

2019

# A Delphi Study Analysis of Best Practices for Data Quality and Management in Healthcare Information Systems

Olivia L. Pollard  
*Walden University*

Follow this and additional works at: <https://scholarworks.waldenu.edu/dissertations>

 Part of the [Databases and Information Systems Commons](#), and the [Health and Medical Administration Commons](#)

---

This Dissertation is brought to you for free and open access by the Walden Dissertations and Doctoral Studies Collection at ScholarWorks. It has been accepted for inclusion in Walden Dissertations and Doctoral Studies by an authorized administrator of ScholarWorks. For more information, please contact [ScholarWorks@waldenu.edu](mailto:ScholarWorks@waldenu.edu).

# Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

Olivia Pollard

has been found to be complete and satisfactory in all respects,  
and that any and all revisions required by  
the review committee have been made.

## Review Committee

Dr. Barbara Turner, Committee Chairperson, Management Faculty

Dr. Jean Gordon, Committee Member, Management Faculty

Dr. Nikunja Swain, University Reviewer, Management Faculty

The Office of the Provost

Walden University  
2019

Abstract

A Delphi Study Analysis of Best Practices for Data Quality and Management in

Healthcare Information Systems

by

Olivia Pollard

MA, College of St. Scholastica, 2007

BS, University of Phoenix, 2001

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management

Walden University

November 2019

## Abstract

Healthcare in the US continues to suffer from the poor data quality practices processes that would ensure accuracy of patient health care records and information. A lack of current scholarly research on best practices in data quality and records management has failed to identify potential flaws within the relatively new electronic health records environment that affect not only patient safety but also cost, reimbursements, services, and most importantly, patient safety. The focus of this study was to current best practices using a panel of 25 health care industry data quality experts. The conceptual lens was developed from the International Monetary Fund's Data Quality Management model. The key research question asked how practices contribute to identifying improvements healthcare data, data quality, and integrity. The study consisted of 3 Delphi rounds. Each round was analyzed to identify consensus on proposed data quality strategies from previous rounds that met or exceeded the acceptance threshold to construct subsequent round questions. The 2 best practices identified to improve data collection were user training and clear processes. One significant and unanticipated finding was that the previous gold standard practices have become outdated with technological advances, leading to a higher potential for flawed or inaccurate patient healthcare data. There is an urgent need for health care leaders to maintain heightened awareness of the need to continually evaluate data collection and management policies, particularly as technology advances such as artificial intelligence matures. Developing national standards to address accurate and timely management of patient care data is critical for appropriate health care delivery decisions by health care providers.

A Delphi Study Analysis of Best Practices for Data Quality and Management in  
Healthcare Information Systems

by

Olivia Pollard

MA, College of St. Scholastica, 2007

BS, University of Phoenix, 2001

Dissertation Submitted in Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Philosophy  
Management

Walden University

November 2019

## Dedication

This is dedicated to God first! Without the many prayers that were answered through this process, I would not have succeeded. I also dedicate this dissertation to my mom, Paulette Pollard, who stuck by my side, checked on me every day to see my progress, and constantly showed me how to weather the storm and plow through obstacles. My sisters, Katjya, Jhyon, and Page, you also made sure I continued to completion and reminded me how much you were proud of me and were counting on me to finish. My friend Tracy and my cousins Eugene and Cordie for keeping me laughing and my spirits high. Also, for checking in on my progress. Pamela Hardy-Shepard, you were in my life for a short period, but you made a huge impact by reinforcing my resilience and reminding me of the role I play in life. Last but not least, Floyd, you left me alone when I needed it and you didn't complain when I was buried in my work. Your support did not go unnoticed. Thank you all for the words of encouragement. I love you so much!

## Table of Contents

List of Tables .....	v
List of Figures .....	vi
Chapter 1: Introduction to the Study.....	1
Introduction.....	1
Background of the Study .....	3
Problem Statement .....	8
Purpose of the Study .....	9
Nature of the Study .....	9
Definitions.....	11
Conceptual Framework.....	12
Data Quality Assessment Framework.....	13
Information Theory.....	14
Merging the Two Theories.....	15
Research Question .....	16
Assumptions.....	16
Scope and Delimitations .....	16
Limitations .....	18
Significance of the Study .....	19
Significance to Management and Technology.....	20
Significance to Theory .....	20
Significance to Social Change .....	21

Summary and Transition.....	21
Chapter 2: Literature Review.....	23
Introduction.....	23
Literature Search Strategy.....	26
Conceptual Framework.....	27
Data Quality Assessment Framework (DQAF) .....	32
Data Quality Dimensions.....	33
Information Theory.....	34
The Evolution of Data Quality in Healthcare .....	34
Gap in Literature.....	36
Typologies of Data Quality Errors.....	37
Tools to Address Poor Data Quality .....	39
Effects of Poor Data Quality.....	43
Summary and Conclusions .....	44
Chapter 3: Research Method.....	45
Introduction.....	45
Research Design and Rationale .....	46
Role of the Researcher .....	46
Method .....	47
Participant Selection .....	48
Instrumentation .....	49
Field Test .....	50

Data Analysis Plan.....	51
Trustworthiness.....	51
Credibility .....	52
Dependability .....	52
Confirmation.....	52
Ethical Procedures .....	52
Preserving Participant Identity.....	53
Summary .....	53
Chapter 4: Results.....	55
Field Test .....	55
Setting .....	56
Demographics .....	56
Data Collection .....	57
Round Details.....	58
Data Analysis .....	59
Round 1 .....	59
Round 2.....	61
Round 3.....	62
Evidence of Trustworthiness.....	67
Credibility .....	67
Dependability .....	67
Confirmation.....	68

Results.....	68
Significance.....	68
Best Practices That Ranked Low .....	69
Comparison of the Best Practices .....	70
Summary.....	71
Chapter 5: Discussion, Conclusions, and Recommendations.....	72
Interpretation of the Findings.....	72
Best Practices to Improve Data Collection .....	73
Best Practices That Apply to Secondary Use of Data.....	74
Best Practices That Apply to Data Quality and Integrity .....	76
Implication and Suggestions for the Future .....	76
Limitations of the Study.....	77
Recommendations.....	78
Implications.....	78
Conclusion .....	79
References.....	80
Appendix A: Round 1 Delphi Survey Questions.....	104
Appendix B: SurveyMonkey and IRB Guidelines.....	105

## List of Tables

Table 1. Literature Search Sources .....	27
Table 2. Round 1 Survey Results Used to Identify Themes for Round 2 Questions.....	60
Table 3. Round 2 Survey Results Used to Identify Themes for Round 3 Questions.....	61

## List of Figures

Figure 1. Conceptual framework of four categories for organizing and analyzing data quality .....	13
Figure 2. Literature review linkages .....	25
Figure 3. Assessment breakdown of how the data quality assessment framework and information theory can be used to evaluate collected data .....	31
Figure 4. Data collection plan.....	49
Figure 5. Breakdown of demographics for recruiting participants .....	57
Figure 6. Single word cloud for all questions .....	60
Figure 7. Round 3 survey results for Q1.....	64
Figure 8. Round 3 survey results for Q2.....	64
Figure 9. Round 3 survey results for Q3.....	65
Figure 10 Round 3 survey results for Q4.....	65
Figure 11. Round 3 survey results for Q5.....	66
Figure 12. Round 3 survey results for Q6.....	66
Figure 13. Round 3 survey results for Q7.....	66
Figure 14. Best practices from Round 3 that ranked high .....	69
Figure 15. Best practices from round 3 that ranked low .....	70
Figure 16. Best practices compared.....	71

## Chapter 1: Introduction to the Study

### **Introduction**

Problems with data quality and data management challenge healthcare systems here and abroad as older paper and pencil record-keeping systems continue to transition to newer electronic environments (Reader, Gillespie, & Roberts, 2014). Countless lawsuits involving mismanagement of data and data errors have cost the U.S. healthcare system billions of dollars, and more importantly, have resulted in numerous accidental deaths. Moges (2014) succinctly outlined the consequences of mismanaged data in healthcare settings when she stated that

[one] practical example is the death of a pediatric patient because of a misplaced decimal point in the medicine prescription and the health care organization which overpaid \$ 4 million per year in claims for patients who were no longer eligible. Similarly, an eyewear company has incurred one million dollars annually because of lens-grinding reworks, which were caused by data errors. Although losses from poor DQ [data quality] vary, they are measured in the billions of dollars in addition to costs measured in lives lost, employee and customer dissatisfactions.

(p. 2)

Rau (2016) recently penned an article cautioning caregivers about patient discharge from healthcare facilities, claiming that patient discharge was perhaps the most dangerous period of patient care. He recounted facts surrounding a lawsuit launched by family members of a woman who was given a highly potent drug called Methotrexate (a cancer drug) instead of a simple diuretic called Metolazone for her diagnosis of heart

failure (Rau, 2016). The amount of Methotrexate the patient received far exceeded the standard dosage; the diuretic that was ordered should have been administered twice per day, but Methotrexate was administered in its place (Rau, 2016). The toxic dose of Methotrexate led to her untimely death, which could have easily been prevented had the entire patient discharge team subscribed to a standard protocol (Rau, 2016). From the pharmacist, to the pharmacy technician, to the home-health nurse, the healthcare system failed the patient (Rau, 2016).

The United States is one of the largest healthcare markets in the world. The World Bank (2014) reported that an estimated 17% of the U.S. gross domestic product, or approximately \$2.9 trillion, was spent on healthcare in 2013. There are aspects of the U.S. healthcare system concerning information management that are of the utmost importance, specifically data quality issues. According to Eckerson (2002), at The Data Warehousing Institute, countless organizations will share \$600 billion a year in costs because of limited understanding of perceived and real data quality issues. Clack and Woepfel (2018) found that organizations that create a data-driven environment saw an increase in revenue by 6% and an increase in productivity by 5% over their industries' competitors. To improve quality outcomes and reduce costs, organizations may find it advantageous to include data analytics that will utilize quality measures and best practices. Information in the form of raw data is often referred to as the lifeblood of healthcare organizations—an essential component of daily, data-driven, decision-making processes across multiple institutional contexts (Foshay & Kuziemy, 2014). The quality

and integrity of healthcare data are essential not only to the overall provision of superior healthcare but also to organizational effectiveness in the delivery of care and services.

The management of data quality and data integrity are central to quality healthcare (Cruz-Correia et al., 2013). As most healthcare organizations transition to electronic health records (EHRs), data quality management has become a critical area of focus within the healthcare industry (Jamoom, Patel, Furukawa, & King, 2014). There are multiple challenges regarding data quality, including the possibility of inaccuracies and data fragmentation (Adler-Milstein & Jha, 2013). The growth of information systems, the demand for information in a real-time format, and an increased volume of raw data gathered from healthcare systems have all led to an increased awareness of and the need for solutions and a set of best practices for data handling (Singh & Singh, 2010). Despite the need for solutions, there is a dearth of information about primary data problems affecting many healthcare organizations and how they are managed.

This chapter includes the background of the study, the problem statement, purpose statement, and research question. Also included is the conceptual framework and an overview of the nature of the study. I conclude the chapter with the limitations, delimitations, and scope of the study.

### **Background of the Study**

Leaders of healthcare organizations have sought avenues for quality improvement in a number of areas. The advent of linking the quality of care for a patient to the quality of data captured has prompted the U.S. government to take notice and to create initiatives for these organizations to invest in improving data quality and/or diagnose areas in their

data workflow that could be enhanced (Reiter et al., 2014). Two incentives for improvement, the Health Technology Information for Economic and Clinical Health Act, in conjunction with federal standards for meaningful use of EHR systems, have prompted healthcare facilities that admitted patients (hospitals) to develop electronic surveillance methods for reporting healthcare-associated infections (Atreja, Gordon, Pollock, Olmsted, & Brennan, 2008; Jha, 2011). Deming (1986) recognized problems with data quality issues in a rapidly changing healthcare system in the mid-1980s and asserted that management was the key to improving data quality and to mitigating errors industry wide. Improvement suggests change. In terms of change, one that has had a significant effect, both on healthcare providers and on recipients of healthcare services, is the treatment and management of data (Deming, 1986). With the emergence of paperless business transactions, the healthcare industry is well into the process of transitioning from paper and pencil data gathering and physical repositories to a new-age electronic environment. This newer environment offers healthcare providers the opportunity to store unlimited data offsite, enhancing security, as well as providing electronic portals for easy document retrieval to any number of healthcare workers and patients.

Technological change is not without its challenges. The extent to which people in the healthcare industry are adversely affected by ongoing change is highlighted in the literature to include the topic of quality improvement. In a sample of six countries that included the United States, the United Kingdom, and France, 91% of organizations are trying to manage data quality errors (Experian Data Quality, 2013). Experian Data Quality (2013) reported that the percentage of inaccurate data has risen from 17% to 22%

in less than a year. Organizational leaders in the United States believe that 25% of their data suffers from inaccuracies that point to human error (at 59%) as the most significant factor for data inaccuracies (Experian Data Quality, 2013). The report revealed that 66% of organizations surveyed have not successfully developed a solid data quality management system (Experian Data Quality, 2013).

Researchers such as Adler-Milstein and Jha (2013), Schultz (2013), Foshay and Kuziemsky (2014), and Chen, Chiang, and Storey (2012) have tried to ameliorate particular data quality issues with business intelligence (BI) tools in the form of big data (a business tool used to take substantial amounts of data and analyze them for relevance and implications). These tools are often integrated into a data warehouse for information that could be extracted from reports for decision-making (Loshin, 2001; Singh & Singh, 2010). A few researchers have tried to mitigate data quality deficiencies with an implementation of big data tools and to meet government requirements, such as meaningful use criteria, in order to receive incentives (Charles, Gabriel, & Furukawa, 2014). In addition, United States government officials have created initiative programs to encourage healthcare organizations that would implement EHR systems (Charles, Gabriel, & Furukawa, 2014). Accreditation programs and organizations such as Centers for Medicare & Medicaid Services seek to offer standardized, quality measures that hold healthcare organizations accountable for the quality of their data and patient care (Charles, Gabriel, & Furukawa, 2014). There is an oversight by those who encourage healthcare leaders to engage in organizational self-assessment and ongoing research to implement best practices for data quality and management. According to Holzinger,

Dehmer, and Jurisica (2014), research is needed on frameworks that identify exceptions in data quality to address the process managed by humans and computers used to manipulate data. Duggirala et al. (2016) posited that information mining of different sources, such as medical writing, EHRs, and online networking, share a significant number of the difficulties identified with security reports information. The nature of information in these sources can be more reliable, contingent upon the structure of the database and the preparation of the individuals who enter the information, shifting from apparently high caliber to a lower grade of information (Duggirala et al., 2016). Logical difficulties will continue to grow with the expansion of new reconnaissance information sources and the advancement of new techniques for submitting unconstrained reports from online and versatile applications (Duggirala et al., 2016). It will be critical for organizations with huge datasets such as the Food and Drug Administration to structure its data innovation frameworks with the goal that information can be submitted, recovered, prepared, and assessed in an institutionalized way. These incorporate EHRs, personal health, claims, models for health information, information from government and private cell phones for following health plans, and information from social sites such as websites, quiet backing bunch locales, and inquiry term logs (Duggirala et al., 2016). Much of the research on data quality has become the mission of independent, private, and government organizations such as the Healthcare Information and Management Systems Society (HIMSS), Experian, National Institutes of Health, American Society for Quality, Agency for Healthcare Research and Quality, American Health Information Management Association, and Gartner. The research from these organizations has demonstrated that

these information sources can be of an incentive in postshowcase healthcare observation and other related fields. Facilitating improvement and usage of a progressed and coordinated secure information mining framework upheld by suitably experienced work force will be fundamental for better educated basic leadership and healthcare administration to continuously monitor for issues (Duggirala et al., 2016; Holzinger et al., 2014). Empirical research is needed to keep the industry aware of current data management practices within healthcare organizations as well as the success rates of an organization's adopted data management practice. In this research, I employed contemporary guidelines and ideas for organizations seeking to implement best practice data management, such as assessment of data quality and data management needs. These helped to ensure that effective decision-making techniques are being applied as technology-driven healthcare modes continue to evolve.

The search for innovative ways to manage the growing amount of data has created opportunities for the introduction of new technology. The introduction of EHRs has allowed healthcare centers access to vast amounts of data, as well as patient access to records through the use of patient portals. As a result, healthcare organizations have made changes to their work practices and processes for using data acquired from EHRs (Tolar & Balka, 2012). At question is whether data quality issues and data integrity of information retrieved from EHRs are being identified and how those data are being managed.

## **Problem Statement**

Data quality and data integrity concerns such as a lack of shared patient history between health care providers, improper coding, insufficient enrollment or insurance information, which has led to the denial of referrals, or data inconsistencies, are not being addressed as often as necessary, all of which adversely affect the delivery of quality care. Multiple challenges pertain to data quality and integrity, including inaccuracies and data fragmentation (Adler-Milstein & Jha, 2013). The general problem is that patients and healthcare organizations are adversely affected by poor data quality (Anderka et al., 2015). Data quality involves the accuracy and completeness of collecting data, consistency of processing the data, and timeliness of data usage (Dixon, Siegel, Oemig & Grannis, 2013).

Errors in demographic information could, and sometimes do, adversely affect key management decisions. This and other aspects of inaccurate data quality for patients and healthcare organizations have been found to be a consequence of using data acquired and processed for later use, resulting in diminished healthcare service to the community (Mawilmada, Smith, & Sahama, 2012). Mawilmada, Smith, & Sahama (2012) expressed that secondary data received from out of an EHR are used primarily for reporting purposes. Reported data cannot be considered usable if the data are bad. According to one Oracle report (2015), as much as \$70.2 million in annual revenue is not sustainable when accurate secondary data are not available for sound decision-making (p. 29). Another study completed by Information Builders (2013) identified that nearly 15% to 20% of

annual revenue from the influx of new and current patients is lost due to inaccurate secondary data (p. 3).

The specific problem is that healthcare leaders need a better understanding of how best practices can be used to effectively manage accurate patient information (Rau, 2016). Bowman (2013) and Weiskopf and Weng (2013) reported that processes and protocols to mitigate data quality errors and to identify their causes or even obtain viable information on EHR data quality and integrity remain scarce in the literature.

### **Purpose of the Study**

The purpose of this qualitative Delphi study was to identify best practices for data quality and data management in healthcare systems. This study has the potential to inform leaders on the implementation of best practice guidelines for data quality and data management.

### **Nature of the Study**

The research method most appropriate for this study was qualitative. According to Baxter and Jack (2008), a qualitative approach allows users and managers of organizational systems an opportunity and provides information based upon their interpretations of ongoing processes activities and data. The research method allowed for virtual interactions between smaller nonrandomly selected participants and their responses to be studied as a whole and for patterns and themes to be identified, where a quantitative study would not. This study consisted of three rounds of questions for a panel of IT and non-IT healthcare data experts who met the selection criteria identified in Chapter 3. The multiperspective survey population consisted of 25 different healthcare

systems data management professionals across the United States who have worked or are working for a healthcare system such as an EHR and have managed the data within this system. From that population, 25 participants were selected using purposive sampling to secure critical perspectives from participants. Twenty participants were necessary to provide a more robust outlook on the topic. An additional five participants created a cushion in case some participants withdrew before the study was completed. With a quantitative method, a study of data handling practices among selected participants would have to be conducted in a controlled environment as an experimental comparison may need to take place, whereas with a qualitative method, participants could be queried in a more natural, comfortable environment.

