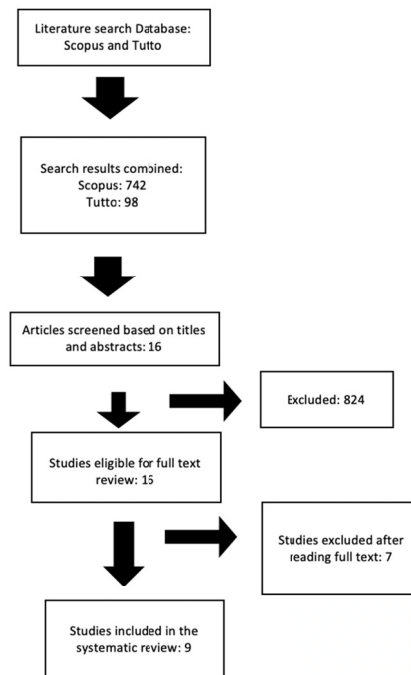


Figure 3. Flow chart of the included and excluded studies: Information flow chart of the different steps of the systematic review

Of the results obtained, 6 present a purely descriptive and qualitative treatment based on data quality processing methods (Bai et al., 2018; Ledikwe et al., 2014; Skyttberg et al., 2016; Verma, 2014; Wills, 2014; Zozus et al., 2014), 2 present a quantitative statistical approach (Mitsunaga et al., 2015; Watson et al., 2017), and finally, in one case both methods exist (Neda Firouraghi et al., 2018).

From a temporal point of view (Figure 4) there is a more marked interest for the topic in 2014 with 4 results (Ledikwe et al., 2014; Verma, 2014; Wills, 2014; Zozus et al., 2014), 2 in 2015 (Mitsunaga et al., 2015; Firouraghi et al., 2018) and only one result per year up to 2018 (Bai et al., 2018; Skyttberg et al., 2016; Watson et al., 2017).

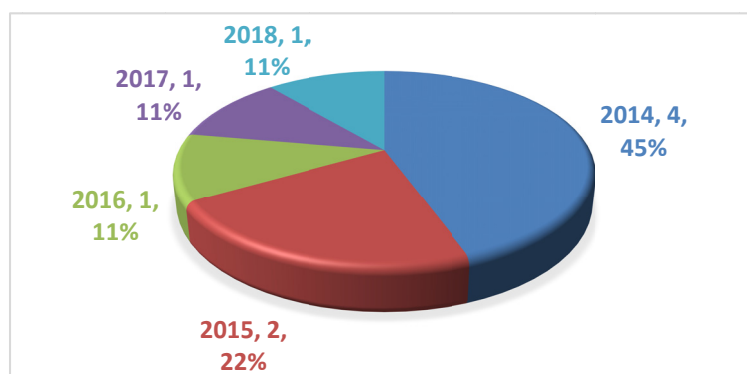


Figure 4. Results per year

The geographical area (Figure 5) more proactive by number of results appears to be the US (Watson et al., 2017; Wills, 2014; Zozus et al., 2014), to follow also Africa with related examples in Botswana (Ledikwe et al., 2014) and Rwanda (Mitsunaga et al., 2015), Sweden (Skyttberg et al., 2016), Australia (Bai et al., 2018), UK (Verma, 2014) and Iran (Firouraghi et al., 2018).

