THE INTERNATIONAL JOURNAL OF BUSINESS & MANAGEMENT

Big Data Quality in Healthcare Organizations in Saudi Arabia

Shaymaa Mousa
Associate Professor, Department of Economics and Administration,
King Abdul-Aziz University, Saudi Arabia
Rabab AlHarbi
MBA student, Department of Economics and Administration,
King Abdul-Aziz University, Saudi Arabia

Abstract:
The adoption of Big Data analytics by organizations has been expanding across numerous industries, bringing both advantages and challenges. The primary challenge stems from the insufficient quality of data used in the analysis process, as well as its impact on the value of the acquired insights. Therefore, the objective of this research is to investigate the factors that influence the quality of Big Data in Saudi Arabian healthcare organizations. These factors include clear business objectives, data source, data analysis, data quality dimensions, customer engagement, and assessment tools. To investigate these aspects, a descriptive and analytical design is used. Additionally, a questionnaire is distributed in both Arabic and English to 350 respondents working within Saudi Arabia's healthcare industry from a variety of occupations and management positions. To analyse the data acquired via the questionnaire, the statistical package for the social sciences (SPSS) is employed. The study's results indicate that the factors analysed have a significant effect on the quality of Big Data in healthcare organizations. The findings of this study will fill the gap in the existing literature by addressing factors that have not been explored in combination previously.

Keywords: *Big data analytics, big data quality, quality dimensions, customer engagement, assessment tools*

1. Introduction

The term 'Big Data' first appeared in the early 1990s; later, Mashey and Laney clarified the definition of the term at the beginning of the twenty-first century. Big Data is defined as complex datasets that are too large for traditional data management systems to store, manage, and process in a timely, cost-effective manner (Nambiar, Bhardwaj, Sethi, &Vargheese, 2013). Big Data has general characteristics, which are referred to as the 5Vs of Big Data, namely volume, variety, velocity, value, and veracity (Rehman, Naz, &Razzak, 2021). The volume of data generated by internet users continues to grow. According to a Deloitte (2017) report, the total amount of digital data in existence is estimated to increase tenfold by 2025. This rapid growth of Big Data comes with rising concerns to the level of Big Data quality that might limit or enhance its use. Data quality may be defined as the fit of data to be used by data consumer (Côrte-Real, Ruivo, & Oliveira, 2020). The quality of Big Data may significantly impact firm performance, efficiency, and decision making. Despite the widespread adoption of Big Data analytics across businesses, only a small number have achieved the desired outcomes on competitive performance. It is important to acknowledge the healthcare industry is both massive and critical but lacks efficiency and the proper use of business analytics in many aspects. Over the next few years, Big Data analytics and technologies are anticipated to revolutionize the way in which humanity employs medicine practices. There are several theoretical different methods for measuring the degree of Big Data quality. This paper adapts a measuring and assessing system in which data quality is evaluated and measured through relevant dimensions. The traditional dimensions used to determine the quality of normal datasets are still relevant and applicable to the concept of Big Data (Juddoo, 2015). Although numerous studies and academic papers have discussed the term 'Big Data' in healthcare organizations, only a small number of studies have addressed the critical nature of data quality in the medical industry. In the context of data quality, there is a significant absence of clarity in terms of the evaluation of Big Data quality in general (Cai& Zhu, 2015). Therefore, the importance of this research may be highlighted in its contribution to the general understanding of the quality of Big Data in healthcare organizations. Moreover, in terms of the geographic framework, no existing study has measured the quality of Big Data in Saudi Arabia. Therefore, this paper varies from previous studies in terms of the scope of the variables applied to examine the quality of Big Data and the location of the research, which is in Saudi Arabian healthcare organizations.

The aim of this work is to study the influence of several aspects on Big Data quality in healthcare institutions in Saudi Arabia. This research is structured as follows: Section 2 contextualizes the research topic within the literature and examines each aspect of the research construct thoroughly. Section3 describes the methodology of the research, including the research design, data collection method, questionnaire design, and analysis techniques. Section4 shows a statistical analysis of the respondents' data and to determine the validity of the hypotheses upon which the results are based. Section

5 discusses the research findings and their implications in relation to the literature. Moreover, this chapter explains the research's limitations and provides the research conclusions.