The research design was a three-round Delphi study that focused on current data quality issues and best practice ideas among healthcare systems experts. Surveys were appropriate for the study, given they provided a focused method for acquiring pertinent information from the selected participants (see A. von der Gracht, & Darkow, 2013; Baxter & Jack, 2008; Jansen, 2010).

Research protocols for this study included a structured open-ended questionnaire in the first round, followed by two rounds of a structured mix of dichotomous, Likert, and scaling questions. A. von der Gracht and Darkow (2013) implied that the results from the first round drive the questions for the second round. After analyzing the results from the first round, the second round of questions were created. The third round of questions were similarly generated. The process of data collection and healthcare data management

best practices was the focus of the data quality evaluation. In this evaluation, I specifically sought to determine if data process errors are consistent among participants.

The two other research designs considered but not used were ethnography and phenomenology. Ethnography would not have been appropriate as a research design for this study, as it would have focused more on organizational culture such as rituals and customs of the population rather than the topic. It would also have required the me to become immersed in the organizational culture of selected research sites rather than to observe participants from outside. Phenomenology would not have been appropriate because it focuses on the meaning of a specific lived experience to a researched group (see Donalek, 2004).

### **Definitions**

The following definitions operationally defined terms that were used in this study:

*Data integrity*: “The extent to which healthcare data are complete, accurate, consistent, and timely” (Data Integrity, 2017).

*Data quality*: “The reliability and effectiveness of data for its intended uses in operations, decision making, and planning” (Data Quality, 2017).

*Data quality assessment framework (DQAF)*: Framework that uses data quality dimensions to create measurements for improvement (Sebastian-Coleman, 2012).

*Data steward*: An individual put in place for business purposes to track, manage, and escalate data quality issues where they arise (Judah & Friedman, 2014).

*Electronic health record (EHR)*: Electronic wallet of health information on a patient that can be used in many different organizations to communicate health status, lab

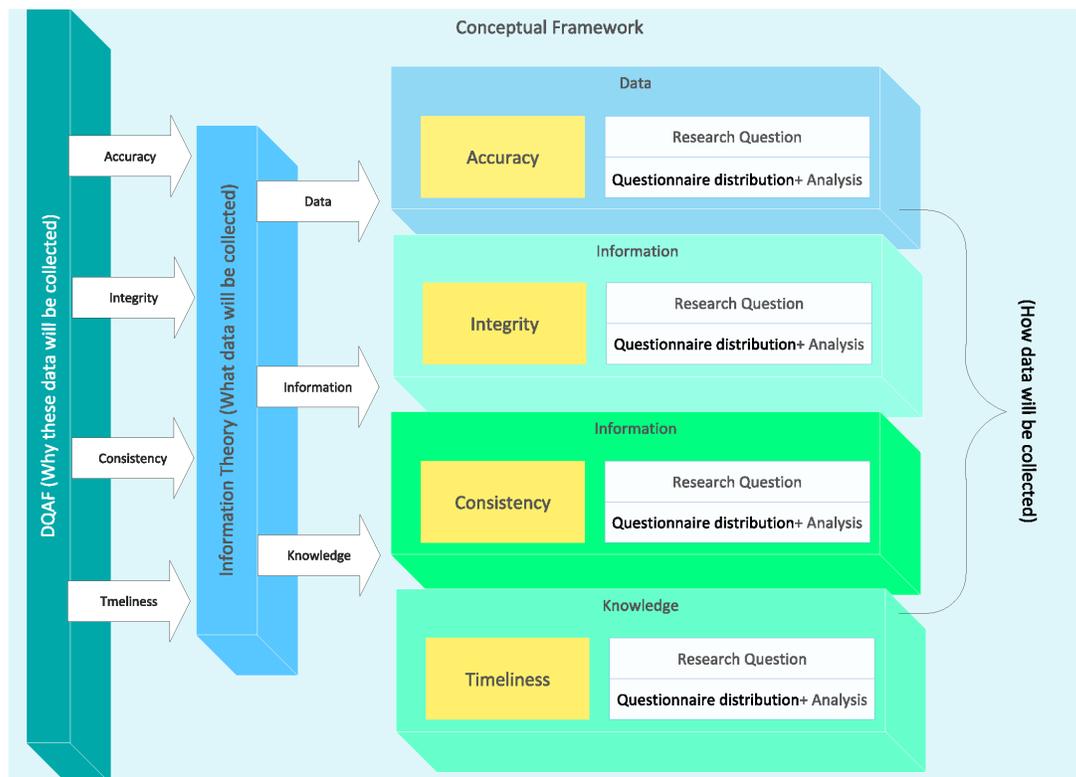
results, and medications. An electronic record used to keep everyone involved in the care of all patients equally aware (Electronic Health RecordEHR, 2017).

*Health information manager:* Health professional who strives to ensure the quality of clinical data for decision-making purposes (Health Information Management, 2017).

*Healthcare Information and Management Systems Society (HIMSS):* Private organization that creates educational services built on improving healthcare information technology (HIMSS, 2015).

### **Conceptual Framework**

Conceptual approaches were applied to guide the evaluation process in assessing data handling practices. The conceptual framework was the data quality assessment framework (DQAF) as the overarching guide for assessment. Figure 1 demonstrates the concepts that ground this study. Chapter 2 addresses this framework in more detail. The DQAF includes elements from information theory, such as the data quality dimension concepts. Information theory was employed to evaluate the extent to which the target panel members selected for this study have developed and used a set of best practices for managing data quality.



*Figure 1.* Conceptual framework of four categories for organizing and analyzing data quality. Adapted from DQAF developed in 2003 by the IMF. Adapted from “Data quality assessment framework and data quality program” by IMF, 2003, <http://www.imf.org/external/np/sta/dsbb/2003/eng/dqaf.htm>. Copyright 2003 by the International Monetary Fund.

## Data Quality Assessment Framework

Originally constructed in 2003 as a qualitative solution to data quality by the International Monetary Fund (IMF), researchers applying the DQAF use specific data quality dimensions such as accuracy and integrity for evaluating statistical data from

countries around the world (Data quality assessment framework and data quality program, 2003). Sebastian-Coleman (2012) asserted that the DQAF does not use accuracy as a dimension for measurement. Instead, DQAF creators have selected validity as a more fitting measurement, owing to their perspective that accuracy would require real world comparisons of data in order to confirm the framework's legitimacy.

I used the DQAF to obtain a perspective of what could be effectively used as a standard guideline for managing data quality. I analyzed the responses of participants to determine Round 2 and Round 3 questions. This analysis was based upon information obtained from the healthcare IT management panel. I distinguished among three practices by noting which type is being used by each panelist: in-house, revised, or derivative, as modified from an accreditation agency or organization. For this study, only certain components were necessary to create the measurements to test. The DQAF model demonstrated the why of the data that were collected and the data quality dimensions. Information theory provided the what of the data collected and addressed the data as data (panel response acquisition), information (panel response), or knowledge (panel response analysis). In a 2010 working paper, Mrkaic used information theory to further categorize the DQAF concepts. Figure 1 shows how the DQAF was used as the overarching framework from which measurements were created to assess the research questions and the data collection method.

### **Information Theory**

Information theory was originally applied as a quantitative mathematic theory. Since its introduction, this theory has been modified and now branches into qualitative

studies and social science. Health informatics, in particular, has used this theory in a number of studies that have launched more than a few different models, to include the Blum model (Nelson & Staggers, 2014). In this study, I applied the Blum model. The model falls under the umbrella of information theory and is one of the most widely used models within health informatics. It can be used to address the processing of data as three elements: data, information, and knowledge. This model uses the processing of elements to explain the creation of descriptive categories and their relationships to each other. This model was employed in the study through surveys in order to disclose how users enter data, manage the data, and translate to information for managers to apply.

### **Merging the Two Theories**

As patient data are entered into an EHR, a margin for error exists—incorrect dates of service, omissions of one or more medications, incorrect measures of laboratory reports, as well as the addition to a patient’s record of surgeries or conditions that were in error yet continued to be carried forward as part of the digital record, to name a few. End users of entered data make treatment decisions based upon information contained in an EHR, often without clarification from the patient regarding the authenticity of their EHR. As a result, the accurate processing of patient records continues to plague the healthcare industry. Literature findings on information theory and the DQAF support the above claim of conflicts in the accurate processing of data (Panahy et al., 2013). Mishaps have been found to occur at the acquisition (data input) stage due to an inadequate understanding of how data will be used in the final output or at the final stage when faulty communication occurs on what output is expected (Panahy et al., 2013). Data

quality dimensions found in the DQAF can be used to find areas of improvement in the process of a system and its information (Panahy et al., 2013).

### **Research Question**

This study was guided by one essential question:

Research Question: What practices contribute to identifying improvements in EHR data collection, secondary use of data, data quality, and integrity?

### **Assumptions**

The research was based on four assumptions. The first assumption was that all the participants' responses were honest, detailed, and straightforward. I assumed that all selected participants would fully participate in all three rounds. I also assumed that the IT managers who were chosen to participate were aware of and actively monitoring data quality or have performed this duty in their past experience. Another assumption was that they had the same level of experience; however, each participant could have their own unique experience with the topic or the field of study.

### **Scope and Delimitations**

The purpose of the study was to investigate best practices for electronic data management through the perspectives and experience of the participants. I sent out surveys to healthcare IT managers. The specific systems managers who were used for this study were those identified as data stewards. Each data steward was selected based on that employee's role as liaison through multiple departments, executing multiple activities, and on that employee's ownership of the data. According to a Mass Storage Systems and Technologies report submitted by Duerr, Parsons, Marquis, Dichtl, &

Mullins (2004), the position of data steward has been in service for more than a decade with the creation of World Data Centers in 1957-1958, but it is newly recognized as a critical position for managing data (p. 48). IT managers with the titles of database administrator, database architect, and data analyst were also considered, as their roles in the organization largely address such design and other technical issues as security, performance, and database backup, while the data steward is responsible for addressing policies, procedures, and data standards (see Loshin, 2001).

Panahy et al. (2013) indicated that managers depend on the quality of data to make useful decisions. Other healthcare management positions such as directors and vice presidents were also considered but not selected, as they represent higher level management positions where decisions do not affect organizational data that directly impact the community.

The United States was selected as the geographical location to study because it represents the third largest populated country in the world from which to gather substantial healthcare information (see Harder, 2013). In addition, it is highly diverse, with an estimated population of 325,719,178 as of July 1, 2017 from which to draw participants (see U.S. Census Bureau, 2017). Smaller more focused pools of participants from metropolitan cities like New York, Philadelphia, or Chicago with large culturally diverse populations were also considered but not selected. Population figures across the United States are much greater.

The selection of the United States with its many IT healthcare managers constituted a delimitation of the study. This geographic location was selected in order to

avoid the potential for an enormous population of participants, such as that from the addition of other countries. A larger geographic location—other countries—limits the researcher's ability to be flexible with participants' schedules and time zone. Another delimitation was the inclusion of only IT managers that manage the data collected. Although expanding the study to include other healthcare systems and data users was considered, the inclusion could well return much irrelevant information.

### **Limitations**

The potential limitations of this study involved time, sample, questions, authentic response, and suggestions. Time constraints and conditions for participants were a cause for concern. The study took place over a 4-month period. Within this time, some participants were not able to dedicate the required time to complete the survey for each round. This limitation was addressed by providing more than one option for survey submit date, system scheduling for surveys, and predeterminations of the process for analysis of the collected responses.

Heterogeneous purposive sampling could produce a sense of bias during participant selection process. I applied maximum variation purposive sampling to secure critical perspectives from participants (see Palinkas et al., 2015). I used a list of participants with a specific expertise identified in Chapter 3.

Some questions may have been difficult for participants to understand, which might have resulted in skewed responses. To lessen the chances of a question being too difficult to answer or the response being skewed, I phrased the questions using industry

standard terminology and provided a space for additional information. A field test of the Round 1 questions was conducted using three experts in the field to refine the questions.

### **Significance of the Study**

This study has the potential to address the gap in literature regarding the lack of empirical studies on best practices to mitigate data quality concerns in healthcare EHR systems. The study's findings could prove significant to the healthcare industry, the public, and practitioners by corroborating prior findings of relevance and by disclosing the need for further research. Moreover, these findings may be significant to the healthcare industry, the public, and practitioners, as they have the potential to provide further evidence of previously reported issues and new problems that may not be reflected in other studies. Accurate data quality measures could strengthen the decision-making process for healthcare services and quality of patient care.

New findings may also extend the knowledge base and usher in newer protocols for data handling. Some researchers (Adler-Milstein & Jha, 2013; Chen et al., 2012; Foshay & Kuziemy, 2014; Schultz, 2013) have tried to ameliorate particular data quality issues with BI tools in the form of big data (a business tool used to analyze large amounts of data for relevance and implications). A few researchers have tried to mitigate data quality issues with an EHR implementation to meet government requirements in order to receive incentives (Charles, Gabriel, & Furukawa, 2014).

This research can aid in the development of best practice guidelines and accountability measures to reduce the number of digital errors EHRs to implement and manage data for optimal results and quality assurance.

**Significance to Management and Technology**

This study will contribute to the knowledge base for the management of data quality in producing information that is readily applicable. With an increasing number of healthcare organizations implementing EHR systems for data management, the need for error-free data will continue to grow. The health industry could benefit from this study by using the data to launch more studies on data quality and data management obstacles. The healthcare industry could be able to manage their data more efficiently by critically identifying, assessing, and implementing some or all identified best practices as a solution.

**Significance to Theory**

This study could provide useful secondary data for efficient decisions and assist in providing the best delivery of care for healthcare organizations and may eventually lead to new theory development. The process to develop this theory would begin with observing current data workflow practices that are identified in the EHR system and the management related to the acquired data. The next step would be to evaluate new data quality errors discovered as a result of this research project and to develop methods of ameliorating them. The last step would be to identify better guidelines for data error prevention that can be consistently applied based upon the most identified data errors or mistakes revealed in the study, and the best practices for quality and error-free data within healthcare organizations.

## **Significance to Social Change**

This study has the potential to be significant, as it may offer insights that could affect the way healthcare organizational leaders and managers view their data and make decisions within their communities based upon said data. Trends and patterns identified in the study could also provide a basis for understanding what data quality deficiencies may be present in the secondary data when making determinations about services. Such trends and patterns could eventually strengthen patient quality of care. The public could benefit from this study's findings, as there could be an increase in needed services in the community. The findings could ameliorate healthcare service reductions or even eliminations that result from insufficient or incorrect data.

## **Summary and Transition**

Chapter 1 addressed the problem of adverse effects within healthcare due to inadequate or questionable quality of data used to make decisions on healthcare delivery. The study could illuminate best practices in healthcare and could meaningfully connect findings to the academic fields of management and technology. The conceptual framework of Blum and DQAF were used to guide the structure of Delphi survey questions and to identify what information to collect. The DQAF provided assessment measures based upon data quality dimensions, and information theory (Blum model) helped to categorize the type of information that fell into each data quality dimension assessed. Chapter 2 is a literature review on the main themes of quality management as it has evolved into data quality management. In Chapter 3, I detail the study's methodology

and design. In Chapter 4, I display research data collections and findings. Chapter 5 contains my analysis of the study's findings and linkages to needed future research.

## Chapter 2: Literature Review

### **Introduction**

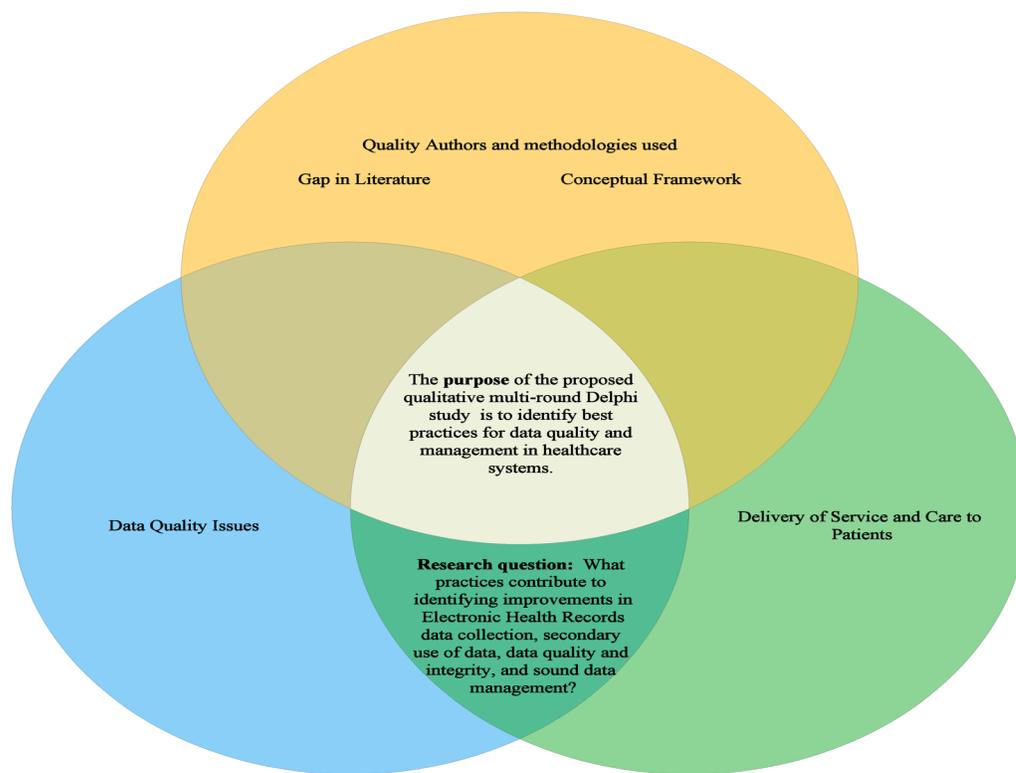
Data workflow practices for identifying data quality and managing data quality deficiencies have yet to be standardized to ensure optimal delivery of care and services to patients. The reasons for poor data quality range from human error to insufficient data, with human error as the leading cause of poor data quality. Moreover, Bowman (2013) suggested human error or user error as one of the leading causes of data quality issues within an EHR. According to a study by the HIMSS (2014), poor data quality becomes evident in healthcare organizations when errors and insufficiencies appear in the process and delivery of care as well as a decline in appropriate decision-making as a result of unreliable data. There are multiple challenges concerning data quality and integrity, including inaccuracies and data fragmentation (Adler-Milstein & Jha, 2013).

Bowman (2013) and Weiskopf and Weng (2013) reported that they found little information regarding poor EHR data quality and integrity. From the available research, Bowman and Weiskopf and Weng concluded that nobody has addressed a strategy to mitigate data quality errors, such as incorrect information found on lab reports, patient enrollment forms, or imported patient records. Another strategy identifying causes of data quality errors is attributed to data input quality such as lack of communication and cross-check (Bowman, 2013). Human error is attributed to potential for data errors, ranging from data input practices of clerks to data users in management positions (Weiskopf and Weng, 2013). As most healthcare organizations transition to EHRs, data quality management has become a critical area of focus within healthcare organizations (Jamoom

et al., 2014). The purpose of this qualitative Delphi study was to identify best practices for data quality and data management in healthcare systems. As the literature has indicated, the healthcare industry lacks a systematic approach to resolving human errors in electronic data handling. Some of these errors have led to untimely deaths, incorrect diagnoses, loss of revenue for healthcare organizations, and other failures that have contributed to substandard health care. The results of this study could assist in the development of a set of guidelines for quality control of data handling practices within U.S. healthcare systems.