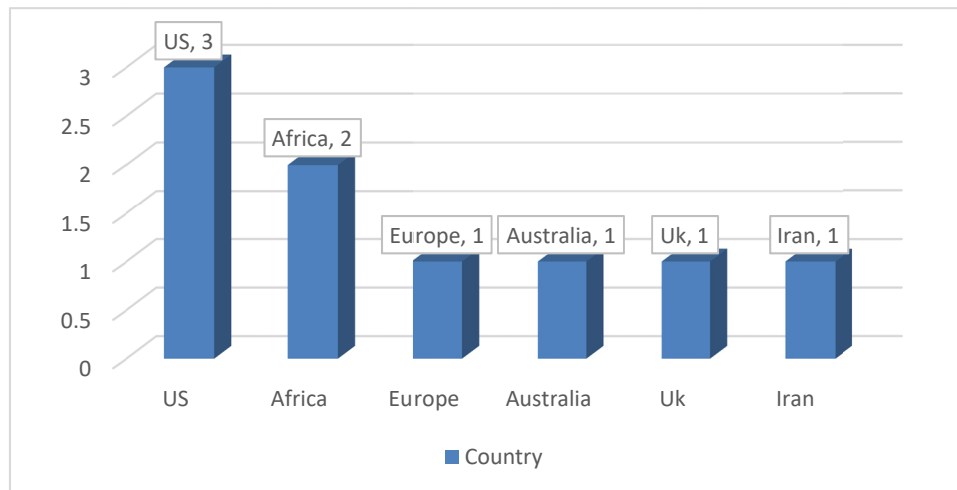


Figure 5. Results per country

### 3.1 Qualitative Approach

The discussion of Bai et al. (2018) presents a specific architecture and a data quality management system based on a 5-step process:

1. the decision of criteria, definitions, and techniques used to measure quality;
2. a database that can memorize and map all available data;
3. the implementation of a virtual data management interface and with governance tasks related to the individual operator;
4. adequate technical visibility;
5. measures and methods can convey strategic choices.

In a first phase the authors applied these results to a specific financial business system, but later this was also extended to the health sector. In this case, the main problems that emerged concern the wide variety of information systems and catalogued information, as well as the need to create systems with an urgent response, an element that characterizes healthcare activity.

The case analysed by Ledikwe et al. (2014) takes into consideration the monitoring and evaluation of data in Botswana. The analysis carried out is first traced back to the analysis of the structure and functions of the internal information systems represented by the health databases. The monitoring took place through interviews with the operators. The process that has been followed is based on the guidelines presented by Hardee K. (2008), and is primarily based on defining the structures, functions, and capacity for monitoring and evaluating data. Secondly, it includes the definition of indicators and guidelines on which to base the analysis, defines the terms of the data collection through cataloguing forms, as well as the management of the data quality process in its entirety; and finally, compatibility with the leading national benchmark comparison system. The study shows that within the analysed geographical area there are several problems concerning sensitivity to data quality. In this context, according to the authors, initiatives aimed at increasing the level of education aimed at data quality are necessary, human resources adequately trained on each level of health both at the level of the structure and the district. What is hoped for is the creation of a single complete source of information for the entire health sector analysed?

The discussion of Skyttberg et al. (2016) focuses on data quality within 9 Swedish emergency departments. The reality was indeed and presents an infrastructure to use the data exchanged both on digital information and on paper. The research areas of the researchers are complete regarding the correctness of the information, completeness, and circulation within the health areas. The methodology used provided semi-structured interviews, observations on the work of medical personnel and documentation analysis. The opportunity to deepen the discussion through the quality examination of the elements, in particular, calls for the need for process standardization, improves digital support for documentation, provides support for the personal workflow, ensures the interoperability of information within the structure, create a quality and performance system that can

improve staff training.

According to Verma (2014), a two-time theoretical health framework is needed. Firstly, on the selection and selection of a particular area of interest on which to focus, on the definition of criteria and standards of measurement, on the management and programming of a method. In a second step, the collection and analysis of the data, the implementation of the changes and their subsequent evaluation of the effects are foreseen. The application and experience are referred to the department of anaesthesia at the Derby City General Hospital.

The qualitative approach provided by Wills M. (2014) is based on three main elements: the use of "small" data, predictive models and real-time analysis. The first element considers that small data are often more easily translated into concrete actions for the benefit of patients. Through predictive models, it is possible to focus on future forecasts and assign a level of risk to each patient in order to provide prompt assistance. Finally, clinical information in real time can benefit concerning data availability speed and even more evidence regarding past clinical history.

The case analysed by Zozus et al. (2014) intends to provide a theoretical framework on the meanings of data quality. The main definitions, as reported, concern completeness, accuracy, and consistency. For completeness is meant the presence of a particular data, this concept can coincide with the presence of particular information requested in a database, calculated for example through a percentage. Accuracy is a property of the data and corresponds to the values assigned arbitrarily by those who manage the data (ISO, 2015). Consistency is understood as an element of uniformity for possible comparable scenarios, for example, coherence between health districts or clinical records.