2. Literature Review

Numerous technological advancements in the field of information systems have evolved across a variety of industries. In the case of quality dimensions, these advancements enable a considerably more targeted and tailored application for each business. In the context of data, the concept of data quality is vast, and there exist several definitions and interpretations. According to the General Administration of Quality Supervision (2008), data quality is described as 'the degree to which a set of inherent characteristics fulfil the requirements' cited by Cai& Zhu (2015, p.2). Historically, data quality has been discussed scholars and industrial practitioners, who have offered specific approaches for managing data quality in the field of information systems (Mecella et al., 2002). Data quality is measured through a variety of dimensions and frequently referred to in the literature as a multi-dimensional subject. However, quality is typically studied under two main categories, namely databases and management. under the databases category, Data Quality is studied from a technical perspective (Galhardas et al., 2001). The management perspective is concerned with different aspects or dimensions of data quality, such as accessibility, credibility, relevancy, interpretability, and judgment (Pipino, Lee, & Wang, 2002). Additionally, Wand and Wang (1996, p.2) state that, 'the choice of these dimensions is based on intuitive understanding, industrial experience, empirical studies, or literature review. However, the literature shows that there is no general agreement on a set of data quality dimensions and their exact definitions exists.'.Moreover, in terms of the quality of big data, the characteristics generally associated with assessing data quality in the domain of conventional datasets are applicable in the context of Big Data. Moreover, while determining the components that are most critical for maintaining the quality of Big Data, the application of Big Data must be considered as it has a considerable influence on the process (Juddoo, 2015).Cai and Zhu (2015) highlight data quality as a challenge while using Big Data analytics in their work. The authors address concerns about the lack of data quality evaluation and assessment tools in the current analytics environment, which may result in the poor use of Big Data insights. Additionally, these authors discuss challenges associated with big data quality, which can be summarized as follows: enormous volume of data makes determining its accuracy within an appropriate timeframe difficult; the heterogeneity of data sources results in an abundance of data forms and complexity; there is complex data implantation; rapid change in data necessitates the use of more sophisticated technologies; and there is a lack of standardized criteria for data quality. The study developed a framework for assessing Big Data quality in three parts, which would identify indicators, aspects of data quality, and specific data quality dimensions. The authors argue that the dimensions of data quality vary depending on the source of Big Data and business requirements and developed a model that standardized the process of evaluating data quality in terms of five critical dimensions: availability; reliability; usability; presentation quality; and relevance. Moreover, there are various factors in measuring data quality under each category. In short, the models presented in this study offer a baseline for evaluating Big Data outputs, creating a solid foundation for future research on various evaluation models.Sidi et al. (2012) argue that there is a widespread belief that data quality is exclusively influenced by accuracy, with little attention for other dimensions, regardless of their usefulness in ensuring improved quality. Measuring data quality across several dimensions enables deeper insights and applicability in a variety of disciplines and industries. To summarize, without recognizing existing relationships among data quality dimensions, driving useful insights cannot be comprehensive. Quality dimensions not only are related to one another, but also play an important role in overall data quality. Côrte-Real et al. (2020) conducted a study that provided insights into the theoretical character of most Big Data quality literature. The study employed a qualitative method to assess firm performance by interviewing one respondent from each organization across Europe and America. As a result, the authors developed a model that investigates the impact of Big Data quality characteristics on IoT and Big Data analytics, which was correlated with firm performance. The model investigated data quality as impacted by four major quality components: completeness; accuracy; currency; and format. This model also includes a review of Big Data analytics and IoT capabilities, as well as their impact on developing competitive advantages for the organization in terms of both financial and strategic performance. This work produced comprehensive models that included many factors as a practical and theoretical framework, combining knowledge-based perspectives with dynamic capabilities theories. In addition, the findings highlight the significance of data accuracy as a measure of quality in the field of Big Data analytics, whereby the quality of data affects companies' performance in a direct way differently from Big Data analytics and IoT capabilities. In short, the results indicate that, when data quality is examined, the influence of Big Data capabilities is seen to be the most efficient.