While a number of research articles have addressed multiple problems concerning data quality within healthcare organizations in general, few researchers have examined the nature of data quality and related data management practices. Poor data quality within EHRs results in lost reimbursements and reduced patient safety as well as a loss of services to communities and increased operating costs (Brown, Weber, & DeBie, 2014; Foot et al., 2014; Todoran, Lecornu, Khenchaf, & Caillec, 2015). The purpose of this literature review was to align information obtained from articles and theories with current data management practices in healthcare systems today and to contribute research findings as a first step in the identification of universal or common practices of healthcare data handlers and users. With the identification of universal or common practices, my next step was to create a systematic approach for identifying and managing data quality issues as a path to the development of a set of data quality best practices. The review in this chapter reveals the need for the study by disclosing the lack of empirical studies available on this topic.

Chapter 2 addresses available literature on the following topics: (a) relevant theories in quality management, (b) trends in applying general quality management theories to assess and improve data quality, (c) healthcare organizations' current data quality concerns, and (d) common practices to address data errors and data quality management. The diagram in Figure 2 depicts the linkage of themes, research question, and purpose of the study.



*Figure 2.* Literature review linkages.

The review of literature supports the research question, selected methodologies, and design of the study. Chapter 2 consist of these major sections: (a) Introduction, (b) Literature Search Strategy, (c) Conceptual Framework, (d) Seminal Authors in Data Quality Management, (e) Gap in Literature, (f) Data Quality issues, (g) Delivery of Service and Care to Patients, and (h) Summary.

## Literature Search Strategy

The following search terms and keywords were used to search scholarly databases and were deemed optimal to acquire information specific to data quality and the healthcare industry: *data quality dimension, data quality, data integrity, electronic health record, data completeness, data timeliness, data consistency, data accuracy, data quality assessment framework, open systems theory, information theory, Blum model, health informatics, data quality issues, data quality problems, poor data quality, data quality characteristics, history of data quality, data quality in healthcare, impact of bad data, bad data, data in healthcare, theories used in healthcare, data quality in EHRs, case studies in healthcare, methodologies and designs used in healthcare, qualitative studies, and delphi method.*

Database searches included the following categories: general databases, technical databases, and health sciences databases. The specific names of these databases included Proquest Dissertations & Theses, Dissertations & Theses at Walden University, ebrary, EBSCO eBooks, Google Scholar, JSTOR, SAGE Encyclopedias, SAGE, Premier, SAGE Research Methods Online, IEEE Xplore Digital Library, Safari Tech Books, Science Journals, MEDLINE with Full Text, CINAHL Plus with Full Text, ProQuest Nursing & Allied Health Source, Health & Medical Complete, CINAHL & MEDLINE Simultaneous Search, PubMed, Science Direct, and Cochrane Database of Systematic Reviews. The New York City library was also used to acquire older books in order to place the topic within its historical context and to complete historical information for other databases. Table 1 displays the sources searched, material found, and material used.

Table 1

*Literature Search Sources*

Source	Peer reviewed articles found	Peer reviewed articles used	Books searched	Books used	Dissertations searched	Dissertations used
General databases	5,680	<b>80</b>	20	<b>2</b>	50	<b>3</b>
Technical databases	2,584	<b>75</b>	10	<b>4</b>	0	<b>0</b>
Health sciences databases	8,393	<b>50</b>	10	<b>2</b>	0	<b>0</b>
Library	80	<b>20</b>	30	<b>5</b>	0	<b>0</b>
<b>Total</b>	<b>8,473</b>	<b>225</b>	<b>70</b>	<b>13</b>	<b>50</b>	<b>3</b>

### Conceptual Framework

There was one conceptual framework applied to this study: the DQAF. This framework was originally created to improve data quality through measurements and data quality dimensions within systems in the economic sector but was later modified and used in other sectors such as education (Sebastian-Coleman, 2012). The DQAF helped to drive the assessment of the data that were collected to support this study. The most frequently used data quality dimensions were also used to steer the collection of the data.

The DQAF allows for the acknowledgement of the assumptions of data quality, data integrity, and data management processes to be addressed by staff who collect the data. In this scenario, measurements were created as a product of the information acquired by data collected from panelists regarding data within EHRs and possible decisions made and manage the actual data. DQAF addressed both the data collection and

data management aspect of the study. Senkubuge, Modisenyane, and Bishaw (2014) used the DQAF as a framework within their methodology to assess the connections between health sector reform processes and reinforcement of health systems. Senkubuge et al. found that there was a lack of government participation in strengthening healthcare systems. They also asserted that there is a need to prioritize between different stakeholders as to what policies should be put in place as well as who will own and manage these systems (Senkubuge et al., 2014).

These findings demonstrate a need for stakeholder contribution and commitment to create a well-managed system. The researchers also addressed availability and accuracy of data problems as well as the absence of sufficiently trained employees in healthcare systems worldwide (Senkubuge et al., 2014). Jerven (2014) determined that errors due to omissions created an accuracy problem with statistical data and indicated that omissions resulted in unreliable data. Jerven used the DQAF, evaluating six fields with five data quality and summarizing dimensions of 16 reports from the IMF on the African trade economy. Three of the five data quality dimensions used in Jerven's study were used in my study: serviceability (timeliness), accuracy, and integrity (completeness). Accuracy relates to information theory (Blum model), which uses the approach of processing data, knowledge, and information (Blum, 1986).

Nelson and Staggers (2014) reported, that since its introduction in the mid-1950s, information theory has been modified from its quantitative origins to now include qualitative studies and social science. The catalyst for information theory was the invention of the computer. Adherents of information theory attempt to link the process of

knowledge acquisition in the human brain to computer operations from data input to output. Health informatics, in particular, has used this theory in countless studies, launching a number of different models such as the Blum model (Holzinger et al., 2014; Nelson & Stagers, 2014). Scott, Sera, and Georgopoulos (2015) explained their use of information theory relative to quantitative evaluations of how decisions are made over knowledge domains. The study consisted of 16 different tasks and questioned what tasks were repeated most often and how they should be categorized (Scott et al., 2015). Scott et al. found that by using a version of information theory (entropy), participants acquired a better understanding of the importance of education.

Todoran, Lecornu, Khenchaf & Caillec (2015) concluded that information theory was the most appropriate theory to use when assessing information quality. They reported that the theory yielded a lens through which to view outcomes used for decision making. The findings from their study a connection provided a justification for my use of information theory in that it informed my study in categorizing results for information quality.

The Blum model falls under the umbrella of information theory and is one of the most widely used models within health informatics to address the processing of data focusing on three elements: data, information, and knowledge. This model uses the processing of elements to explain the building and linkage of elements to render a final product much like the design of a flow chart depicting the steps involved with playing sports. This model was fundamental to my study to show how the data received is digitally entered by users to create information and for managers to use as knowledge in

decision making. Within this process, there is a breakdown that occurs with the quality of data entered into the electronic record keeping system. This breakdown could occur at the data acquisition level due to the lack of understanding of how the data will be used in the final output, or there could be a breakdown in communication on what final output is expected.

Figure 3 shows the assessment breakdown of how the data quality assessment framework and information theory can be used to evaluate collected data. The assessment of the internal and external validity of the data input by personnel who engage in digital record entry and management are the focus of data quality during this research. Information theory will specifically address the data collection aspect of this study by applying consistency, integrity, accuracy, and timeliness to data, knowledge, and information.

ASSESSMENT BREAKDOWN				
DQAF Dimensions →	Timeliness	Accuracy	Consistency	Integrity
Information Theory ↓				
Data/Observation	Observe and note the time between when the questionnaire goes out and respnses as well as the time between each round.	Observe and note the reports of errors to compare against common practices, policies, and procedures as outlined by other panelists.	Observe and note the information on the questionnaires to record how the information may differ from round to round.	Observe and note how the process of data input is carried out to see if it meets the common practices, policies, and procedures as outlined by the panelists.
Data/Documentation	Collect questionnaires from each round to analyze for trends. This will help to create the second and third round of	Collect questionnaires with information on thresholds set to compare to what is	Collect questionnaires that contain data that have been identified to have an "issue."	Collect questionnaires with information at different points in the process to catch errors.
Information/Observation	Observe and note when the information was requested compared to when it was received.	Observe and note how the information is stored and managed to common practices, policies, and procedures as outlined by other panelists.	Observe and note the information provided by data warehouse to support their knowledge to compare to what will actually be done with the	Observe and note how managers compare their inputte dinformation to its output.
Information/Documentation	Collect the request for information and compare it to a time stamp acknowledging receipt.	Collect questionnaires from panlists that are industry experts such as managers of data and decision makers of the data as a source of original documentation.	Collect samples of information from questionnaires in the form of exmaples provided in responses to questions.	Collect notes on process procedures from questionnaires to idientfy trends in common practices, policies, and procedures as outlined by panelists.
Knowledge/Observation	Observe and note the time it takes from receipt of the questionnaire to receipt of responses.	Observe and note how the data is managed and extracted compared to what is received by management.	Observe and note the knowledge of different panelistsbased on the type of position and daily interaction with the data.	Observe and note the process of managing the data collected to the use of the data.
Knowledge/Documentation	Collect questionnaires from IT to compare to data receipt of questionnaires by data decision makers.	Collect example information from the questionnaires to identify the trends to common practices, policies, and procedures as outlined by other panelists.	Collect exmaples from the questionnaires of information that is available through the responses of the panelists.	Collect reponses in the questionnaires that will show the managed raw data and how it is processed and provided to decision makers.

*Figure 3.* Assessment breakdown of how the data quality assessment framework and information theory can be used to evaluate collected data. Adapted from DQAF developed in 2003 by the IMF. Adapted from “Data quality assessment framework and data quality program” by IMF, 2003, <http://www.imf.org/external/np/sta/dsbb/2003/eng/dqaf.htm>. Copyright 2003 by the International Monetary Fund.

## **Data Quality Assessment Framework (DQAF)**

The DQAF is quickly becoming a useful tool for assessing data quality in industries other than finance. Its main use is to reveal quality issues by finding anomalies in data patterns within a system (Sebastian-Coleman, 2012). Although utilization of the framework by Yang et al. (2014) was for a financial system, the assessments and measurements created allowed researchers to improve standards for the system and use of big data tools. Senkubuge et al. (2014) validated the use of the DQAF in their study on health systems, health sector reform, and delivery of services as an assessment tool to analyze data collected. Their study focused on two data quality dimensions, availability and accuracy. They used accuracy as one of the data quality dimensions to qualify the use and demand of information within the system under study. Although the creators of the DQAF use validity in place of accuracy in their framework, accuracy was used in my study.

In addition to the four data quality dimensions selected for this study, the DQAF includes completeness. This dimension is a component of the dimension of Integrity (see Appendix A) in my study as it is based upon the existence of data. Moreover, the foundation of this study is the existence of the data, which is the general focus; therefore, the dimension of completeness, rather than appearing to be omitted, is subsumed under the dimension of Integrity.

Abdul-Hamid (2014) agreed that the use of the DQAF would yield structured assessments to evaluate systems information. Abdul-Hamid used the following five data quality dimensions: integrity, accessibility, accuracy, reliability, and serviceability. His

study found varying degrees of strengths and weaknesses in the quality of data and decisions based on the data. Abdul-Hamid extended his study further than Senkubuge et al. by using a situational analysis tool SWOT (Strengths, Weaknesses, Opportunities, and Threats), and benchmarking, in conjunction with the DQAF to secure stronger results. SWOT was developed in the 1960s as a technique to unleash an individual's career potential, but its principles are applied today as a problem-solving tool.

### **Data Quality Dimensions**

Data quality dimensions and characteristics of data quality, which are often called information quality dimensions, are used to create measurements for data quality. There are 16 data quality dimensions, which are discussed in research (Wang & Strong, 1996). However, just four of them are most often used by researchers as a guide to assess the quality of data within an organization. Shankaranarayanan and Cai (2006), along with Bray and Parkin (2009), conducted studies in which three of the most commonly used dimensions of data quality, completeness, accuracy, and timeliness were used to test the quality of data production. Although there were multiple ways to view the data from a quality perspective, the authors found that decisions related to data processes influence how data quality is perceived, thereby introducing subjectivity rather than consensus into the results (Shankaranarayanan & Cai, 2006); Bray & Parkin, 2009). Data quality dimensions provide a starting point for organizations to begin identifying and building data quality measurement categories. Although there are over 200 data quality dimensions addressed in research, in this study I will address the four most widely discussed data quality dimensions among researchers on this topic. Abdul-Hamid (2014),

and Moghaddasi, Rabiei, & Sadeghi (2014) identified integrity, accuracy, consistency, and timeliness as the four specific data quality dimensions used most frequently.

### **Information Theory**

Ramírez-Gallego, Mouriño-Talín, Martínez-Rego, Bolón-Canedo, Benítez, Alonso-Betanzos, & Herrera, (2018) found information theory to be a broad theory that encompasses many different aspects and versions of what a theory is or could be. One of the information theory models frequently used in health information/informatics systems is the Blum Model. Blum (1986) demonstrated how medical informatics or clinical information can be processed in a system as three categories: data, knowledge, and information. This model eventually became the data, knowledge, information model to which some researchers have added wisdom. Goossen (2014) used the model to draw a connection in his study from human thinking and systems thinking with information from an EHR.

### **The Evolution of Data Quality in Healthcare**

The topic of quality in all of its dimensions is broad. From quality assurance to data quality, countless scholars have written volumes on the nature of quality and quality issues in their respective fields. However, a few might be considered as pioneers whose writings are fundamental to most fields of study that embrace quality issues such as Deming, Juran, and Shewhart.

Deming (1986) established the idea that system and employee performance quality first arises from management, then progresses to other employees in the organization with Total Quality Management. His approach leaned more toward a theory

of employee management than one of product and focused on incentives to improve employee performance. This assumed that employee performance improvement would refine the quality of the product.

One of the reasons his 14 points of managing were adapted to the healthcare industry was to introduce management to a new way of thinking for assessing acceptable levels of errors for ever-revising assessment protocols. Juran (1951) approached quality, and the management of quality, through different areas of management processes, such as planning, control, and improvement. Juran believed that proper training and management of processes could result in lowering the cost of quality.

Shewhart (1939) believed that methods that yielded accuracy and precision of quality outcomes should be deployed. Ivanov extended Shewhart's theory regarding minimizing errors, and the connection between data and information (Sebastian-Coleman, 2012). One assessment method that emerged from the work of these early authors on topics related to quality was Total Quality Management (TQM) developed in the 1950's by Deming and later expanded on by Shewhart (Joshi, Ransom, Nash & Ransom, 2014).

In the 1980s, these ideas were extended with the introduction of Lean Six Sigma, which sought to remove inefficiency by addressing timeliness and to solve business process problems by reducing variations (Nave, 2002). In the 1990s, Harvard Business School professors, Norton and Kaplan (2016), developed a strategy called Balanced Scorecard to measure business processes, customer satisfaction, new ideas, and training of personnel. All these approaches have been applied to data quality today. Their various

aspects have been used to create tools designed to address data quality problems currently being discovered.

### **Gap in Literature**

This study has the potential to address the gap in the lack of empirical studies on data quality deficiencies or errors in health care EHR systems. Organizations recognize the need for usable research to strengthen data quality. According to Holzinger, Dehmer & Jurisica (2014), there is a need for future research for creating frameworks to identify exceptions in data quality and to address the process used by humans and computers in data manipulation. Shankaranarayanan & Blake (2017) posited that future data quality research would require Using a predictive algorithm data mining model in a dataset to assess the performance and manage quality within an EHR. This conclusion was determined by utilizing a case study approach with supporting data from datasets. Researchers have advanced vigorous and valid arguments to address unanswered questions regarding the question “what are considered universal practices for handling data quality problems within EHRs” (Joglekar, Anderson, & Shankaranarayanan, 2013)

Much of the research on data quality has become the mission of independent, private, and government organizations, such as Healthcare Information and Management Systems Society (HIMSS), Experian, National Institutes of Health (NIH), American Society for Quality (ASQ), Agency for Healthcare Research and Quality (AHRQ) and Gartner. The dearth of academic literature that addresses poor data quality suggests a need for scholarly empirical research on this topic in order to inform healthcare

management in charge of digital record keeping of effective practices regarding records' management and error prevention pertaining to EHRs.

### **Typologies of Data Quality Errors**

Experian (2013) categorized data errors by developing a list of typologies. It conducted a corporate study that identified duplicate entries as one of the major data quality errors (p. 18). Hedt-Gauthier et al. (2012) identified duplicate entries and missing information as data errors or flaws. McCoy, Wright, Kahn, Shapiro, Bernstam & Sittig (2013) used a quantitative method with two hospitals and five locations in their search to find an accurate number of duplicate records. Their study findings demonstrated a variation in the percentage rates of input data errors, ranging from 16.49% to 40.66%, with something as simple as the first and last name. When more variables were added, errors decreased. Managing these duplications revealed differences in the process as well. The purpose of this study was to identify duplicate records and find solutions to manage them.

McCoy et al.'s (2013) addressed areas of concern that can be found in healthcare records systems. The key finding in this article was that assessing practices for managing duplicate information, managers could help to prevent duplications. McCoy et al also addressed the importance of misspelled data as another deficiency of data quality. Researchers reported only duplicate patient information yet failed to make a distinction between obvious duplications and information that had the potential to become a duplicate. This distinction was not based upon information that was used, which was the demographic information of participants.

Berndt, McCart, Finch & Luther (2015) examined the data quality of all related documents from an EHR system. The research used a case study method, which included four hospitals with a dataset of 26,010 clinical progress notes. Researchers found that error rates contained in the documents remained at the target rate of 30%, with a small sample of 1,000. As the sample size grew, the error rate rose to 40%. The study was conducted to test the text mining approach that is being used with free form text documents to see how well it performs. The results of this study indicate weaknesses in the data-mining approach. Simply identifying rates of errors by percentages fail to yield meaningful information on patient outcomes. Results also revealed that a small sample size is not a good indicator of data mining as a solution to data errors. Berndt, McCart, Finch & Luther (2015) sought to find what common practices could be derived from text mining. Their discovery of the number of errors that can be found in sample sizes contributes to my question about a data quality threshold that an organization uses and why. Because of this study, the authors “began to establish some guidelines when choosing sample sizes in the face of target quality issues. These findings have particular significance with regard to annotation efforts” (pp. 1,14).

Berndt’s et al (2015) study determined where the errors would start to appear in the data. Among issues uncovered, misspelled words appeared with regular frequency. This indicated a decline in quality but not the data quality characteristics that the authors claimed had been addressed. A specific acknowledgment of data quality characteristics May have highlighted those areas. Altwaijry, Kalashnikov & Mehrotra (2013) identified missing values as one of the common data quality problem types. Their study consisted

of a quantitative analysis to show that a query-driven approach to data cleaning was more effective than other commonly accepted practices or approaches. Kuehl et al. (2015) also identified missing values as one of the most important data quality problems.

### **Tools to Address Poor Data Quality**

De Almeida et al. (2013) described data as a critical part of any organization (p. 19). Data is a fundamental element of most organizations, whether small or large, that keeps business moving in a forward direction. Data may come in many forms and its importance will increase depending on the nature of a particular business. For example, in healthcare, electronic patient data is highly important as it promises to streamline general office practices such as setting up patient records, ordering tests, medications, and treatments, among other tasks. Gunawan (2016) asserted that data within an EHR must be protected and private according to HIPAA regulations (p. 7). Despite the importance that the U.S. healthcare industry places upon the need for quality, poor and insufficient data continues to be an obstacle for the delivery of superior healthcare.

