

The Evolution of Medical Informatics as a Coherent Academic Discipline

by

Fred Karl Weigel

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 12, 2011

Keywords: medical informatics, electronic medical records, mixed methods, diffusion of
innovations, content analysis

Copyright 2011 by Fred Karl Weigel

Approved by

Casey G. Cegielski, Chair, Associate Professor of Management Information Systems
R. Kelly Rainer, Privett Professor of Management Information Systems
F. Nelson Ford, Associate Professor of Management Information Systems

Abstract

The purpose of this content analysis was to ascertain the prevalent themes, the challenges that exist, and future directions for the medical informatics (MI) discipline. Seven scholarly publications from the ten-year period of 2002 – 2011 provided the data. The sample included article texts collected from the MI publications *Journal of the American Medical Informatics Association* and *International Journal of Medical Informatics*. Additional data came from the related fields of medicine—the *Journal of the American Medical Association* and the *New England Journal of Medicine*—and management information systems—*Management Information Systems Quarterly*, *Information Systems Research*, and *Communications of the Association for Computing Machinery*. All published article texts were collected from the medical informatics publications for the ten-year period. For those publications outside the MI mainstream, advanced Boolean queries identified articles for collection. A total of 2,315 (2,188 retained) article texts were collected.

The first phase of the mixed methods approach was quantitative and applied *Centering Resonance Analysis* (CRA) to identify themes in the data. The second, qualitative phase consisted of manually coding the data against categories developed from the literature review. CRA identified the following 10 themes emerging from the literature: Analytics, Healthcare Operations and Standards (with sub-themes: Operations, Project Management, and Information Assurance), Aspects of Healthcare Research, Knowledge Transfer/Communication (with sub-themes: Extending beyond the Organization, Internal to the Organization, and Patient-Provider),

Perceptions and Managing Expectations of Information Technology, and Software as a Service.

The manual coding identified that 34.5% (755) of the articles addressed Information Architecture, 34.1% (746) addressed Direct Patient Care, 10.7% (235) addressed Relative Advantage, and 1.9% (41) addressed Compatibility.

The themes discovered indicate the discipline consists of information systems, healthcare, operations, communication, and research. With continuing legislation emphasizing digital health records, dramatic and rapid improvements in technology, and the ever-pressing need to reduce healthcare costs, the demand for medical informatics is great. Although medical informatics is young, the field has established deep roots and a strong foundation. We can expect to see persistent growth and maturity in the field as scholars, practitioners, and researchers continue to provide value to the healthcare of the ever-increasing population.

Acknowledgments

I am grateful for my dissertation committee: Dr. Casey Cegielski, Dr. Kelly Rainer, and Dr. Nelson Ford. I appreciate your constant assistance, guidance, enlightenment, and encouragement. I owe you a debt of gratitude. I have learned a lot about myself and I am very glad I chose this path. Thank you for the hours spent entertaining my ideas, counseling me, and directing my path.

To my Air Force office mates, Ben Hazen and Rob Overstreet, thank you for your friendship and encouragement. I appreciate the banter, the lunches, your help, and your work ethic. You set the academic example and scholarly pace I would have liked to have from the beginning.

Dr. Dianne Hall, thank you for welcoming me even before I received the offer here and thank you for the work you do in helping the military. Dr. Allison Jones-Farmer, I am grateful for your opening the world of statistics to me. To the faculty and staff of the AU College of Business, thank you for providing a warm, encouraging environment.

To my Army peers who led the way, Mark Mellott, Ken Jones, Al Hamilton, and Pete Marks, each of you has encouraged me and provided guidance along the way. You have gone out of your way and put hours of time into helping me. Thank you.

Last in the acknowledgement, but first in my heart is my family. To my ten wonderful children: Ashley, Emily, Gideon, Adam, Grace, Gabriella, Blondina, Beth, Eliahna, and Evangeline – you make it worthwhile. Thank you for your patience, your tolerance of all my

time away from you, and for your enthusiasm when I am home. Finally, to my best friend, my wife, and love of my life, Elissa, thank you for being my partner in yet another exciting journey as we share our lives together. You are the Proverbs 31, Titus 2 woman who kept the family functioning and ensured the children received enough love. Thank you for your patience and encouragement. Without your sacrifices, I would have not completed this degree.

Table of Contents

Abstract	ii
Acknowledgments.....	iv
List of Tables	xi
List of Figures	xii
Chapter 1. Introduction	1
Problem Statement.....	3
Purpose.....	5
Research Questions.....	6
Worldview	6
Significance of the Study.....	8
Summary of Chapter One and Outline of Chapters.....	9
Chapter One	9
Chapter Two.....	10
Chapter Three.....	10
Chapter Four	11
Chapter Five.....	11

Chapter 2. Literature Review	12
Origins of Medical Informatics.....	12
Chronology of Medical Informatics	15
1970s and Earlier	15
1980s	17
1990s	20
2000 to Present.....	22
On the Importance of Theory.....	34
Theory Development	35
Adopter Categories	37
Stages of the Innovation-Decision Process.....	38
Perceived Characteristics of Innovations.....	39
Chapter Summary	41
Chapter 3. Methods	42
Content Analysis.....	43
Process	44
Mixed Methods Analysis	46
Sampling Units	46
Recording Units	47
Observational Units	48

Sampling	48
Medical Informatics Publications	50
Medical and Management Information Systems Publications.....	50
Query Method	51
Data Preparation	53
Centering Resonance Analysis	55
Data Analysis – Centering Resonance Analysis.....	58
Operationalized Variable Definitions	60
Qualitative Data Analysis - Manual Coding.....	65
Coders and Coding Procedures.....	65
Inter-coder Reliability	67
Chapter Summary	68
Chapter 4. Findings.....	70
Descriptives	71
Data Preparation	73
Mixed Methods Analysis	78
Phase I: Quantitative Data Analysis –Centering Resonance Analysis	78
Noun Phrase Network Information.....	79
Influential Words	85
Exploratory Factor Analysis	86

Final Theme Solution.....	91
Phase II: Manual Coding	94
Inter-Coder Reliability	94
Qualitative Results - Manual Coding.....	97
Final Coding Results.....	104
Chapter Summary	107
Chapter 5. Discussion and Conclusions.....	110
Summary	110
Interpretation of the Findings	113
Healthcare Operations and Standards: Operations.....	114
Healthcare Operations and Standards: Project Management.....	114
Healthcare Operations and Standards: Information Assurance	115
Knowledge Transfer/Communication: Extending Beyond the Organization.....	116
Knowledge Transfer/Communication: Internal to the Organization	117
Knowledge Transfer/Communication: Patient-Provider	118
Perceptions and Managing Expectations of Information Technology.....	118
Analytics	119
Software as a Service	120
Aspects of Healthcare Research.....	121
Diffusion of Innovations Perceived Characteristics of Innovations	122

Limitations	124
Recommendations for Future Research.....	126
Concluding Remarks.....	129
References.....	130
Appendix A: Sample SQLite Query Code.....	141
Appendix B: Coding Procedures Guide.....	146

List of Tables

Table 3.1: Journals Sampled	49
Table 3.2: Operationalized Variables	62
Table 4.1: Number of Journal Articles by Publication	75
Table 4.2: Noun Phrase Network Information - All Years, MI Pubs, and non-MI Pub	81
Table 4.3: Noun Phrase Network Information by Year	83
Table 4.4: Variance Explained.....	87
Table 4.5: Factor Loadings for Exploratory Factor Analysis with Varimax Rotation	89
Table 4.6: Final Theme Solution	92
Table 4.7: Inter-coder Reliability.....	96
Table 4.8: Hierarchy of Articles to Themes for Publications 2002-2011 (raw results).....	98
Table 4.9: Hierarchy of Articles to Themes for Publications by Year (raw results)	101
Table 4.10: Hierarchy of Articles to Themes for Publications Spanning 2002-2011.....	104
Table 4.11: Hierarchy of Articles to Themes for Sampled Publications by Year	106

List of Figures

Figure 1: Sigmoid Adoption Curve.....	37
Figure 2: Adopter Categorization	38

Chapter 1. Introduction

It is remarkable that the first personal computers did not appear until the late 1970s, and the World Wide Web dates only to the early 1990s. This dizzying rate of change, combined with equally pervasive and revolutionary changes in almost all international health care systems during the past decade, makes it difficult for health care planners and institutional managers to try to deal with both issues at once (E. H. Shortliffe & Cimino, 2006, p. 4)

Providing adequate healthcare to the populace is an increasingly difficult task.

Healthcare providers contend with ever-increasing costs, restrictive regulations and medical identity theft. Other challenges facing healthcare providers include patients with access to a virtually limitless supply of frequently contradictory (Damman, 2010) medical information available via the internet and minimal standardization across healthcare facilities (Weigel, Landrum, & Hall, 2009). Healthcare professionals, government agencies, and other entities recognize that much of healthcare providers' time is spent on activities related to information management (E. Shortliffe, Perreault, Wiederhold, & Fagan, 2001). As a result, these bodies embrace health information systems, as a partial panacea, if not a complete cure-all, to the aforementioned healthcare challenges (Angst and Agarwal, 2009; Carroll, et al., 2002; Koonitz and Powner, 2007). The recently enacted American Recovery and Reinvestment/Health Information Technology for Economic and Clinical Health Act (ARRA/HITECH; "American Recovery and Reinvestment Act of 2009,"), Patient Protection and Affordable Care Act (PPACA; ("Patient Protection and Affordable Care Act," 2010), and Health Care and Education Reconciliation Act (HCERA; "Health Care and Education Reconciliation Act ", 2010) provide

examples of the magnitude of emphasis the United States government places on improving healthcare for U.S. citizens.

The ARRA/HITECH Act, in particular, provides specific guidelines that healthcare providers and healthcare facilities will need to meet to qualify for payment under the Medicare and Medicaid Electronic Health Record incentive program. Deemed *meaningful use*, the act specifically provided criteria that require the use of electronic/digital healthcare information systems. For example, the act requires the use of a computerized physician order entry (CPOE) system for medication orders and requires that patients be provided an electronic copy of their health information (e.g., electronic health record), upon request. Other, less explicitly described objectives related to electronic information systems include maintaining active medication lists for patients, maintaining active medication allergy lists for patients, and recording patient demographics.

Of the three laws listed above, the ARRA/HITECH Act has the most dramatic direct effect on the use of healthcare information systems. Although not as specifically focused on information systems, objectives of PPACA such as improving patient safety and reducing medical errors ("Patient Protection and Affordable Care Act," 2010, p. SEC. 2717. Ensuring the Quality of Care), also relate to the use of healthcare information systems and provide material for academic research and practitioner attention.

In defining healthcare information systems, information technology artifacts and their nomological networks are included. Healthcare information systems are comprised of systems that support the healthcare mission and include systems such as electronic medical/health records, computerized physician order entry systems, clinical decision support systems, and personally-controlled health records—a health record accessible by both physicians and patients

through varied means (Halamka, Mandl, & Tang, 2008). These healthcare information systems are tools used—in an ideal world—to improve the efficiency and effectiveness of providing healthcare to patients.

The field of research perhaps in the best position to analyze the benefits and intricacies of healthcare information systems is the field of medical informatics, which is at the intersection of healthcare and computing. At its most general level, medical informatics can be described as the fusion of medicine and information systems applied to improve and enhance patient care.

Although research involving healthcare, medicine, and technology can be found predating their 1974 book, Anderson, Gremy, and Pages (1974) and Collen (1986) provided one of the earliest mentions of the term *medical informatics*, borrowing from the French term *informatique*, a term frequently used regarding medical information science. Medical informatics as defined by Shortliffe, Perreault, Wiederhold, and Fagan is “the study of biomedical information and its use in decision-making” (2001, p. xi). Collen shared a broader view of medical informatics that includes:

medical computing, medical data processing, medical information processing, medical computer science, medical information science, medical information systems, healthcare information systems, computer hardware and software, computer and information technology, applications of computers and data processing to the health services and basic concepts of computer science fundamental to medicine (Collen, , p. 779).

Haux, in his reflection on the medical informatics discipline, provides his own self-assessed definition of medical informatics, “the discipline, dedicated to the systematic processing of data, information and knowledge in medicine and health care (Haux, 2010, p. 600).

Problem Statement

Scholars and practitioners of medical informatics may be capable of providing solutions to the challenges in healthcare required by ARRA/HITECH, HCERA, and PPACA. While some

fields of research have existed for centuries (e.g., Philosophy, Mathematics; Cajori, 1991; Dunn, 2007; Marias, 1967) medical informatics research is still in comparative infancy. With origins dating back only to the early 1960s, the field of medical informatics can be considered to be in its early stages.

Much of the research in medical informatics has attempted to answer the question of what occurred (describing the phenomenon) while leaving the answer to the question of why it occurred (applying theory) to the readers' speculation. If one considers research literature as a continuing discussion about a field or topic, there are occasions when a pause in the conversation is necessary during which participants reflect on earlier discussions and reassess future directions for the conversation. The current study provides an important opportunity to pause in the medical informatics conversation and address aspects of the research in field.

“The goal of any science is the production of cumulative knowledge” (Hunter & Schmidt, 2004, p. 17). In the sciences, the goal of research is to expand and accumulate knowledge in the field (Cooper, 2010). Scholars attempt to learn more about a particular topic by analyzing previously performed studies and by performing additional studies. This iterative and repetitive process has led to the current state in the field of medical informatics. However, to provide for the continued growth of a field, it is important to summarize a body of knowledge, lest the number of studies performed in the field becomes too large for an individual scholar to consume. Without such summaries, the sheer volume of studies becomes a morass for an individual scholar to wade through and slows the progress of new knowledge. For example, a simple query for the Boolean phrase “*medic* AND comput**” (the asterisk equating to a wild card value) in publications between 1950 and 1980—a period with substantially fewer computers than today—yields 5,594 citations. Update the search to include the years 1980 through 1990

and the number of citations jumps to over 26,000. It is unreasonable to expect that each medical informatics researcher would read and absorb this huge volume of material.

A more manageable approach is to combine studies into single consolidated reviews based on a particular theory, topic, or theme. This study analyzes medical informatics from the perspective of the medical informatics literature and the related fields of the management information systems and the field of medicine. Taking a sequential mixed methods approach to the analysis, this study synthesizes medical informatics literature from two leading journals in the field of medical informatics, three leading journals from the closely related field of management information systems, and two leading medical journals. Using these publications, the study extracts themes from the literature and identifies challenges and future directions for the discipline. The next section elaborates on this purpose in detail.

Purpose

The purpose of this content analytic inquiry is to ascertain what has been written about medical informatics since its inception in the 1960s and to extend the discussion based on current themes of discussion in the literature. The quantitative first phase of the investigation takes a semi-automated approach through noun phrase identification. The second, qualitative phase is manual coding using categories developed from the literature review. Investigating a broad spectrum of seven scholarly publications including medical informatics, management information systems, and medical publications, the study attempts to discern the state of the field over the past decade—from 2002 to 2011—and discover directions for future research in the academic and practitioner-based areas of the medical informatics field.

Research Questions

This study elicits from the literature the themes and topics that are the most frequent topics of conversation in the professional literature. This study examines several aspects of the journal publications and the articles that focus on medical informatics to ascertain what difficulties, if any, have arisen in the field, and what is the context of the current trends in medical informatics. Specifically, given that the field of medical informatics continues to grow and appears to have a bright future, this study attempts to answer the following research questions and sub-questions:

Research Question 1: What themes have emerged in the medical informatics discipline since its inception in the 1960s?

Sub-question 1: What themes emerge from the medical informatics literature?

Sub-question 2: What medical informatics themes emerge from the related fields of medicine and management information systems?

Research Question 2: What challenges exist for the medical informatics discipline?

Research Question 3: What future directions does the literature suggest for the field?

These questions provide the basis for the coding and analysis of the sample of articles collected. The diffusion of innovations theory provides the theoretical framework for the research. Using both qualitative and quantitative content analysis methods, the research obtains and analyzes the data. Subsequent chapters explain in detail both the theoretical framework and analytic methods.

Worldview

The approach to this study is guided by a pragmatic worldview; rather than being tied to a specific approach, the methods used are appropriate based on the research problem or the

research questions. In essence, pragmatists believe that what works at the time of the research study is appropriate. The focus remains on the research questions and the methods used maybe quantitative, qualitative, or mixed methods. By focusing on the research questions and not the methods in particular, pragmatic researchers have greater freedom to choose what works at the time of the study and what is necessary for the particular study. The pragmatic worldview allows a focus on the problem at hand and uses as many available techniques and approaches as necessary in an attempt to answer the problems and the research questions (Rossman and Wilson, 1985, from Creswell & Clark, 2007, p. 10).

While mixed methods studies are gaining in popularity, there are detractors of this research approach. The primary argument against the mixed method design lies in what is considered *incompatibility theory*. This theory asserts that qualitative methods and quantitative methods “cannot and should not be mixed” (Brannen, 2005; Howe, 1988, p. 14; Onwuegbuzie & Leech, 2005). Evaluating the opposing reasoning approaches (logic) to the methods, deductive for quantitative and inductive for qualitative, one might consider the premise behind incompatibility theory plausible. Likewise, the positivist worldview behind quantitative studies seems to contrast starkly that of the interpretivist and constructivist worldviews found in qualitative research. However, “the goal of mixed methods research is not to replace either of these approaches but rather to draw from the strengths and minimize the weaknesses of both in single research studies and across studies” (Johnson & Onwuegbuzie, 2004, pp. 14-15). Likewise, Brannen (2005) argues that there is more overlap than dissimilarity between the two methods and encourages readers to reconcile the paradigms that associate, for example, numbers with quantitative methods and words with qualitative methods. Taking the mixed methods approach, researchers can use qualitative data to support quantitative results and vice versa.

Scholars synthesizing the two methods are able to provide a rich understanding of phenomena of interest.

Significance of the Study

The study is significant for at least three reasons. First, this study constitutes a starting point from which to understand the evolution of the medical informatics. Starting with a literature review to examine the field from its start in the 1960s, the study analyzes the literature of the last decade to identify themes and future directions the field may take. Further, despite several calls to study medical informatics discussions outside the boundaries of journals focused specifically on medical informatics, there is a void in such research. This study identifies emerging themes related to medical informatics by spanning the medical, management information systems, and the medical informatics literature.

Another contribution of this study lies in its methods. This study introduces a quantitative method of text analysis, *centering resonance analysis* (Corman, Kuhn, McPhee, & Dooley, 2002), to the field of medical informatics. Centering resonance analysis (CRA) is a “text analysis method that has broad scope and range and can be applied to large quantities of written text and transcribed conversation. It identifies discursively important words and represents these as a network, then uses structural properties of the network to index word importance” (Corman, et al., 2002, p. 157). As may be inferred from the previous quote, CRA can handle voluminous bodies of text that would be extremely difficult for manual, human coding.

Those who work in the information systems or medical informatics areas will find this study useful because it provides current, viable information about the field that enables the practitioners to be more effective. Summarizing the medical informatics literature to date offers

a concise, meaningful understanding of the state of the discipline. Without such summaries, the sheer volume of studies becomes unwieldy for an individual scholar to consume. The results of this study provides a single consolidated review as a reference point for the field and its current state.

As previously stated in the purpose of the study, the three-fold substantive objectives of this study are (a) to understand the themes and topics that have received heightened attention in the literature, (b) to discern the state of the medical informatics field from 2002 to 2011, and (c) to discover possible directions for future research in the academic and practitioner-based field of medical informatics. While there are several definitions of medical informatics, the prevailing definitions include aspects of information systems/information technology used to provide more effective and/or efficient medical care of patients (Anderson, et al., 1974; Collen, 1986; E. Shortliffe, et al., 2001). For the purpose of this study, medical informatics is defined as *the discipline dedicated to the systematic processing, analysis, and dissemination of health-related data through the application of digital information systems (computers) to various aspects of healthcare, research, and medicine.*

Summary of Chapter One and Outline of Chapters

Chapter One

Chapter One introduced the research study, provided an overview of the context of the study, introduced the research questions, the sub-questions, and the research problem. The chapter discussed the objectives of the study and situated the topic within the literature of medical informatics, medicine, and management information systems. Chapter One introduced the significance of the study and explained the pragmatic worldview used in the study.