Another study conducted by Wamba et al. (2018) in France included a survey with 150 respondents from IT and business analysis personnel and managers. The study built a framework based on the resource-based view (RBV) and IT quality theories. The paper studied the effects of four major factors of Big Data analytics quality, namely technology, information, and talent qualities, making strategic alignment a moderator in relation to firm performance. Moreover, the IT quality theory addressed four factors related to information quality itself, namely accuracy, completeness, currency, and format. The study concluded that a strong correlation between Big Data quality and firm performance took into consideration strategic alignment as a component. However, the results of the study highlight the fundamental functions of quality of technology, information, and talents in enhancing Big Data analytics quality (BDAQ) in a BDA ecosystem. The model of this paper presented a holistic examination of several factors that influence the quality of Big Data analytics and their effect on firm performance. Moreover, this model may serve as the base for future work to evaluate the implementation of big data analytics. In conclusion, the study delivers a list of criteria that may be utilized as a management manual and monitoring system, though it may be useful to also look at additional factors that may have an impact on the BDAQ framework. Completeness, accuracy, format, and currency are the four quality dimensions that have

been widely used to define big data quality in the previous two studies (Côrte-Real et al., 2020; Wamba et al., 2018). However, there are no commonly agreed-upon definitions for these dimensions in existing literature. Nelson et al. (2005, p.6) offers the following definitions: accuracy is 'the degree to which information is correct. unambiguous, meaningful, believable, and consistent.'; completeness is 'the degree to which all possible states; relevant to the user population are represented in the stored information.'; currency is 'the degree to which information is up-to-date, or the degree to which the information precisely reflects the current state of the world that it represents.'; and format is 'the degree to which information is presented in a manner that is understandable and interpretable to the user and thus aids in the completion of a task.'.Admittedly, though a number of studies have examined quality dimensions in the context of Dig Data around the world in various industries, such as Wamba et al. (2018) in France and Côrte-Real et al. (2020) in both Europe and America, there have been no studies in the healthcare industry worldwide that have examined these quality dimensions in the context of Big Data.

3. Research Methodology

3.1. Research Model and Hypothesis

This research aims to measure the quality of Big Data in the healthcare institutions in the Kingdom of Saudi Arabia. Based on the literature review of previous studies, authors formulate the following hypotheses for this research:

- H_1. Clear business goals have a significant impact on Big Data quality.
- H_2. Data source has a significant impact on Big Data quality.
- H_3. Data quality dimensions have a significant impact on Big Data quality.
- H_4. Data analysis has a significant impact on Big Data quality.
- H_5. Customer engagement has a significant impact on Big Data quality.
- H_6. Assessment tools have a significant impact on Big Data quality.

In this research, a descriptive and analytical design was used. This type of design is based on the use of descriptive and quantitative statistical methods. A descriptive design is one that is designed to describe the distribution of one or more variables, without regard to causal or other hypotheses (Ranganathan& Aggarwal, 2018).