Qualitative cases only partially respond to the data quality measurements provided by the World Health Organization. For completeness, it is useful to refer to Ledikwe et al. (2014) which shows how the characteristic is directly related to the presence of health guidelines outlined. For Skyttberg et al. (2016) completeness is instead associated with the fullness of the compilation fields of electronic health records (EHR's). Finally, for Zozus, et al. (2014) can be associated with four main elements: 1. completeness of the synthesis elements (for example the columns of a database), 2. the values inserted inside, 3. completeness of the values row and 4. the completeness understood as an opportunity to extract data by column and row by starting procedures to create operating percentages useful for healthcare.

Data timeliness is defined by Ledikwe et al. (2014) as a level of timeliness in the loading of data by health professionals, according to Verma, (2014) this meaning depends on the type of data and use.

Internal and external data coherence is considered in the case of Ledikwe et al. (2014) as the data management system in various health districts in Botswana (and therefore internally) is compared but also using an external "Monitoring and Evaluation" (M & E) system at national level. Also in the study by Skyttberg et al. (2016), it is possible to find elements of coherence, in this case, the consistency is compared between the databases and the paper documentation available to health personnel, the external consistency comes from the examination of 9 emergency departments in Sweden.

### *3.2 Quantitative Approach*

Among the appropriately quantitative approaches, it is useful to analyse the treatment of Mitsunaga et al. (2015), which provides a focus on the accuracy of data in household registries by health workers in Rwanda. The study envisaged the possibility of creating a data quality assessment system based on indicators related to demography and women's health. For each category a register was created, in this sense, the assessment of the quality or non-quality was done through the direct observation of researchers and interviews. Each category then had an accuracy rating ranging from 0 to 5, and precise estimates were given based on the 95% confidence interval. The qualifying element for the research team, in this case, is the division of the data quality process by patient type and the statistical approach used.

The case treated by Watson et al. (2017) was born within a health program aimed at children suffering from pneumonia. The first data quality action was the creation of a centralized database that eliminates the logistics problem of transporting information between different workplaces. In particular, the sources included concern the results of diagnostic tests such as radiographs and digital listening files (for example thorax).

The data and information processed to correspond to qualitative parameters for the diagnosis of pneumonia. The analysis, in this sense, has had several objectives:

1. monitoring of input and data quality in databases almost in real time;

2. monitoring of clinical trial operations in real time (e.g., through sample volume and time from sample collection to laboratory reception);
3. the insertion of data at several collection points in each site (e.g., Laboratory, clinic, first aid);
4. rapid implementation of any changes to the information system required by the clinical staff.

The validation of the system has provided for a training activity aimed at homogenizing data insertion in the database and monthly monitoring systems on the lack of data.

The quantitative cases also examined partially meet the criteria outlined by the World Health Organization. Completeness and timeliness is analysed by Mitsunaga et al. (2015) as an essential element during the home visits of health personnel and the compilation of digital folders, in the case of Watson et al. (2017) refers exclusively to the completeness referred to demographic data, clinical vaccinations, environmental elements, and risk factors.

Both the aforementioned cases also respond to the principle of internal coherence, in the case of Mitsunaga et al. (2015) this refers to the consistency between the information in the registers and the interviews made by the medical staff, finally, for Watson et al. (2017) represents the percentage of data present and corrected in the 9000 clinical cases analyzed.

### 3.3 Quantitative & Qualitative Approach

The approach described by Firouraghi et al. (2018) provides for the analysis of data quality within the haemodialysis database. The database analysed contained 2367 patients with 72 different types of variables. The activity carried out envisaged the removal of redundant data, the detection, and removal of exceptional values, as well as the management measures for missing values. The statistical methods used envisaged the use of variance and standard deviation measurements, quartiles and frequency histograms to detect anomalous values. The methodology includes the analysis of data with the verification of missing information or anomalies as completeness and timeliness elements required by the World Health Organisation (2017a). The variables taken into consideration show a lack of data between 0 and 19.73%. The primary objective in this case too is represented by the implementation of strategies able to limit errors in the management of patient data; by identifying the increase in the quality of data entry, it is indeed possible to reduce the risk of an adverse event, thus increasing efficiency.