Chapter Two

Chapter Two addresses the literature review. Because the study is a content analysis of the medical informatics literature, the literature review will differ from standard research designs. Chapter Two provides an overview of the medical informatics literature, develops the context in which the issue resides, the types of questions that have been asked and answered, and the types of research methods previously applied. The research concludes the review of the literature with a discussion of theory development and the overriding theory for this study. *Diffusion of Innovations* is an appropriate theory for understanding the growth in the medical informatics literature and Chapter Two provides a detailed discussion of this theory in the context of this study.

Chapter Three

Chapter Three describes the research design for this study. The chapter includes an overview of content analysis, a discussion of the method of data collection, an explanation of the two methods of analysis—quantitative and qualitative—and a discussion of the applicability of the methods to this study. In discussing the data collection, Chapter Three addresses the body of texts sampled and the inclusion and exclusion criteria. From the data collection discussion, the chapter explains the coding categories, and the reliability to be achieved. The chapter provides an explicit enumeration of the steps in the content analysis procedure, including the data cleaning and the identification of the themes and topics identification, and the inferential techniques used.

Chapter Four

The findings chapter, Chapter Four, presents the results of the quantitative and qualitative analyses performed. This chapter shares the manifest findings, those that are explicitly visible, presented in the form of themes derived from a quantitative semi-automated approach, centering resonance analysis and exploratory factor analysis. Proceeding from the qualitative analysis—a manual coding of the texts—the chapter develops the findings and presents them as primary themes of research that were identified during the literature review and adapted during the subsequent methodological stage.

Chapter Five

Chapter Five summarizes and concludes the dissertation by providing the interpretation and discussion of the findings. This chapter addresses the strengths and limitations of this study and elaborates on future directions for research on the topic. The chapter closes by summarizing the chapter and study with general comments about the contributions to both the academic field and the practitioner aspects of medical informatics.

Chapter 2. Literature Review

The literature review is used for, among other things, building a cohesive understanding of previous research in a particular field, filling gaps in the scholarly discussion of the topic, and setting the framework for situating a study in a context in which the issue resides (Cooper, 2010; Creswell, 2009; Krippendorff, 2004). The chapter begins the chapter with a discussion of the origins of digital information systems. The discussion proceeds to integrate healthcare and information systems history, thereby providing an understanding of the origins of the field of medical informatics. Following this discussion, the chapter overviews several key analyses of medical informatics literature and the outcomes of those studies to allow the reader to gain a historical perspective of the development and the growth of the field since inception.

Origins of Medical Informatics

ENIAC, the *Electronic Numerical Integrator And Computer*, was the first electronic digital computer—able to solve a wide variety of general-purpose problems. Although originally designed to calculate artillery firing tables for the United States Army during World War II (Fritz, 1996), ENIAC arguably started an information systems explosion and changed our world forever. While some pundits thought businesses were not interested in the ENIAC, it was only a few years from ENIAC's public operational date in 1946 that the digital computer entered the business world when the International Business Machines Corporation—more commonly known as IBM—began developing computers for commercial and government use (Cortada, 2006).

Thus began the entry of information systems into the world of business as tools to aid managers in making better decisions. On the medical front, in those early years of digital computing, the sheer costs of maintaining the large systems and staff required to support the systems was prohibitive for all but very large medical facilities of first world countries (Tan and Global, 2009).

Interestingly, while ENIAC was developed for use by the U.S. Army for computing artillery trajectories and subsequent digital computer applications extended into the business world, the concept for the electronic computer can be traced back to the U.S. Army Medical Department in 1879 (Collen, 1986; Ginn, 1997). Major John Shaw Billings, then Assistant Surgeon General of the U.S. Army, was asked to assist the U.S. Census Bureau for the 1880 and 1890 censuses. At that time, the census was performed manually and required tedious hand sorting. A rapid growth in the U.S. population made the task overwhelming. Billings conceptualized the idea of using cards with notches in them to represent the census data. Billings shared the concept with a statistical engineer, Herman Hollerith, who took the concept from idea to reality with the invention of punch cards. Following the 1890 U.S. census, Hollerith's punch cards became a more widely used tool, being adapted for use in public health surveys and eventually converted for use in electro-mechanical tabulating systems. In 1896, Hollerith incorporated his company as the Tabulating Machine Company. Over many years and through multiple mergers, the company ultimately evolved into IBM (Ginn, 1997). Eventually, the mechanical computing devices were replaced with digital computers and the punch cards became a source of data tabulation by these digital computers (Collen, 1986; E. H. Shortliffe & Cimino, 2006).

During the 1950s and 1960s, the growth of digital computer use continued. Many programming languages were developed including FORTRAN (*FORmula TRANslation*), COBOL (*COmmon Business-Oriented Language*), BASIC (*Beginners All-purpose Symbolic Instruction Code*), and LISP (*LISt Processing*) among others (Collen, 1986). As indicated by the name, one language in particular—MUMPS (*Massachusetts General Hospital Utility Multi-Programming System*)—was developed with a healthcare focus in mind (Collen, 1986). The MUMPS language was designed to provide database applications to multiple users simultaneously and is still in use in healthcare applications today. As examples, the U.S. Department of Defense and the U.S. Department of Veterans Affairs (an early adopter of MUMPS), still use the language as part of their electronic medical record systems. As was the case with most languages developed in the early age of digital computing when memory and storage were at a premium, MUMPS was designed for terse coding and was not *user-friendly*. Instead, one might consider MUMPS an *expert-friendly* language in that it requires extensive training to learn to program it.

The medical field has existed for hundreds of years and management literature has existed for over a century. Since the early days of digital computing, managers have applied information systems to solve problems and support decision-making. However, the fields of management information systems and medical informatics are in relative infancy. Since the start of the first formal university-level management information system (MIS) program in 1968 (Nolan & Wetherbe, 1980), scholars have attempted to evaluate information systems and the value these systems provide organizations. As one such attempt, DeLone and McLean (1992) developed the Information Systems Success model. In their model, the authors analyzed and derived six categories defining information systems success from previous research, created an

interdependent model of these categories toward success outcomes, and initially labeled the outcomes as individual impact and organizational impact. Their purpose was to provide a general framework from which scholars could better understand the construct of information systems success (DeLone & McLean, 1992).

In his 1980 paper from the International Conference on Information Systems, Keen explained the necessity of discussing the field of management information systems and its development as a standard field of scientific research (1980). Although directed toward the management information systems discipline, his exhortations readily apply to the field of medical informatics and suggest several questions for the young discipline. What are the origins of medical informatics? What do we see emerging in medical informatics research? In what directions do we see the field heading? How does the field affect patient healthcare? This study attempts to answer these questions. A chronology of studies from medical informatics literature will prove useful in establishing the context of this study.

Chronology of Medical Informatics

1970s and Earlier

An indicator of the growth of a scientific field is the creation of scholarly research in the field. Although books are included in the definition of scholarly research, the primary means of sharing academic research is through the publication of peer-reviewed journals. The earliest peer-reviewed journals in medical informatics included *Computers in Biomedical Research*, *Computers in Biology and Medicine*, *Journal for Clinical Computing*, and *Computers in Medicine*, started in 1967, 1970, 1972, and 1972, respectively (Collen, 1986). These journals

provided a forum of discussion for sharing conceptual, theoretical, and empirical knowledge to stimulate further growth of knowledge in the field.

As early as 1961, with the publication of Dixon's *BIMD Computer Program Manual, Volume 1*, books and compilations expanding on the concept of medical informatics began to appear (Collen, 1986). Other early publications Collen (1986) discusses include Sterling, Pollack, and Center's *MEDCOMP: Handbook of Computer Applications in Biology and Medicine* (1964), Proctor and Adey's *Proceedings, Automatic Data Processing of Electrocardiography and Electroencephalography* (1964), and Atkins' *Proceedings of the Symposium, Progress in Medical Computing* (1965). Although these publications themselves did not demonstrate any significant turning points in the field of medical informatics, they helped to establish the foundation for this new scientific paradigm of medical informatics.

As previously stated, it was not long for the fields of medicine and information systems to integrate. Following the operational date of 1946 for ENIAC, references to medicine and computers together began to appear in the 1950s in the fields of "biophysics, bioengineering, and biomedical electronics publications" (Collen, 1986, p. 778). The number of articles published on topics that included both computing and medicine began to grow and in the early 1970s, some believed there should be a new name generated for the emerging domain of knowledge. Although there was little debate over using the term *medical*, there was some concern over what term to use to encompass and describe the aspects of technology, education, and engineering that were developing in the field.

In his essay, Collen expressed that medical informatics was "a new knowledge domain of computer and information science, engineering and technology in all fields of health and medicine, including research, education and practice" (Collen, 1986, p. 778). Citing Anderson in

written communication from May 1986, Collen described the origin of the term medical informatics as developing from a combination of the French terms, *informatique* and *automatique*—terms that were used in Europe in describing medical information science or data processing—and the English word, *medical*. Thus, perhaps the earliest mention of the term *medical informatics* came from Anderson, Gremy, and Pages' Education in Informatics of Health Personnel (1974). Despite not defining the term “medical informatics” in their book, Anderson et al. use the term frequently throughout the book and provided a foundation structure for developing a curriculum to teach medical informatics (1974). Collen inferred from Anderson, et al. that the authors believed that medical informatics included all of the following:

medical computing, medical data processing, medical information processing, medical computer science, medical information science, medical information systems, health care information systems, computer hardware and software, computer and information technology, applications of computers and data processing to the health services and basic concepts of computer science fundamental to medicine (Collen, 1986, p. 779).

The discussion over the term *medical informatics* and its definition continued. In 1977, as the program chair for the Third World Conference on medical informatics, Collen defined medical informatics as “the application of computer technology to all fields of medicine-medical care, medical teaching and medical research” (Collen, 1986, p. 779). Others felt that medical informatics was broader in scope. The discussion continued into the 1980s.

1980s

Members of the medical informatics community met at the Symposium on Medical Informatics in 1985 and further discussed the definition of the term, *medical informatics*. The outcome of the discussion was that *medical informatics* was expanded beyond the computing device and the information it processes, to include “all medical research and development,

education and medical practice, including physician assistance functions such as clinical decision support and expert consultant models” (Collen, 1986, p. 780).

In his historical look at the field, Collen provided a descriptive narrative discussion of the field rather than an analysis. In his defense, his goal was to provide a historical chronology of the evolution of the field of medical informatics. He organized the paper by discussing the origin of the term *medical informatics* and followed it with a discussion of the origins of the field itself, discussed previously in this chapter.

While at first thought one might not recognize it, an early adaptation of medical informatics was in the area of the medical library. As Collen explained, “A major contribution to medical informatics occurred when [the National Library of Medicine] initiated computerizing the *Index Medicus* with the printing of the 1964 edition and implemented the Medical Literature Analysis and Retrieval System (MEDLARS)” (Collen, 1986, p. 781; emphasis in the original). In 1977, MEDLARS expanded with a networked connection of medical libraries and evolved into MEDLINE (for MEDLARS online). The MEDLINE database is still widely in use in 2011.

Although more of a methodology paper than one of medical informatics literature analysis, Greenes and Siegel recognized the rapidly changing field of medical informatics and other emerging fields (1987). Further, they noted the problems that the changes represented for the National Library of Medicine because the field was “in a state of flux, and there [was] a lack of generally agreed upon definitions of the boundaries and structures of the field” (Greenes & Siegel, 1987, p. 411). Aggravating the problem, the medical informatics field is one that crosses multiple disciplines—not only those connected with healthcare. These characteristics made it difficult to track articles that should be included in the *medical informatics* category of the National Library of Medicine. In an effort to create a method to evaluate and aggregate literature

in emerging fields, Greenes and Siegel used both objective and subjective quantitative methods to establish the scope and define the content of the medical informatics field (1987). Combining citation analysis and survey methods, Greenes and Siegel used the multi-factorial method to develop one of the first academic rankings of the medical informatics journals and proceedings. By polling the membership of the American College of Medical Informatics, Greenes and Siegel also determined which proceedings and journal publications were considered important to the members of the medical informatics field. While their self-assessed poor survey compliance rate of 45% achieved may have been a weakness of their study, Greenes and Siegel provided substantial recommendations to the medical informatics discipline and methods to evaluate the discipline. Among their recommendations and conclusions were that impact factors—measures reflecting the mean number of citations of articles published and often considered a proxy for the relative importance of a publication—“do not reflect well the specific considerations about the importance of the publications to the field of medical informatics” (Greenes & Siegel, 1987, p. 414). Greenes and Siegel suggested that co-citation maps and tracing citation frequencies might produce results closer to that of the peer group survey. Perhaps another indicator that the debate over the scope of medical informatics was ongoing can be found in their final recommendation, which was a suggested definition for medical informatics:

the field concerned with the cognitive, information processing, and information management tasks of medical and health care, and biomedical research, and with the application of information science and technology to these tasks (Greenes & Siegel, 1987, p. 414).

What is notable is the specific inclusion of *information management* in their definition. In the 1980s, the related field of *Management Information Systems* was also evolving and developing. Despite parallel paths, the two fields did not appear to share much knowledge. That situation

would change in the 1990s as the rate of information technology growth became more exponential and computers started appearing on virtually every desktop.

1990s

During the 1990s, researchers began analyzing the body of medical informatics literature at a level beyond that solely of journal rankings. They produced analyses that used specific article citations as the units of analysis. However, the debate over the scope and definition of the field continued. As evidence, Morris and McCain began their essay about “the disciplinary nature and internal structure of the [medical informatics] field” with an extended discussion about previous definitions of the field (Morris & McCain, 1998, p. 448). Perhaps their own suggestion of the two primary characteristics that are encompassed in most medical informatics definitions should suffice for the definition itself: “references to health sciences, biomedicine, and the healing arts; and reference to the use of information management techniques and technologies in support of those pursuits” (Morris & McCain, 1998, p. 448). Despite the lengthy initial definitions of their medical informatics discussion, Morris and McCain’s actual objective was to assess and understand the multidisciplinary structure of the medical informatics field and its relation to similar fields. Morris and McCain wanted to determine which articles defined the core set of journal literature in medical informatics and they used inter-citation network analysis as the method to make their determination. In their initial data inclusion phase, Morris and McCain included two non-medical informatics journal titles that they subsequently removed based on the lack of a substantial number of citations made and received among the candidate “core journals” they used as their starting point. As a reminder, one of Morris and McCain’s goals was to identify the core medical informatics literature and that supported the removal of

the non-medical informatics journals despite the loss of potentially interesting information. In their defense Morris and McCain admitted,

only a subset, or core set, of journals are being considered—not all scientific journals that might contribute or receive medical informatics citations, or all journals relevant to medical informatics overall, or even all journals that might be of interest to medical informatics researchers (Morris & McCain, 1998, p. 454).

This is important to note because an objective of the current research is to identify what potentially informative medical informatics knowledge can be found outside the core medical informatics literature.

Using cluster analysis and multidimensional scaling, Morris and McCain identified five unified groupings of research in the journals of their final data set. Of these five themes, *General Medical Informatics*, *Decision Making*, *Biomedical Computing*, *Computing in Biomedical Engineering*, and *Education*, they found that the *education* grouping was relatively isolated from the other four. This isolation indicated that the researchers writing the articles in Morris and McCain's sample demonstrated a clear demarcation between education and the other four groupings (i.e., those that are more related to practitioner-based aspects of medical informatics and academic research focus). Morris and McCain's findings were interesting in that they focused on extracting thematic categories, which differed from earlier studies such as Sittig and Kaalaas-Sittig's biomedical informatics journal ranking study (1995).

Deeming citation analysis alone insufficient to answer the question of publication rankings, Sittig and Kaalaas-Sittig developed their study as a more encompassing, comprehensive look (1995). Sittig and Kaalaas-Sittig analyzed the medical informatics literature using a multitude of evaluation criteria that included impact factors, total citations, survey data of the American College of Medical Informatics Fellows, and interlibrary loan requests to the U.S. National Library of Medicine. They determined that the publications *Computers and Biomedical*

Research, MD Computing, Methods of Information in Medicine, Medical Decision Making, and *Computers in Biology and Medicine* comprised the top publications in the field of biomedical informatics at the time of the publication (Sittig & Kaalaas-Sittig, 1995). While Morris and McCain focused on research themes in their study, one outcome of their study was a ranking of 20 core medical informatics journals that included many publications that were in Sittig and Kaalaas-Sittig's study (Morris & McCain, 1998).

2000 to Present

After the turn of the century, concerns arose that the field of health care needed to improve on the effective use of information (W. R. Hersh, 2002). Despite the technological advances resulting in cheaper cost of ownership, information systems have become more complex, and the cost of training users and maintaining large fleets of individual systems has increased. Concerns about patient safety, patient privacy, and medical errors are issues that some believe information systems may be able to reduce or eliminate (W. R. Hersh, 2002). Electronic medical records (EMR) and computerized physician order entry (CPOE) systems are possible solutions to some of these concerns. For example, the CPOE frequently have built-in error detecting software to alert healthcare providers of potential harmful medication interactions. However, using information systems in healthcare often requires more time from the healthcare provider than using paper methods. Although the costs may be made up in other areas such as improving documentation, more accurate coding for insurance billing, or error reduction, it is difficult to measure the values gained versus the costs expended (W. R. Hersh, 2002).

Despite the almost thirty years that had passed since the first use of the term *medical informatics*, the discussion over the definition of the field continued into the 21st Century (W. R.

Hersh, 2002). In his paper about the then current status of the medical informatics field, Hersh discussed the views that medical informatics is a “service (e.g., helping clinicians implement informatics applications),” and his preferred view that medical informatics is “a science that addresses how best to use information to improve health care” (W. R. Hersh, 2002, p. 1955). The latter perspective seems to be the viewpoint of a preponderance of the scholars in the field. More editorial in nature than the previous reviews of the field, Hersh’s essay suggests the future of medical informatics will move toward addressing the “imperatives of improving documentation, reducing error, and empowering patients” (W. R. Hersh, 2002, p. 1957) While his study was not an empirically based analysis, Hersh declared a number of core themes for the field of medical informatics. Of those he discussed, he detailed the value of standardization in terminology for aggregating and comparing data across different healthcare facilities and entities. An example of standardizing can be found in the U.S. Department of Defense’s electronic medical record, AHLTA, which provides healthcare providers a “drill-down” approach of entering pre-selected diagnoses instead of free-text entry. In addition to providing a means of standardizing the entry, the drill-down method aids in proper insurance account coding. Hersh also believed that the systems should be usable - they should be integrated into the healthcare workflow and produce an appreciable overall benefit. In a prophetic view, Hersh pointed out that healthcare providers would have more direct interaction with the medical informatics systems; the need to have more accurate documentation for billing and for evidence to support adequate medical error avoidance will be driving forces behind the increased interaction (W. R. Hersh, 2002).

In a methodological paper focused toward the library and information science researchers, Andrews’ (2003) study of the medical informatics literature provided useful insight

into the medical informatics discipline. Using the bibliometric¹ method of author co-citation analysis similar to that of Morris and McCain (1998), and Sittig and Kaalaas-Sittig (1995), Andrews evaluated the medical informatics literature using *authors* as the units of analysis. By counting the number of times authors were cited together by a third author, Andrews measured the “distances” of medical informatics scholars from one another; the “underlying assumption of [author co-citation analysis] is that the more two authors are cited together, the closer the relationship between them” (Andrews, 2003, p. 47). Andrews recognized that despite the studies previously reviewing medical informatics literature, none had studied the relationships among the leading authors in the medical informatics discipline. As previously mentioned, Andrews’ (2003) primary focus was on librarians and information scientists and therefore, his goal was for his study to be a tool that the librarians and information scientists could use to serve the medical informatics discipline.