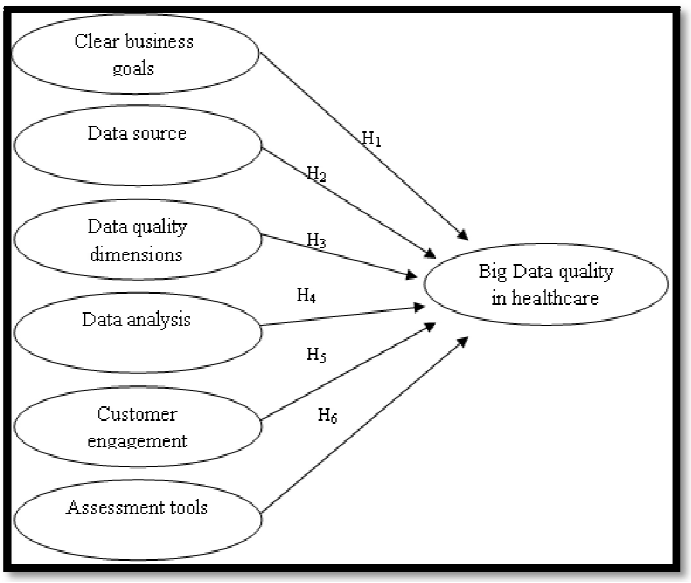

Figure1: Research Model

3.2.Instrument and Measurement

In this work, a questionnaire was the research tool employed to collect sampling data. The information and statistics used for analysis are primary data. Primary data was collected through questionnaires distributed to respondents via email and social media. According to Roopa and Rani (2012, p.56), 'a questionnaire is simply a list of mimeographed or printed questions that is completed by or for a respondent to give his opinion'. Furthermore, 'a questionnaire is the main means of collecting quantitative primary data.' (Roopa& Rani, 2012, p.56).Following a literature review in the field of Big Data in the healthcare institutions, the researcher updated a questionnaire that consisted of two parts. The first part collected the demographic information of the respondents, regarding gender, age, and occupation. The second part consisted of several statements included in six dimensions (business goals, data source, data quality, data analysis, customer engagement, and assessment tools).

The questionnaire consisted of group of dimensions and statements to check research hypotheses and answer questions. The statements were distributed into two levels of measurement scales (nominal and ordinal). The 'nominal scale organizes data into mutually exclusive (nonoverlapping), exhausted groups where no order or rating of the data can be imposed, or one can say, 'there is no meaningful order or ranking' (Bluman, 2019, p 8). Demographical information (gender, age, and occupation) was included under the nominal scale. On the other hand, 'the ordinal scale: 'classifies data into categories that can be ranked; however, precise differences between the ranks do not exist' (Bluman, 2019, p 8). In

this research, all statements for the six dimensions were included under ordinal scale. To verify respondents' opinion about this research, a five-item Likert scale was used.

3.3. Sampling and Population

In this research, the target population covers all healthcare institution employees in Kingdom of Saudi Arabia. The sample size of this research was determined using the Steven Thompson formula (Krejcie& Morgan, 1970):

$$n = \frac{N * p(1-p)}{[(N-1) * (\frac{d^2}{Z^2}) + p(1-p)]}$$

where, N: Population size = 467,650, p: probability value = 0.5, d: error = 0.05, Z, standard normal value =1.96. Using the Thompson formula, the sample size of 350 healthcare institutions employees in Kingdom of Saudi Arabia were chosen using the available method.

4. Results

4.1. Pilot Study

In this research, a pilot study was constructed from outside of the main sample of the research and included 30 healthcare institution workers in the Kingdom of Saudi Arabia. The pilot study included validity and reliability tests, the findings of which are presented in the following section. However, based on the results of the validity and reliability tests, no adjustments to the questionnaire were needed.

4.2. Validity and Reliability Measurement

To check the construct validity of the current research, the researcher computed the Pearson correlation coefficients between statements and it dimensions using pilot sample data. Results show that all Pearson correlations are statistically significant at level (α=0.01); furthermore, all statements correlated with the total score of it is the dimension. The Pearson correlation coefficients ranged between 0.645-0.968 with a high degree of validity, which means the measure established the desired measurement goals. The researcher also verified the validity of the study tool with a constructed Pearson correlation of each dimension according to the questionnaire total score. In addition, results show that all dimensions of the questionnaire have a high degree of correlation with the questionnaire total score. This correlation has a statistically significant result at the level of significance (0.01). Also, results show the range of correlation coefficients range between 0.951-0.974.

Results reveals that the overall reliability is (98%) is sufficient to guarantee the reliable internal consistency of the questionnaire. This table also shows that 'data quality' has the highest Cronbach's Alpha value of 95%, and 'customer engagement' dimension has the lowest Cronbach's Alpha value of 74.5%. These values show that the study tool was reliable.