Big Data Analytics was introduced as an early intervention tool to solve a range of data insufficiencies in the business sector. Its application to data problems in the healthcare industry provided an opportunity to address insufficiencies attributed to processing health information in the form of raw data. Raw data are referred to as the *lifeblood* of healthcare organizations, a collectively essential component of daily, data-driven, decision-making processes across multiple institutional contexts (Foshay & Kuziemy, 2014). The quality and integrity of healthcare data are essential not only to the overall provision of superior healthcare, but also to organizational effectiveness in the

delivery of care and services. Researchers in more than a few studies explored ideas on the identification and management of data quality issues. Some researchers (Adler-Milstein & Jha, 2013; Schultz, 2013; Foshay, & Kuziemy, 2014; Chen, Chiang, & Storey, 2012) attempted to remedy some data quality issues with BI tools in the form of big data. A few researchers sought to mitigate data quality issues with an EHR implementation and to meet government requirements in order to receive incentives (Charles, Gabriel, & Furukawa, 2014).

Superior data quality and integrity remains an important goal within the U.S. healthcare system. As most healthcare organizations transition to EHRs, data quality management has become a critical area of focus (Jamoom et al., 2014). The many challenges regarding data quality include potential inaccuracies and fragmentation (Adler-Milstein & Jha, 2013). Holzinger, Dehmer, & Jurisica (2014) conducted an overview of data quality errors that currently exist and offered suggestions for handling them now and in the future. Rather than list errors and solutions, Holzinger et al addressed what frameworks should be considered as most useful when determining how to solve data and management errors. Holzinger, Dehmer, & Jurisica (2014) stated, “We need machine intelligence to deal with the flood of data, but at the same time we must acknowledge that humans possess certain problem solving and cognition abilities far beyond computation” (p. 7). They showed areas of data workflow process managers find most challenging regarding data quality deficiencies and their solutions. In the article, it was demonstrated that universal workflow guidelines are considered and/or used as solutions for data quality problems.

The application of Big Data analytics as a solution to poor data quality was followed by the introduction of several other textual analysis tools that looked promising. QNOTE was one such tool. Burke et al. (2014) conducted a quantitative study to determine if a tool called QNOTE (according to the National Institutes of Health QNOTE is not an actual acronym, but rather a valid name for a tool that evaluates EHRs) was an accurate gauge to measure the quality of data in clinical notes. Burke's (2014) study used 300 clinical notes from 100 outpatients and applied QNOTE, which had 12 variables, to score the quality of clinical notes. Researchers found that QNOTE excelled in measuring the quality of data within the study's clinical notes. This article (Burke et al., 2014) is relevant to my study because it supports the lack of empirical studies on data quality tools and solutions.

Chen, Chiang & Storey (2012) used a quantitative study to find challenges to using BI and analytics. The purpose of the study was to introduce "next steps" and inform the academic community of the need for research on BI and analytics. The study commenced with 6,187 papers and ended with 3,602 after removing IT publications and journal articles retrieved from databases such as *Science Direct* and *PubMed*.

Researchers discovered that even with a substantial increase in the use of EHRs and the introduction of big data analytics, a lack of scholarly research remains on data analytics.

Cusack, Hripesak, Bloomrosen, Rosenbloom, Weaver, Wright, & Mamykina (2013) examined how redundant data could be considered a data quality error. Cusack et al., 2013 conducted a study during a professional meeting that engaged 100 participants in

breakout sessions to create action items for discussion. These breakout sessions resulted in multiple findings and recommendations.

The purpose of data capture and documentation, in conjunction with business areas within healthcare organizations to be supported, was acknowledged as important to understanding data workflow practice (Cusack, et al., 2013). Cusack, et al (2013) also found that benchmarks are necessary to evaluate data capture and documentation processes. They concluded that data capture and documentation should not result in diminished workflow activities and delivery of care but should only support the processes that take place daily. These findings affect judgments of end-users (management) and their abilities to render quality patient care based on EHR data. Study benchmarking and assessing information for data quality and data quality management, highlights what is currently being discussed as data quality problems within the healthcare industry. They also addressed decision-making as an outcome of captured data within an EHR system. One key finding was that different areas within a healthcare organization have different workflow patterns and daily activities to consider when using an EHR system, given that patient safety is the overarching goal of a record keeping system. To that end, a core connection from the article was the relationship of electronic data capture and documentation processes and practices to patient safety and quality of care.

Quality of care related to healthcare data has implications beyond data input, patient and physician portals, and record keeping. Healthcare providers routinely access digital records for diagnostic purposes, specialists' referrals, and changes to patient prescriptions, and other reasons. Records access by healthcare providers anticipates

action for the benefit of patients and constitutes what is commonly called *secondary use* of patient information (Church, 2002). Kwon, Lee & Shin (2014) used a survey driven quantitative study to discover if the advantages of data quality use and management has been a considered factor for adopting big data analytic tools. Researchers applied the research-based view theory to a survey that was disseminated to 969 organizations (306 respondents). They found that organizations that were ready to make well-planned decisions with their secondary data were more apt to invest in resources like big data analytics. This article supports my study because of its focus on the secondary data that were used to make decisions without needed tools for managing data quality.

Taggart, J., Liaw, S., Dennis, S., Yu, H., Rahimi, A., Jalaludin, B., & Harris, M. (2012) extracted patient data from four practices to complete their quantitative study. The purpose of that study was to identify data quality problems within an electronic healthcare system. Using a registry from which they had extracted data, researchers found that data regarding gender and date of birth were substantially incomplete and even that high levels of inconsistencies and incomplete records appeared to be routine occurrences. Further, it suggested the possibility of establishing best practice guidelines to address these obstacles to good data. They also examined the perspective of data capture and the use of the information within the system as secondary use.

### **Effects of Poor Data Quality**

Healthcare organizations are negatively affected by poor data quality. Deciphering the impact of poor data quality within the field of healthcare suffers from a lack of agreement on the part of academics (researchers) as to which dimension of

healthcare is most adversely affected. Foot et al. (2014) emphasized that community services were affected by the poor quality of healthcare information (p.2), while Debbarma, Nath & Das (2013) found that the greatest negative impact to superior healthcare delivery was poor decision making resulting from bad data. Judah & Freidman (2014) agreed that poor decision-making was a factor and extended the discussion of impact to include organizational costs.

### **Summary and Conclusions**

Chapter 2 contained a review of literature related to an analysis of information theory and DQAF. Based on the literature review, there appears to be a breakdown in the process that occurs with data quality. Major themes in the literature base were the evolution of data quality from quality and improvement management, data quality dimensions and the methods to assess data quality, a data quality assessment framework to measure common or routine data-related practices in use by healthcare organizations, data information, Blum Knowledge model used in health informatics.

## Chapter 3: Research Method

### **Introduction**

The purpose of this Delphi study was to identify best practices in data quality and integrity in healthcare. This study took place over a period of 4 months and was conducted with three rounds of questions for a panel of IT and non-IT data experts. These panel of experts were chosen based upon their ability to provide information based on their current connection to working with and managing healthcare data. A sample of 25 IT and non-IT data managers who met the inclusion criteria identified in this chapter was used. Those data managers were system managers responsible for oversight of the manipulation and maintenance of healthcare. The experts were recruited from sources such as LinkedIn and Trade associations.

Research protocols for this study included a structured open-ended questionnaire in the first round, followed by two rounds of mixed structure questions using dichotomous, Likert, and scaling ranking strategies. The results from the first round drove the questions for the second round. After analyzing the results from the first round, the second round of questions were created. The third round of questions were similarly generated. The process of data collection and healthcare data management best practices was the focus of the data quality evaluation.

This chapter on the research design consists of six major sections: (a) Introduction, (b) Research Design and Rationale, (c) Role of Researcher, (d) Methodology, (e) Issues of Trustworthiness, and (f) summary. In Section 1, I provide a synopsis of the research study and its purpose. Section 2 addresses how the research

design and methodology applied to the research question. In Section 3, I provide an idea of the role I played within the study as well as an overview of my prior experiences within the field of healthcare practice. In Section 4, I will outline the methodology— participants, instruments, pilot study and recruitment. In Section 5, I examined issues of trustworthiness such as ethical procedures and credibility. In Section 6, I provide a summary.

### **Research Design and Rationale**

This study was guided by one essential question:

Research Question: What practices contribute to identifying improvements in EHR data collection, secondary use of data, data quality, and integrity?

### **Role of the Researcher**

My role as the researcher was to distribute and collect surveys from the participants and to analyze collected surveys. In the Delphi study, I developed the subsequent round questions based on the responses of the previous round (see A. von der Gracht & Darkow, 2013). A. von der Gracht and Darkow (2013) stated rules that a researcher should follow when conducting a Delphi study: eliminate obscure questions and use correct industry terminology and clearly stated guidelines.

Throughout the research process, I performed all actions necessary to acquire the data that were used to formulate each round of questions. I had no personal connection to any of the panelists selected for this study. Potential researcher biases and/or power relationships were managed through the anonymity of all study participants. Once selected, participants were given identification numbers, and all their responses were kept

confidential. Other potentially ethical concerns addressed would be the offer of incentives for participation and/or conflict of interest among participants.

### **Method**

A Delphi approach was the most appropriate methodology for this study. Ponto (2015) discussed the use of open-ended questions to allow respondents to provide a distinct response to a question. Respondents have the freedom to state their most honest opinion on a topic based on experience, which will produce exploratory data that may reveal unpredicted opportunities and/or issues.

Open-ended questions generate information that is useful in suggesting ways to format questions and select study participants as might prove appropriate (Ponto, 2015). Qualitative research surveys have been used to gain a more thorough understanding of a person's underlying thinking and incentives of a topic, issue, or problem (Ponto, 2015). The focus of this study was on the issues that precipitate data quality errors and a consensus of best practices used to manage them as seen through the eyes of the IT data stewards within the healthcare industry.

The Delphi method was developed in 1944 and later refined in 1950s by the RAND Corporation. The method was created as a technological forecasting tool. It is applied by gathering a group of unidentified experts to complete surveys through a series of three or more rounds (Sekayi & Kennedy, 2017). The objective is to create a consensus in thinking among the panelists (Sekayi & Kennedy, 2017). This study followed the accepted Delphi method practice and protocol. McMillan, King, & Tully (2016) asserted that data collected from a survey in a Delphi method are strengthened by

the ability of the researcher to assist in creating robust guidelines; although conversely, Brady (2015) found that using this method yielded inconsistent data from surveys due to how an expert is determined when creating a sample, questions on bias, and the lack of perceived correct answers or opinions acquired from the consensus. Despite controversial perspectives regarding the value of Delphi method results, findings may reveal information about participants that could be overlooked in a quantitative method.

### **Participant Selection**

The multiperspective population consisted of 25 different healthcare systems data management professionals, also known as data stewards, across the United States who have worked or are working for a healthcare system such as an EHR and have managed the data within this system over a period of 5 years. Participants were recruited from sources such as LinkedIn, trade associations, and networking. The panelists were chosen using a purposive sampling technique. The criteria included professional level experience with healthcare data and IT or non-IT position with data quality. I preferred that the panelist was currently working with data.

The rationale for using a purposeful sample in a qualitative Delphi method study was to obtain results that provided a general picture of the research problem, such as what data quality issues have been identified and how they are managed in their respective organizations. Potential research panelists were contacted and provided with a complete outline of the study. Members of the sample population were selected based upon their current position and knowledge of data process and best practices currently being used. All participants received a consent to participate form.

During the survey, participants were asked a series of open-ended questions and a mix of dichotomous, Likert, and scaling questions that allowed me to obtain data about their assumptions of quality issues and current data handling practices. This form of questioning produced reliable comparative data. Open-ended questions have been shown to be the most appropriate form of questioning when encountering complex or convoluted answers (Fowler, 1995). This line of questioning is also advantageous in measuring the knowledge of participants (Fowler, 1995). All questionnaires were collected, analyzed, and augmented through note taking. The resulting information was reviewed for themes or patterns in order to produce raw data for analysis.

### Instrumentation

The participants were asked to answer four open-ended questions (See Appendix A) in the first round of the study. Surveys were administered via Survey Monkey. Figure 4 shows the data collection plan.

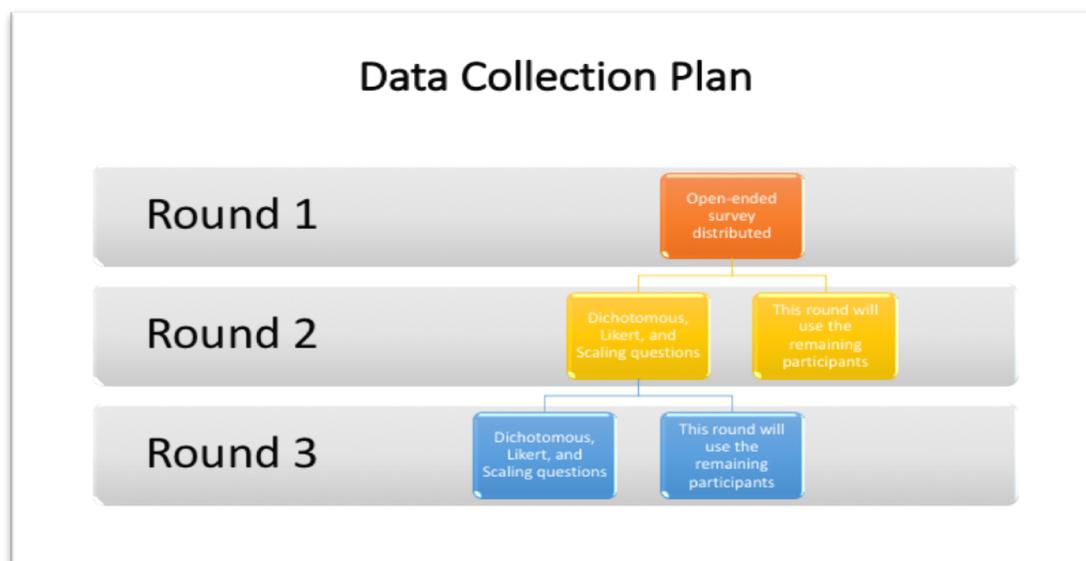


Figure 4. Data collection plan.

Survey Monkey was used to distribute and collect the survey responses. I took notes during my analysis and coding of the data using Microsoft Excel and OneNote document. Collected surveys were compared and cross-checked. A unique survey number and code was used to identify each participant and survey. One unforeseen challenge that would have required a follow-up request might be responses that are not clear or responses that are after thoughts. Another unforeseen challenge would have been participants not wanting to answer certain questions. I noted and described the missing responses in my analysis.

### **Field Test**

Field test are integral components of research studies. Clow and James (2014) asserted that a field test can produce results that can assist the researcher in finding ways to strengthen their measurement tool for gathering necessary data. A field test of the open-ended questions to strengthen the responses received from participants took place with three healthcare IT managers not included in the overall study.

A field test was conducted with the proposed Round 1 questions to test their validity. Field test of interview/Round 1 questions are a common practice. Three healthcare IT professionals evaluated the questions for efficacy. These experts were not included in the study. Initial drafts of the questions were revised based on the expert's input.

### **Data Analysis Plan**

Each participant was assigned an identifier code to protect their anonymity. A Survey Monkey link including the first round of open-ended questions was emailed to the selected panelists. This survey included the directions, the due date, and how to handle any questions the panelists might have incurred.

Round 1 responses were entered into an Excel spreadsheet and Microsoft OneNote document to look for trends and patterns. Once a significant number of themes were found, the Round 2 questions were created. These questions were sent to the panelists through Survey Monkey with directions. These questions did not include a space for additional information. The responses were collected. In addition to trend spotting, this round of questions consisted of a Likert scale, and a simple frequency distribution statistical method was used to analyze the responses (see Sullivan & Artino, 2013). A third round of questions was created and distributed in the same manner as the first two rounds. The final analysis consisted of a feasibility scale to determine the consensus among the group (see A. von der Gracht & Darkow, 2013).

### **Trustworthiness**

In qualitative research, issues of trustworthiness include credibility, transferability, dependability, confirmation, and ethical procedures. These areas are fundamental to the concept of validity, as they help to ensure that areas of conflict or inconsistency may be eliminated or at the least minimized. Triangulation was used to assure the validity of information. I also keep detailed notes on all steps of my research process.

**Credibility**

Credibility refers to the extent that the results are valid and reliable and confirmable. Several approaches were used such as data saturation (Padgett, 2012), triangulation and transparency using field notes. Triangulation was accomplished using methodological triangulation, which includes the use of document, such as the literature review, and field notes. Triangulation is useful in minimizing procedural and sampling bias (Rowe, Frantz, & Bozalek, 2013).

**Dependability**

The process of the study was reported in detail to allow for the attempt at recreating the same results. This will also allow the future researchers to assess the validity of the methods used in this study. These actions will also establish trustworthiness.

**Confirmation**

Although all data will be analyzed objectively, other researchers would not be able to arrive at the same results using an equivalent methodology. I presented the data in the same order in which they were collected and analyzed. I documented the details of the process of data collection, data analysis, and interpretation of the data. I also documented what topics were found to be distinctive and interesting during the data collection, as well as any specific thoughts and/or rationale about coding.

**Ethical Procedures**

Ethical concerns relating to non-identification of persons contained in surveys were managed by assigning codes to participants. An informed consent letter containing

the IRB approval number (10-12-18-0070966) was sent to each participant. Appended also to the request forms were the documents to allow participant refusal of participation or early withdrawal from the study. Conflicts of interest, as well as relationship boundaries, were addressed in the consent form for each participant. All participants were allowed to withdraw from the study at any time. Participation were based upon a genuine desire of panelists selected to be a part of the healthcare industry's electronic data improvement plan.

Research documents were secured on my password-protected computer. The documents will be destroyed in three years. Data collected by Survey Monkey will be stored on their servers. Please see Appendix B for an outline of SurveyMonkey information for the IRB.

### **Preserving Participant Identity**

The participants were assigned a code by SurveyMonkey to protect their identity, which maintains robust data security with restricted research account policies. Only the researcher will know who participated and any connecting background information. The participant were referred to by their assigned code throughout the study and on any reports generated form the student. During analysis the participants code was used to identify their responses. Anonymity remained throughout the study.

### **Summary**

This chapter has described the proposed research methodology to be used to collect and analyze data obtained from participants in order to examine and evaluate routine practices for EHR data quality and integrity management in healthcare

organizations. A qualitative approach for this study will entail a field test of the first round of questions with due refinements to help ensure usefulness and validity of the final study questions. A data analysis plan was developed to ensure a workable timeline for data to be collected and analyzed. Chapter four will disclose the findings from the study, with conclusions and recommendations for further research addressed in chapter five.

## Chapter 4: Results

The purpose of this qualitative Delphi study was to identify best practices for data quality and data management in healthcare systems. I further identified the best practices to include in guidelines for data quality and data management. This study was guided by one essential question: What practices contribute to identifying improvements in EHR data collection, secondary use of data, data quality, and integrity? In this chapter, I present the results of the study in seven sections.

### **Field Test**

The field test was conducted using three participants who were not included as a part of the study. The participants were presented with the questions I planned to use in the first round of my Delphi study and had 5 days to respond. The expected scope of participation for the field test was explained to the participants, and prior agreement to participate was acquired. The purpose of the study was explained as well as my expectations from the field test. The following questions were provided for the field test participants to answer as part of their review:

- Are these questions written clearly?
- Would you feel comfortable answering the questions?
- Is the time reference clear to the respondents?
- Would any of the questions require the respondent to think too long or hard before responding? If so, which ones?
- Do any of the questions generate response bias? If so, which ones?
- Is the survey too long?

- Have any other important issues been overlooked?

The first round of open-ended questions were revised to address responses such as specificity, assumptions of poor data quality, the use of techniques vs. strategies, removal of certain words such as *some*, and clarifying terms used such as *fitness of data*.