Using the 196 American College of Medical Informatics fellows as his population data source, Andrews distilled the author list to 50 authors after initial analysis and determined that the five-year period from 1994 to 1998 provided an adequate sample of articles. Looking solely at the frequency counts of number of times cited, the top five authors in the medical informatics discipline (for Andrews’ sample) were *Beck J.*, *Cimino J.*, *Pauker S.*, *McDonald D.*, and *Clayton P.*, names likely to be recognized by many medical informatics researchers (2003). The Andrews study parallels that of Morris and McCain with the obvious distinction that Morris and McCain used publication titles as the units of analysis while Andrews used the authors’ names (Andrews, 2003; Morris & McCain, 1998). Both sets of authors (Morris and McCain, and

¹ Bibliometrics is “the use of statistical methods to analyze a body of literature to reveal historical development and as the scientific and quantitative study of publications” (DeShazo, et al., 2009, p. 7)

Andrews) performed cluster analysis, factor analysis, and multi-dimensional scaling analyses on their data. While Morris and McCain determined five generally cohesive categories of medical informatics literature with their cluster analysis, Andrews identified both a six-cluster and a three-cluster solution. The strength of Andrews' study can best be expressed in his own words:

it can be one of several tools used to help individuals access and visualize scholarly communication within the field. For instance, while those familiar with the medical informatics community and its literature will already know that, say, McCray and Campbell work in similar areas and are often cited together, those who are not well oriented with the field, particularly new researchers ... could find such information useful (Andrews, 2003, p. 55).

As Hersh (2002) discussed, Andrews also identified the need for language standardization in the medical informatics literature. Andrews, citing a Cimino (1998) article, suggested that there is some ambiguity in the terms used in the field and recommended more research along those lines. Additionally, Andrews suggested that future research ideas might include a look at medical informatics literature as an inter-disciplinary perspective - not relying solely on the medical informatics journals, but including other disciplines' publications to explore and share knowledge across fields (2003).

As one might expect in any fast-growing field of research, the rate of growth in academic publication output grows dramatically. As such, there tends to be a greater need or desire for reviews of said field to provide clearer understanding and greater data-reduction of the ever-expanding body of literature. The field of medical informatics is not an exception. Since the turn of the century, there are more reviews of the medical informatics literature than in the combined decades since the discipline's inception.

Completing the discussion of the first five years of the 21st century, we consider Eggers, Huang, Chen, Yan, Larson, Rashid, Chau, and Lin's (2005) discussion of the literature. Analyzing citation and literature data from the years 1994-2003, Eggers, et al. identified

prominent authors, primary topics in the field, and the relationships among them. Eggers, et al. used basic analysis, content map analysis, and citation network analysis in an attempt to reduce the vast literature available to a usable summary. As suggested by Andrews, Eggers, et al. attempted to expand their analysis beyond the confines of the medical informatics literature by including a MEDLINE search for the term *medical informatics* in building their data sample. Additionally, in collecting their sample, Eggers, et al. included 22 medical informatics journals—18 of which are found in Andrews’ (2003) study—medical informatics keywords in a MEDLINE search, and articles authored by American College of Medical Informatics fellows. As Andrews did, Eggers, et al. focused their analysis at the author level for their basic analysis. Eggers, et al. identified *Cimino, J., Hasman, A., Greenes, R., Miller, P., and Haux, R.*, as the five most prolific authors in the field by publication count alone (Eggers, et al., 2005, p. 45). However, looking at the citation counts—the measure Andrews used for frequency counts—we see the top five authors were *Bates, D., Cimino, J., McDonald, C., Patel, V., and Hripcsak, G.* (Eggers, et al., 2005, p. 46).

To identify trends in the medical informatics literature over the ten years of their literature sample, Eggers, et al. (2005) segregated the literature into three periods, 1994-1997, 1998-2000, and 2001-2003. Using the *Arizona Noun Phraser* software to extract medical noun phrases, the team of researchers built content maps of the topics derived for each year group segregation and compared the results. The authors identified newly-emerging topics of discussion in the literature (e.g., Human Genome, Medical Imaging, Neural Networks, etc.) and those areas of the literature that were not growing rapidly (e.g., Hospital Information Systems; Eggers, et al., 2005). It is important to note that while other noun phrase identifying software is available, the Arizona Noun Phraser software was produced by University of Arizona

researchers and is no longer available online, thus making the study less replicable. An interesting outcome of the research is that Eggers, et al. (2005) identified many pattern changes among the three content maps, supporting earlier research about the fast growth and change in the medical informatics discipline (Andrews, 2003; Collen, 1986, p. 779; Eggers, et al., 2005; Greenes and Siegel, 1987; Morris and McCain, 1998; Sittig and Kaalaas-Sittig, 1995).

Using an approach similar to Andrew's (2003) approach and Morris and McCain's (1998) multi-dimensional scaling approach, Eggers, et al. (2005) created an author map analysis to group individual authors based on their research interests. While the author maps are not reproduced in this study, it is interesting to note that as computer technology has improved over the years, we are seeing a trend toward improved visualization techniques and tools that aid the researcher in understanding the data under investigation.

As of 2009, the discipline of medical informatics still seemed to be struggling with its identity (DeShazo, LaVallie, & Wolf, 2009). Despite the established MEDLINE definition for medical informatics—"the field of information science concerned with the analysis and dissemination of medical data through the application of computers to various aspects of health care and medicine"—other definitions continue to appear in the literature (DeShazo, et al., 2009, p. 7). As previously mentioned, most definitions include references to an inter-disciplinary field and include scientific research as an aspect of the medical informatics field. However, DeShazo, et al. suggested that some scholars question where in science medical informatics should be positioned (2009). Further, some question whether it should be considered a distinct field at all (DeShazo, et al., 2009).

To categorize medical informatics literature in their study, DeShazo, et al. used the MEDLINE definition of medical informatics and included literature assigned by MEDLINE to

the medical informatics category. While the DeShazo, et al. study focused on the medical informatics discipline and specifically the journal and article units of analysis, the authors performed it using bibliometric methods similar to those of aforementioned studies (Andrews, 2003; Morris and McCain, 1998; Sittig and Kaalaas-Sittig, 1995) - the difference being primarily of the data sample rather than of the methodology applied (DeShazo, et al., 2009). DeShazo, et al. selected MEDLINE publications over the twenty-year period from 1987 to 2006 for articles categorized as medical informatics publications. As in the previous studies, DeShazo, et al. performed frequency counts and citation analyses. Additionally, they evaluated their sample of 77,023 articles to determine an exponential average growth rate curve of 12% each year and the exponential curve appeared to explain 97% of the variance versus a linear curve that explained only 79% (DeShazo, et al., 2009).

DeShazo, et al. generated, among other output, two journal ranking lists - one based on journal citation reports and one based on MeSH (Medical Subject Headings from the MEDLINE database) medical informatics terms. The top five publications based on the medical informatics MeSH index were *Proceedings/IEEE Engineering in Medicine and Biology Society Conference*, *IEEE Transactions on Image Processing*, *Medical Physics*, *Proceedings of AMIA Annual Symposium*, and *Studies in Health Technology and Information*. The top five publications based on the journal citation reports were *IEEE Transactions on IT in Biomedicine*, *Journal of the American Medical Informatics Association*, *International Journal of Medical Informatics*, *Methods of Information in Medicine*, and *Biomedizinische Technik*. Of note, the top three publications on the journal citation reports list, *IEEE Transactions on IT in Biomedicine*, *Journal of the American Medical Informatics Association*, and *International Journal of Medical*

Informatics, also appeared on the top medical informatics MeSH indexed list at rankings 23, 25, and 35 respectively, indicating high citation strength of each of the publications.

Through their analysis, DeShazo, et al. found that medical informatics topics are found in both medical informatics-specific publications and non-medical informatics-specific publications (2009). While that statement may seem overly simplified, the relevance is that DeShazo, et al. found little evidence of clearly demarcated lines between medical informatics literature and non-medical informatics literature.

DeShazo, et al. did not identify clusters of topics that were solely focused on medical informatics. In practically no cases did MEDLINE indexers categorize articles as only medical informatics articles. In addition, DeShazo, et al. found that a substantial number of articles are published in journals not typically identified as medical informatics specific journals. They found that over 100 journals publish at least 20 medical informatics MeSH-indexed articles per year—a stark contrast to the 20 publications Morris and McCain (1998) identified as core medical informatics publications two decades earlier. The DeShazo, et al. article revealed that while they focused their research on the MeSH term, “medical Informatics,” a broader term of “informatics” would have included literature in the categories of *dental informatics*, *nursing informatics*, *public health informatics*, and *medical informatics*, and thus, would likely have altered the citation counts and study outcomes (2009).

One of the conclusions DeShazo and colleagues determined is that medical informatics articles can be found with greater frequency than in the past in non-medical informatics journals. This finding supports the idea that medical informatics is multidisciplinary and that medical informatics is becoming more established as a discipline. It suggests, too, that future researchers

should include non-medical publications when collecting data for assessing the medical informatics discipline.

While most of the authors of previous reviews of the medical informatics literature have used bibliometric analytic methods in describing the discipline, Schuemie, Talmon, Moorman, and Kors (2009) chose a semantic approach, extracting *n-grams* from the combined titles and abstracts of all 6,000,000 MEDLINE records for the years 1993 - 2008. *N-grams* are “sequences of words that occur in the text,” and in the case of the Schuemie, et al. (2009, p. 77) paper, includes sequences of one word (*unigrams*), two words (*bigrams*), and three word (*trigrams*). Once they completed their n-gram extraction and created profiles of the document sets to categorize them, the authors cluster analyzed the n-grams and generated a two-dimensional depiction of the clusters’ journal titles—another example of increased use of visualization techniques—with the distance between journals indicating approximate dissimilarity between the n-gram profiles of each of the journals. Within the visualization, the authors further demarcated the journals with a large shaded circle indicating a distinct set of coherent journals. They concluded from their results that the domain of medical informatics literature encompassed 16 journals, the top five of which include: *Medinfo*, *International Journal of Medical Informatics*, *Proceedings of the Medical Informatics Association Symposium*, *Proceedings of the MIE conferences*, and *Methods of Information in Medicine*. Interestingly, the *Journal of the American Medical Informatics Association*, a journal originating in 1994—and ranking sixth in the Schuemie, et al. study—continues to appear near the top of the journal rankings of the literature reviews. This is particularly noteworthy since many of the other journals are much older than JAMIA. Using their top 16 journal publications, the authors used cluster analysis to identify three predominant categories within the medical informatics domain:

- 1) the organization, application, and evaluation of health information systems,
- 2) medical knowledge representation, and
- 3) signal and data analysis (Schuemie, et al., 2009, p. 76).

Using the clusters of journals they extracted in the earlier phase of the study, Schuemie et al. (2009) categorized the article counts over five three-year time periods. Basing their conclusions on this categorization and in contradiction to earlier studies, they suggested that the medical informatics discipline has remained relatively stable. Further analyzing the topic clusters, they concluded that while most medical informatics journals have changed focus (i.e., moved into different topic clusters) over the fifteen year period of their study sample, four of the publications maintained a stable and consistent focus: *Computers, Informatics, Nursing*, the *Proceedings of the AMIA Symposium*, *Computers in Biology and Medicine*, and the *Journal of Medical Internet Research*.

Another contribution of the Schuemie, et al. (2009) paper is found in their discussion of the subjective bias of journal selection found in other literature studies. Schuemie et al. believed their semi-automatic approach of journal selection was superior to subjective manual methods, but it is not without its own bias. By selecting the MEDLINE database as their source of data, they have added subjective bias to their own study and excluded publications that address medical informatics topics but are not MEDLINE publications.

Like Collen (1986), Haux (2010) provided an overview and perspective of the medical informatics literature rather than an in-depth literature analysis. Haux provided his insight and perspective of the past, present, and possible future directions for medical informatics (2010). As with virtually all the aforementioned essays, the question of a medical informatics definition arises—and Haux' discussion is not different in that aspect. However, he provides an admittedly

simple definition: “the discipline, dedicated to the systematic processing of data, information and knowledge in medicine and health care” (Haux, 2010, p. 600). For the purpose of this study, all definitions are synthesized to define medical informatics as *the discipline dedicated to the systematic processing, analysis, and dissemination of health-related data through the application of digital information systems (computers) to various aspects of healthcare, research, and medicine.*

Haux confirmed that medical informatics is a young field, particularly when compared to other medical fields, and it is directly related to the development of digital computers and other information and communication technology (2010). Another aspect that recurs in the literature is that medical informatics, especially in the last ten years, is a maturing field. In Haux’ view, medical informatics, in its early years, was not a necessity, but a “‘nice-to-have’ discipline” (Haux, 2010, p. 601). In today’s environment, however, that situation has changed and medical informatics is increasingly relied on as one of the foundational healthcare fields.

Haux referred to previous papers in his discussion of the medical informatics literature in describing the discipline, arguing that the field is defined by its *methodology and technology*, *application domain*, and by its *practical aims* (Haux, 2010, p. 604). In Haux’ studies with Hasman (2006, 2007), they determined that three *methodological* categories existed in which the majority of the medical informatics literature sifted out: *decision modeling*, *engineering modeling*, and *communication processes*. These results conflict with those found in Schuemie, et al. (2009), although Haux offered a caveat that “communication as well as decision processes may primarily be in research on the organization, application, and evaluation of health information systems” (2010, p. 604). Further categorizing the literature based on the International Medical Informatics Association Yearbook of Medical Informatics (2006), Haux

offered that the *application* sub-domains of medical informatics can be aligned with the following:

- medical informatics contributing to good medicine and good health for the individual,
- medical informatics contributing to good medical and health knowledge, and
- medical informatics contributing to well-organized health care (Haux, 2010, p. 604).

Finally, Haux proposed two practice-facing goals: “to contribute to progress in the sciences and to contribute to high-quality, efficient health care” (Haux, 2010, p. 604). It is important to note that Haux was addressing the practical side of the medical informatics discipline. As mentioned previously, academicians have frequently excluded the quasi-academic/practitioner publications in the field. One can argue that with less stringent peer-review processes and therefore, faster times to press, these publications may have a perspective that is closer to the true actions of the healthcare community than the academic publications that have longer periods between research study and publication. The longer wait time for the academic publication process increases the possibility that the information may be stale if not obsolete altogether.

While too lengthy for a complete discussion here, Haux continued his discussion with two possible future perspectives for medical informatics. The first point of view he shared was the more conservative, evolutionary one. It includes a view that changes will be small and made in incremental steps. On the other hand, his second perspective predicted that the field will make dramatic and revolutionary changes that will have drastic, positive effects on healthcare. Haux did claim that neither of the two lists is meant to be exhaustive (2010).

On the Importance of Theory

In producing scientific research, one should define what constitutes scientific research. One view defines scientific research as a systematic series of actions focused toward generating knowledge and that the knowledge “can be seen as generating belief statements about the actual world ...” (Pawar, 2009, p. 17). In this process, one of the actions is the theory-building, a step in which the researchers develop an idea or belief about the actual world and specify how their belief or idea applies to the real world. In this specification, researchers develop empirically testable and verifiable aspects of their theoretical model. There lies a fine line between explaining data and developing theory. Data describe empirical patterns that were observed, while theory explains why the patterns were observed or expected to be observed (Sutton & Staw, 1995). References, on the contrary, provide the reader the logic on which the authors base their theory and provide the reader direction for further investigation; they give the reader an indication of what led the authors’ thoughts to their conclusions and provide a means of checking the author’s accuracy (Sutton & Staw, 1995). Theory is developed from logical reasoning using findings to substantiate hypotheses. It is through the theory development process that the body of scientific knowledge continues to grow (Pawar, 2009).

Gregor takes a broad approach to the definition of theory compared with Whetten (Gregor, 2006; Whetten, 1989). While Whetten states that theory is comprised of the answers to the questions, *what, how, who, where, when, and why*, Gregor is willing to accept as theory the answers to some of the questions individually. For example, Gregor provides Davis and Olson’s 1985 textbook, *Management Information Systems: Conceptual Foundations, Structures, and Development*, as an example of theory that explains how something should be done in practice; it is these instructions about how information systems should be “designed, implemented, and

managed,” that is considered theory building (Gregor, 2006, p. 4). Of the five types of theory Gregor proffers - *analysis, explanation, prediction, explanation and prediction, and design and action*, only *explanation*, and *design and action* meet Whetten’s definition of theory because they include answering the question of *why* a phenomenon occurs. The *why* may be the most fertile question because it makes the theorist and the reader challenge the way he or she thinks about established norms and concepts; it causes him or her to extend the mind beyond a method he or she is accustomed to understanding.

A cursory look at the medical informatics literature indicates that there is a lack of theory development and the focus is more on descriptive reporting. Many studies have described experiments that medical informatics systems researchers have performed, but are lacking in answers to the question of why the phenomena occurred. For example, Halamka, Mandl, and Tang (2008) provide a thorough discussion of the experiences and challenges of personal health records, their study is bereft of theoretical analysis or development related to the challenges identified. While it is the literature itself that warrants a discussion of theory development, the focus of this study is not solely on theory. See the following papers for an extensive discussion on the topic of theory and theory development: (Dubin, 1969; Gregor, 2006; K. G. Smith and Hitt, 2005; Sutton and Staw, 1995; Weick, 1989, 1995; Whetten, 1989).

Theory Development

The theory, *Diffusion of Innovations*, is the primary theoretical lens for the literature review. Diffusion of innovations is a frequently-studied and accepted theory, having few changes since its inception over 60 years ago (Rogers, 2003). Diffusion of innovations is useful for this study because it takes into account the process of implementing an innovation and deploying, or diffusing, the innovation throughout organizations or group. Following is a brief

explanation of diffusion of innovations (for an exhaustive discussion of Diffusion of Innovations, refer to Brown, 1981 and Rogers, 2003).

Rogers explains diffusion as “the process in which an innovation is communicated through certain channels over time among the members of a social system” (Rogers, 2003, p. 5). He describes an innovation as “an idea, practice, or object that is perceived as new by an individual unit of adoption” (Rogers, 2003, p. 475). While at one point in the evolution of the diffusion literature, innovations were generally accepted as having a positive social impact, this is no longer the case and may be a cause for the varying adoption rates (Brown, 1981). Rogers explains adoption as “a decision to make full use of an innovation as the best course of action available” (Rogers, 2003, p. 473).

Healthcare information systems can create interesting challenges for adoption. In addition to creating a burden of learning a new system for already overworked healthcare providers, healthcare information systems are held to more stringent standards than information technology systems in other fields. These higher standards affect the adoption of healthcare innovations.

Innovations include ideas, practices, and objects. This study, however, confines innovations to consist of information systems artifacts. For certain healthcare information systems to serve their functions properly, it is imperative that healthcare providers use the information system. For example, if a hospital adopts an electronic medical record system and members of the radiology department do not use the electronic medical record, they may not be aware of an order for an x-ray a physician ordered; in this example, all providers need to use the system for the system to function properly. On the other hand, if not all healthcare providers use a hospital’s automated supply chain system, the effect on patient care may be minimal.

Adopter Categories

In developing his model, Rogers (2003) offered that people adopt innovations at different rates in a manner that approximates an *S* shaped curve (Figure 1). To make comparisons possible, Rogers suggested classifications of ideal categories based on where individuals fall along the adoption sigmoid. *Innovators*, *early adopters*, *early majority*, *late majority*, and *laggards* comprise the five adopter classifications, and in Diffusion of Innovations theory, these ideal categories follow a standard normal curve as depicted in Figure 2. Although the number of adopters may be few until word spreads about a new innovation, *innovators* are quick to employ new innovations. As time passes and knowledge of the new innovation spreads, the rate of adoption increases sharply as the sharp upward turn in Figure 1 indicates. During this period, the *early adopters* and *early majority* individuals employ the innovation. Later, the rate of adoption decreases as the *late majority* adopts the innovation. The curve in Figure 1 levels off toward the top—indicating the rate of adoption is decreasing—the *laggards* adopt the innovation and adoption reaches saturation. In the diffusion model, saturation does not necessarily mean 100% adoption. There may be individuals who never adopt a particular innovation.