Table 2. Studies characteristics for data quality health in the world

Author	Type of study	Features	Measures	Application	Data collection	Country	Year
Bai et al.	Qualitative	Creation of a data quality management system architecture.	Accuracy, Completeness, Accessibility, Consistency, Non-redundancy, Redability, Usefulness, Trust.	The first application was financial, the decision is to extend the functionality of the application to the health sector.	Through the upload of information in an innovative tool from health business managers.	Australia	2018
Ledikwe J. H.	Qualitative	Analysis of the structure and functions for the databases. Creation of indicators and drafting of guidelines. Data collection and process management. Classified level of data accuracy as "good" or "poor" with a statistical method and create a list of data useful for the data quality process.	Accuracy, Reliability, Precision, Completeness, Timeliness, Integrity and Confidentiality (Hardee K., 2008).	Monitoring and evaluation (M&E) clinical data in Botswana.	Interviews	Botswana	2014
Mitsunaga et al.	Quantitative		Accuracy	Household registrers by community health workers.	Interviews, observations and databases	Rwanda	2015



Firouraghi et al.	Qualitative and quantitative	Analyse the different sources and then: removing the duplicated data, detecting outliers and correcting or removing them, and handling the missing values.	Accuracy, Correctness, completeness.	Hemodialysis Database	Electronic databases in which to catalog missing information or outliers.	Iran	2015
-------------------	------------------------------	--	--------------------------------------	-----------------------	---	------	------

Table 2. Continued

Author	Type of study	Features	Measures	Application	Data collection	Country	Year
Skyttberg et al.	Qualitative	Data quality improvement: standardize the care process, improving digital documentation support, provide workflow support, ensure interoperability, perform quality control.	Correctness, Completeness, Currency.	Emergency departments	Interviews, observations, analysis of documentation templates.	Sweden	2016
Verma R.	Qualitative	In a data quality process should include this step: Select or re-select topic, define criteria and Set standards, Plan Methodology, Collect and analyse data, Evaluate effects of change.	Accuracy, Accessibility, Comprehensiveness, Consistency, Currency, Definition, Granularity, Precision, Relevancy, Timeliness.	Department of anaesthesia	Electronic databases but at the same time analyze the process in the case of data collection on paper.	Uk	2014
Watson et al.	Quantitative	Creation of a central database (EDC), transfer and evaluating processes such as: Chest Radiographic Images and Digital chest Auscultation Sound files.	Quality indicators particular for Pneumonia	Pneumonia Etiology Research for Child Health (PERCH).	Through Electronic Data Capture system (EDC).	US	2017
Wills M. J.	Qualitative	Small data - predictive modeling expansion - real-time analytics.	Integration of small-scale registry - using algorithms to identify patients at highest risk.	Hospital units using data in real time and in streaming.	Through the technological interface best known by healthcare personnel.	US	2014
Zozus et al.	Qualitative	Theoretical background about data quality dimensions	Completeness, accuracy, consistency	Health sector	-	US	2014

The table summarizes the main characteristics of the studies regarding the author, the year of publication, the type of study, the main features included, the main measures, the applications, the method of data collection and the country.

Table 3. Selection of qualitative and quantitative results for consistency with the World Health Organization Framework