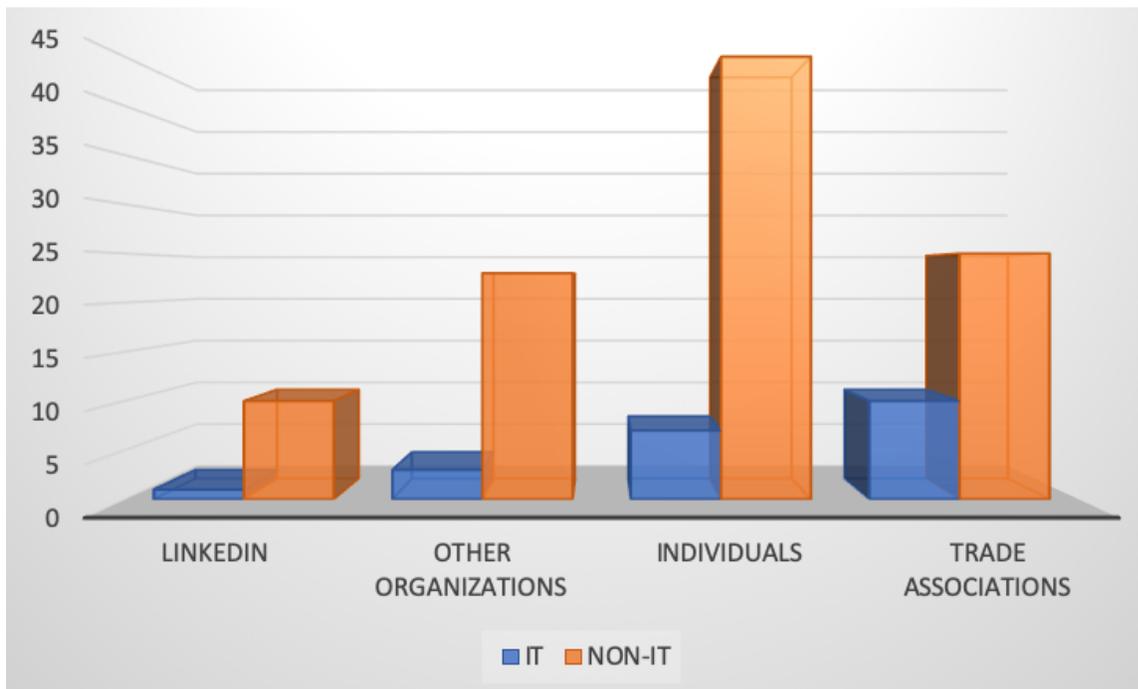
### **Setting**

The study was administered in the United States with participants in IT and non-IT data positions and considered experts in their field. All communication with participants was by email, such as survey distribution, reminders, follow up reminders, and thank yous. Agreement and confirmation of participation were sent and received between 10/15/18 and 1/16/19 to a combination of 90 organizations and individuals of which 24 agreed to participate by completing an informed consent document in reply to my invitation email. Not all 24 participants completed each of the three rounds of data collection. There were no personal or organizational conditions that influenced participants or their experiences at the time of the study that would influence the interpretation of the results because all participants completed the survey at their leisure where they felt they were in a comfortable environment and where they could concentrate and take their time. This setting was highly encouraged in the survey invite.

### **Demographics**

Participants involved in the study were selected based on professional level experience with healthcare data and IT or non-IT position with data quality. I preferred that the panelist was currently working with data within the healthcare setting. The virtual panel of IT and non-IT data experts was chosen based upon their ability to provide

information based on their current connection to working with and managing healthcare data. The data managers were system managers responsible for the oversight of the manipulation and maintenance of healthcare data. The experts were mostly recruited from word of mouth colleagues found on LinkedIn. Trade associations and other organizations as well as individuals were contacted. Figure 5 shows the breakdown of the demographics for recruiting participants. There is a bar to show the IT and the non-IT potential participants who were contacted from the four different categorical groups.



*Figure 5.* Breakdown of demographics for recruiting participants.

### **Data Collection**

This study consisted of three rounds of survey questions distributed to a panel of IT and non-IT healthcare data experts who met selection criteria identified in Chapter 3.

Initially, I sought to include 25 participants, and the first round of questions was distributed to 24 participants, of which 17 responded. Round 2 yielded 14 responses of the 24 surveys sent. The same number of surveys were completed in Round 3, which was 14 out of 24. The survey was monitored daily once it was sent out by installing the SurveyMonkey application on my mobile phone and tablet and turning on notifications. I also had my inbox notifications on so that I would receive email alerts whenever a participant had a question or completed a survey.

### **Round Details**

Round 1 included four open-ended questions distributed via SurveyMonkey to 24 participants. One email invitation collector was created to capture which participants completed the survey (17), opened the survey (22), and had not opened the survey (2). There were no partial surveys included. In Round 2, a matrix of dropdown menus question type was used that included four pages of general questions and a list of choices ranging from *strongly disagree* to *strongly agree* for the participant to choose from. Each page contained 18 to 22 statements to be ranked. This survey was distributed via SurveyMonkey to 23 participants. One email invitation collector was created to capture which participants completed the survey (14), opened the survey (18), and had not opened the survey (5). There were no partial surveys included. In Round 3, a ranking question type was used that included seven pages with general directions. There was one partial survey included. The participants were given a 2-week deadline date explained at the beginning of the survey and the survey email. A reminder email was scheduled to go out 7 days after the start of the survey clock. A day before the survey was scheduled to

close, I went in to see which participant had not completed the survey and sent him or her second reminder. This sparked a few participants to look in their spam or request the survey be resent as they could not find the original survey. The IRB received and approved all questions for each round before the survey was distributed.

## **Data Analysis**

### **Round 1**

In Round 1, the participants were asked to answer four open-ended questions. A list of words that emerged as themes throughout the responses was derived by using the SurveyMonkey word cloud feature, which allowed for a search through all the responses to create the word cloud based on the words found most often in each response for each question. Table 4 illustrates the questions, words found, and the percentages. Figures 6 to 9 are the four questions asked in the survey and the resulting word cloud. Table 4 includes a breakdown of the Round 1 survey results and the percentage for the words found the most from the results.



Round 2

To extract best practices that were most representative of the data, statements with the highest ranking in *strongly agree* at 50% or higher was used. Table 5 shows the statements, the percentages, and the best practices derived from the statements for Round 3.

Table 3. Round 2 Survey Results

Satatement	Strongly agree %	New Eeement
1.In most organizations, data quality is not defined.	50.00%	Data quality definitions
2.There is a lack of data governance practice implementations and data management tools.	57.14%	Data governance
3.Communication is key among systems, staff.	71.43%	Data management tools
4.There is a lack of training among the staff on how the data will be used.	57.14%	Usage Training
5.Reduce manual processing of data to reduce errors and increase timeliness of output for better decision making.	57.14%	Automated processes Timeliness
6. Key stakeholders such as clinicians, patients, and administrators are impacted by poor data quality.	92.86%	Stakeholder transparency
7. Analytics and workflows are impacted by poor data quality.	71.43%	Stronger Analytics Stronger Workflows
8. Quality of data aligns with quality of care for patients.	50.00%	Quality alignment
9. Patient outcome are impacted by poor data quality.	50.00%	Improved outcomes
10. Organizations are impacted in the area of decision making regarding poor data quality	50.00%	Stronger decision making
11. Untrustworthy data leads to mistakes.	64.29%	Trustworthy data
12. Patient billing are impacted by poor data quality.	57.14%	Accuracy
13. Patient costs are impacted by poor data quality.	64.29%	Integrity
14. Organizations are impacted in the area of ROI regarding poor data quality	57.14%	ROI
15. Organizations are impacted in the area of planning regarding poor data quality	57.14%	Consistency
16. Organizations are impacted in the area of forecasting regarding poor data quality	64.29%	Forecasting
17. A solid data quality plan will ensure communication and an understanding of the business process among groups that handle the data.	50.00%	Clear process
18. Roles, responsibilities, and ownership can be established with a data management plan.	50.00%	Roles, responsibilities, and ownership
19. A data management plan can help to create standards set for data.	71.43%	Set standards
20. A data management plan can help to eliminate redundancy.	57.14%	Eliminate redundancy
21. A solid data management plan should include best practices for fir	69.23%	Find root causes
22. A solid data management plan should include best practices for is:	53.85%	Issue Management
23. A solid data management plan should include best practices for id	69.23%	Identify errors
24. A solid data management plan should include best practices for pr	61.54%	Improvement plan
25. A solid data management plan should include best practices for a	76.92%	Data steward
26. Best practices to eliminate data quality errors and issues in the fut	85.71%	Corrections
27. Best practices to eliminate data quality errors and issues in the fut	57.14%	Validation
28. Best practices to eliminate data quality errors and issues in the fut	50.00%	Maintenance Metadata
29. Best practices to eliminate data quality errors and issues in the fut	64.29%	Audits Best practice

There were no statements in which the participants unanimously strongly disagreed. There were statements in which the participants ranked high in slightly disagree and neutral. They were as follows:

- Best practices to eliminate data quality errors and issues in the future includes implementation of lean Six Sigma continuous improvement, 35.71% Neutral
- Information generated from the poor data is often considered "unusable" because of fragmentation, 35.71% Neutral
- Healthcare data quality is missing regulatory input, 42.86% Neutral
- Data sources should be limited for consistency and to avoid duplications, 21.43%

**Slightly and strongly disagree.** There were 29 statements used to transform into 33 best practices. Some statements formed two best practices. Any statement that resulted in a duplicate best practice was not included. Two participants emailed after they had completed their survey indicating they wished this round was also open-ended as they had more to say. For example, participant DQS012 stated:

I would have loved to engage qualitatively around why I answered some of these questions the way I did – some of the best practices you touch on is very valid for a best practice, but some of it is also very much linked to the maturity of the system in question.

### **Round 3**

In round 3, the participants were asked to rank the five best practices provided in each question on a scale of 1-5, with 5 being the highest. Some participants indicated that it was difficult to rank because there were so many, they felt should be ranked the

highest. The themes found to rank closest to 5 were: usage training, timeliness, clear processes; eliminate redundancy, corrections, and audit. There was a tie between quality alignment and stronger decision making in question 4 (3.71) in which the lowest standard deviation was used (1.28 and 1.03) to break the tie. The themes found to rank closest to 1 were: data quality definitions, automated processes, trustworthy data, integrity, Roles, responsibilities, and ownership, validation, and metadata. Below you will find figures 7-13 showing each question along with the weighted average, and standard deviation for each ranked best practice.

Answered: 14 Skipped: 0

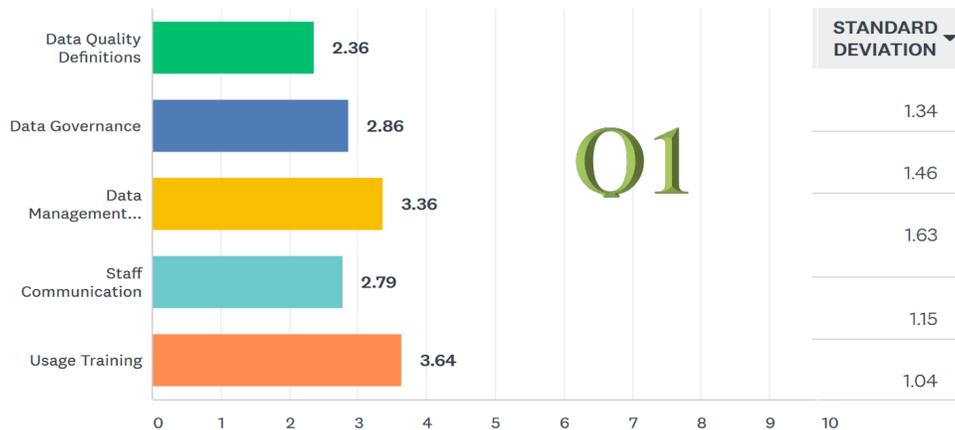


Figure 7. Round 3 survey results for Q1.

Answered: 14 Skipped: 0

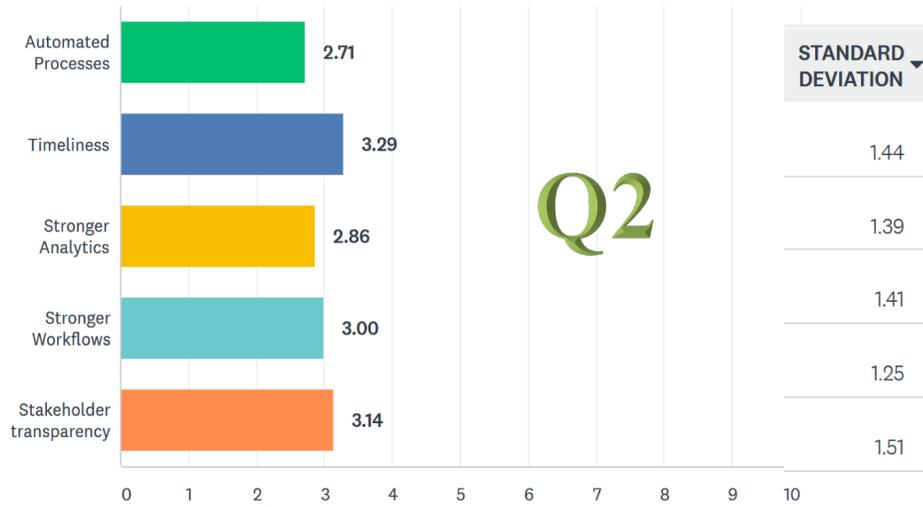


Figure 8. Round 3 survey results for Q2.

Answered: 14 Skipped: 0

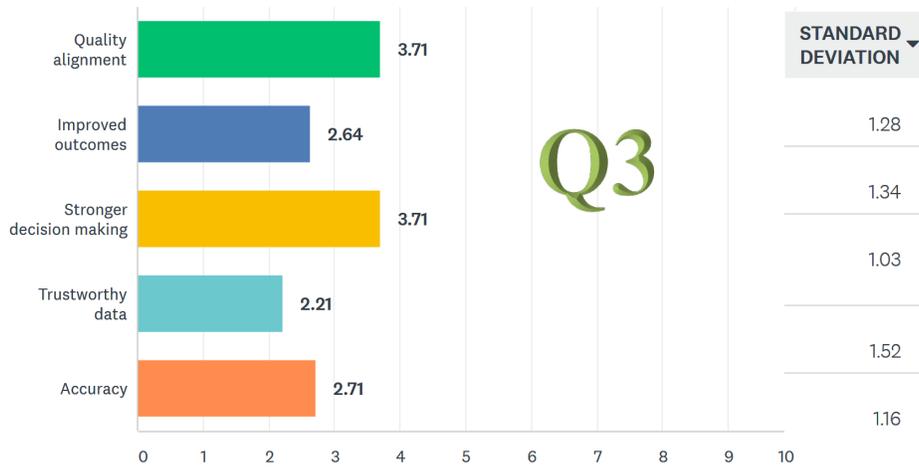


Figure 9. Round 3 survey results for Q3.

Answered: 14 Skipped: 0

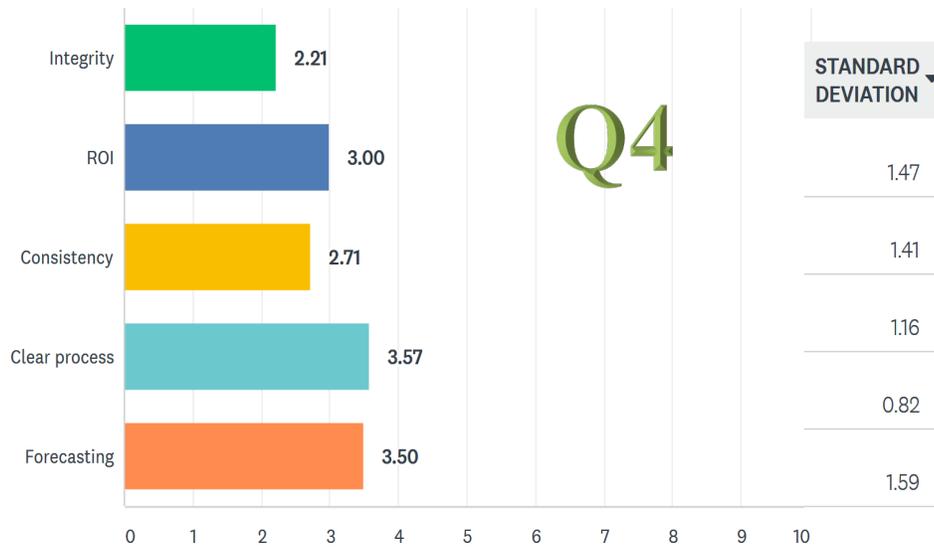


Figure 10. Round 3 survey results for Q4.

Answered: 13 Skipped: 1

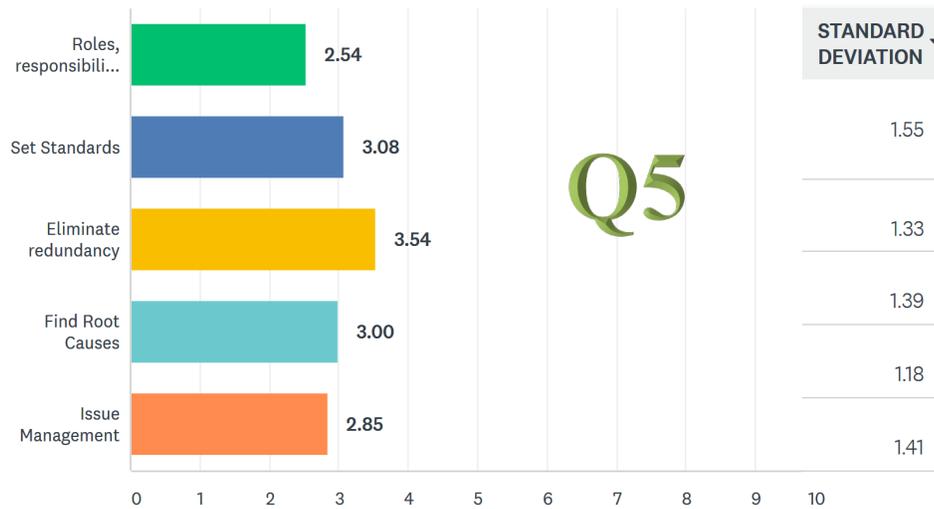


Figure 11. Round 3 survey results for Q5.

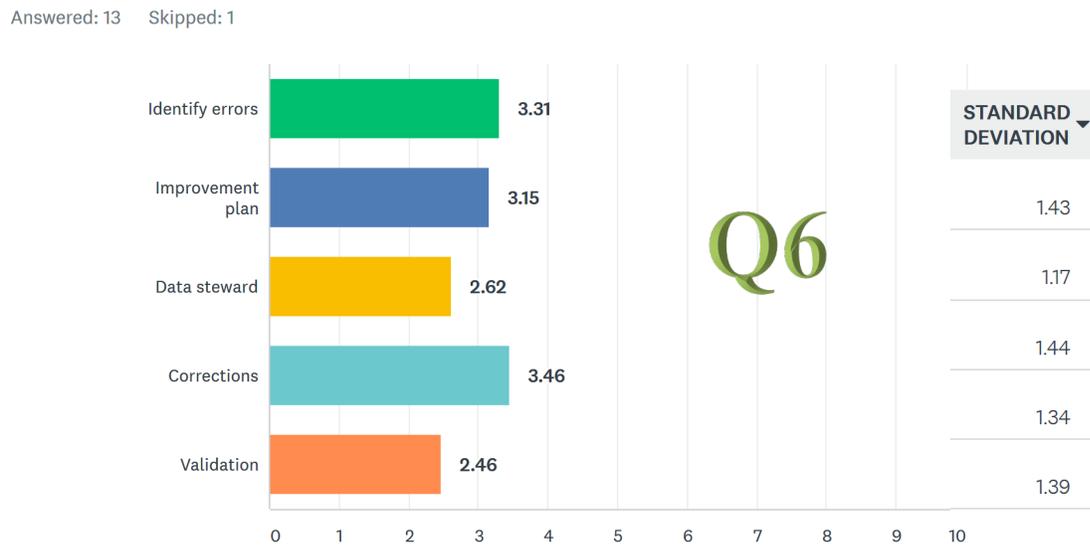


Figure 12. Round 3 survey results for Q6.