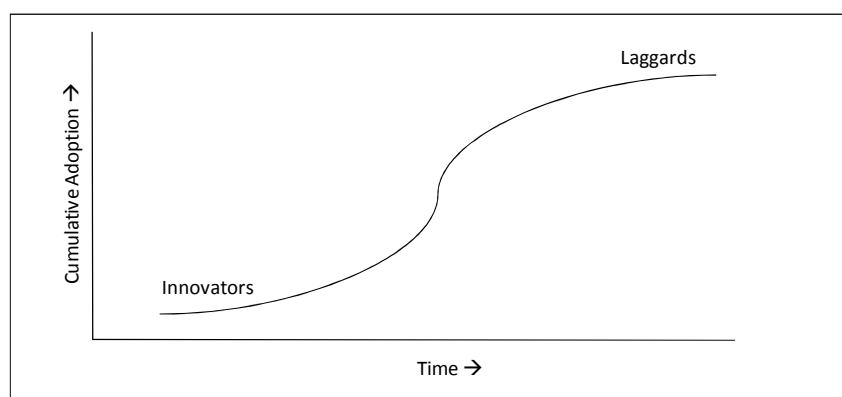


Figure 1. Sigmoid Adoption Curve (adapted from Rogers, 2003)

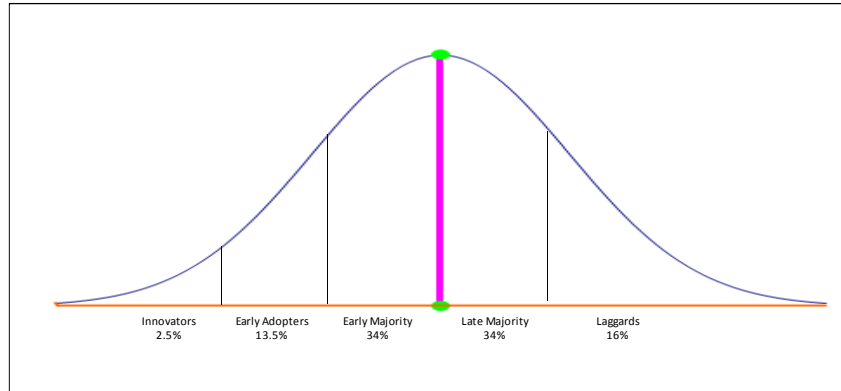


Figure 2. Adopter Categorization (adapted from Rogers, 2003)

Individuals who fall in the Innovators category tend to pursue new ideas with passion and are adventurous, almost to the point of compulsion. One finds socially influential individuals in the *early adopter* category. While the early adopters are the people who wield the most opinion capital in the social system, somewhat ironically, the *early majority* group members are seldom opinion leaders in the social system (Rogers, 2003). Frequently, it is the desire to remain competitive that brings later users to adopt the innovation (Brown, 1981). Skeptical individuals can be found in the *late majority* category and frequently they adopt an innovation to relieve the social (peer) pressures of the system. As the last members to adopt an innovation, the *laggards* generally are found to control virtually no opinion capital and frequently are fearful of the innovation (Rogers, 2003).

Stages of the Innovation-Decision Process

There are five stages through which individuals progress when evaluating an innovation for possible adoption (Figure 3 and in Rogers, 2003, p. 170). The stages are *knowledge*, *persuasion*, *decision*, *implementation*, and *confirmation*. This progression from initial knowledge of an innovation to confirmation of the adoption decision is what Rogers (2003) referred to as the innovation-decision process. It is within the innovation-decision process that

we find the five perceived characteristics of the innovation, and the theoretical analysis of the literature focuses on these characteristics. Specifically, the five characteristics of the innovation affect the persuasion stage of the innovation decision process. During the persuasion stage of evaluation, potential adopters develop either a positive or negative opinion toward the innovation (Rogers, 2003). In diffusion of innovations theory, the perceived characteristics of innovations influence the adopters' attitudes toward the innovation.

Perceived Characteristics of Innovations

Rogers explained the perceived characteristics of innovations as *relative advantage*, *compatibility*, *complexity*, *trialability*, and *observability* (2003). Rogers defined *relative advantage* as “the degree to which an innovation is perceived as better than the idea it supersedes” (Rogers, 2003, p. 15). The emphasis in the quote might well be placed on *perceived* because the importance is not in whether the innovation is truly better than the superseded idea so much as the individual's perception of the value of the new idea. The advantage of the new idea may be measured by various means: financial, level of convenience, or satisfaction levels, to name a few.

In the diffusion of innovations model, *compatibility* is “the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters” (Rogers, 2003, p. 15). Social norms play a substantial role in people's decisions whether to adopt an innovation. Those innovations that are comprised of values or beliefs determined by the individuals to be incompatible with their subjective norms will be rejected and the individuals will not adopt the innovation.

Rogers defines *complexity* as “the degree to which an innovation is perceived as difficult to understand and use” (2003, p. 16). Because some innovations are easier to comprehend than

others, they may be more eagerly adopted within a social structure. The literature review showed that the variable *ease of use* was used as an alternative to complexity in some studies (Cruz, Neto, Muñoz-Gallego, & Laukkanen, 2010; Manns, 2002). *Ease of use* and *complexity* can be viewed as parallel, while opposite constructs (Davis, 1989).

Trialability, in the Rogers model, involves the degree to which an innovation may be experimented with on a limited basis (2003). Software companies frequently take advantage of this characteristic by allowing potential customers to download and use a limited version of their software. For medical informatics systems, *trialability* may include the ability to use the system at other workers' computers, at vendors' locations, or vendors may bring systems to the potential customers.

The diffusion of innovations model includes *observability* and *observability* is “the degree to which the results of an innovation are visible to others” (Rogers, 2003, p. 16). The more visible an innovation is to potential adopters, the more likely they are to adopt it. In healthcare, *observability* may be restricted because of concerns over the security of personal health information. For example, patients may fear the compromise of their sensitive personal health information if they show others their new personally controlled healthcare record. Despite the benefits others may receive from a personally controlled healthcare record, the patients' security concerns will limit word-of-mouth advertising (i.e., *observability*). Although this personal health information security concern may be overcome by something as simple as providing a “dummy” account for potential adopters of the system to use, it does affect on *observability*.

Chapter Summary

This chapter presented an overview of the medical informatics literature over the past half century and established the foundation and context for further inquiry into the themes of the medical informatics literature. Outlining the pioneering works of medical informatics authors, the chapter addressed the types of questions that have been asked and answered, the types of research methods previously applied, and what has worked and what has not worked. The chapter established the definition of medical informatics for the purpose of this study: *the discipline dedicated to the systematic processing, analysis, and dissemination of health-related data through the application of digital information systems (computers) to various aspects of healthcare, research, and medicine*. This chapter established the importance of theory development for medical informatics literature and outlined the theoretical model, *Diffusion of Innovations*, for this study.

The next chapter provides an overview of the content analytic method and discusses the research design for the study. The chapter continues with an outline of the methods of data collection including an explanation of the journals sampled. The chapter includes the sampling, recording, and observational units and a discussion of the variables selected for the analysis. The procedures section explains the semi-automated method of analysis and the manual coding analysis that comprise this mixed methods study.

Chapter 3. Methods

The previous chapter established the context of this study by addressing the history of medical informatics and discussing theoretical development in the medical informatics field. This chapter explains the methods used to assess emergent medical informatics themes. The chapter is divided into three main sections. The first provides an overview of content analysis, the general procedures used with the method, introduces the two phases of the analysis—quantitative and qualitative—and describes the data collection and preparation.

The second section of the chapter explains the quantitative analysis method. This method uses centering resonance analysis and exploratory factor analysis to extract the themes from the journal texts. Centering resonance analysis is an automated technique that identifies influential nouns and noun phrases from a body of text. The influential words are used in the subsequent exploratory factor analysis to develop the preliminary themes. The preliminary themes are further analyzed through latent coding to develop the final themes list.

The final section of the chapter explains the qualitative analysis performed. This section describes the manual coding method used to further analyze the text. The section begins with an explanation of the thirteen operationalized definitions used for this portion of the analysis and a description of the coders involved. An explanation of inter-coder reliability and the minimums accepted for this study follows.

Content Analysis

Content analysis is one method of extracting various themes and topics from text. Although aspects of content analysis date back as early as the 1600s in efforts by the church to seek out threats to its authority and the printing of materials, the first mentions of the term content analysis are not found until three centuries later (Krippendorff, 2004). Content analysis can be understood as, “an *empirically grounded method*, exploratory in process, and predictive or inferential in intent (Krippendorff, 2004, p xvii; italics from the original). While the literature yields several definitions of content analysis, the generally accepted definition of content analysis is that the method is a means of making valid, reliable inferences from a textual data source (Weber, 1990). “The content analyst views data as representations not of physical events but of texts, images, and expressions that are created to be seen, read, interpreted, and acted on for their meanings, and must therefore be analyzed with such uses in mind” (Krippendorff, 2004, pxiii).

In summary, content analysis provides us a means by which we can derive understanding of a body of text. While there are nuances to the various definitions available, most agree that content analysis is a research technique that uses data reduction primarily with textual data to make the text more manageable for inference and analysis (Krippendorff, 2004; Weber, 1990). One of the main ideas behind content analysis is that large bodies of text are grouped into a relatively small number of categories based on some criteria so that the large bodies of text can be managed and understood. Content analysis includes the process of going from words, to numbers, back to words. In essence, the content analyst takes the words and turns them into numbers through word frequencies, factor analyses, and other statistical measures. Then, the

analyst takes the numbers and interprets them and shares the results using words in the study results and discussion.

Process

Performing a content analysis can be a relatively straightforward process. The general procedure in performing content analysis include the following steps (Corman, et al., 2002; Krippendorff, 2004; Neuendorf, 2002):

1. Develop research questions and/or hypotheses. This answers the question of why the study is performed and guides the researchers in the study. Although many studies include a guiding theory, exploratory studies frequently are performed without a guiding theory.

2. Perform a review of the literature. The literature review situates the study in the ongoing discussion for the discipline; it identifies the context in which the study is performed. The literature review identifies questions that have been asked and answered, and what works and does not work.

3. Conceptualize. The researchers determine for what variables will they have to collect data to answer the research questions/evaluate the hypotheses. For example, the researcher may collect data on author names, year of publication, abstract information, and so on.

4. Operationalize. Based on the variables selected, the researchers define measures that match the conceptualization to ensure internal validity. The researchers define the (a) *sampling units* - the bodies of text that will be included in the study (e.g., from what publications and what years were the sample of articles drawn), (b) the *recording units* - the parts of the text that will be categorized and/or described that are usually contained in the sampling units (e.g., words, noun phrases), and (c) the *context* or *observational units*, defined as the limits of the text that will be categorized and/or described (e.g., sentence, paragraph, abstract).

5. Collect data. The researchers justify their data selection conceptualization and collect the data based on their justification.

6. Develop a *coding protocol*. For computer coding, the coding protocol includes an explanation of the method of content analysis the software performs. For researchers performing human coding, the coding protocol is usually comprised of a *codebook* that provides instructions to the coders and explains all the variables and measures, and a *coding form*, the form used by the coders to record coding data. If human coding is employed, the human coders are trained on the protocol and they conduct an inter-coder reliability assessment. This reliability assessment is the degree to which the coders agree on their coding assessment of the text, correcting for chance agreement whenever possible.

7. Code the main body of text. For computer coding, the researchers execute the software analysis and spot check for validation. For human coding, the coders code the text—they should have at least a 10% overlap to evaluate reliability—and the researchers evaluate inter-coder reliability for each variable (e.g., Scott's *pi*, Krippendorff's *alpha*).

8. Discuss results. The researchers tabulate the results and counts, and explain their discoveries.

These steps provide an overview and are general. Other factors have to be addressed based on issues such as the researcher's specific topic, sampling decisions, and coder training. The steps of content analysis mirror those of empirical positivist research. That is by design. Over the evolution of content analysis, researchers using the method have striven to refine the method to increase reliability and validity of the results. The aforementioned general steps provide the basis for sound research that can provide rich results that are replicable, objective, and systematic.

Mixed Methods Analysis

The study employs mixed methods analysis in two phases. The first phase is a semi-automated quantitative method, so named because the study used software for the automated portion and follows the automated result with a manual analysis component. The purpose of the first phase is to identify themes emerging from the article texts of the seven journal publications, to compare the texts by year, and contrast the texts of the medical informatics publications against those of the non-medical informatics publications.

The second phase of the study is a qualitative analysis. The purpose of the second phase is to determine how the literature of this study compares to previous studies and the Diffusion of Innovations model. Using the DiscoverText web-based software application, three researchers qualitatively coded the data based on criteria developed in the literature review and phase one of the study. The three coders read the article texts in detail and coded in the appropriate theme and/or characteristic classification(s) based on the procedures in the coding procedures guide.

Sampling Units

This exploratory analysis identifies, describes, and explains the medical informatics themes that arise in the literature. Using semi-automated computerized content analytic methods and manual methods, the study parses texts of journal articles to derive an understanding of what the medical informatics field is discussing and where the field is headed. Based on the multidisciplinary focus of the medical informatics field, and both the explicit and implicit calls in the literature discussed in the previous chapter, the research samples articles published in medical informatics journals and articles related to medical informatics that are published in non-medical informatics.

Although many of the reviews discussed previously addressed the multidisciplinary focus of the field and the need for understanding the disciplines related to medical informatics, no authors analyzed publications outside those specifically designated as medical informatics publications - with one caveat. Morris and McCain (1998) did discuss publications that were outside the scope of medical informatics in their initial data inclusion phase by including two non-medical informatics journal titles in their sample set. However, they subsequently removed them in their exclusion phase based on a lack of a substantial number of citations made and received among the candidate “core journals” they used as their starting point (Morris & McCain, 1998). As a reminder, one of Morris and McCain’s goals was to identify the core medical informatics literature and thereby they had justification for removing the non-medical informatics journals despite the loss of potentially interesting information.

Recording Units

In the first phase of the study, the semi-automated quantitative content analysis, the researcher analyzed the noun phrases of the article texts to identify themes in the literature. Noun phrases are the subject or object of sentences and are each comprised of a noun and possibly more adjectives or nouns (Corman, et al., 2002). Although verb phrases, another linguistic model, provide the action in texts or conversation, the noun phrases are the only elements that “can be unambiguously classified as entities in discourse” (Corman, et al., 2002, p. 174) , and therefore, are of greater explanatory value than are the verb phrases.

For the second phase of the study, the manual coding analysis, the recording units are the words from which the article texts are comprised. Based on preliminary readings of the article texts, there was sufficient evidence to determine meaning from either the sentences or the individual words in the texts. While sentences in the texts are rife with meaning, the general

goal in defining recording units is to define them as the smallest unit from which meaningful content can be drawn (Krippendorff, 2004). Therefore, the recording unit selected for the second phase was the individual word.

Observational Units

While recording units are generally the smallest units of analysis in a content analysis study, the observational units range in size depending on the content being analyzed. In a play, for example, the observational unit may be the entire play if the researcher is trying to distinguish the core themes of the play. On the other hand, if the researcher seeks to understand the basis for the formation of the characters in the play, the individual acts of the play may be better observational units for the study. The size of observational units is generally larger than that of recording units and may or may not include the recording unit. The observational unit should be determined to achieve a balance between making it large enough to provide adequate meaning and hence, validity, while being as small as is possible to maintain reliability. For the purpose of this study, the entire article abstract text satisfies that balance.

Sampling

The goal of this study is to determine what themes arise from the medical informatics discipline. This study uses seven publications from which article abstracts and texts from the period of 2002 - 2011 were drawn for analysis. To arrive at the sample of articles, the sampling units, several criteria were used. From those publications that are considered medical informatics publications, all articles were used in the inclusion phase of the selection process. The publications that are mainstream medical informatics publications include the *Journal of the*

American Medical Informatics Association (JAMIA), and the International Journal of Medical Informatics (IJMI).

As stated in Chapter Two, several authors expressed the multidisciplinary perspective of medical informatics and hence, recommend including articles from outside the mainstream medical informatics publications for subsequent research. Taking that advice, several publications outside the prevailing medical informatics literature were included: the *Journal of the American Medical Association (JAMA)*, the *New England Journal of Medicine (NEJM)*, *Management Information Systems Quarterly (MISQ)*, *Information Systems Research (ISR)*, and *Communications of the Association for Computing Machinery (CACM)*. The complete listing of the publications included and their respective foci are in Table 3.1.

Table 3.1
Journals Sampled

Abbreviation	Title	Focus
CACM	Communications of the Association for Computing Machinery	Management Information Systems
IJMI	International Journal of Medical Informatics	Medical Informatics
ISR	Information Systems Research	Management Information Systems
JAMA	Journal of the American Medical Association	Medical
JAMIA	Journal of the American Medical Informatics Association	Medical Informatics
MISQ	Management Information Systems Quarterly	Management Information Systems
NEJM	New England Journal of Medicine	Medical

Medical Informatics Publications

To determine which mainstream medical informatics journals to include in the sample, the recent review articles mentioned previously in Chapter Two that provided journal rankings were referenced. The articles referenced include *The structure of medical informatics journal literature* (Morris & McCain, 1998), *Mapping medical informatics research* (Eggers, et al., 2005), *Publication trends in the medical informatics literature: 20 years of 'Medical Informatics' in MeSH* (DeShazo, et al., 2009). and *Mapping the domain of medical informatics* (Schuemie, et al., 2009). In each of these studies, the following journals received top rankings: the *Journal of the American Medical Informatics Association* (JAMIA) and the *International Journal of Medical Informatics* (IJMI). Although each of the journal ranking studies evaluate the publications using different measures, both JAMIA and IJMI consistently rank at the top of medical informatics journals. As these publications are part of the core medical informatics literature, all articles for the ten-year period from these journals are included in the data set.

Medical and Management Information Systems Publications

Based on the repeated calls for analysis that extend beyond the core medical informatics journals (see Chapter Two), the study includes articles from the *Journal of the American Medical Association* (JAMA), the *New England Journal of Medicine* (NEJM), *Management Information Systems Quarterly* (MISQ), *Information Systems Research* (ISR), and *Communications of the ACM* (CACM). Using JAMA extends the data set into the medical discipline to gain a perspective of what the medical field is discussing about medical informatics. The publication, JAMA, is an international, peer reviewed medical journal that has been published continuously since 1883; it is the most widely circulated medical journal in the world (retrieved April, 12,

2011, from <http://jama.ama-assn.org/site/misc/aboutjama.xhtml>). The journal has an impact factor for 2009 of 28.9. The impact factor is a measure that is often used as a proxy for the relative importance of a scientific journal. For the medical discipline, JAMA's score is high. On its recently revamped website, JAMA provides a specific section, *Informatics/Internet Medicine*, from which all article citation, abstract, and text information was collected for the ten year period.

While the New England Journal of Medicine is another renowned publication in the medical discipline, the publication's website does not have a distinct section for medical informatics research like that found at JAMA's website. Therefore, the study employed an advanced query in Thompson Reuter's *Institute for Scientific Information (ISI) Web of Science* to identify articles. The *ISI Web of Science* provides for simultaneous searching of the *Science Citation Index - Expanded*, the *Social Sciences Citation Index*, and the *Arts & Humanities Citation Index* databases.

Query Method

Using the *ISI Web of Science* query service to collect articles for the NEJM, the following Boolean query was performed:

TS=(“information system” OR comput* OR technol* OR informatic) AND SO=(New England Journal of Medicine)”

where TS is the ISI code for defining the topics to search in the query, SO is the ISI code for defining the publications to include in the search, and the asterisks are wildcards, representing any group of characters, including no character (e.g., health* = health, healthy, healthcare, etc.).