Author	Type of study	Completeness and timeliness of data	Internal coherence of data	External coherence of data	Presence of external comparisons
Bai et al.	Qualitative	Presence of the completeness of the data.	Na	Na	Na
Ledikwe J. H.	Qualitative	The research highlights how there are no internal guidelines able to delineate the concept of completeness of data. Timeliness is related to the number of free operators able to load data on databases.	Yes, analysis between data management systems between districts.	Definition of national guidelines, in particular "Monitoring and Evaluation (M&E) data process.	Na
Mitsunaga et al.	Quantitative	The research highlighted data completeness problems during home visits by health professionals. Furthermore, issues concerning the timeliness of data are emphasized.	According to the definition of WHO, this can be seen as internal consistency between the information in the registers and the interviews as well as the civil records.	Na	Na
Firouraghi et al.	Qualitative and quantitative	The study highlighted problems in the timeliness and completeness of the data reported in the Hemodialysis Database.	Na	Na	Na
Skyttberg et al.	Qualitative	Completeness is associated with the fullness of the electronic health records (EHR's); among the measures suggested for the increase, there is the provision of standardized documentation and able to provide all the necessary information. This measure is referred to as particularly tricky in the context of the emergency department.	The internal coherence is indirectly mentioned when the differences between the electronic database and paper documentation are highlighted.	The study examines nine emergency departments of Sweden.	Na
Verma R.	Qualitative	Timeliness is directly related to the type of use of the data.	Na	Na	Na
Watson et al.	Quantitative	The completeness of the data refers to the entirety of the demographic data, vaccination clinics, environmental and risk factors.	Internal consistency is represented by the% of data present and corrected in the cases 9000 cases and clinical controls analyzed.	Na	Na
Wills M. J.	Qualitative	Na	Na	Na	Na

#### 4. Findings

There are many applications related to the data quality process, in general, in all the research included in the systematic review there is the affirmation that the process of analysis and evaluation of data quality represents an essential point. This aspect is complementary to the clinical care that is delivered to patients. The identification of errors and incorrect practices allows to improve the quality of the databases used and, indirectly, avoids errors

or improper practices (Firouraghi et al., 2018).

One of the elements common to all research is the need to set up data platforms; cataloging in databases is, therefore, the first preparatory step at the beginning of the data quality process.

In the cases reported by Ledikwe et al. (2014), Mitsunaga et al. (2015) and Skyttberg et al. (2016) the construction of a digitized space able to host qualitative and quantitative data could be done through the assessment of working methods through the use of interviews and observations aimed at the activity of health workers.

In other cases, this could be done directly on already operational databases to which it was possible to apply specific indicators that take into consideration the accuracy and quality of the content (Firouraghi et al., 2018).

The keys to reading and commenting are made, in most cases, through the selection of variables related to the clinical data that are available to researchers. Some examples refer to weight, age, body mass, the level of systolic pressure or values of blood test results which instead provides a model of data quality divided by type of patient, for example, children under the age of 5 or women between 15-49 years.

From compatibility with the requirements of the World Health Organisation (2017a), it is outlined how both qualitative and quantitative cases pay much attention to the description of completeness and, in part, also to the timeliness of data. External coherence appears to be present exclusively in qualitative studies (Ledikwe et al., 2014; Skyttberg et al., 2016). Finally, the presence of external comparable was not found in any study.

## 5. Conclusion

The systematic review focuses on the issue of data quality within the health sector. The use of performance data cataloging systems places the need to reduce the risks and errors that are of particular importance in the healthcare sector.

Nowadays the tools that technology reserves allow us a greater simplification and greater control of the quality.

The analysis conducted focuses on some relevant elements:

- the construction or use of databases functional to the analysis of data quality;
- the presence of both qualitative methods (through interviews and cataloging of data), and quantitative (through the use of functional statistics to verify the quality of data);
- the presence in all the results of processes that include, in addition to the time of cataloging the data, also the verification of indicators and the consequent modification of how health professionals carry out data retention operations;
- the completeness of the data considered important and analysed in a transversal way in the results;
- the absence of cases of comparison but rather the exclusive analysis of their cases as suggested by the World Health Organisation (2017a).

Regarding the first point, as it was possible to analyse, the practical cases reported consider the construction of the virtual working space through databases, as the first functional step for the start of the data quality process and the collection of health data (Bai et al., 2018; Mitsunaga et al., 2015; Firouraghi et al., 2018; Watson et al., 2017; Wills, 2014).

At this point in the discussion, the approach to data quality appears twofold: on the one hand results of process and method, among them interviews with health personnel as in the case of Ledikwe et al. (2014) and on the other statistical analysis on medical proxies as in the case of Watson et al. (2017).