Dimension	Mean	Standard Deviation	Rank
Customer engagement	3.86	.866	1
Assessment tools	3.78	.888	2
Clear business goals	3.75	.952	3
Data analysis	3.65	.894	4
Data quality	3.64	.888	5
Data source	3.59	.958	6

Table 1: The Dimensions in Descending Order According to Their Means

In table 1, the customer engagement dimension comes first with a mean of 3.86 and a standard deviation of 0.866, whereas data source comes at last with mean (3.59) and standard deviation (0.958).

Based on Table 2, there is a statistically positive correlation between all independent variables (clear business goals, data source, data quality, data analysis, customer engagement, and assessment tools) with the dependent variable (Big Data quality). Pearson correlation coefficients (r) ranged from 0.883 – 0.912. All of these coefficients have a positive strong correlation among independent and dependent variables.

		Independent Variables					
Dependent Variable		**Clear Business Goals**	**Data Source**	**Data Quality**	**Data Analysis**	**Customer Engagement**	**Assessment Tools**
Big Data Quality	Pearson Correlation	.883**	.902**	.912**	.898**	.869**	.887**
	Sig. (2-tailed)	0	0	0	0	0	0
	N (Sample Size)	350	350	350	350	350	350

Table 2: Pearson Correlation

***Correlation Is Significant at the 0.01 Level (2-Tailed)*

4.3. Test of Research Hypotheses

4.3.1. Therefore, Clear Business Goals Have a Significant Impact on Big Data Quality

- H_1. Clear business goals have a significant impact on Big Data quality.

Based on results, clear business goals have a mean of 3.75 ('agree') and standard deviation of 0.952. It shows that there is a positive correlation between clear business goals and Big Data quality since (r=0.883). This relationship has a statistically significant impact on Big Data quality since (*p*-value=0.000), which is less than (0.05).

4.3.2. Therefore, Data Source Has a Significant Impact on Big Data Quality

- H_2: Data source has a significant impact on Big Data quality.

Based on results, data source has mean of 3.59 ('agree') and standard deviation of 0.958). It shows that there is a meaningful positive correlation between data source and Big Data quality since (r=0.902). The correlation has a significant impact on Big Data qualitysince (*p*-value=0.000), which is less than (0.05).

4.3.3. Therefore, Data Quality Has a Significant Impact on Big Data Quality

- H_3: Data quality has a significant impact on Big Data quality.

Data quality has a mean of 3.64 ('agree') and standard deviation of 0.888. It shows that there is a meaningful positive correlation between data quality and Big Data quality since (r=0.912). This correlation has a significant impact on Big Data qualitysince (*p*-value=0.000) which less than (0.05).

4.3.4. Therefore, Data Analysis Has a Significant Impact on Big Data Quality

- H_4: Data analysis has a significant impact on Big Data quality.

Data analysis has a mean of 3.65 ('agree') and standard deviation of 0.894. It shows that there is a meaningful positive correlation between data analysis and Big Data quality since (r=0.898). This correlation has a significant impact on Big Data qualitysince(*p*-value=0.000), which is less than (0.05).

4.3.4. Therefore, Customer Engagement Has a Significant Impact onBig Data Quality

- H_5: Customer engagement has a significant impact on Big Data quality.

Customer engagement has mean of 3.65 ('agree') and standard deviation of 0.894. It shows that there is a meaningful positive correlation between data analysis and Big Data quality since (r=0.898). This correlation has a significant impact on Big Data qualitysince (*p*-value=0.000), which is less than (0.05).

4.3.5. Therefore, Assessment Tools Have a Significant Impact on Big Data Quality

- H_6. Assessment tools have a significant impact on Big Data quality.

Assessment tools have a mean of 3.78 ('agree') and standard deviation of 0.888. Results show that there is a positive correlation between assessment tools and Big Data quality since (r=0.887). This correlation has a statistically significant impact on Big Data quality since (*p*-value=0.000), which is less than (0.05).

5. Conclusion

As a result of the exponential expansion of data generated by the Internet, the IoT, and other sources inside and outside organizations across multiple industries, the significance of Big Data has emerged as a noteworthy topic that poses numerous challenges and obstacles, as well as an area of both potential and opportunity for organizations.