MISQ, CACM, and ISR are the top three management information systems publications based on an analysis that synthesized nine previous studies' journal (Rainer & Miller, 2005). Of the three journals, CACM is the only management information systems publication that has a more practitioner based target audience. The other two, MISQ and ISR, target the academic audience. Including these journals in the study provides a good basis for understanding the management information system discipline's perspective on medical informatics. The reason for using CACM is to garner the perspective of practitioners who are applying the technologies. In other words, what are the information technology professionals discussing? Again, this selection is based on the call for a broader scope from the authors of the review articles in Chapter Two of this thesis.

For each of the MIS publications, a query method similar to the one used for the NEJM was employed. Articles were included based on the query to identify those related to healthcare and medicine. For the publications in the management information systems discipline, MISQ, ISR, and CACM, another *ISI Web of Science* search was performed using the following Boolean query:

“TS=(pharm* OR informatic* OR drug* OR health* OR medic* OR bio*) AND SO=(MIS Quarterly)”

Note: For the CACM and ISR queries, the SO=(MIS Quarterly) were replaced with SO=(Communications of the ACM) and SO=(Information Systems Research) respectively.

As previously mentioned, all abstract and citation information was input into a citation manager database. The citation manager database for this study is the freely available software, *Zotero*, which can be downloaded at <http://www.zotero.org/>. *Zotero* integrates with the Firefox browser—also free, and available from the non-profit organization *Mozilla*—and *Zotero* stores

citation information, abstracts, researcher notes, etc., in an *SQLite* database format. The source code for *SQLite* is in the public domain; hence, there is much freely available software to manipulate *SQLite* databases. Once in the *SQLite* database, the free application, *SQLite Manager*, was used to query the database and extract the abstract text, author information, publication date, and other citation information for each record in the database. An example of the code used to extract the JAMIA data from the *SQLite* database is found in Appendix A. Similar code was used to extract the data from the other publications.

It is important to note some particulars about the data collected for this study. While previous researchers have excluded letters to the editor, this study included them because they exemplify the views of members of the academic community and the editors believed they were important enough to be included in the journal. Although the letters do not fall under the scope of academic research and are not peer-reviewed, they do reflect views of the field. The letters to the editor included in this study are comprised of the full text of the letters. Likewise, while the majority of the publications sampled are peer-reviewed, some of the articles from CACM, a refereed journal, are not peer-reviewed. The study kept these articles in the overall dataset, deferring to the judgment of the editors of these publications in adding the articles to CACM.

Data Preparation

After extracting the data from the *SQLite* database, it was imported it into Microsoft Excel™ to perform manual data preparation and cleaning procedures. The *date* field from the *SQLite* database included the year, month, and day. The four-digit year was extracted to use as a categorization field. Many of the abstracts included standardized formats including keywords such as *objective*, *study design*, etc. It was necessary to remove the keywords to prevent their skewing the quantitative analysis, centering resonance analysis, which is based on noun phrases.

Other minor modifications that did not affect the analysis included changing journal publication titles to abbreviations and temporarily removing the author information.

If the article text met the exclusion criteria, it was removed, or excluded, from further analysis. The exclusion criteria were defined as any article texts that were clearly unrelated to medical informatics—they did not meet this study's definition of medical informatics nor did they have both information systems technology and healthcare related content. For the purpose of this study, medical informatics is defined as *the discipline dedicated to the systematic processing, analysis, and dissemination of health-related data through the application of digital information systems (computers) to various aspects of healthcare, research, and medicine* (Chapter Two). Unless the text met the exclusion criteria, it was included for further analysis. This approach was necessary to ensure that references were captured that some may consider to be on the periphery of medical informatics, yet are still a part of the field. Keep in mind, one goal of this study is to extend the analysis of the medical informatics literature beyond the medical informatics journals to develop a broader understanding of what encompasses the field. When in doubt, the preference for the study was to include articles, thereby deferring to the journal keyword technicians for each of the ISI Web of Science databases, *Science Citation Index - Expanded*, the *Social Sciences Citation Index*, and the *Arts & Humanities Citation Index*. This is because researchers searching for medical informatics literature using terms similar to that used in this study will receive similar results; they will base their analyses off said results. Again, the intent is to include articles and understand the discussion in the field of medical informatics.

Upon completion of the data cleaning, the article text data was exported as text files by year and publication for the quantitative phase of the study. The centering resonance analysis

software requires input in the form of text that complies with the American Standard Code for Information Interchange (ASCII) standard. The ASCII text format is a standardized character-encoding scheme based on the English alphabet. It has been a standard format since its origin in 1963, it is very common, and most software applications can import data that is in the ASCII format.

No special preparation was necessary to prepare the dataset for the qualitative phase of the study. DiscoverText web-based software was used for the manual coding process. DiscoverText is able to use Excel™ files in their proprietary format without modification.

Centering Resonance Analysis

In the past two decades, as technology has improved at a tremendous pace, the amount of data produced has similarly grown at a rapid pace. As greater volumes of data become available for analysis, content analysis “by hand” becomes increasingly difficult and unreliable. Coder fatigue increases the likelihood for unreliable coding. Manpower costs can become prohibitive because of the hours required for analyzing large texts. Word counts can be nearly impossible to perform on massive texts. Decades ago, content analysts were limited to relatively small bodies of text because of these limiting factors. Today, however, with the advent of faster processors, larger storage devices, greater memory, and superior graphics capabilities, computers are useful tools for the content analyst. With computers, analysts are able to parse massive volumes of text and analyze the texts in ways that were impossible just a few decades ago. As an example of the voluminous texts for analysis, the accumulated texts for this study from only a ten-year period equate to about 2,600 pages of text. While content analysts have been using computers to perform frequency counts and provide word lists for years, newer methods of analysis requiring greater computing power are now affording scholars the opportunity to delve into texts with a

more artificially intelligent approach. One such method, centering resonance analysis, is a means to analyze text for noun phrases and provide analysis based on the noun phrases. A noun phrase is “a noun plus zero or more additional nouns and/or adjectives, which serves as the subject or object of a sentence” (Corman, et al., 2002, p. 174).

Centering resonance analysis is a “text analysis method that has broad scope and range and can be applied to large quantities of written text and transcribed conversation. It identifies discursively important words and represents these as a network, then uses structural properties of the network to index word importance” (Corman, et al., 2002, p. 157). As inferred from the previous quote, centering resonance analysis can handle large bodies of text that would be virtually impossible for human coding. As an example, Corman and colleagues’ (2002) estimated that if they were to analyze the communications of a small, 50 person organization for one week by recording all discussions of the employees, they would generate about 18,750 transcribed pages of text. Aside from the challenge of getting coders even willing to try to code that much, there would be problems in preventing coder fatigue, and the likelihood that the coders’ viewpoints would change between the first page coded and the last page coded would increase. That is where the benefit of using computational models for analysis lies.

Centering resonance analysis “finds and maps concepts linking diverse chains of discussion and reasoning in and across conversations, then can compare maps between different groups and organizations” (Mcphee, Corman, & Dooley, 2002, p. 275). Centering resonance analysis is grounded in *centering theory* (Corman, et al., 2002). Centering theory derives its name from the idea that authors focus their written statements around *centers*, words or noun phrases that form the subjects of the discussion - what the author is writing about. These centers connect to previous centers and subsequent centers to form a cohesive network of speech. In

other words, these centers connect with other, subsequent centers to create a flow of ideas that the reader is able to comprehend. For each grouping of text after the first, there is a center that links backward toward the best forward-looking center from the previous grouping of text.

A simple nursery rhyme provides a basic example of the concept: “Jack and Jill went up the hill to fetch a pail of water. Jack fell down and broke his crown, and Jill came tumbling after.” In the first sentence, the noun phrase “Jack | Jill” is linked to the “hill” and the “hill” is a backward-looking noun phrase to the initial “Jack | Jill” noun phrase center. In the second sentence, “Jack” is a backward-looking center that links to the “Jack | Jill” center from the previous text (sentence) and it is that link that creates the network of coherent ideas that helps the reader understand the flow from the first sentence to the second. These networks of centers represent the main concepts, influence, and interrelationships of the text; they are predictable and make the texts comprehensible and relevant. The word networks created with centering resonance analysis illuminate *influential* words, words that “facilitate the connection of meaning among many different words, across very different parts of the overall word network” (Mcphee, et al., 2002, p. 278), and are “very rich data structures that preserve significantly more information about a text than keywords or word frequency statistics” (Corman, et al., 2002, p. 172). Referring back to the Jack and Jill example, centering resonance analysis measures the *influence* of the word as how often the word is between other words, the “betweenness centrality” of the word and indicates the “likelihood of being on the shortest path in the network connecting any other two words” (Corman, et al., 2002, p. 172). As illustrated in Chapter Two, there have been analyses of the literature using bibliometric methods, surveys, and non-empirical methods. Using centering resonance analysis extends current medical informatics knowledge both in methodology and in understanding of the structure and content of the discipline.

Data Analysis – Centering Resonance Analysis

The purpose of the quantitative first phase of the study was to extract medical informatics themes from the article texts collected from the seven sample journals. When working with large bodies of text, manual thematic analysis can be time-consuming and is open to errors from coder fatigue. Drawing from seven publications over a ten-year period is likely to result in a large volume of text and, therefore, is a good candidate for automated coding. Using a software application can aid in reducing the information in the texts to a more manageable size. The software selected for use in this study was specifically developed to support centering resonance analysis and is named *Crawdad Text Analysis System* (Corman & Dooley, 2006).

The first stage of the centering resonance analysis included developing noun phrase network information for each of the separate years of texts. Centering resonance analysis generates meaningful networks of nouns and noun phrases that represent the main concepts of a body or bodies of text. The influence and interrelationships of these networks are developed during the analysis. Using the *Crawdad* software, the researcher analyzed the article data files and created three statistics for each year of the study: *nodes*, *density*, and *group influence*. The measure produced—*node*—indicates the number of centering points, or noun and noun phrases the network contains; a node is a point of connection within a centering resonance analysis network. The *density* is a ratio of the number of network connections that exist among nodes compared to the number of network connections that could possibly exist and is an indicator of how tightly connected the network is. The *group influence* score is an indicator of how coherent the entire network—in this case, each of the article texts file—is within itself. A high group influence score indicates that the network is highly focused and centralized. Both the density

and group influence scores are standardized measures with minimum scores of zero and maximum scores of one.

In addition to the node, density, and group influence properties, the top 50 influential words for each of three stratifications of article texts was collected. The first stratification was by year to assess variations over time. The second and third were to assess variations between the medical informatics publications and the non-medical informatics publications. The more influential the word is, the more it “ties other words together in the text network and facilitates meaning” (Tate, Ellram, & Kirchoff, 2010, p. 24). Influence scores range from zero to one; influence values of .01 or greater are considered significant, and influence values of .05 or greater are considered very significant (Corman & Dooley, 2006). The influence values and the top influential words provide an overview of the general themes throughout each year and can provide a basis to identify consistency of themes over the years.

Finally, the article texts were consolidated into separate medical informatics and non-medical informatics files; the article texts that came from JAMIA and IJMI went into the medical informatics file and the texts that came from CACM, JAMA, MISQ, ISR, and NEJM went into the non-medical informatics file. As with the article texts by years, the node, density, and group influence information, and the top 50 influential words for the medical informatics article texts and the non-medical informatics texts were collected. The articles from mainstream medical informatics publications and those from publications of the closely related fields of medicine and management information systems were assessed for differences.

An exploratory factor analysis (EFA) using the 50 most influential words of the texts of all years and all publications as variables with the influence values as score values for each of the variables was performed. The EFA used principal components analysis with Varimax rotation to

assess the underlying thematic structure of the body of abstract texts. The number of factors to extract was based on an eigenvalue cutoff of one, with all factors greater than one extracted. The result of the EFA provided a first look at emerging themes.

The EFA provides a good foundation to develop themes using the factors identified. However, some scholars suggest a further step referred to as latent coding via manual human coding to discern the unobservable content for refining the themes (Tate, et al., 2010). Principal Components EFA is used to “mathematically derive a relatively small number of variables to use to convey as much of the information in the observed/measured variables as possible” (Leech, Barrett, & Morgan, 2008, p. 58; underlining from the original removed). The EFA is a data reduction method, but identifying the themes from the reduced data set requires human interaction. This human interaction, latent coding, reveals a richer depth of understanding of the underlying constructs.

There are disadvantages to latent coding methods, however. When subjectivity is introduced during the latent coding, reliability begins to deteriorate. Although this argument may discourage latent coding, centering resonance analysis only develops networks of words (specifically, nouns and noun phrases) and not networks of theoretical constructs or concepts. Thus, the secondary latent coding analysis is warranted and can “logically connect words to themes and strengthen the face validity of the theme” (Tate, et al., 2010, p. 25) . Starting with the rotated factor solution, descriptive names for each of the factors were developed. The names indicated the themes determined to be inherent in the texts associated with each of the factors.

Operationalized Variable Definitions

As stated, the purpose of this study was to analyze the medical informatics discipline for emerging themes and topics. Based in the previous discussion of the literature, the coding

categories are developed. As a point of reference, it is important to restate the research questions and sub-questions before clarifying the coding categories.

Research Question 1: What themes have emerged in the medical informatics discipline since its inception in the 1960s?

Sub-question 1: What themes emerge from the medical informatics literature?

Sub-question 2: What medical informatics themes emerge from the related fields of medicine and management information systems?

Research Question 2: What challenges exist for the medical informatics discipline?

Research Question 3: What future directions does the literature suggest for the field?

The preceding questions provide the basis for the manual coding and analysis of the sample of articles collected. The diffusion of innovations theory serves as the theoretical framework for this study. Qualitative and quantitative content analysis methods are used to obtain and analyze the data. Rogers' perceived characteristics of innovations and Haux' view of the application sub-domains of medical informatics were used to created the variables for this study (Table 3.2).

Table 3.2
Operationalized Variables

Variable	Description
Title	The title of the article to identify and delineate specific articles from one another
Author	The full name of the article's author to provide further identification and specification.
Publication Year	The year that article was published to provide article identification and to delineate by year.
Journal Title	The journal title to provide article identification and to delineate between medical informatics and non-medical informatics publications.
Remove	The article text is unrelated to medical informatics and should be removed from the dataset
Direct Patient Care	Information system/technology that contributes to good medicine and good health for the individual and aids organizational leaders in making decisions (e.g., electronic medical records)
Business Analytics	Exchange, creation, distribution, analysis, or adoption of medical and health knowledge ideas, insights, and experiences within and across organizations (e.g., data mining for health reporting)
Information Architecture	Contributes to well-organized, patient-centered health care and appropriate information management (e.g., intra-organizational health information systems architecture; health information data exchange standardization)
Relative Advantage	The degree to which an innovation is perceived as better than the idea it supersedes
Compatibility	The degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters
Complexity	The degree to which an innovation is perceived as difficult to understand and use
Trialability	The degree to which an innovation may be experimented with on a limited basis
Observability	The degree to which the results of an innovation are visible to others

Note. Perceived Characteristics of Innovations are from Rogers, 2003, pg 170.

For both phase I and phase II of the analysis, four variables were applicable to the research questions, sub-questions, and theoretical foundation of the study. These are the manifest variables. The four variables related to the descriptive aspects of each article abstract and include the following manifest variables:

Variable 1: Title. The *title* of the article to identify and delineate specific articles from one another.

Variable 2: Author. The full name of the article's *author* to provide further identification and specification.

Variable 3: Publication Year. The year that article was published to provide article identification and to delineate by year.

Variable 4: Journal Title. The *journal title* to provide article identification and to delineate between medical informatics and non-medical informatics publications.

The second group of variables, variables five through 13, focus on two areas, both of which are more inferential, or latent, in nature than they are manifest. The data collection used an approach with an inclusive focus when selecting articles for analysis. With a conservative inclusive approach, there is a greater likelihood that articles collected will not pertain to the subject matter. Therefore, it was necessary to include an option for the coders to identify an article as one that should not be included in further analysis. Variable five, *Remove*, is the variable used to indicate the coder's recommendation to withdraw the article from further analysis. In phase II, the manual coding phase of the study, the variables were focused toward the coding of the articles and are based more on the coders inference of the article. The variables can be considered the latent variables. The latent variables focus on the ideas around each article's fit in the perceived characteristics of innovations and its applicability to the three

overarching sub-domains as suggested by Haux (2010) and enhanced previously in this study. In Haux' review, he offers the three application sub-domains of medical informatics:

- medical informatics contributing to good medicine and good health for the individual,
- medical informatics contributing to good medical and health knowledge, and
- medical informatics contributing to well-organized health care (Haux, 2010, p. 604).

Using Haux' three sub-domains as a basis, the study generated the variables, *Direct Patient Care*, *Business Analytics*, and *Information Architecture* (variables six, seven, and eight, respectively). Variables nine through 13, *Relative Advantage*, *Compatibility*, *Complexity*, *Trialability*, and *Observability*, are drawn from the perceived characteristics of innovations from Rogers' Diffusion of Innovations theory (2003). Coding the article texts for the presence or absence of the latent variables in the two groups provides a representation of the pattern and frequency of the concepts in other scholars' works. The following codes and definitions were appropriate for the analysis.

Variable 5: Remove. As stated, the articles in this category are those that unrelated to medical informatics and should be *removed* from the dataset and not used for further analysis.

Variable 6: Direct Patient Care. An article coded as being *direct patient care* will discuss information system/information technology that contributes to good medicine and good health for the individual and aids organizational leaders in making decisions (e.g., electronic medical records).

Variable 7: Business Analytics. The *business analytics* code includes articles that discuss the exchange, creation, distribution, analysis, or adoption of medical and health knowledge ideas, insights, and experiences within and across organizations (e.g., data mining for health reporting).

Variable 8: Information Architecture. An article coded as *information architecture* includes a contribution to well-organized, patient-centered health care and appropriate information management (e.g., intra-organizational health information systems architecture; health information data exchange standardization).

Variable 9: Relative Advantage. In the first of the perceived characteristics of innovations, an article coded as *relative advantage* is one that discusses the degree to which an innovation is perceived as better than the idea it supersedes.

Variable 10: Compatibility. A *compatibility* article addresses the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters.

Variable 11: Complexity. To be included in the *complexity* code, the article must consider the degree to which an innovation is perceived as difficult to understand and use.

Variable 12: Trialability. A discussion of the degree to which an innovation may be experimented with on a limited basis is the requirement for an article to be coded with *trialability*.

Variable 13: Observability. An article coded *observability* is one that addresses the degree to which the results of an innovation are visible to others.

Qualitative Data Analysis - Manual Coding

Coders and Coding Procedures

The second phase of the study, the manual coding phase, is primarily focused toward gaining an understanding of the sub-domains of the medical informatics discipline and the appropriateness of diffusion of innovations theory in the medical informatics discipline. The

coders included this researcher and two fellow doctoral students in the management information systems program at a major university. At the time of the study, both of the secondary coders had completed their doctoral coursework and their comprehensive exams. The coders were selected because of their experience with diffusion of innovations theory through independent research and a focused seminar. In addition, the coders were familiar with this study and brought with them previous qualitative content analysis research experience.

The manual coding phase of article theme variables coding spanned four weeks. The coders reviewed and discussed the coding procedures until all fully understood the directions and intent. The coders participated in a practice coding session of about 10-15 article texts from the data set after which adjustments to the coding procedures were made to ensure reliability and understanding. After unanimous agreement on the coding procedures, each coder independently coded their data set using the online text analytics application, DiscoverText (www.discovertext.com).

DiscoverText provides a means for users to rapidly code electronic data through a novel interface. Users can use mouse clicks or keyboard shortcuts to select the codes for their text; however, the speed of the system is found when users use the keyboard for coding. Keyboard shortcuts were assigned for selecting the appropriate code (e.g., “X” for Complexity, “D” for Direct Patient Care) and a keyboard return key is used to move to the subsequent text. Arrows can move to previous or subsequent texts. Although DiscoverText provides additional analysis tools, none were used for this study.