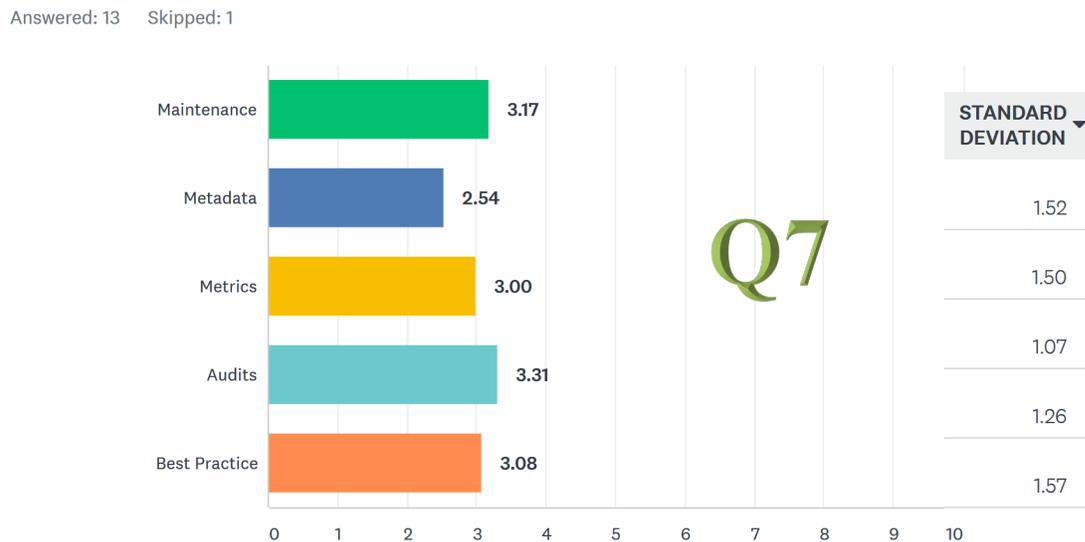


Figure 13. Round 3 survey results for Q7. Where is the chart

### **Evidence of Trustworthiness**

Issues of trustworthiness were addressed by adhering to credibility, transferability, dependability, confirmation, and ethical procedures. These areas were important to ensure validity of the data collected and to eliminate chances of conflict or inconsistency. Triangulation was used to also assure the validity of information by using open-ended questions, Likert scale ratings of statements, and best practice rankings. Detailed notes were kept on all steps of my research process.

#### **Credibility**

Credibility was achieved by identifying data saturation in each round. When duplicate information started to emerge, the search for themes was ended. In addition, triangulation was used by implementing different survey types. Participants were from different parts of the country and held different IT/Data positions. Each survey type yielded the same or similar conclusions about the data. Transparency was achieved by transferring my field notes to the study. Participants were from different parts of the country and held different IT/Data positions.

#### **Dependability**

The process of the study was reported in detail to allow for the attempt at recreating the same results. This will also allow future researchers to assess the validity of the methods used in this study. These actions will also establish trustworthiness.

**Confirmation**

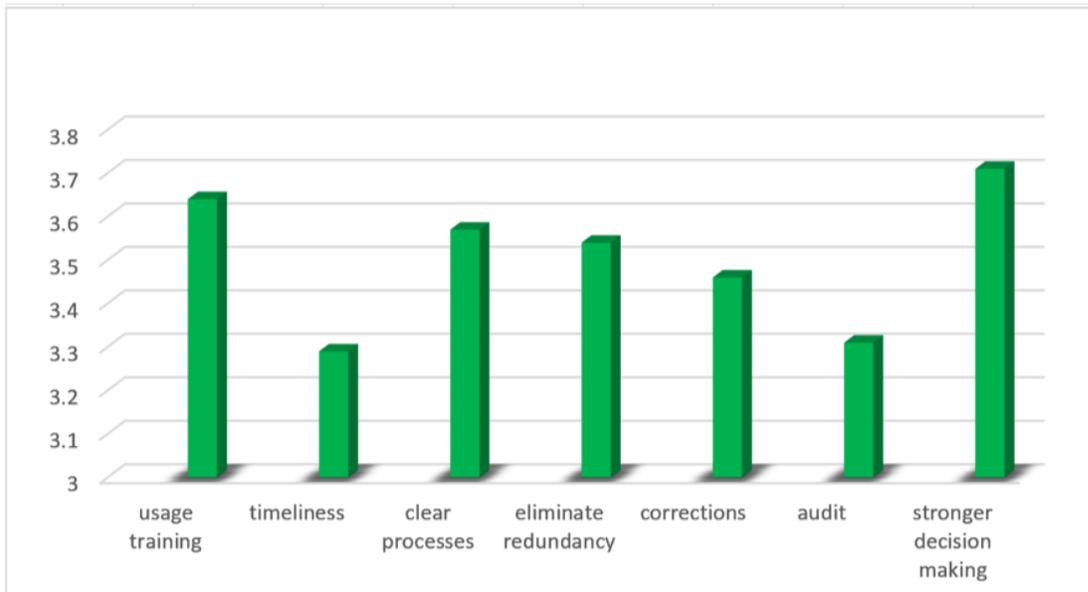
All data collected is documented along with the details of the process of data collection, data analysis, and interpretation of the data. Topics that were found to be distinctive and interesting during the data collection, as well as any specific thoughts and/or rationale about coding were documented.

**Results**

Using the Data Quality Assessment Framework (DQAF), there were some patterns identified of best practices that should be included to maintain or create best practices for data quality.

**Significance**

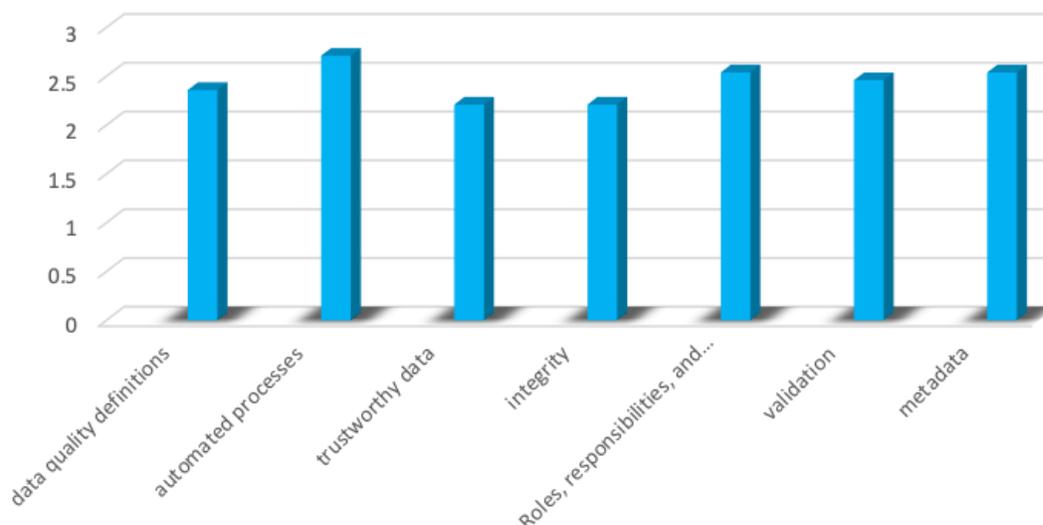
Five best practices ranked highest throughout all the rounds. They were usage training, timeliness, clear processes, eliminate redundancy, corrections, and audit. There was a tie between quality alignment and stronger decision making in question 4 (3.71) in which the lowest standard deviation was used (1.28 and 1.03) to break the tie. A bar chart of the best practices that ranked highest in round three can be found in Figure 14.



*Figure 14.* Best practices from Round 3 that ranked high.

### **Best Practices That Ranked Low**

The themes found to rank closest to 1 were: data quality definitions, automated processes, trustworthy data, integrity, Roles, responsibilities, and ownership, validation, and metadata. A bar chart of the best practices that ranked lowest in round three can be found in Figure 15.



*Figure 15.* Best practices from round 3 that ranked low.

### **Comparison of the Best Practices**

A view of the best practices side by side to compare how they ranked visually was necessary to understand how much of a difference the best practices were from one to the other. For example, although most participants discussed data quality definitions quite a bit in the beginning, in the end, usage training was found to be more important to best practices. Participants scored timeliness as a more important best practice against automated processes. There was a need to have clear processes over trustworthy data. The two rankings that were the most surprising was the elimination of redundancy over integrity of the data and audits over validation. Validation was also a common theme in the beginning. The need for corrections and stronger decision making over

roles/responsibilities and metadata was not surprising. The comparison of the best practices can be found in Figure 16.

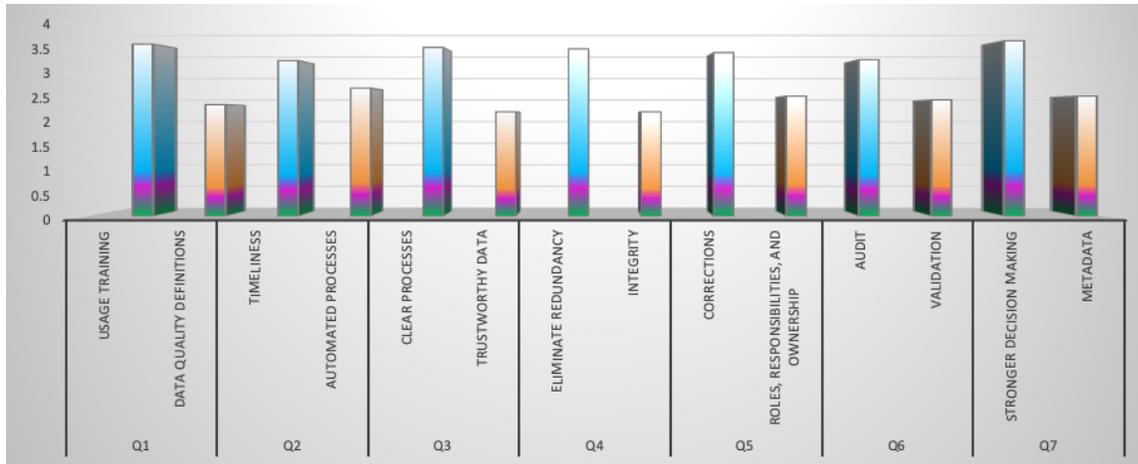


Figure 16. Best practices compared.

### Summary

In chapter 4, the following was discussed: field test, setting, demographics, data collection, data analysis, and results for all three rounds of my Delphi study. Steps taken to ensure evidence of trustworthiness was explained. Best practices that ranked high, low and a comparison of the resulting final best practices was presented and explained. Charts and diagrams were included for a visual representation of the data collected. In chapter 5, the interpretation of the findings will be presented as well as the limitations of the study, implications, and recommendations.

## Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this qualitative Delphi study was to identify best practices for data quality and data management in healthcare systems. In this study, the best practices guidelines to include for data quality and data management were further identified. The best practices in healthcare data quality management and a possible connection to the findings in the academic fields of management and technology were revealed. The conceptual framework of Blum and the DQAF were used to guide the structure of the Delphi survey questions and to identify what information to collect.

The DQAF was used to provide assessment measures based upon data quality dimensions, and information theory (Blum model) assisted in categorizing the type of information into each data quality dimension to be assessed. The key findings identified what the known best practices in the data quality industry were, what best practices helped to improve data collection in healthcare systems, what the most important best practices in data quality and integrity were, and what best practices address the secondary use of data.

Chapter 5 includes interpretation of the findings, limitations of the study, recommendations, implications, and a conclusion. The interpretation of findings section is broken down into four subsections to provide more detail.

### **Interpretation of the Findings**

The key findings were grouped into three categories: improvement of data, secondary data criteria, and data quality and integrity. The best practices identified from

the patterns from the data were turned into professional procedures found to be the most effective through this research.

### **Best Practices to Improve Data Collection**

Technology is constantly changing. However, the best practices for data integrity and quality have not changed with the technological advancements. These best practices, according to current experts in the field, in contrast to former best practices of data quality definitions, automated processes, roles, responsibilities, and metadata actually scored very low when assessed for efficacy and effectiveness.

The best practices identified and categorized to improve data collection were user training and clear processes. Bowman (2013) attributed user error as one of the leading causes of data quality issues within an EHR. The panelists agreed with researchers such as Bowman and Weiskopf and Weng (2013) in concluding that training, clear communication, and cross-checking the data as they move through the process mitigate errors and provide useful and accurate data. Although cross-checking did not rank high on the list of best practices, it was discussed earlier in Round 1 as a potential best practice and should be included to catch any errors. Human error will happen; this was identified in all of the studies. While human error is inevitable, mechanisms can be put in place to mitigate these types of data error.

Training, as a best practice, could include such procedures as training focused on data entry, the purpose of the data, how the data should be handled, and the data checking processes. Training on clear processes will help the user understand how the data will flow from one department to another and eventually help in making decisions about

services that can be offered or are necessary for the community. Many end-users who input the data are unaware of the interrelatedness of multiple systems and how the data they input are used. This level of transparency could benefit the entire organization by making the end user accountable for the accuracy and integrity of the data they enter.

Foshay and Kuziemy (2014) posited that improving data quality should start with the inclusion of best practices that contain data analytics and quality measures. In as little as 5 years, experts now consider data analytics and quality measures not as effective in ensuring data quality improvement. They felt that addressing data quality the endpoint was less effective. In contrast to the relative importance of data analytics today, analytics ranked at the bottom of the list of most important best practices by the end of Round 3. Based on the findings, it appears that experts have shifted towards addressing data integrity at the point of entry rather than retrospectively at the end of the data flow.

### **Best Practices That Apply to Secondary Use of Data**

The highest-ranking best practices were corrections, audit, and stronger decision-making matrices in contrast to the accepted best practices focused on trustworthy data and ownership. Corrections, such as more frequent audits, appeared important and a significant action item to improve data quality, particularly for using secondary sources for information. Secondary use of data is common throughout the healthcare industry.

Altwaijry et al. (2013) and Kuehl et al. (2015) advanced the idea that missing values is a very common data quality issue. Holzinger et al. (2014) contended that utilization of machines to deal with the massive amounts of data would be ideal and human observation and audits are critical to ensuring data integrity. This finding suggests

that BI and Artificial Intelligence (AI) may play an increasingly important role in aiding human data monitoring activities. Kwon et al. (2014) suggested that big data analytics would enhance decision-making from secondary data. With the advent of Big Data, data analytics for secondary data is becoming increasingly critical. Missing values could be addressed and corrected during data integrity audits, using statistical methods to assess the potential for standard deviation to identify outliers, patterns, and null values. This approach would then apply business rules to data audits rather than relying solely on human observation. The application of business rules using software that applies business rules and AI algorithms would enable a deeper dive into the data than humans can achieve alone. Complimentary data analytics process with a human interface could more accurately verify the veracity as well as the strength of the data before they are sent for decision making.

Mawilmada, Smith, & Sahaman (2012) stated that healthcare services have suffered because of inaccurate or unusable secondary data. Oracle (2015) reported that inaccurate secondary data are very costly to the organization, often resulting in missed opportunities and less than optimum decision making. While neither Mawilmada, Smith, & Sahama nor Oracle suggested any definitive best practices to ensure the validity of the secondary data, it is clear that best practices taken must be more robust than those in the past. This can only happen by combining multiple audit procedures using AI, business rules, and human assessment.

## **Best Practices That Apply to Data Quality and Integrity**

Dixon et al. (2013) pointed out the best practices they found to be most important for quality and integrity: accuracy, completeness, consistency, and timeliness. The top-ranked best practices are in line with the reported best practices from Dixon et al. This would mean that the best practices that were considered important in the past in this area are currently as important within the industry. The current experts concurred in part, highlighting timeliness and eliminating redundancy. This approach focuses on addressing data quality and integrity proactively at the onset rather than retrospectively. This shift in focus aligns with Jerven's (2014) observations that timeliness as is now a critical dimension of data quality. For example, both Experian (2013) and Hedt-Gauthier et al. (2012) identified duplicate entries as one of the major data quality errors and flaws. Redundancy in data can come from improper data entry as well as flawed or outdated processes, which indicates that synchronization will also become increasingly important in addressing data quality accuracy and integrity by eliminating redundancy earlier in the audit process. Without synchronization, the likelihood of duplicate records and database compromise are virtually certain. Redundancy can have unpredictable cascading effects, depending on how the data flow to other integrated systems within the network (Cusack et al., 2013).

## **Implication and Suggestions for the Future**

The consensus among the IT and non-IT healthcare systems data quality participants identified principle best practices that directed towards timeliness, eliminating redundancy, on-going rather than retrospective corrections, enhanced audits

combining both human interface audits along with business plan software and possibly AI, earlier decision making, user training, and the need for clear, transparent data management processes. Inclusion of these best practice guidelines will enhance existing accountability measures to reduce the number of digital errors in EHRs, resulting in an increased likelihood of optimal decisions and quality assurance. Senkubuge et al. (2014) asserted that there is a need for stakeholders buy in to set in place the policies and take ownership of the management of their data. Outdated and obsolete best practices should be abandoned as ineffective and inefficient to address the rapid changes in technology in favor of current best practices or not included at all. While integrity, validation, trustworthy data, ownership, data quality definitions, automated processes, roles, responsibilities, and metadata are still important, how they are addressed and managed has changed dramatically even within the past decade. Processes and best practices should be reviewed frequently for currency and monitored, particularly when a significant software update or new technology for data management is introduced. Data quality and integrity are key to appropriate patient management and treatment. Best practices that ensure that quality and integrity are not only essential to the success of the organization but also to the overall healthcare delivery systems and practices throughout the health care industry.

### **Limitations of the Study**

One of the major limitations of this study was time. Not all participants who started in Round 1 completed all three rounds. Time constraints and conditions for participants were a concern from the beginning of the study because of the commitment

to the 4-month period in 2-week intervals for each round. Some participants were not able to dedicate the required time to complete the survey for each round as they anticipated when signing up. There is a possibility that not all factors provided in the study were considered.

### **Recommendations**

As technology changes, the policies and practices that ensure data quality and integrity must keep pace. Traditionally, there has been a significant lag between the time of technology implementation and the evaluation of its effectiveness. Where data integrity, accuracy, and quality are critical, such as the patient medical record, healthcare organizations must adopt effective and continuous monitoring practices that address data quality. Along with those practices, leaders should impanel an in-house panel of experts that includes all stakeholders from the individual user, IT application specialists, along with a cross-functional representation of all affected by data flows. There should be an allowance for modifications and a review for the currency of the best practices in place or any standardized best practices created as a result of this study.

### **Implications**

Data quality and integrity are integral to the successful operation of any organization, regardless of the industry. As people and organizations become increasingly reliant on technology and the massive amounts of data now available, the findings of this study suggest practices are important to IT and non-IT participants. Service to the customer and consumer of services and products is dependent on the quality and integrity of data that the organization's practitioners and leaders use to make decisions, some of

which may be life-altering for a customer. Establishing a list of best practices that are constantly evaluated for effectiveness would create a logical set of practices that may be applicable across industries and different healthcare settings.