What appears clear and standard is that the start of data evaluation activities presupposes the meticulous selection of the analysis area. Subsequently it is necessary to carry out the cataloguing of the data, almost as if it were a photograph of the operational methods of data management; the calculation of indicators such as completeness, internal and external consistency, and finally, the adoption of corrective actions coordinated, for example, by highly standardized modules and preventive training courses aimed at health professionals (Ledikwe et al., 2014)

Among the most widely used measures, completeness emerges, both in qualitative and quantitative results only; according to what emerged is a standard of analysis from which to start the analysis. Even in the light of what has been learned, we can state that the data quality process should not stop at a single indicator. In this sense, the completeness element represents a quantitative value referring to the fullness of data within a database; the overall judgment on data quality should, however, be extended, combining observations, internal points of view

and external comparisons.

## References

- Bai, L., Meredith, R., & Burstein, F. (2018). A data quality framework, method and tools for managing data quality in a health care setting: an action case study. *Journal of Decision Systems*, 27, 144-154. <https://doi.org/10.1080/12460125.2018.1460161>
- Batini, C., & Scannapieco, M. (2006). *Data Quality: Concepts, Methodologies and Techniques*. Springer-Verlag New York, Berlin. Retrieved from <https://www.springer.com/us/book/9783540331728>
- Biancone, P. P., Secinaro, S., & Brescia, V. (2018). The innovation of local public-sector companies: Processing big data for transparency and accountability. *African Journal of Business Management*, 12(15), 486-500. <https://doi.org/10.5897/ajbm2018.8596>
- Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A Review of Data Quality Assessment Methods for Public Health Information Systems. *International Journal of Environmental Research and Public Health*, 11, 5170-5207. <https://doi.org/10.3390/ijerph110505170>
- Davenport, T. H. (1998). *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press. Retrieved from [https://www.researchgate.net/profile/Thomas\\_Davenport2/publication/229099904\\_Working\\_Knowledge\\_How\\_Organizations\\_Manage\\_What\\_They\\_Know/links/53db939e0cf216e4210bf841/Working-Knowledge-How-Organizations-Manage-What-They-Know.pdf?origin=publication\\_detail](https://www.researchgate.net/profile/Thomas_Davenport2/publication/229099904_Working_Knowledge_How_Organizations_Manage_What_They_Know/links/53db939e0cf216e4210bf841/Working-Knowledge-How-Organizations-Manage-What-They-Know.pdf?origin=publication_detail)
- DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: a literature review. *Quality & Quantity*, 52, 1523-1559. <https://doi.org/10.1007/s11135-017-0533-4>
- Duvier, C., Neagu, D., Oltean-Dumbrava, C., Dickens, D. (2018). Data quality challenges in the UK social housing sector. *International Journal of Information Management* 38, 196-200. <https://doi.org/10.1016/j.ijinfomgt.2017.09.008>
- Firouraghi, N., Jahromi, S. E., Ashkan, S., Morvaridi, M. R., & Sharifian, R. (2018). Data Quality Assessment and Recommendations to Improve the Quality of Hemodialysis Database. *Journal of Health Management & Informatics*, 5, 21-26. Retrieved from <http://jhmi.sums.ac.ir/index.php/JHMI/article/view/441>
- Frost & Sullivan (2012). *Reducing Information Technology Complexities and Costs for Healthcare Organizations*. Retrieved from <https://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf>
- Hardee, K. (2008). *Data Quality Audit Tool: Guidelines for implementation*. Retrieved from <https://www.measureevaluation.org/resources/publications/ms-08-29>
- Hoeren, T. (2017). Big data and the legal framework for data quality. *International Journal of Law and Information Technology*, 25, 26-37. <https://doi.org/10.1093/ijlit/eaw014>
- ISO (2015). Data quality-Part 8: Information and data quality: Concepts and measuring. Available at: <http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/08/60805.html>
- Ledikwe, J. H., Grignon, J., Lebelonyane, R., Ludick, S., Matshediso, E., Sento, B. W., Sharma, A., Semo, B. W. (2014). Improving the quality of health information: a qualitative assessment of data management and reporting systems in Botswana. *Health research Policy and Systems*, 12(7). <https://doi.org/10.1186/1478-4505-12-7>
- Mitsunaga, T., Hedt-Gauthier, B., Ngizwenayo, E., Farmer, D., Gaju, E., Drobac, P., Basinga, P., Hirschhorn, L., Rich, M., Winch, P., Ngabo, F., & Mugeni, C. (2015). Data for Program Management: An Accuracy Assessment of Data Collected in Household Registers by Community Health Workers in Southern Kayonza, Rwanda. *Journal of Community Health*, 40, 625-632. Available at: <https://doi.org/10.1007/s10900-014-9977-9>
- O'Brien, T. (2015). Accounting' for Data Quality in Enterprise Systems. *Procedia Computer Science*, 64, 442-449. <https://doi.org/10.1016/j.procs.2015.08.539>
- Skyttberg, N., Vicente, J., Chen, R., Blomqvist, H., & Koch, S. (2016). How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments. *Bmc Medical Informatics and Decision Making*, 16, 61. <https://doi.org/10.1186/s12911-016-0305-4>