The coders analyzed article abstracts, letters to the editors, and brief introductory statements (hereafter, *texts*) of journal articles, searching for medical informatics themes in the article texts. The coders received explicit instructions on using DiscoverText to code the article

texts (Appendix B). The medical informatics definition—*the discipline dedicated to the systematic processing, analysis, and dissemination of health-related data through the application of digital information systems (computers) to various aspects of healthcare, research, and medicine*—was provided to the coders to give them a common understanding and starting frame of reference. The coders were advised of the research questions and sub-questions for the study and were instructed to keep the research questions in the forefront of their minds during coding. The coding instructions included a direction for the coders to familiarize themselves with the coding variable definitions as mentioned in the previous section and to use the definitions for identifying articles with the two groups of latent variable codes. Further, the coders were advised to take a ten minute break after coding 25 texts or after one hour. After breaks, the coders were to re-read the operational definitions and coding instructions before restarting the coding process.

Inter-coder Reliability

Given the large size of the dataset, minimum coder overlap was set at 15% for the second phase of the study, the manual coding phase. Although the greater the overlap, the greater the opportunity to recognize coder and codebook discrepancies, 15% ensures adequate coverage, particularly when using three or more coders. With three coders, it is less likely that chance agreement will occur. For chance agreement to occur when inter-coder reliability is measured, the coders would have to select the same code by random selection. To address the potential for chance agreement, the researcher calculated simple percentages of agreement and Krippendorff's alpha for each of the secondary coders agreement with the primary coder. Krippendorff's alpha, with its origin in content analysis, is appropriate because it scales well across any number of coders (greater than one) and sample sizes, and it can be used when the data set includes missing

data (Hayes & Krippendorff, 2007; Krippendorff, 2007). Krippendorff recommends heuristics for the use of alpha with the caveat that “the choice of reliability standards should always be related to the validity requirements imposed on the research results, specifically to the costs of drawing wrong conclusions” (Krippendorff, 2004, p. 242). In content analysis research,

- variables with alpha values below $\alpha = .667$ are generally considered unreliable and should not be accepted,
- variables with alpha values between $\alpha = .667$ and $\alpha = .800$ should be used only for drawing tentative conclusions, and
- variables with alpha values greater than $\alpha = .800$ can be considered reliable (Krippendorff, 2004, p. 241).

Because the cost of drawing wrong conclusions in this study is low, the listed heuristics are acceptable. For each applicable variable coded manually in the second phase of the study, the study reported the simple percentages of agreement and the Krippendorff’s alpha.

Chapter Summary

This chapter explains the two content analytic methods chosen for and implemented in this study. Seven academic publications that have either a mainstream medical informatics focus or a focus in the related fields of medicine or management information systems comprised the population from which the sample data were collected. The journals from which article texts were collected were the *Journal of the American Medical Informatics Association*, and the *International Journal of Medical Informatics*, the *Journal of the American Medical Association*, the *New England Journal of Medicine*, *Management Information Systems Quarterly*, *Information Systems Research*, and *Communications of the Association for Computing Machinery*. All article texts from the medical informatics publications for the years spanning 2002 – 2011 were

collected. Queries were used to collect articles for the same period from the non-medical informatics publications.

After collecting and preparing the data, the study applied two content analytic methods to identify emerging themes in the literature. The first, centering resonance analysis, is a quantitative method used to identify influential noun phrases from a body of text. As a complement to the centering resonance analysis, the study performed an exploratory factor analysis using the most influential words identified in the centering resonance analysis to extract themes from the texts. The second phase of the study performed manual coding of the article texts using categories identified from the literature review and the steps established in the coding procedures. The coders documented the data using an online text analysis tool, *DiscoverText* and an inter-coder reliability analysis was performed. The next chapter presents the findings of the two content analytic approaches.

Chapter 4. Findings

This research study is an exploratory analysis to ascertain the prevalent themes, the challenges that exist, and future directions for the medical informatics discipline. To derive the data for the study, seven scholarly publications were sampled from the ten-year period of 2002 - 2011. The sample population included abstract and article texts drawn from the mainstream medical informatics publications *Journal of the American Medical Informatics Association* (JAMIA), and *International Journal of Medical Informatics* (IJMI). Additional samples were collected from publications in the related fields of medicine and management information systems: the *Journal of the American Medical Association* (JAMA) the *New England Journal of Medicine* (NEJM), *Management Information Systems Quarterly* (MISQ), *Information Systems Research* (ISR), and *Communications of the Association for Computing Machinery* (CACM). From the medical informatics publications, JAMIA and IJMI, all the published articles for the ten year period were collected. For those publications outside the medical informatics mainstream, JAMA, NEJM, MISQ, ISR, and CACM, the study used queries focused on medical informatics terms to identify articles to collect for analysis. This chapter discusses the results of the two phases of the study. Phase I, the quantitative analysis, includes the discussion of the centering resonance analysis technique for identifying themes—an approach focused on the manifest content—and Phase II, the qualitative analysis, includes discussion of the manual coding approach used for further theme identification and categorization based on the latent content of the texts.

Descriptives

Several criteria were used to select articles in this study. Using a decade as the basis for the study, for the journals published from 2002 - 2011, a broad spectrum of journal articles were included that spanned the medical, medical informatics, and management information systems disciplines. To cover the medical informatics field, articles were sampled from two mainstream, highly rated, medical informatics publications: the *Journal of the American Medical Informatics Association* (JAMIA), and the *International Journal of Medical Informatics* (IJMI). To incorporate the medical discipline, the *Journal of the American Medical Association* (JAMA) and the *New England Journal of Medicine* (NEJM) were sampled. Finally, to ensure the field closely related to medical informatics, management information systems, was represented, articles from *Management Information Systems Quarterly* (MISQ), *Information Systems Research* (ISR), and *Communications of the ACM* (CACM) were included.

While previous researchers have excluded letters to the editor, this study included them because they exemplify the views of members of the academic community and the editors felt they were important enough to be included in the journal. The letters reflect the views of the field despite their not being considered academic research (i.e., not peer-reviewed). For those letters that are included, the full text of the letter is incorporated into the dataset. Further, some of the articles from CACM, a refereed journal, are not peer-reviewed; yet, they are included for reasons similar to those for letters to the editor. Each of the letters to the editor and non peer-reviewed articles take valuable publication space and if the editors believe they are important enough to include, this researcher accepts their judgment. These article texts remained in the

overall dataset, deferring to the judgment of the editors of these journals in adding the articles or letters to their publications.

Based on the different types of publications, varying means of collecting the articles were used. For the medical informatics publications, JAMIA and IJMI, the sample includes all articles for the ten-year period determined for the study. For the medical journals, the sample used two different techniques: for JAMA, the study collected all the articles from the specific section of their online journal devoted to medical informatics and internet-based medicine. For NEJM, the study used a Boolean query in Thomson Reuter's ISI Web of Knowledge collection to extract articles that involved the relation between medicine and information technology. Likewise, to ensure a broad view of the medical informatics discipline, the researcher queried the ISI Web of Knowledge collection for medically-related articles in MISQ, ISR, and CACM (see Chapter Three for sample query code). Using the various means of data collection, the study gathered all the article texts in a citation manager software application.

Using the citation manager software, Zotero, all article information was captured in an SQLite database—the SQLite format is in the public domain and, therefore, free. Once collected in the SQLite database, the abstract text, author information, publication date, and other citation information for each record in the database were extracted in a comma separated value format, a common spreadsheet format, for the data preparation and cleaning process. An example of the code used to extract the JAMIA data from the SQLite database is found in Appendix A. Similar code was used to extract the data for the other publications.

Data Preparation

The manual data preparation and cleaning procedures began after using the comma separated value formatted files to import the data into Microsoft Excel™. As part of the cleaning, the four-digit year and the publication name were extracted for use in identifying and discerning each article text. Some of the publications included relatively standardized abstract formats including frequently used keywords such as *purpose*, *objective*, *study design*, etc. To prevent skewing of the quantitative analysis, which is based on noun phrases, these high-frequency keywords were removed.

The next step of the data preparation process consisted of evaluating the data for exclusion from further analysis. One exclusion criterion was defined as any article text that was clearly unrelated to medical informatics. To be clearly unrelated to medical informatics, the article texts could not meet this study's definition of medical informatics. In addition, the article texts were excluded if they did not have content that addressed both information systems technology and healthcare. The exclusion criteria were used conservatively to ensure a broad perspective of the medical informatics discipline was captured. Unless the text specifically met the exclusion criteria and therefore, was removed, the text was retained for further the analysis. One of the objectives of this study was to garner an understanding of the field of medical informatics from a wide perspective. For example, articles discussing computed tomography (CT) scans were included even when the discussion in the article was focused more toward the medical aspects of the scans than the CT technology.

The results of the data collection and cleaning process are found in Table 4.1. The manifest content—that which is observable—is expressed as quantitative descriptive data

representing the ten years of journal publications sampled. The original dataset based on the inclusion criteria totaled 2,315 article texts and the subsequent total after data processing equaled 2,188. These values are represented in Table 4.1. Table 4.1 includes the number of articles per journal from the original collection phase, the number of articles removed based on the exclusion criteria, the total remaining articles, and the number of remaining articles represented as a percentage of the original dataset. The numbers represent a baseline from which frequencies and percentages are generated for the themes found in the analyses.

Table 4.1
Number of Journal Articles by Publication

Journal Abbreviation	Journal Title	Journal Focus	Met Inclusion Criteria	Excluded	Total Remaining	% of Original Remaining
CACM	Communications of the Association for Computing Machinery	Management Information Systems	70	18	52	74.29%
IJMI	International Journal of Medical Informatics	Medical Informatics	966	17	949	98.24%
ISR	Information Systems Research	Management Information Systems	10	0	10	100.00%
JAMA	Journal of the American Medical Association	Medical	238	0	238	100.00%
JAMIA	Journal of the American Medical Association	Medical Informatics	812	33	779	95.94%
MISQ	Management Information Systems Quarterly	Management Information Systems	19	0	19	100.00%
NEJM	New England Journal of Medicine	Medical	200	59	141	70.50%
Total			2,315	127	2,188	94.51%

Despite the conservative exclusion approach, 59 (29.50%) articles were removed from the original NEJM collection before subsequent analysis. This may indicate an overuse/improper use of terms such as “information system” in the medical literature. Or perhaps, this is an illustration of the overlap and multidisciplinary focus of the medical informatics field, and thus illustrates the difficulties of comprehensive analysis in the discipline. Regardless, the 59 articles removed did not recount research or discussions related to medical informatics as defined for this study. After data cleaning, 141 (70.50%) of the original 200 articles remained for subsequent analysis.

Difficulties lie with journals that are not specifically focused toward academia. While in academic publications there is a relatively standard format for articles which include an abstract. Such is not the case with the quasi-practitioner/academic focused publication, CACM. For example, with CACM, of the original 70 articles returned from the query, only 23 (32.86%) included abstracts. For the remaining 47 (67.14%) articles, it was necessary to manually extract the abstract information from the article text. After assessing the CACM articles against the exclusion criteria, 18 (25.71%) articles were removed from the data set; two contained errata data and 16 did not meet the criteria of discussing medicine and technology. Thus, the total of usable CACM articles after data cleaning was 52 (74.29% of the original).

The large number of IJMI articles (966) produced rich results. Of the 966 articles collected in the inclusion phase, only 17 were removed, retaining 949 (98.24%) for further analysis. Of the 17 (1.76%) articles excluded, three were errata data, three were calls for papers, one was a list of submission instructions for authors, and one was a tribute to a scholar who had passed away. The remaining nine excluded articles included publication announcements, a list

of conference participants, an editorial board member listing, and similar non-medical informatics specific information.

Another large number of articles came from JAMIA. Because it is a mainstream medical informatics publication, all JAMIA article texts were collected for the period 2002-2011. From the original 812 articles collected, 33 (4.06%) articles were removed, leaving 779 (95.94%) for further analysis. Unlike IJMI article texts, JAMIA texts did not include any extraneous texts such as errata or submission instructions. All 33 article texts were removed based on their meeting the exclusion criteria.

No article texts from the original data collection for each of the three remaining publications were removed. Based on the previously stated collection method for the publications and the exclusion criteria, it was not necessary to exclude any texts from them. The full complement of article texts were retained for subsequent analysis from the publications, ISR, JAMA, and MISQ ($N = 10, 238, \text{ and } 19$, respectively).

To prepare the data for use in the quantitative portion of the study with the centering resonance analysis software, the data were exported from Microsoft Excel™ in the character-encoding scheme, ASCII. The American Standard Code for Information Interchange (ASCII) is a commonly used text format that is readable by most software applications. The ASCII format is based on the English alphabet and has been in use for decades. The centering resonance analysis software requires data input in the ASCII text format.

For the manual coding qualitative phase of the study, the Excel™ files were used without modification from the way they were upon completion of the cleaning process. The DiscoverText web-based software was used for the coding process is able to import, among

many other formats, Excel™ files in their proprietary format. The import process in DiscoverText is straightforward and streamlined.

Mixed Methods Analysis

The study is comprised of two phases, the semi-automated quantitative method, and the qualitative method. To perform the quantitative phase, software was used for the automated portion and the automated result was followed with a manual analysis component, hence, the term *semi-automated method*. To analyze the data in the second phase of the study, the qualitative phase, three researchers coded the data based on criteria derived during the literature review (see Chapter Two). The coders read the article texts that remained after the data preparation/cleaning process and coded them for categorization in the appropriate theme and/or characteristic classification(s) using the procedures in the Coding Procedures handout (Appendix B).

Phase I: Quantitative Data Analysis –Centering Resonance Analysis

The first phase of the study, the quantitative analysis phase, was comprised of analyzing the prepared dataset using centering resonance analysis methods to extract medical informatics themes from the article texts collected from the seven sample journals. Because the focus of the study is on the emergent themes in the medical informatics discipline over time, the citation information and abstracts were collated by year and saved as text files for further analysis. The result of the data cleaning and collation yielded 2,686 pages of text to analyze. Computer-assisted analysis is beneficial when working with large volumes of text. The software used in this study was specifically designed to support centering resonance analysis and is named *Crawdad Text Analysis System* (Corman & Dooley, 2006).

Noun Phrase Network Information

The first stage of centering resonance analysis consisted of generating noun phrase network information for the entire dataset as a whole, the individual years, the aggregated medical informatics publications, and the aggregated non-medical informatics publications. The results for all years combined and the individual are found in Table 4.2 (table limited to the top twenty words for clarity). With all years and publications combined, the number of nodes—the number of noun or noun phrases in the network—is 17,501 (Table 4.2 “All Years Combined”). The density score was .0018 (the density score is a ratio of the number of network connections that exist among nodes compared to how many could possibly exist). The group influence score, an indicator of the coherency of the entire network of noun phrases, was .0596. Both the density and group influence scores are standardized measures with minimum scores of zero and maximum score of one. The weak density and influence scores indicate that the network of noun phrases among all publications is not tightly connected and the texts are not very coherent within the network; i.e., there is much diversity in the content of the journal article texts.

When segregating the results by mainstream medical informatics publication versus non-mainstream, the noun phrase network information shows similar, loosely connected networks of noun phrases, but a higher level of coherency. For those article texts that came from the medical informatics mainstream publications, the density was .00242 (Table 4.2 “MI Pubs - All Years”). While the density is greater than that of the combined dataset, it is not a indication that the network is tightly connected. In fact, the network is loosely connected. The density of the non-medical informatics publications texts was not substantially better (density = .00225; Table 4.2 “non-MI Pubs - All Years”). Likewise, the group influence values for the medical informatics publication texts and the non-medical informatics publication texts are low (.07606 and .09553,

respectively). Both of these group influence scores indicate that the article texts, once segregated by those that are from the mainstream medical informatics journals and those that are not, are slightly more coherent than when they are aggregated. The higher coherency indicates that the discussions in each of these segmented groups of texts have a greater propensity toward common themes than when combined.

Table 4.2

Noun Phrase Network Information - All Years, MI Pubs, and non-MI Pubs

	All Years Combined		MI Pubs - All years		Non-MI Pubs all years	
	Nodes	Density	Words	Influence	Words	Influence
Years	17501					
Nodes	17501		12726		9346	
Density	0.0018		0.00242		0.00225	
Group Influence	0.0596		0.07606		0.09553	
	Word	Influence	Word	Influence	Word	Influence
	patient	.060	system	.076	patient	.096
	system	.059	patient	.049	use	.040
	information	.034	information	.043	health	.039
	health	.033	data	.040	system	.035
	data	.030	health	.036	study	.032
	use	.026	medical	.029	information	.024
	study	.025	study	.025	medical	.022
	medical	.024	clinical	.023	care	.020
	clinical	.020	care	.021	clinical	.020
	care	.018	use	.019	disease	.020
	physician	.015	method	.018	physician	.019
	hospital	.015	model	.017	data	.018
	model	.013	hospital	.017	risk	.017
	method	.013	physician	.016	new	.016
	time	.012	al.	.013	hospital	.015
	al.	.011	time	.013	year	.014
	disease	.011	technology	.012	CT	.014
	technology	.011	analysis	.011	case	.013
	analysis	.010	research	.011	rate	.013
	case	.010	approach	.011	level	.012

Note. MI = medical informatics

The results of the article texts separated by year are similar to those for the “All Years Combined” dataset (Table 4.3). The density scores range from a high of .0052 for 2011 to a low of .0031 for both 2004 and 2007 ($M = .0036$, $SD = .000629$). The Group influence values range from a high of .1207 for 2011 and a low of .0773 for 2007 ($M = .0947$, $SD = .014179$). These density and group influence scores suggest that the article texts included in this study are comprised of fairly diverse discussions.

Table 4.3 (2002-2005)

		2002		2003		2004		2005	
		Word	Influence	Word	Influence	Word	Influence	Word	Influence
Nodes	4,604			4,406		4,851		4,236	
Density	0.0034			0.0034		0.0031		0.0037	
Group Influence	0.1029			0.0999		0.1207		0.0890	
	Word	Influence	Word	Influence	Word	Influence	Word	Influence	
	patient	.103	system	.100	patient	.121	patient	.089	
	system	.083	patient	.088	system	.077	system	.071	
	information	.049	data	.050	information	.060	health	.054	
	data	.049	information	.046	health	.048	information	.052	
	clinical	.049	health	.042	medical	.042	clinical	.050	
	health	.043	use	.034	clinical	.033	use	.045	
	medical	.041	clinical	.034	care	.032	care	.035	
	study	.038	medical	.033	data	.029	data	.035	
	care	.031	care	.031	physician	.026	study	.034	
	use	.029	study	.030	study	.025	medical	.032	
	disease	.026	hospital	.026	use	.025	hospital	.021	
	case	.019	model	.018	process	.023	physician	.021	
	informatics	.018	process	.017	hospital	.021	method	.018	
	error	.016	physician	.015	aortic	.018	time	.017	
	time	.015	diagnosis	.015	guideline	.017	analysis	.016	
	hospital	.015	risk	.014	university	.015	communicatio	.016	
	physician	.013	case	.013	computer	.015	development	.016	
	program	.012	evaluation	.013	time	.013	result	.015	
	percent	.012	previous	.013	order	.013	support	.013	
	level	.012	term	.013	year	.013	research	.012	

Note . MI = medical informatics

Table 4.3 (continued; 2010-2011)
Noun Phrase Network Information by Year

Years	2010		2011	
Nodes	4,166		2,226	
Density	0.0039		0.0052	
Group Influence	0.0780		0.0962	
	Word	Influence	Word	Influence
	system	.078	system	.097
	health	.067	health	.093
	patient	.063	patient	.086
	data	.055	information	.075
	information	.047	care	.066
	medical	.038	data	.058
	use	.037	medical	.044
	study	.030	use	.035
	care	.027	study	.032
	technology	.021	technology	.030
	clinical	.020	clinical	.029
	implementatio	.019	user	.028
	physician	.018	physician	.026
	analysis	.018	method	.025
	method	.017	hospital	.019
	new	.017	model	.017
	research	.015	research	.016
	hospital	.015	potential	.015
	record	.015	analysis	.015
	time	.014	case	.014

Note . MI = medical informatics

Influential Words

In addition to the node, density, and group influence properties, the top 50 influential words for each of the groups of texts were generated (see Tables 4.2 and 4.3; tables limited to top 20 influential words for clarity). The influence score is normalized and ranges from zero to one. The higher the influence score, the more influential the word is. The influence scores of .10 or greater indicate that the words are significantly influential. The more influential the word is, the more it is the central point, tying together the meanings and thoughts of the text. Some of the

most influential words across all years and categories (i.e., medical informatics mainstream vs. non-mainstream) include *patient, system, health, data, information, medical, clinical, care, and hospital*. The influence of these words should be expected in a discipline founded in the fields of medicine and information technology. While the top influential words give a hint toward the themes of the medical informatics literature, further analysis can extract a refined perspective on the field.