### **Conclusion**

The quality and integrity of the data begin with policies and practices that start at the point of user entry or acquisition of secondary data. These policies and practices should be succinct, comprehensible, uniform, and generalizable. The findings of this study demonstrate the importance industry experts place on addressing the quality and integrity of data in the era of Big Data and rapid technological advancement, particularly with the inclusion of AI that is becoming more ubiquitous in decision-making software. According to those experts, continual monitoring and audit of not only how data is collected, used, and managed but also the policies and processes that will ensure the quality and integrity of that data is essential to overcome the past and current challenges.

In this chapter, an interpretation of the findings was explored. The details were broken down further with subchapters such as Best Practices to Improve Data Collection, Best Practices That Apply to Secondary Use of Data, Best Practices That Apply to Data Quality and Integrity, and What Should be Included as Best Practices. The other major sections of this chapter were limitations which discussed the issues that occurred beyond the control of the researcher, recommendations, and implications.

## References

- A. von der Gracht, H., & Darkow, I. L. (2013). The future role of logistics for global wealth—scenarios and discontinuities until 2025. *Foresight, 15*(5), 405-419. doi:10.1108/FS-05-2012-0031
- Abdul-Hamid, H. (2014). What matters most for education management information systems: A framework paper (Working Paper no.7). Retrieved from [http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting\\_doc/Background/EMIS/Framework\\_SABER-EMIS.pdf](http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/Background/EMIS/Framework_SABER-EMIS.pdf)
- Adler-Milstein, J., & Jha, A. K. (2013). Healthcare's "big data" challenge. *American Journal of Managed Care, 19*(7), 537-538. Retrieved from <https://www.ajmc.com/journals>
- Agency for Healthcare Research and Quality. (2015). *About AHRQ*. Washington, DC: United States. Retrieved from <http://www.ahrq.gov>
- Allen, J. I., Igo, M., & Robin, D. P. (2006). Best practices: community-based gastroenterology practices. *Clinical Gastroenterology and Hepatology, 4*(10), 1292-1295. doi.org/10.1016/j.cgh.2006.07.023
- Al-Khwilani, M. N., Shmsan, M., & Al-akhram, N. (2013). Evaluating the quality of databases instances based on completeness and accuracy of data. *International Journal of Computer Science & Network Security, 13*(1), 35-38. Retrieved from [http://paper.ijcsns.org/07\\_book/201301/20130107.pdf](http://paper.ijcsns.org/07_book/201301/20130107.pdf)
- Alshawi, S., Missi, F., & Eldabi, T. (2003). Healthcare information management: The integration of patients' data. *Logistics Information Management 16*(3/4), 286-295.

doi.org/10.1108/09576050310483772.

- Altwaijry, H., Kalashnikov, D. V., & Mehrotra, S. (2013). Query-driven approach to entity resolution. *Proceedings of the VLDB Endowment*, 6(14), 1846-1857.
- American Society for Quality. (2015). Who we are. Retrieved from <http://asq.org/about-asq/who-we-are/index.html>
- Ancker, J., Shih, S., Singh, M., Snyder, A., Edwards, A., & Kaushal, R. (2011). *Root causes underlying challenges to secondary use of data* (pp. 57-62). Proceedings from AMIA '11: Improving health: Informatics and IT changing the world. Washington D.C. Retrieved from <http://asq.org/about-asq/who-we-are/index.html>
- Andel, C., Davidow, S. L., Hollander, M., & Moreno, D. A. (2012). The economics of health care quality and medical errors. *Journal of Healthcare Finance*, 39(1), 39. Retrieved from <https://www.healthfinancejournal.com/index.php/johcf>
- Anderka, M., Mai, C. T., Romitti, P. A., Copeland, G., Isenburg, J., Feldkamp, M. L.... & Stanton, C. (2015). Development and implementation of the first national data quality standards for population-based birth defects surveillance programs in the United States. *BMC Public Health*, 15(1), 925, 1-7 doi:10.1186/s12889-015-2223-2
- Ansari, S. (2005). Teaching note: Systems theory and management control. Retrieved from <http://faculty.darden.virginia.edu/ansaris/systems%20theory%20and%20mcs-tn.pdf>

- Aragon Penoyer, D., Cortelyou-Ward, K. H., Noblin, A. M., Bullard, T., Talbert, S., Wilson, J., & Briscoe, J. G. (2014). Use of electronic health record documentation by healthcare workers in an acute care hospital system. *Journal of Healthcare Management, 59*(2), 130-146. Retrieved from <https://journals.lww.com/jhmonline/pages/default.aspx>
- Atreja, A., Gordon, S. M., Pollock, D. A., Olmsted, R. M., & Brennan, P. J. (2008). The healthcare infection control practices advisory committee. Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. *American Journal of Infection Control, 36*(suppl), S37-S46.
- Babbie, R. (2016). *The basics of social research (6<sup>th</sup> ed)*. E-book Boston, MA: Cengage Learning.
- Baškarada, S., & Koronios, A. (2014). A critical success factor framework for information quality management. *Information Systems Management, 31*(4), 276-295. doi: 10.1080/10580530.2014.958023
- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR), 41*(3), 16, 1-52. doi: 10.1145/1541880.1541883
- Batini, C., & Scannapieco, M. (2016). *Data and information quality: dimensions, principles and techniques*. New York, NY: Springer International Publishing.

- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13(4), 544-559. Retrieved from [https://nsuworks.nova.edu/cahss\\_journals/](https://nsuworks.nova.edu/cahss_journals/)
- Beasley, J. W., Holden, R. J., & Sullivan, F. (2011). Electronic health records: research into design and implementation. *British Journal of General Practice*, 61(591), 604-605. doi:10.3399%2Fbjgp11X601244
- Berndt, D. J., McCart, J. A., Finch, D. K., & Luther, S. L. (2015). A case study of data quality in text mining clinical progress notes. *ACM Transactions on Management Information Systems*, 6(1), 1-21. doi:10.1145/2669368
- Blomberg, J., & Karasti, H. (2013). Reflections on 25 years of ethnography in CSCW. *Computer Supported Cooperative Work (CSCW)*, 22(4), 373-423. doi:10.1007/s10606-012-9183-1
- Blum, B. (1986). Clinical information systems—a review. *Western Journal of Medicine*, 145(6), 791. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/journals/183/>
- Boslaugh, S. (2010). *Encyclopedia of research design*, Thousand Oaks, CA: SAGE Publications.
- Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary use of EHR: Data quality issues and informatics opportunities. *Summit on Translational Bioinformatics*, 2010, 1–5. Retrieved from <https://www.amia.org>
- Bowers, S., McPhillips, T., & Ludäscher, B. (2012, June). Declarative rules for inferring fine-grained data provenance from scientific workflow execution traces. *In*

- International Provenance and Annotation Workshop 7525* (pp. 82-96). Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-34222-6\_7
- Bowman, S. (2013). Impact of electronic health record systems on information integrity: Quality and safety implications. *Perspectives in Health Information Management*, 10(Fall), 1c. Retrieved from <https://perspectives.ahima.org/>
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679. doi:10.1080/1369118X.2012.678878
- Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Services Research*, 42(4), 1758-1772. doi:10.1111/j.1475-6773.2006.00684.x
- Braunstein, M. L. (2014). *Contemporary health informatics*. Chicago, IL: AHIMA Press.
- Brady, S. R. (2015). Utilizing and adapting the Delphi method for use in qualitative research. *International Journal of Qualitative Methods*, 14(5), doi.org/10.1177/1609406915621381.
- Bray, F., & Parkin, D. (2009). Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *European Journal Of Cancer*, 45(5), 747-755. <http://dx.doi.org/10.1016/j.ejca.2008.11.033>
- Brown, G., Weber, D., & DeBie, K. (2014). Is PPGIS good enough? An empirical evaluation of the quality of PPGIS crowd-sourced spatial data for conservation planning. Elsevier, (43), 228-238. doi:10.1016/j.landusepol.2014.11.014

- Burke, H. B., Hoang, A., Becher, D., Fontelo, P., Liu, F., Stephens, M., & Gimbel, R. W. (2014). QNOTE: an instrument for measuring the quality of EHR clinical notes. *J Am Med Inform Assoc*, *21*(5), 910-916. doi:10.1136/amiajnl-2013-002321.
- Byrd, L. W., & Byrd, T. A. (2012). Developing an instrument for information quality for clinical decision making. *System Science (HICSS), 2012 45th Hawaii International Conference* (pp. 2820-2829). Manoa, HI. doi: 10.1109/HICSS.2012.210
- Cabitza, F. (2013). An information reliability index as a simple consumer-oriented indication of quality of medical web sites. *Quality issues in the management of web information* (pp. 159-177). Berlin, DE: Springer International. doi: 10.1007/978-3-642-37688-7\_8
- Centers for Medicare & Medicaid Services (CMS). (2015). About CMS. Washington, DC: United States. Retrieved from CMS website: <https://www.cms.gov>.
- Charles, D., Gabriel, M., & Furukawa, M. F. (2014). Adoption of electronic health record systems among US non-federal acute care hospitals: 2008-2013. *ONC data brief*, *16*, 1-9.
- Chatfield, M., & Kocher, K. E. (2015). Advancing the use of administrative data for emergency department diagnostic imaging research. *Academic Emergency Medicine*, *22*(12), 1417-1426. doi: <http://dx.doi.org/10.1111/acem.12827>
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics:

- from big data to big impact. *MIS quarterly*, 36(4), 1165-1188. Retrieved from <https://www.misq.org/>
- Church, R. M. (2002). The effective use of secondary data. *Learning and Motivation*, 33(1), 32-45. doi:10.1006/lmot.2001.1098
- Clack, L., & Woepfel, J. (2018). Data analytics drives strategic planning in healthcare. *Journal Of AHIMA*, (June), 42-43. Retrieved form <https://journal.ahima.org/>
- Clow, K. E., & James, K. E. (2014). The marketing research process. In *Essentials of Marketing Research: Putting Research into Practice*. (pp. 25-61). Thousand Oaks, CA: SAGE.
- Collins, S. A., Fred, M., Wilcox, L., & Vawdrey, D. K. (2012). Workarounds used by nurses to overcome design constraints of electronic health records. *NI 2012: Proceedings of the 11th International Congress on Nursing Informatics* (Vol. 2012). American Medical Informatics Association. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24199061>
- Cruz-Correia, R., Boldt, I., Lapão, L., Santos-Pereira, C., Rodrigues, P. P., Ferreira, A. M., & Freitas, A. (2013). Analysis of the quality of hospital information systems audit trails. *BMC medical informatics and decision-making*, 13(1), 1. doi: 10.1186%2F1472-6947-13-84
- Cusack, C. M., Hripcsak, G., Bloomrosen, M., Rosenbloom, S. T., Weaver, C. A., Wright, A. & Mamykina, L. (2013). The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. *Journal of the*

- American Medical Informatics Association*, 20(1), 134-140. doi:10.1136/amiajnl-2012-001093
- Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 903-920. doi:10.1111/bjet.12230
- Data Integrity. (2017). In *Pocket glossary of health information management and technology (5th ed.)*. Chicago, IL: American Health Information Management Association.
- Data Quality. (2017). In *Pocket glossary of health information management and technology (5th ed.)*. Chicago: American Health Information Management Association.
- Data quality assessment framework and data quality program. (2003). Retrieved from <http://www.imf.org/external/np/sta/dsbb/2003/eng/dqaf.htm>
- Davis, N. A., & LaCour, M. (2014). *Health information technology*. Houston, TX: Elsevier Health Sciences.
- de Almeida, W. G., de Sousa, R. T., de Deus, F. E., Nze, G. D. A., & de Mendonça, F. L. L. (2013, June). Taxonomy of data quality problems in multidimensional Data Warehouse models. In 2013 8th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-7). IEEE. Retrieved from <https://ieeexplore.ieee.org/Xplore/home.jsp>
- Debbarma, N., Nath, G., & Das, H. (2013). Analysis of data quality and performance issues in data warehousing and business intelligence. *International Journal of*

*Computer Applications*, 79(15), 20-26. doi:10.5120/13818-1862

- Deming, W. E. (1986). *Out of the crisis*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.
- Donalek, J. G. (2004). Phenomenology as a qualitative research method. *Urologic nursing*, 24(6), 516-517. Retrieved from <https://www.sun.org/unj>
- Dixon, B., Siegel, J., Oemig, T., & Grannis, S. (2013). Electronic health information quality challenges and interventions to improve public health surveillance data and practice. *Public Health Reports*, 128(6), 546-553. (Washington, D.C.: 1974). doi:10.1177/003335491312800614
- Duerr, R. E., Parsons, M. A., Marquis, M., Dichtl, R., & Mullins, T. (2004, April). Challenges in long-term data stewardship. In *MSST* (pp. 47-67). Retrieved from <http://storageconference.us/2004/Papers/05-Duerr-a.pdf>
- Duggirala, H. J., Tanning, J. M., Smith, E., Bright, R. A., Baker, J. D., Ball, R., ... & Boyer, M. (2016). Use of data mining at the Food and Drug Administration. *Journal of the American Medical Informatics Association*, 23(2), 428-434. doi:10.1093/jamia/ocv063
- Dumbleton, J. (2013). 2014: Key trends driving the change in data quality technology. Retrieved from [http://d2ox65wlcvkof.cloudfront.net/media/marketing/downloads/pdf/research-papers/uk/Gartner\\_Experian\\_UK\\_issue1.pdf](http://d2ox65wlcvkof.cloudfront.net/media/marketing/downloads/pdf/research-papers/uk/Gartner_Experian_UK_issue1.pdf)
- Eckerson, W.W. (2002). Data quality and the bottom line: achieving business success through a commitment to high quality data. *TDWI report series*. Seattle, WA: The

- Warehousing Institute. Retrieved from  
<http://download.101com.com/pub/tdwi/Files/DQReport.pdf>
- Electronic Health Record. (2017). In *Pocket glossary of health information management and technology (5th ed.)*. Chicago: American Health Information Management Association.
- Eriksson, P., & Kovalainen, A. (2008). 9 Case Study Research. In *Qualitative methods in business research*. (pp.115-137). London: SAGE. doi:  
10.4135/9780857028044.d66
- Experian Data Quality. (2013). The state of data quality [White paper]. Retrieved from  
<https://www.experian.com/assets/decision-analytics/white-papers/the%20state%20of%20data%20quality.pdf>
- FAIR Health. (2017, March 15). Nation's largest private healthcare database acquires all Medicare claims data [Press release]. Retrieved February 19, 2018, from  
<https://www.prnewswire.com/news-releases/nations-largest-private-healthcare-database-acquires-all-medicare-claims-data-300423650.html>
- Foot, C., Sonola, L., Bennett, L., Fitzsimons, B., Raleigh, V. S., & Gregory, S. (2014). *Managing quality in community health care services*. King's Fund. Retrieved from <https://www.kingsfund.org.uk/>
- Foshay, N., & Kuziemy, C. (2014). Towards an implementation framework for business intelligence in healthcare. *International Journal of Information Management*, 34(1), 20-27. doi:10.1016/j.ijinfomgt.2013.09.003
- Fowler, F. J. (1995). *Improving survey questions: design and evaluation* (Vol. 38).

Thousand Oaks: Sage.

Gartner. (2015). About Gartner. Retrieved from

<http://www.gartner.com/technology/about.jsp>.

Goossen, W. T. (2014). Detailed clinical models: Representing knowledge, data and semantics in healthcare information technology. *Healthcare informatics research*, 20(3), 163-172. doi: 10.4258%2Fhir.2014.20.3.163

Gunawan, J. (2016). Electronic health records in nursing practice: a concept analysis.

*International Journal of Innovations in Medical Education and Research*, 2(1), 5-8. Retrieved from doi:10.5455/ijimer.2015.19082015012

Harder, B. (2013, July 16). 5 Cities with tons of top hospitals. *U.S. News & World*

*Report*. Retrieved from <http://health.usnews.com/health-news/best-hospitals/articles/2013/07/16/5-cities-with-tons-of-top-hospitals>.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications.

*International Journal of Production Economics*, 154, 72-80. doi: 10.1016/j.ijpe.2014.04.018

Healthcare Information and Management Systems Society (HIMSS). (2015). About

HIMSS. Retrieved from

[http://action.sierraclub.org/site/MessageViewer?em\\_id=177341.0](http://action.sierraclub.org/site/MessageViewer?em_id=177341.0).

Health Information Management. (2017). *Pocket glossary of health information*

*management and technology (5th ed.)*. Chicago: American Health Information Management Association.

Hedt-Gauthier, B., Tenthani, L., Mitchell, S., Chimbwandira, F., Makombe, S., Chirwa,

Z., & Jahn, A. (2012). Improving data quality and supervision of antiretroviral therapy in Malawi: an application of Lot Quality Assurance Sampling. *BMC*

*Health Services Research*, 12, 196. doi:10.1186/1472-6963-12-196

Holzinger, A., Dehmer, M., & Jurisica, I. (2014). Knowledge discovery and interactive

data mining in bioinformatics state-of-the-art, future challenges and research

directions. *BMC bioinformatics*, 15(6), 11. doi:10.1186/1471-2105-15-S6-11

Hoque, Z. (2014). 20 years of studies on the balanced scorecard: Trends,

accomplishments, gaps and opportunities for future research. *The British*

*Accounting Review*, 46(1), 33-59. doi: 10.1016/j.bar.2013.10.003

Hyett, N., Kenny, A., & Virginia Dickson-Swift, D. (2014). Methodology or method?

A critical review of qualitative case study reports. *International Journal of*

*Qualitative Studies on Health and Well-being*, 9.

doi:10.3402/qhw.v9.23606

Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). *Data quality and record linkage*

*techniques*. New York: Springer Science & Business Media.