Initially, the objective of this research was to examine the factors that affected the quality of Big Data analytics in Saudi Arabian healthcare organizations with the commonly used approach of measuring quality across specific dimensions. These factors are the clarity of business objectives, data sources, data analysis, Big Data quality dimensions, customer engagement, and assessment tools. However, these variables have been investigated in this combination in the literature, but no existing research on Big Data and healthcare organizations operating in Saudi Arabia have been undertaken. Furthermore, the study used a questionnaire that was provided in both Arabic and English and completed by 350 respondents from the healthcare profession, representing a range of occupations and managerial positions. However, an analysis of the results reveals that all the criteria analysed have a significant effect on the quality of Big Data insights, allowing for more efficient application, which has a favourable impact on total performance and competitive advantage for a business. The findings corroborate those of studies undertaken in different industries and countries (Côrte-Real et al., 2020; Wamba et al., 2018). As a result, the research will contribute to the body of knowledge by filling in gaps in the understanding of how to successfully apply Big Data. Therefore, by addressing new dimensions and areas of Big Data quality, it will be possible to test these variables in a variety of enterprises and industries that seek to adopt a quality-oriented framework for Big Data analytics.

In conclusion, using industry-specific successful components of Big Data quality enables organizations and consumers to benefit from effective solutions as high-quality Big Data insights provide an opportunity to optimize and successfully implement this data in the current business environment.

6. References

i. Bluman, A. G. (2019). *Elementary statistics: A step by step approach: A brief version* (No. 519.5 B585E.). McGraw-Hill.
ii. Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data science journal*, *14*.
iii. Deloitte (Ed.). (2017). (rep.). *The Data Landscape A Report for Facebook - Deloitte*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-tmt-the-data-landscape.pdf
iv. General Authority for statistics. (2019). *Fifty-fifth issue of the Statistical Yearbook*. https://www.stats.gov.sa/en/258.
v. Juddoo, S. (2015). Overview of data quality challenges in the context of Big Data. In *2015 International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-9). IEEE.
vi. Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, *30*(3), 607-610.
vii. Côrte-Real, N., Ruivo, P., & Oliveira, T. (2020). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business
viii. Galhardas, H., Florescu, D., Shasha, D., Simon, E., &Saita, C. (2001). *Declarative data cleaning: Language, model, and algorithms* (Doctoral dissertation, INRIA).
ix. Mecella, M., Scannapieco, M., Virgillito, A., Baldoni, R., Catarci, T., &Batini, C. (2002,October). Managing data quality in cooperative information systems. In *OTM Confederated International Conferences' On the Move to Meaningful Internet Systems'* (pp. 486-502). Springer, Berlin, Heidelberg.
x. Nambiar, R., Bhardwaj, R., Sethi, A., &Vargheese, R. (2013). A look at challenges and opportunities of big data analytics in healthcare. In *2013 IEEE international conference on Big Data* (pp. 17-22). IEEE.
xi. Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211-218.
xii. Rehman, A., Naz, S., &Razzak, I. (2021). Leveraging big data analytics in healthcareenhancement: trends, challenges and opportunities. *Multimedia Systems*, 1-33.
xiii. Nelson, R. R., P. A. Todd and B. H. Wixom (2005). 'Antecedents of information and system quality: an empirical examination within the context of data warehousing.' Journal of Management Information Systems 21(4): 199-235.
xiv. Sidi, F., Panahy, P. H. S., Affendey, L. S., Jabar, M. A., Ibrahim, H., & Mustapha, A. (2012, March). Data quality: A survey of data quality dimensions. In *2012 International Conference on Information Retrieval & Knowledge Management* (pp. 300-304). IEEE.
xv. Wamba, S. F., Akter, S., & De Bourmont, M. (2018). Quality dominant logic in big data analytics and firm performance. *Business Process Management Journal*.
xvi. Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86-95.
xvii. WHO. (2013). (rep.). *Country Cooperation Strategy for WHO and Saudi Arabia 2012–2016*. Retrieved fromhttps://applications.emro.who.int/docs/CCS_Saudia_2013_EN_14914.pdf?ua=1&ua=1