Exploratory Factor Analysis

To assess the underlying thematic structure of the medical informatics literature, an exploratory factor analysis (EFA) was performed. The 50 most influential words were the variables and the influence values were score values for each of the corresponding variables. The EFA was performed using principal components analysis with Varimax rotation. Based on an eigenvalue greater than one cutoff, the resulting factor solution includes 17 factors and explains 85.85% of the variance (Table 4.4). After rotation, the first factor explained 9.64% of the variance, the second factor accounted for 6.80% of the variance, the third factor accounted for 6.61% of the variance, the fourth factor explained 5.91% of the variance, the fifth factor accounted for 5.83% of the variance, the sixth factor accounted for 5.71% of the variance, the seventh factor explained 5.66% of the variance, the eighth factor accounted for 5.30% of the variance, the ninth factor accounted for 4.43% of the variance, the tenth factor explained 4.08% of the variance, the eleventh factor accounted for 4.03% of the variance, the twelfth factor accounted for 3.91% of the variance, the thirteenth factor explained 3.87% of the variance, the fourteenth factor accounted for 3.86% of the variance, the fifteenth factor accounted for 3.85% of the variance, the sixteenth factor explained 3.52% of the variance, and the seventeenth factor accounted for 2.86% of the variance.

Table 4.4
Variance Explained

Component	% of Variance	Cumulative %
1	9.64	9.64
2	6.80	16.43
3	6.61	23.04
4	5.91	28.95
5	5.83	34.78
6	5.71	40.49
7	5.66	46.15
8	5.30	51.45
9	4.43	55.88
10	4.08	59.96
11	4.03	63.99
12	3.91	67.89
13	3.87	71.76
14	3.86	75.62
15	3.85	79.47
16	3.52	82.99
17	2.86	85.85

The rotated factor loadings provide a starting point for evaluating the themes (Table 4.5; loadings less than .40 are omitted to improve clarity). The first factor received strong loadings on the influential words *confidentiality*, *surveillance*, *protocol*, *provider*, *time*, and *disease* (.983, .982, .965, .781, .725, and .551, respectively). The second factor had strong loadings on the influential words *maker* (.979), *HIT* (.968), *process* (.744), and *research* (.735). The third theme had strong loadings on the influential words, *service* (.929), *user* (.911), *technology* (.901), and *risk* (.540). From the fourth group, the influential words *computing*, *computer*, *health*, and *medical* materialized (.870, .743, .615, and .590, respectively). The fifth factor received strong loadings on the influential words, *peer*, *internet*, and *individual* (.956, .935, and .841, respectively). The sixth factor had strong loadings on the influential words, *quality*, *physician*, *hospital*, *care*, *medical*, and *clinical* (.873, .735, .702, .514, .493, and .422, respectively). The

seventh factor had strong loadings on the influential words, *CT*, *risk*, *method*, *system*, *patient*, *disease*, and *use* (-.778, -.612, .562, .540, -.543, -.507, and -.499, respectively). Of the seventh group of influential words, five had negative loadings, *CT*, *risk*, *patient*, *disease*, and *use*. The eighth theme had strong loadings on the influential words, *firm*, *IT*, and *knowledge* (.951, .888, and .651, respectively). The ninth factor had strong loadings on the influential words, *result*, *study*, and *analysis* (.842, .792, and .434, respectively). Out of the tenth group, the influential words, *attitude* (.929), *information* (.699), and *individual* (.401) arose. The eleventh factor received strong loadings on the influential words, *conference* (.801) and *data* (.796). The twelfth factor had strong loadings on the influential words, *case*, *process*, and *project* (.924, .566, and .564, respectively). The thirteenth factor had strong loadings on the influential words, *performance*, *analysis*, and *knowledge* (.833, .774, and .423, respectively). From the fourteenth group, the influential words, *communication*, *patient*, and *provider* materialized (.856, .582, and .542, respectively). The fifteenth factor received strong loadings on the influential words, *management*, *project*, and *knowledge* (.973, .739, and .412, respectively). The sixteenth factor had strong loadings on the influential words, *trust* and *research* (.890 and .423, respectively). Both of the influential words for the seventh and final factor had negative loadings, *community* (-.855) and *use* (-.449).

Table 4.5
Factor Loadings for the Rotated Factors

Factor	Component	Loading
1	confidentiality	.983
	surveillance	.982
	protocol	.965
	provider	.781
	time	.725
	disease	.551
2	maker	.979
	HIT	.968
	process	.744
	research	.735
3	service	.929
	user	.911
	technology	.801
	risk	.540
4	computing	.870
	computer	.743
	health	.615
	medical	.590
5	peer	.956
	internet	.935
	individual	.841
6	quality	.873
	physician	.735
	hospital	.702
	care	.514
	medical	.493
	clinical	.422
7	ct	-.778
	risk	-.612
	method	.562
	system	.540
	patient	-.543
	disease	-.507
	use	-.499

Table 4.5 (continued)
Factor Loadings for the Rotated Factors

Factor	Component	Loading
8	firm	.951
	it	.888
	knowledge	.651
9	result	.842
	study	.792
	analysis	.434
10	attitude	.929
	information	.699
	individual	.401
11	conference	.801
	data	.796
12	case	.924
	process	.566
	project	.564
13	performance	.833
	analysis	.774
	knowledge	.423
14	communication	.856
	patient	.582
	provider	.542
15	management	.973
	project	.739
	knowledge	.412
16	trust	.890
	research	.423
17	community	-.855
	use	-.448

Note 1 : The top influential words removed from the final factor solution include: *biology* and *new*

Note 2 : The top influential words that loaded on more than one factor include: *analysis, disease, knowledge, individual, medical, patient, process, project, provider, research, risk, use,*

Final Theme Solution

While the EFA provides a good foundation for developing themes from the factors identified, some scholars suggest a further step—latent coding—to refine the themes (Tate, et al., 2010). Coding latent content is likely to introduce variation in results among researchers because the content is less directly observable and researchers use more subjective measures in their analyses. When subjectivity is introduced during the latent coding, reliability begins to deteriorate. On the other hand, Crowdad develops networks of words—specifically, nouns and noun phrases—and not networks of theoretical constructs or concepts. The secondary latent coding analysis can “logically connect words to themes and strengthen the face validity of the theme” (Tate, et al., 2010, p. 25). Starting with the rotated factor solution, the researcher re-read texts from the dataset that referenced the influential words in each of the factors. Based on a thorough reading of the article texts and an examination of the factor loadings, the two data sources were synthesized to express the themes of the literature. The final theme solution and the factors for each theme are presented in Table 4.6.

Table 4.6
Final Theme Solution

Theme	Factors
Analytics	12, 13
Healthcare Operations and Standards: Operations	6
Healthcare Operations and Standards: Project Management	15
Healthcare Operations and Standards: Information Assurance	1
Aspects of Healthcare Research	2, 7, 9, 16
Knowledge Transfer/Communication: Extending beyond the Organization	11
Knowledge Transfer/Communication: Internal to the Organization	8
Knowledge Transfer/Communication: Patient-Provider	5, 14
Perceptions and Managing Expectations of Information Technology	10, 17
Software as a Service	3, 4

The first theme emanating from the article texts was *analytics* and combined both factors 12 and 13. The article texts associated with factors 12 and 13 referenced analytics—the application of information systems to combine operational research and technology to solve hospital business problems. Several factors combined to create the major theme of *Healthcare Operations and Standards* (factors 1, 6, and 15). After reading several of the texts associated with the influential words from factor 6, the sub-theme *Operations* became apparent within the major theme of *Healthcare Operations and Standards*. Another sub-theme that seemed to reveal itself from the *Healthcare Operations and Standards* was *Project Management*. The texts surrounding the influential words, *management*, *project*, and *knowledge* from the fifteenth factor

appear to index project management. *Security/Privacy* is the third sub-theme of the *Healthcare Operations and Standards* theme and the words found in factor one of the rotated factor solution seem to address patient security and privacy discussions in the texts. Rotated factors 2, 7, 9, and 16 all provided perspectives about research; some of the articles were directly related to research in the healthcare environment and some provided discussions of laboratory related research. The words associated with rotated factors 5, 8, 11, and 14 all appear to reference a major theme, *Knowledge Transfer/Communication*, and the texts that the influential words of each factor point to support it. Breaking down the *Knowledge Transfer/Communication* theme further, the texts associated with the influential words of factor 11 support a sub-theme of *Inter-organization*. Although the theme, *Knowledge Transfer/Communication: Inter-organization*, addresses articles discussing the communication of the healthcare organization with individuals or organizations outside its own, the sub-theme *Knowledge Transfer/Communication: Intra-organization* recognizes articles discussing healthcare organizations and the communication that goes on within the organization. The *Knowledge Transfer/Communication: Inter-organization* theme has factor 11 as its source. The sub-theme, *Knowledge Transfer/Communication: Patient-Provider* (factors 5, 14), focuses on those article texts that discuss the communications and/or communications processes that occur between a patient and a healthcare provider. The *Perceptions and Managing Expectations of Information Technology* theme is based on readings of the article texts associated with the influential words of factors 10 and 17. The final theme to emerge from the article texts was comprised of two factors—three and four—and addressed studies that analyzed the use of software centralized on a server and executed on a remote computer desktop. The influential words from factor three (*service, user, technology, risk*) and

factor four (*computing, computer, health and medical*) combined to define the final emerging theme of *Software as a Service*.

Phase II: Manual Coding

The goal of this study is to describe the themes emerging in the medical informatics discipline, as represented by the abstracts and texts of the medical informatics publications and publications of closely related disciplines: management information systems and medicine. The second phase of the study, the manual coding phase, is primarily focused toward gaining an understanding of the sub-domains of the medical informatics discipline and the appropriateness of diffusion of innovations theory in the medical informatics discipline.

Inter-Coder Reliability

Before making inferences based on manually-coded data, it is essential to ensure the data is reliable based on the *a priori* standards of the study. Simple agreement was calculated as a percentage for all latent variables in the second phase of the study. The variables each had finite responses, either present (variable = 1) or absent (variable = 0). Krippendorff's alpha was calculated to correct for chance agreement and further reduce the possibility of coder agreement based on chance. Correction for chance agreement was based on Krippendorff's heuristic (Krippendorff, 2004). Those variables with an alpha value below $\alpha = .667$ were considered unreliable and, therefore, not used for further analysis. Variables with alpha values between $\alpha = .667$ and $\alpha = .800$ were considered reliable only for drawing tentative conclusions and variables were considered reliable if their alpha values exceeded $\alpha = .800$. Because the cost of drawing wrong conclusions in this study is low, the listed heuristics are acceptable. Using the Krippendorff heuristic as a reference for simple agreement, the *a priori* simple agreement

standard was set at a minimum 80% for reliability, and 68% as the minimum standard for drawing tentative conclusions.

While each of the secondary coders received about 250 article texts to code (coder A received 251; coder B received 250), the primary researcher coded the complete set of article texts (2,188). The coder overlap was 23%, exceeding the 15% set as the minimum *a priori*. The minimum simple agreement for all variables was met between the primary researcher and coder A, agreeing 89.2% on texts using the variable *Information Architecture*, which exceeds the minimum 80% standard. Therefore, all variables were agreed upon by at least 89.2% and were considered reliable based on simple agreement. Once corrections were introduced using Krippendorff's alpha calculation to adjust for chance agreement, all variables met the acceptable reliability requirements except *Business Analytics*, *Complexity*, *Trialability*, and *Observability*. While the primary researcher and coder A produced marginally unacceptable agreement on variable *Remove*, ($\alpha = .6642$), the primary researcher and coder B produced strong agreement ($\alpha = .880$). After analysis, it was agreed that the results of the variable, *remove*, despite the strong agreement between the primary coder and coder B, would be considered unacceptable. Table 4.7 depicts the inter-coder reliability between the primary coder and coders A and B for the manual coding phase of the study. The table includes the variable names, the simple agreement stated as a percent, the Krippendorff's Alpha, the number of agreements and disagreements between each of the two coders and the primary coder, the number of cases coded, and the number of agreement decisions made. The variable, *remove*, will be discussed further in the discussion section.

Table 4.7
Inter-coder Reliability

Variable	Inter-coder Reliability of Primary Coder and Coder A (no. Cases = 251; no. Decisions = 502)				Inter-coder Reliability of Primary Coder and Coder B (no. Cases = 250; no. Decisions = 500)			
	Percent		α	Disagree- ments	Percent		α	Disagree ments
	Simple Agreement	Agree- ments			Simple Agreement	Agree ments		
Remove	93.2271	0.6642	234	17	94.4	0.8804	236	14
Direct Patient Care	95.2191	0.9046	239	12	93.6	0.8551	234	16
Business Analytics	89.6414	0.7125	225	26	91.2	0.5957	228	22
Information Architecture	89.2430	0.7342	224	27	96.4	0.9071	241	9
Relative Advantage	93.6255	0.8116	235	16	95.6	0.8617	239	11
Compatibility	100.0000	1.0000	251	0	99.6	0.9072	249	1
Complexity	95.6175	0.7517	240	11	98.8	0.3952	247	3
Trialability	96.8127	0.4846	243	8	98	0.4354	245	5
Observability	98.4064	0.6592	247	4	98.8	0.3952	247	3
Not Listed	100		251	0	100		250	0

Note . Variables with alpha values between $\alpha = .667$ and $\alpha = .800$ should be used only for drawing tentative conclusions; variables with alpha values greater than $\alpha = .800$ can be considered reliable.

Qualitative Results - Manual Coding

For each article, the coders identified the latent variables *Direct Patient Care, Business Analytics, Information Architecture, Relative Advantage, Compatibility, Complexity, Trialability, and Observability*. The variable, *Remove*, was used as an indicator that the article text was unrelated to medical informatics as defined as *the discipline dedicated to the systematic processing, analysis, and dissemination of health-related data through the application of digital information systems (computers) to various aspects of healthcare, research, and medicine*. The results of the analysis were captured in several tables to address the research questions and sub-questions. To provide an overview of the entire dataset, Table 4.8 depicts the results of the coding as an aggregate of all sampled article texts from all publications spanning the years 2002 - 2011. Also, Table 4.8 includes a breakdown of the coding based on the article texts sampled as segregated by the mainstream medical informatics publications (i.e., the *International Journal of Medical Informatics* and the *Journal of the American Medical Informatics Association*) and non-medical informatics publications (i.e., *Communications of the Association for Computing Machinery, Information Systems Research, the Journal of the American Medical Association, Management Information Systems Quarterly, and the New England Journal of Medicine*).

Table 4.8

Hierarchy of Articles to Themes for Sampled Publications Spanning 2002-2011 (raw results)

Theme	All Years, All Journals		Medical Informatics Journals		non-Medical Informatics Journals	
	Frequency	% of Articles	Frequency	% of Articles	Frequency	% of Articles
Remove	225	10.3%	113	6.5%	112	24.3%
Direct Patient Care	746	34.1%	608	35.2%	138	30.0%
Business Analytics	885	40.4%	699	40.5%	186	40.4%
Information Architecture	755	34.5%	648	37.5%	107	23.3%
Relative Advantage	235	10.7%	186	10.8%	49	10.7%
Compatibility	41	1.9%	38	2.2%	3	0.7%
Complexity	108	4.9%	102	5.9%	6	1.3%
Triability	32	1.5%	26	1.5%	6	1.3%
Observability	24	<i>1.1%</i>	22	<i>1.3%</i>	2	<i>0.4%</i>
Total Article Texts Analyzed	2,188		1,728		460	
Total Themes Identified	3,051	139%	2,442	141%	609	132%

Note. Percent of Total Themes Identified >100 because of the possibility that multiple themes are found in a single article text.

Total theme frequency for each journal grouping exceeds the total number of articles for each journal grouping because of the possibility that multiple themes are found in a single article text. Bold text indicates the most frequently identified theme. Italicized text indicates the least frequently identified theme.

The raw results of the coding of the overall dataset showed similarities among the most frequently identified variables in the various segregated groupings. When looking at the entire dataset as a whole (see *All Years, All Journals* columns of Table 4.8), Business Analytics was the most frequently identified theme. Of the 2,188 article texts in the dataset, Business Analytics was identified 885 times (40.4%). Likewise, *Business Analytics* was found to be the most frequently identified theme when the dataset was segregated into two groupings based on whether the journal is a mainstream medical informatics publication or a publication in the related fields of either medicine or management information systems (Business Analytics frequency = 699 (40.5%) and 186 (40.4%), respectively; see Table 4.8, columns labeled *Medical Informatics Journals*, and *non-Medical Informatics Journals*, respectively).

In like manner, the raw results of the dataset indicated similarities among the least frequently identified variables in the separate groups. Again referring to the *All Years, All Journals* columns of Table 4.9, we see that the least frequently observed variable was Observability (frequency = 24; 1.1%). The same variable, Observability, was found to be the least frequently identified variable in each of the other groupings, medical informatics journals and non-medical informatics journals (Table 4.8, columns labeled *Medical Informatics Journals*, and *non-Medical Informatics Journals*, respectively). In the medical informatics journal grouping, observability was observed only 22 times (1.3%) in the 1,728 article texts of the group. In the non-medical informatics group, the frequency was even lower (frequency = 2; 0.4%). The raw results of the analysis stratified by year are depicted in Table 4.9 and show a greater diversity than when that of the dataset when observed as a whole. However, the diversity was somewhat limited. For example, three themes consistently revealed themselves as the most identified themes. Direct Patient Care was identified as the most frequent theme in 2008 and

2011 (frequency = 97 (37.6%) and 30 (46.0%), respectively). For the years 2004, 2005, 2006, 2007, and 2009, Business Analytics was most frequently identified (frequency = 77 (37.2%), 99 (43.0%), 93 (45.8%), 145 (46.0%), and 125 (43.7%), respectively). For the remaining years of 2002, 2003, and 2010, Information Architecture was most frequently identified (frequency = 73 (40.3%), 78 (38.2%), and 88 (40.6%), respectively). In reviewing the least frequently identified variables, we see that there were four variables that were consistently the lowest. For 2005 and 2006, Compatibility came up least frequently, being identified only three times (1.3%) in 2005, and only one time (0.5%) in 2006. Complexity was another variable identified infrequently; in fact, in 2004, there were zero observations of Complexity. The least frequently identified theme for the years 2003, 2008, 2009, and 2011 was Trialability (frequency = 2 (1.0%), 2 (0.8%), 3 (1.0%), and 0 (0.0%), respectively). In the 2008 dataset, however, there was another theme identified as a tie with Trialability, Observability. Observability was the least frequently identified theme in 2004, 2007, 2008, and 2010 (frequency = 0 (0.0%), 1 (0.3%), 2 (0.8%), and 0 (0.0%), respectively).