- Verma, R. (2014). Data quality and clinical audit. *Anaesthesia & Intensive Care Medicine*, 15, 392-394. <https://doi.org/10.1016/j.mpaic.2014.04.018>
- Wang, X., Williams, C., Liu, Z. H., & Croghan, J. (2017). Big data management challenges in health research-a literature review. *Briefings in bioinformatics*. <https://doi.org/10.1093/bib/bbx086>
- Watson, N. L., Prospero, C., Driscoll, A. J., Higdon, M. M., Park, D. E., Sanza, M., Deluca, A. N., Awori, J. O., Goswami, D., Hammond, E., Hossain, L., Johnson, C., Kamau, A., Kuwanda, L., Moore, D. P., Neyzari, O., Onwuchekwa, U., Parker, D., Sapchokul, P., Seidenberg, P., Shamsul, A., Siazeele, K., Srisaengchai, P., Sylla, M., Levine, O. S., Murdoch, D. R., O'Brien, K. L., Wolff, M., & Deloria Knoll, M. (2017). Data Management and Data Quality in PERCH, a Large International Case-Control Study of Severe Childhood Pneumonia. *Clinical infectious disease: an official publication of the Infectious Diseases Society of America*, 64, S238-S244. <https://doi.org/10.1093/cid/cix080>
- Wills, M. J. (2014). Decisions through data: Analytics in healthcare. *Journal of Healthcare Management*, 59, 254-262. Retrieved from [https://ache.org/Faculty\\_Students/Studentessay\\_Undergrad\\_Wills.pdf](https://ache.org/Faculty_Students/Studentessay_Undergrad_Wills.pdf)
- World Health Organisation. (2017a). *Data quality review: a toolkit for facility data quality assessment. Module 1. Framework and metrics*. Retrieved from [http://www.who.int/healthinfo/tools\\_data\\_analysis/dqr\\_modules/en/](http://www.who.int/healthinfo/tools_data_analysis/dqr_modules/en/)
- World Health Organisation. (2017b). *Data quality review: a toolkit for facility data quality assessment. Module 2. Desk Review of Data Quality*, 56. Retrieved from [http://www.who.int/healthinfo/tools\\_data\\_analysis/dqr\\_modules/en/](http://www.who.int/healthinfo/tools_data_analysis/dqr_modules/en/)
- World Health Organisation. (2017c). *Data quality review: a toolkit for facility data quality assessment. Module 3. Data Verification and System Assessment*, 56. Retrieved from [http://www.who.int/healthinfo/tools\\_data\\_analysis/dqr\\_modules/en/](http://www.who.int/healthinfo/tools_data_analysis/dqr_modules/en/)
- Zozus, M. N., Hammond, W. E., Green, B. B., Kahn, M. G., Richesson, R. L., Rusincovitch, S. A., Simon, G. E., & Smerek, M. M. (2014). Assessing Data Quality for Healthcare Systems Data Used in Clinical Research. Retrieved from [https://www.nihcollaboratory.org/Products/Assessing-data-quality\\_V1%200.pdf](https://www.nihcollaboratory.org/Products/Assessing-data-quality_V1%200.pdf)

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).