Information Builders. (2013). The real cost of bad data: Six simple steps to address data

quality issues (White paper). Information Builders Corporate headquarters, New

York, NY. Retrieved from <http://www.insurance->

[canada.ca/ebusiness/canada/2013/Information-Builders-whitepaper-Bad-Data-](http://www.insurance-canada.ca/ebusiness/canada/2013/Information-Builders-whitepaper-Bad-Data-)

- International Monetary Fund. (2006). Albania: Report on the observance of standards and codes—Data module, response by the authorities, and detailed assessments using the data quality assessment Framework. (Rep. No. 06/393, pp. 1-176). Washington, DC: International Monetary Fund Publication Services. Retrieved from <https://www.imf.org/en/Publications/CR/Issues/2016/12/31/Albania-Report-on-the-Observance-of-Standards-and-Codes-Data-Module-Response-by-the-20059>
- Jack, W., Balijepally, V., & Tanniru, M. (2014). Integrating strategic and operational decision making using data-driven dashboards: the case of St. Joseph Mercy Oakland Hospital. *Journal of Healthcare Management/American College of Healthcare Executives*, 60(5), 319-330. Retrieved from <https://www.ache.org/learning-center/publications/journals>
- Jamoom, E. W., Patel, V., Furukawa, M. F., & King, J. (2014). EHR adopters vs. non-adopters: Impacts of, barriers to, and federal initiatives for EHR adoption. *Healthcare*, 2(1), 33-39. doi: 10.1016/j.hjdsi.2013.12.004
- Jansen, H. (2010). The logic of qualitative survey research and its position in the field of social research methods. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 11, No. 2). doi: 10.17169/fqs-11.2.1450
- Jerven, M. (2014). On the accuracy of trade and GDP statistics in Africa: Errors of commission and omission. *Journal of African Trade*, 1(1), 45-52. doi: 10.1016/j.joat.2014.08.001

- Jha, A.K. (2011). The promise of electronic records around the corner or down the road?  
*JAMA*, 306, 880-881. doi:10.1001/jama.2011.1219
- Joglekar, N., Anderson, E. & Shankaranarayanan. (2013). Accuracy of aggregate data  
in distributed project settings: Model, analysis and implications. *ACM Journal of  
Data and Information Quality* 4(3), 13:1-13:22. doi: 10.1145/2458517.2458521
- Jones-Farmer, L. A., Ezell, J. D., & Hazen, B. T. (2014). Applying control chart methods  
to enhance data quality. *Technometrics*, 56(1), 29-41. doi:  
10.1080/00401706.2013.804437
- Joshi, M., Ransom, E. R., Nash, D. B., & Ransom, S. B. (2014). *The Healthcare Quality  
Book: Vision, Strategy, and Tools* (3<sup>rd</sup> edition). Chicago, IL: Health  
Administration Press.
- Judah, S., & Friedman, T. (2014, February 17). Twelve ways to improve your data  
quality. Retrieved from [https://www.gartner.com/doc/2666816/ways-improve-  
data-quality](https://www.gartner.com/doc/2666816/ways-improve-data-quality).
- Jugulum, R. (2014). *Competing with high quality data: concepts, tools, and  
techniques for building a successful approach to data quality*. Hoboken, NJ: John  
Wiley & Sons.
- Katz, D., & Kahn, R. (1966). *The social psychology of organizations*. New York:  
John Wiley & Sons.
- Kerr, K., Norris, T., & Stockdale, R. (2008). The strategic management of data  
quality in healthcare. *Health Informatics Journal*, 14(4), 259-266. doi:  
10.1177/1460458208096555

- Kloss, L. (2015). *Implementing health information governance: Lessons from the field*. Chicago: AHIMA Press.
- Kuehl, D. R., Berdahl, C. T., Jackson, T. D., Venkatesh, A. K., Mistry, R. D., Bhargavan-Chatfield, M., & Kocher, K. E. (2015). Advancing the use of administrative data for emergency department diagnostic imaging research. *Academic Emergency Medicine*, 22(12), 1417-1426. doi: 10.1111/acem.12827
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387-394. doi: 10.1016/j.ijinfomgt.2014.02.002
- Liaw, S. (2013). Clinical decision support systems: Data quality management and governance. *Health Information Governance in a Digital Environment*, 193, 362. Retrieved from <https://www.iospress.nl/bookserie/studies-in-health-technology-and-informatics/>
- Liaw, S. T., Rahimi, A., Ray, P., Taggart, J., Dennis, S., de Lusignan, S., & Talaei-Khoei, A. (2013). Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *International journal of medical informatics*, 82(1), 10-24. doi: 10.1016/j.ijmedinf.2012.10.001
- Liyanage, H., Liaw, S., & De Lusignan, S. (2012). Accelerating the development of an information ecosystem in health care, by stimulating the growth of safe intermediate processing of health information (IPHI). *Informatics in Primary Care*, 20(2), 81-86. Retrieved from <https://hijournal.bcs.org/index.php/jhi>

- Loshin, D. (2001). *Enterprise knowledge management: The data quality approach*. New York: Morgan Kaufmann.
- Loshin, D. (2010). *The practitioner's guide to data quality improvement*. Houston: Elsevier.
- Mawilmada, P., Smith, S., & Sahama, T. (2012). Investigation of decision-making issues in the use of current clinical information systems. *Studies in Health Technology And Informatics*, 178136-143. Retrieved from <https://www.iospress.nl/bookserie/studies-in-health-technology-and-informatics/>
- Maydanchik, A. (2007). *Data quality assessment*. Bradley Beach, NJ: Technics Publications.
- McCoy, A. B., Wright, A., Kahn, M. G., Shapiro, J. S., Bernstam, E. V., & Sittig, D. F. (2013). Matching identifiers in electronic health records: implications for duplicate records and patient safety. *BMJ Qual Saf*, 22(3), 219-224. doi: 10.1136/bmjqs-2012-001419
- McGilvray, D. (2008). *Executing data quality projects: ten steps to quality data and trusted information*. Houston: Elsevier.
- McMillan, S. S., King, M., & Tully, M. P. (2016). How to use the nominal group and Delphi techniques. *International journal of clinical pharmacy*, 38(3), 655-662. doi.org/10.1007/s11096-016-0257-x
- Moges, H.T. (2014). *A contextual data quality analysis for credit risk management in financial institutions*. (Unpublished dissertation). University of Leuven, Belgium.
- Moghaddasi H, Rabiei R, & Sadeghi N. (2014). Improving the quality of clinical coding:

- a comprehensive audit model. *Journal of Health Management & Informatics* 1(2): 36-40. Retrieved from <http://jhmi.sums.ac.ir/index.php/JHMI>
- Morbey, G. (2013). *Data quality for decision makers: A dialog between a board member and a DQ expert*. Wiesbaden, DE: Springer Fachmedien Wiesbaden.
- Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks: Sage.
- Mrkaic, M. (2010). IMF working paper: Information content of DQAF indicators-empirical entropy analysis. USA: International Monetary Fund. doi: 10.5089/9781455205356.001.A001
- National Institutes of Health (NIH). (2015). Institutes, centers, offices. Washington, DC:
- Nave, D. (2002). How to compare six sigma, lean and the theory of constraints. *Quality Progress*, 35(3), 73. Retrieved from [www.asq.org](http://www.asq.org)
- Needham, D. M., Sinopoli, D. J., Dinglas, V. D., Berenholtz, S. M., Korupolu, R., Watson, S. R., & Pronovost, P. J. (2009). Improving data quality control in quality improvement projects. *International Journal for Quality in Health Care*, 21(2), 145-150. doi: 10.1093/intqhc/mzp005
- Nelson, R., & Staggers, N. (2014). Theoretical foundations of health informatics. *Health informatics online for Nelson and Staggers health informatics: An inter-professional approach* (1st ed., pp. 20-38). St. Louis, MO: Mosby/Elsevier.
- Norton, D. & Kaplan, R. (2016). Balanced scorecard basics. Retrieved from <http://balancedscorecard.org/Resources/About-the-Balanced-Scorecard>
- Olson, J. E. (2003). *Data quality: The accuracy dimension*. Miamisburg, OH: Elsevier Science.

- Oracle. (2015). Risk and reward: Why getting your risk data into shape will lead to business rewards. Retrieved from Oracle website: <http://www.cfp-events.com/wp-content/uploads/2015/07/Risk-and-Reward.pdf>
- Padgett, D. K. (2012). Strategies for rigor. *Qualitative and mixed methods in public health*. (pp. 203-221). Thousand Oaks: SAGE Publications, In.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health*, 42(5), 533–544. doi: 10.1007/s10488-013-0528-y
- Panahy, P. H. S., Sidi, F., Affendey, L. S., Jabar, M. A., Ibrahim, H., & Mustapha, A. (2013). A framework to construct data quality dimensions relationships. *Indian Journal of Science and Technology*, 6(5), 4422-4431. Retrieved from [www.indjst.org](http://www.indjst.org) .
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Thousand Oaks: SAGE Publications
- Paygude, P., & Devale, P. R. (2013). Automation of data validation testing for QA in the project of DB migration. *International Journal of Computer Science*. 3(2), 15-22. doi:10.1.1.416.1668&rep=rep1&type=pdf
- Phillips, J., & Davidson, P. (2010). 16 Focus group methodology: Being guided on a journey from novice to expert. V. Minichiello, & Jeffrey A. Kottler (Eds.), *Qualitative journeys: Student and mentor experiences with research*. (pp. 255-277). Thousand Oaks: SAGE Publications, In. doi: 10.4135/9781452272078.n16

- Ponto, J. (2015). Understanding and evaluating survey research. *Journal of the Advanced Practitioner in Oncology*, 6(2), 168–171.  
doi.org/10.6004/jadpro.2015.6.2.9
- Puhan, M. A., Akl, E. A., Bryant, D., Xie, F., Apolone, G., & Riet, G. (2012). Discussing study limitations in reports of biomedical studies- the need for more transparency. *Health and Quality of Life Outcomes*, 10, 23. doi: 10.1186/1477-7525-10-23
- Ramírez-Gallego, S., Mouriño-Talín, H., Martínez-Rego, D., Bolón-Canedo, V., Benítez, J. M., Alonso-Betanzos, A., & Herrera, F. (2018). An information theory-based feature selection framework for big data under apache spark. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(9), 1441-1453.  
doi:10.1109/TSMC.2017.2670926
- Rau, J. (2016). Hospital discharge: It's one of the most dangerous periods for patients. *Kaiser Health News*. Retrieved from <https://www.washingtonpost.com/news/to-your-health/wp/2016/04/29/from-hospital-to-home-a-dangerous-transition-for-many-patients/>
- Reader, T. W., Gillespie, A., & Roberts, J. (2014). Patient complaints in healthcare systems: a systematic review and coding taxonomy. *BMJ quality & safety*, 23(8), 678-689. doi:10.1136/bmjqs-2013-002437
- Redman, T. C. (2008). *Data driven: profiting from your most important business asset*. Cambridge, MA: Harvard Business.
- Reiter, K. L., Halladay, J. R., Mitchell, M. C., Ward, K., Lee, S. Y. D., Steiner, B., &

- Donahue, K. E. (2014). Costs and benefits of transforming primary care practices: a qualitative study of North Carolina's Improving Performance in Practice. *Journal of Healthcare Management*, 59(2), 95-108. doi: 10.1097/00115514-201403000-00004
- Rowe, M., Frantz, J., & Bozalek, V. (2013). Beyond knowledge and skills: the use of a Delphi study to develop a technology-mediated teaching strategy. *BMC Medical Education*, 13, 51. doi: 10.1186/1472-6920-13-51
- Samitsch, C. (2014). *Data quality and its impacts on decision-making: How managers can benefit from good data*. New York: Springer.
- Scott, N. M., Sera, M. D., & Georgopoulos, A. P. (2015). An information theory analysis of spatial decisions in cognitive development. *Frontiers in Neuroscience*, 9, 14. doi: 10.3389/fnins.2015.00014
- Sebastian-Coleman, L. (2012). *Measuring data quality for ongoing improvement: a data quality assessment framework*. Burlington, CT: Morgan-Kaufmann.
- Sekayi, D., & Kennedy, A. (2017). Qualitative Delphi method: A four round process with a worked example. *The Qualitative Report*, 22(10), 2755-2763. Retrieved from [https://nsuworks.nova.edu/cahss\\_journals/](https://nsuworks.nova.edu/cahss_journals/)
- Senkubuge, F., Modisenyane, M., & Bishaw, T. (2014). Strengthening health systems by health sector reforms. *Global health action*, 7. doi: 10.3402/gha.v7.23568
- Sharique, H., Padman, R. (2006). *Analyzing the Effect of Data Quality on the Accuracy of Clinical Decision Support Systems: A Computer Simulation Approach*. Paper presented at the American Medical Informatics Association, Hilton Towers,

- Shewhart, W. A., & Deming, W. E. (1939). *Statistical method from the viewpoint of quality control*. North Chelmsford, MA: Courier Corporation
- Shultz, K. S., Hoffman, C. C., & Reiter-Palmon, R. (2005). Using archival data for IO research: Advantages, pitfalls, sources, and examples. *The Industrial-Organizational Psychologist*, 42(3), 31. Retrieved from [http://network.bepress.com/social-and-behavioral-sciences/psychology/industrial-and-organizational-psychology/?utm\\_campaign=PDFCoverPages&utm\\_medium=PDF&utm\\_source=digitalcommons.unomaha.edu%2Fpsychfacpub%2F5](http://network.bepress.com/social-and-behavioral-sciences/psychology/industrial-and-organizational-psychology/?utm_campaign=PDFCoverPages&utm_medium=PDF&utm_source=digitalcommons.unomaha.edu%2Fpsychfacpub%2F5)
- Siemens, G. (2014). Supporting and promoting learning analytics research. *Journal of Learning Analytics*, 1(1), 3-5. doi.org/10.18608/jla.2014.11.2
- Singh, R., & Singh, K. (2010). A descriptive classification of causes of data quality problems in data warehousing. *International Journal of Computer Science Issues*. *IJCSI*, 7(4). doi: 10.1.1.403.9008&rep=rep1&type=pdf
- Shankaranarayanan, G. G., & Blake, R. (2017). From content to context: The evolution and growth of data quality research. *J. Data and Information Quality* 8(2): 1-28. doi.org/10.1145/2996198
- Shankaranarayanan, G. G., & Cai, Y. (2006). Supporting data quality management decision-making. *Decision Support Systems*, 41(1), 302. doi:10.1016/j.dss.2004.12.006
- Spath, P. (2013). *Introduction to Healthcare Quality Management*. Chicago: Health Administration Press.

- Spencer, C. S., Roberts, E., & Gaskin, D. J. (2015). Differences in the rates of patient safety events by payer: implications for providers and policymakers. *Medical care*, 53(6), 524-529. doi.org/10.1097/mlr.0000000000000363
- Struijs, P., Braaksma, B., & Daas, P. J. (2014). Official statistics and big data. *Big Data & Society*, 1(1), 1-6. doi:10.1177/2053951714538417
- Sullivan, G. M., & Artino, A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*, 5(4), 541–542. doi: 10.4300/JGME-5-4-18
- Sun, G., & Khoo, C. S. G. (2017). Social science research data curation: Issues of reuse. *Libellarium: journal for the research of writing, books, and cultural heritage institutions*, 9(2). doi: 10.15291/libellarium.v9i2.291
- SurveyMonkey and IRB Guidelines. (n.d.). Retrieved September 02, 2018, from [https://help.surveymonkey.com/articles/en\\_US/kb/How-does-SurveyMonkey-adhere-to-IRB-guidelines](https://help.surveymonkey.com/articles/en_US/kb/How-does-SurveyMonkey-adhere-to-IRB-guidelines)
- Taggart, J., Liaw, S., Dennis, S., Yu, H., Rahimi, A., Jalaludin, B., & Harris, M. (2012). The University of NSW electronic practice-based research network: Disease registers, data quality and utility. *Studies in Health Technology and Informatics*, 17(8). 219-227. Retrieved from <https://www.iospress.nl/bookserie/studies-in-health-technology-and-informatics/>
- Thompson, N. D., Edwards, J. R., Bamberg, W., Beldavs, Z. G., Dumyati, G., Godine, D., ... & Wilson, L. (2013). Evaluating the accuracy of sampling to estimate central line-days simplification of the national healthcare safety network

- surveillance methods. *Infection Control & Hospital Epidemiology*, 34(03), 221-228. doi: 10.1086/669515
- Todoran, I. G., Lecornu, L., Khenchaf, A., & Caillec, J. M. L. (2015). A methodology to evaluate important dimensions of information quality in systems. *Journal of Data and Information Quality (JDIQ)*, 6(2), 11. doi: 10.1145/2744205
- Tolar, M., & Balka, E. (2012). Caring for individual patients and beyond: Enhancing care through secondary use of data in a general practice setting. *International Journal of Medical Informatics*, 81(7), 461-474. doi.org/10.1016/j.ijmedinf.2012.01.003
- U.S. Census Bureau. (2017). *Quick facts: United States (Report No. PST045217)*. Washington, Retrieved from U.S. Department of Commerce:  
<https://www.census.gov/quickfacts/fact/table/US/PST045217>
- van Fenema, E., Giltay, E., van Noorden, M., van Hemert, A., & Zitman, F. (2017). Assessing adherence to guidelines with administrative data in psychiatric outpatients. *Journal of evaluation in clinical practice*, 23(1), 5-13.  
[doi.org/10.1111/jep.12414](https://doi.org/10.1111/jep.12414)
- van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, 35,1-4. doi: 10.7748/ns2002.06.16.40.33.c3214
- Wagner, E. H., Austin, B. T., Davis, C., Hindmarsh, M., Schaefer, J., & Bonomi, A. (2001). Improving chronic illness care: Translating evidence into action. *Health Affairs*, 20(6), 64-78. doi: 10.1377/hlthaff.20.6.64

- Wang, R. W., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal Of Management Information Systems*, 12(4), 5-33.  
Retrieved from <https://www.jmis-web.org/issues>
- Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, 20(1), 144-151. doi: 10.1136/amiajnl-2011-000681
- Wills, M. J. (2014). Decisions through data: Analytics in healthcare. *Journal of Healthcare Management*, 59(4), 254-262. Retrieved from <https://journals.lww.com/jhmonline/pages/default.aspx>.
- World Bank. (2014). Health expenditure, total (% of GDP) [Data file]. Retrieved from <http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS>
- Yang, S., Li, J., Cai, J., Guo, K., Gao, X., & Meng, F. (2014). Data-oriented method to big data standard system creation: A case of Chinese financial industry. *Annals of Data Science*, 325-328. doi.org/10.1007/s40745-014-0024-6
- Zozus, M. N., Hammond, W.E., Green, B.B., Kahn, M.G., Richesseon, R.L., Rusincovitch, S.A., Simon, G.E., & Smerek, M.M. (2014). Assessing data quality for healthcare systems data used in clinical research *NIH Health Care Systems Research Collaboratory* [White paper]. Durham, NC, NIH. Retrieved from [https://www.nihcollaboratory.org/Products/Assessing-data-quality\\_V1%200.pdf](https://www.nihcollaboratory.org/Products/Assessing-data-quality_V1%200.pdf)

### Appendix A: Round 1 Delphi Survey Questions

1. What challenges and reasons for poor data quality in a healthcare system exist today?
2. How does poor data quality affect the efficiency and fitness of data quality with the data in healthcare systems?
3. How can a data management plan mitigate data quality errors?
4. What are some of the best practices or techniques used today to eliminate data quality errors and issues in the future?

## Appendix B: SurveyMonkey and IRB Guidelines

Students are certainly permitted to conduct research via the SurveyMonkey platform, if they abide by our Terms of Use. Many students use SurveyMonkey to conduct research for their dissertations or graduate work. The guidelines for using SurveyMonkey as a tool to survey research participants are as follows. These are criteria that most university IRB's recommend when using an online survey tool to collect data. It is important to engage your Institutional Review Board to approve.

### **Obtaining Written Permission to Conduct Research Using SurveyMonkey**

We are happy to assist you with getting the approvals you need to perform your student research. Here is a letter on SurveyMonkey letterhead that you can provide to your IRB to evidence permission to use the SurveyMonkey platform to conduct your research: [Permission to Conduct Research Using SurveyMonkey \(PDF\)](#)

### **Secure Transmission**

- It is important to enable SSL encryption. Sensitive data must be protected as it moves along communication pathways between the respondent's computer and SurveyMonkey servers.
- Disable IP address tracking to make the survey anonymous.

### **Informed Consent**

- Include a consent form on the first page of your survey.
- SurveyMonkey records the respondent time stamp. This is important especially for respondents that consented to taking your survey.

- The survey should allow for “no response “or “prefer not to respond” as an option for every survey question. A survey where a respondent cannot proceed without answering the question is in violation of the respondent’s right to withhold information.
- At the end of the survey, the respondent should be given an option to withdraw from survey.

### **Database and Server Security**

- SurveyMonkey Privacy Policy
- Security Statement

### **HIPAA Compliance**

If you're a covered entity regulated by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and want to collect protected health information in your health and well-being surveys, please see HIPAA Compliance at SurveyMonkey for more details.

SurveyMonkey and IRB Guidelines. (n.d.)