Table 4.9 (2002-2006)
Hierarchy of Articles to Themes for Sampled Publications by Year (raw results)

Theme	2002		2003		2004		2005		2006	
	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles
Remove	25	13.8%	18	8.8%	23	11.1%	25	10.9%	20	9.9%
Direct Patient Care	44	24.3%	70	34.3%	71	34.3%	71	30.9%	66	32.5%
Business Analytics	62	34.3%	74	36.3%	77	37.2%	99	43.0%	93	45.8%
Information Architecture	73	40.3%	78	38.2%	70	33.8%	76	33.0%	69	34.0%
Relative Advantage	23	12.7%	29	14.2%	17	8.2%	33	14.3%	35	17.2%
Compatibility	4	2.2%	6	2.9%	3	1.4%	3	1.3%	1	0.5%
Complexity	2	1.1%	15	7.4%	9	4.3%	15	6.5%	12	5.9%
Triability	5	2.8%	2	1.0%	1	0.5%	8	3.5%	4	2.0%
Observability	3	1.7%	3	1.5%	0	0.0%	5	2.2%	4	2.0%
Total Article Texts Analyzed	181		204		207		230		203	
Total Themes Identified	241	133%	295	145%	271	131%	335	146%	304	150%

Note. Percent of Total Themes Identified >100 because of the possibility that multiple themes are found in a single article text. Total theme frequency for each journal grouping exceeds the total number of articles for each journal grouping because of the possibility that multiple themes are found in a single article text. Bold text indicates the most frequently identified theme. Italicized text indicates the least frequently identified theme.

Table 4.9 (continued; 2007-2011)
Hierarchy of Articles to Themes for Sampled Publications by Year (raw results)

Theme	2007		2008		2009		2010		2011	
	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles
Remove	35	11.1%	26	10.1%	24	8.4%	23	10.6%	6	6.9%
Direct Patient Care	90	28.6%	97	37.6%	113	39.5%	84	38.7%	40	46.0%
Business Analytics	145	46.0%	95	36.8%	125	43.7%	85	39.2%	30	34.5%
Information Architecture	105	33.3%	88	34.1%	76	26.6%	88	40.6%	32	36.8%
Relative Advantage	31	9.8%	29	11.2%	23	8.0%	17	7.8%	3	3.4%
Compatibility	2	0.6%	5	1.9%	10	3.5%	5	2.3%	2	2.3%
Complexity	20	6.3%	15	5.8%	10	3.5%	8	3.7%	2	2.3%
Trialability	5	1.6%	2	0.8%	3	1.0%	2	0.9%	0	0.0%
Observability	<i>1</i>	<i>0.3%</i>	2	0.8%	5	1.7%	0	0.0%	1	1.1%
Total Article Texts Analyzed	315		258		286		217		87	
Total Themes Identified	434	138%	359	139%	389	136%	312	144%	116	133%

Note. Percent of Total Themes Identified >100 because of the possibility that multiple themes are found in a single article text. Total theme frequency for each journal grouping exceeds the total number of articles for each journal grouping because of the possibility that multiple themes are found in a single article text. Bold text indicates the most frequently identified theme. Italicized text indicates the least frequently identified theme.

Despite the varied results, it is important to note that the frequency of the themes identified remained relatively constant over the years. There were no stark contrasts indicating that one year a particular variable was identified most frequently and another year in which the same variable was identified least frequently. This lends to the idea that, for the article texts sampled, the content has been fairly consistent over the years.

Removing the themes deemed unreliable based on the inter-coder reliability analysis, we see a different outcome of the most frequently and least frequently identified themes. Based on the reliability results, the variables *Business Analytics*, *Complexity*, *Trialability*, and *Observability* were removed, leaving the themes *Direct Patient Care*, *Information Architecture*, *Relative Advantage*, and *Compatibility*. The category *Remove* failed to meet the acceptable reliability criteria and was also withdrawn from the list. An overview of the entire dataset spanning all years and all publications, after removing unreliable data, is provided in Table 4.10. Table 4.10 also includes a breakdown of the final dataset of the coding based on the article texts sampled as segregated by the mainstream medical informatics publications (i.e., the *International Journal of Medical Informatics* and the *Journal of the American Medical Informatics Association*) and non-medical informatics publications (i.e., *Communications of the Association for Computing Machinery*, *Information Systems Research*, the *Journal of the American Medical Association*, *Management Information Systems Quarterly*, and the *New England Journal of Medicine*).

Table 4.10

Hierarchy of Articles to Themes for Sampled Publications Spanning 2002-2011

Theme	All Journals		Medical Informatics Journals		non-Medical Informatics Journals	
	Frequency	% of Articles	Frequency	% of Articles	Frequency	% of Articles
Direct Patient Care	746	34.1%	608	35.2%	138	30.0%
Information Architecture	755	34.5%	648	37.5%	107	23.3%
Relative Advantage	235	10.7%	186	10.8%	49	10.7%
Compatibility	<i>41</i>	<i>1.9%</i>	<i>38</i>	<i>2.2%</i>	<i>3</i>	<i>0.7%</i>
Total Article Texts Analyzed	2,188		1,728		460	
Total Themes Identified	1,777	81%	1,480	86%	297	65%

Note. Bold text indicates the most frequently identified theme. Italicized text indicates the least frequently identified theme.

Final Coding Results

The final results of the coding of the overall dataset provide insight to the article texts sampled. When looking at the entire final dataset as a whole (column *All Journals*, Table 4.10), Information Architecture was the most frequently identified theme (frequency = 755; 34.5%). Likewise, in the medical informatics publications, Information Architecture was the most frequently identified theme (frequency = 648 (37.5%); column *Medical Informatics Journals*, Table 4.10). However, in the non-medical informatics journals, Direct Patient Care was identified 138 times (30.0%), which is more frequent than the other themes (Table 4.10, column *non-Medical Informatics Journals*).

The final results of the dataset coding indicated similarities among the least frequently identified variables in the separate groups. Again referring to the *All Journals* columns of Table 4.10, we see that the least frequently observed variable was Compatibility (frequency = 24; 1.9%). The same variable, Compatibility, was found to be the least frequently identified variable in each of the two other groupings, medical informatics journals and non-medical informatics journals (Table 4.10, columns labeled *Medical Informatics Journals*, and *non-Medical*

Informatics Journals, respectively). In the medical informatics journal grouping, Compatibility was observed only 38 out of the 1,728 article texts (2.2%). In the non-medical informatics group, the frequency was substantially lower (frequency = 3; 0.7%).

In looking at the stratified final results of the analysis in Table 4.11, we see that the most frequently identified themes over the years are limited to Direct Patient Care and Information Architecture. Eliminating the unreliably coded themes yields different results. For example, while Direct Patient Care was identified as the most frequent theme in only 2008 and 2011 in the raw dataset, Direct Patient Care was identified most frequently in the final dataset in years 2004, 2008, 2009, and 2011 (frequency = 71 (34.3%), 97 (37.6%), 113 (39.5%), and 30 (46.0%), respectively). Once the unreliable data was removed, Information Architecture became the most frequently identified theme in the remaining years of 2002, 2003, 2005, 2006, 2007, and 2010. The frequencies for Information Architecture for 2002 was 73 (40.3%), for 2003 was 78 (38.2%), for 2005 was 76 (33.0%), for 2006 was 69 (34.0%), for 2007 was 105 (33.3%), and for 2010 was 88 (40.6%). In reviewing the least frequently identified variables in the final dataset, we see that Compatibility was consistently the lowest over all years 2002 - 2011. The Compatibility frequency for 2002 was 4 (2.2%), for 2003 was 6 (2.9%), for 2004 was 3 (1.4%), for 2005 was 3 (1.3%), for 2006 was 1 (0.5%), for 2007 was 2 (0.6%), for 2008 was 5 (1.9%), for 2009 was 10 (3.5%), for 2010 was 5 (2.3%), and for 2011 was 2 (2.3%).

The consistency in the high frequency themes and low frequency themes identified seem to indicate that the articles have been fairly coherent in their content over the years and among the both the medical informatics publications and the non-medical informatics publications.

Table 4.11 (2002-2006)
Hierarchy of Articles to Themes for Sampled Publications by Year

Theme	2002		2003		2004		2005		2006	
	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles
Direct Patient Care	44	24.3%	70	34.3%	71	34.3%	71	30.9%	66	32.5%
Information Architecture	73	40.3%	78	38.2%	70	33.8%	76	33.0%	69	34.0%
Relative Advantage	23	12.7%	29	14.2%	17	8.2%	33	14.3%	35	17.2%
Compatibility	4	2.2%	6	2.9%	3	1.4%	3	1.3%	1	0.5%
Total Article Texts Analyzed	181		204		207		230		203	
Total Themes Identified	144	80%	183	90%	161	78%	183	80%	171	84%

Note. Bold text indicates the most frequently identified theme. Italicized text indicates the least frequently identified theme.

Table 4.11 (continued; 2007-2011)

Theme	2007		2008		2009		2010		2011	
	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles	Freq.	% of Articles
Direct Patient Care	90	28.6%	97	37.6%	113	39.5%	84	38.7%	40	46.0%
Information Architecture	105	33.3%	88	34.1%	76	26.6%	88	40.6%	32	36.8%
Relative Advantage	31	9.8%	29	11.2%	23	8.0%	17	7.8%	3	3.4%
Compatibility	2	0.6%	5	1.9%	10	3.5%	5	2.3%	2	2.3%
Total Article Texts Analyzed	315		258		286		217		87	
Total Themes Identified	228	72%	219	85%	222	78%	194	89%	77	89%

Note. Bold text indicates the most frequently identified theme. Italicized text indicates the least frequently identified theme.

Chapter Summary

This research study is an exploratory analysis to ascertain the prevalent themes in the medical informatics discipline spanning the years 2002 - 2011 using seven scholarly publications to derive the data. This chapter discussed the results of the quantitative and qualitative analysis components. While the original dataset included 2,315 articles, over the course of the study, 127 article texts were removed from the study because they (a) clearly did not reference medicine or medically-related research, (b) clearly did not reference information technology, and/or (c) were unobtainable. The total number of *Communications of the ACM* articles originally collected was 70, of which 53 (74.29%), were retained for further analysis. The original set of *International Journal of Medical Informatics* articles numbered 966; 17 were removed, leaving 949 (98.24%) of the original for further analysis. Of the original 812 articles initially collected from the *Journal of the American Medical Informatics Association*, 779 (95.94%) met the requirements for retention and further study. Of all the publications, the *New England Journal of Medicine* had the greatest number of articles removed from the study, retaining only 141 (70.50%) of the original 200 included articles. In contrast, all (100%) of the original dataset articles from the publications *Information Systems Research* (10), *Management Information Systems Quarterly* (19), and the *Journal of the American Medical Association* (238) were retained for the two analysis phases of the study.

Phase I was the quantitative analysis, and included the centering resonance analysis (CRA) technique for identifying themes. The CRA method was used to focus on the manifest content. The results of Phase I yielded ten themes relevant to medical informatics. Of the ten themes, six were found to aggregate around the major themes of *Healthcare Operations and Standards* and *Knowledge Transfer/Communication*. In the major theme, *Healthcare Operations*

and Standards, the sub-themes of *Operations*, *Project Management*, and *Security/Privacy* were prevalent in the article texts. Sub-themes of the major theme, *Knowledge Transfer/Communication*, included *Inter-organization communication*, *Intra-organization communication*, and *Patient-Provider communication*. The remaining four themes found in the final theme solution included *Analytics*, *Healthcare Research*, *Perceptions and Managing Expectations of Information Technology*, and *Software as a Service*.

In Phase II, the analysis was qualitative and focused on the manual coding approach used for further theme identification and categorization based on the latent content of the texts. One goal was to explore differences between medical informatics and non-medical informatics mainstream article texts. An additional goal was to explore differences in article texts over the span of the study, 2002 - 2011. The dominant theme found in the aggregated body of texts (all years and all publications included) was *Information Architecture*, identified in 755 article texts of a total of 2,188 texts (34.5%). The least frequently identified theme was *Compatibility*, having been identified in only 41 of the 2,188 texts (1.9%). The most and least frequently identified themes in the mainstream medical informatics literature were *Information Architecture* (frequency = 648 (37.5%)) and *Compatibility* (frequency = 38; 2.2%), respectively. The non-mainstream medical informatics literature yielded the same variable with the lowest frequency, *Compatibility* (frequency = 3; 0.7%), but a different theme with the highest frequency, *Direct Patient Care* (frequency = 138; 30.0%). The difference between *Information Architecture* and *Direct Patient Care* in the medical informatics publications was 2.0% and 6.7% in the non-medical informatics publications, suggesting only a small difference between the two groupings. The analysis of the texts grouped by year indicated that *Direct Patient Care* and *Information Architecture* were identified most frequently, with *Direct Patient Care* identified most frequently

in years 2004, 2008, 2009, and 2011. *Information Architecture* was the most frequently identified theme in years 2002, 2003, 2005, 2006, 2007, and 2010. For each year, *Compatibility* was identified least frequently. The articles in the dataset included field studies, literature reviews, surveys, and case studies related to using information technology to assist in research, operations, healthcare operations, and social conditions. The next chapter discusses the implications of these findings within the context of the guiding theory, the research questions and sub-questions, and the objectives of this study.

Chapter 5. Discussion and Conclusions

The goal of this study was to identify emerging themes from the medical informatics literature among publications published from 2002 - 2011. Centering resonance analysis, exploratory factor analysis, and manual coding were the methods employed to collect and analyze the article texts and describe the derived themes. The purpose of this chapter is to present the findings in the context of the medical informatics discipline and the current healthcare environment. The chapter begins with a summary of the research study followed by the interpretations of the findings. Following the summary, the chapter continues with limitations of the study and suggestions for future research. The chapter concludes with general comments about the medical informatics discipline.

Summary

The overall purpose of this study was to discover what themes have emerged in the medical informatics discipline since the inception of the field in the 1960s, what challenges exist for the discipline, and what future directions does the literature suggest for the field. Heeding the call of previous researchers, the study looked at the medical informatics discussions from perspectives of articles published in both medical informatics focused journals and those published about medical informatics topics in the related fields of medicine and management information systems. Other subordinate focus areas for the study were to identify challenges and suggest future directions for the medical informatics discipline.

To meet the goals of the study, 2,188 usable article texts were collected from articles published between 2002 - 2011 from seven journals. The mainstream medical informatics publications used were the *Journal of the American Medical Informatics Association*, and the *International Journal of Medical Informatics*. The publications used that are outside the prevailing medical informatics literature were from the neighboring fields of medicine and management information systems: the *Journal of the American Medical Association*, the *New England Journal of Medicine*, *Management Information Systems Quarterly*, *Information Systems Research*, and *Communications of the Association for Computing Machinery*.

Due to the diversity of the article sources, three methods of collecting the article texts were used. All article texts for the decade were collected from each of the two medical informatics-specific journals. The *Journal of the American Medical Association* has a specific collection of medical informatics articles and from that collection, all the articles for the specified time frame were pulled. To collect article texts from the remaining publications, advanced queries in Thompson Reuter's Institute for Scientific Information Web of Science were performed to identify articles that met the inclusion criteria for the study.

A mixed methods approach was used to analyze the data. For the quantitative portion, centering resonance analysis and exploratory factor analysis were used to discover ten major themes emerging from the literature. Of the ten, three of the themes, *operations*, *project management*, and *information assurance*, exhibited an intertwining relationship under the major area of *healthcare operations and standards*. In similar manner, three themes coalesced strongly to define the theme of *knowledge transfer/communications*. The three themes in this grouping are *knowledge transfer/communications: extending beyond the organization*, *knowledge transfer/communications: internal to the organization*, and *knowledge transfer/communications:*

information assurance. The remaining four independent themes to emerge were *analytics*, *aspects of healthcare research*, *perceptions and managing expectations of information technology*, and *software as a service*.

In the qualitative phase of the study, the researcher manually coded the article texts as they applied to the perceived characteristics of diffusion of innovations theory and the thematic categories developed from the review of the literature. The initial categories of the characteristics of innovation were *relative advantage*, *compatibility*, *complexity*, *trialability*, and *observability*. The thematic categories were *direct patient care*, *business analytics*, and *Information Architecture*. The categories *relative advantage*, *compatibility*, *direct patient care*, and *Information Architecture* met the minimums established for inter-coder reliability. Of the characteristics of innovations categories, aspects of *relative advantage* were identified in 10.7% of the articles and traits of compatibility were identified in 1.9% of the article texts. Characteristics of the themes *direct patient care* and *information architecture* were identified in 34.1% and 34.5% of the article texts, respectively.

While empirical medical informatics studies are abundant, the field will benefit from more devotion to theoretical grounding in studies. Theories provide a common perspective of why phenomena occur and from which others can conceptualize phenomena. Likewise, theory provides metaphorical pictures and can incite greater insight for readers than that from solely an explanation of what occurred. Examples in existing literature wherein researchers could apply strong theory were provided. Diffusion of Innovations is a strong model in which there are many facets that can aid understanding of why certain phenomena we are seeing occur in the medical informatics literature.

The future of medical informatics appears to be in the direction of patient centered healthcare. It seems clear that in the future, information technologies will have a greater impact on the ways in which healthcare providers provide care to their patients and the way in which patients are involved in their own healthcare. As more and more facilities turn toward patient centered electronic records and patient controlled healthcare records, the way providers tend to their patients will ultimately change. Indications appear that there will be better patient care as a result (Tan & Global, 2009).

Interpretation of the Findings

With Moore's law and other factors that affect the rate of technology growth, it is difficult to determine what the future holds for medical informatics. However, the themes of medical informatics emerge from the literature in six general areas: *Analytics, Healthcare Operations and Standards, Aspects of Healthcare Research, Knowledge Transfer/Communication, Perceptions and Managing Expectations of Information Technology,* and *Software as a Service*. Of these, all but one, *Healthcare Research*, align with the medical informatics sub-domains, *Direct Patient Care, Business Analytics,* and *Information Architecture*, modifications of categories suggested by Haux (2010). Although *Healthcare Research* did not align with Haux' sub-domains as adapted to this study, its emergence in the centering resonance analysis does illustrate the importance of the theme and it will be discussed further. The themes, *Healthcare Operations and Standards* and *Knowledge Transfer/Communication* align well with the *Direct Patient Care* sub-domain in that they address information systems use that contributes to healthcare organizational leaders' ability to make good decisions and aids in the providing of good medicine and good health for the individual. Although during manual coding the *Business Analytics* theme failed to meet the minimum reliability requirements, the centering resonance

analysis in the quantitative phase produced the *Analytics* theme, which has the similar characteristics of focusing on the analysis and management of medical health knowledge ideas, insights, and experiences. The remaining themes, *Perceptions and Managing Expectations of Information Technology*, and *Software as a Service* coincide with the application sub-domain, *Information Architecture*.

Healthcare Operations and Standards: Operations

Healthcare Operations and Standards can be further sub-divided to delineate the areas of *Operations*, *Project Management*, and *Information Assurance*. The *Operations* sub-category, for this study, focuses on the literature that discusses the use of information systems in the functions of day-to-day operations at the micro-level of patient health care as opposed to the organizational focus. Aronsky and colleagues provide us with an example of a study involving *Healthcare Operations and Standards: Operations*, in their work evaluating the use of a computerized emergency department census board, a central location for patient and operational information, as a replacement to a non-digital dry erase board (Aronsky, Jones, Lanaghan, & Slovis, 2008).

Healthcare Operations and Standards: Project Management

Considering that merely 15 years ago, it was rare to find more than one or two computers in an office—let alone on every desk—and data centers were a fraction of the size they are today, it is not unexpected to see the necessity of professional project management in developing and implementing information systems projects in the healthcare environment. The sub-theme *Project Management* that has emerged in the *Healthcare Operations and Standards* area is an indicator of this greater need. While implementing tools and services in any aspect of the healthcare process can be daunting, the rate of technology growth, concerns about information

security, patient safety, patient privacy, and healthcare provider/patient relations are a few factors that exacerbate the problem for information systems professionals. Proper project management techniques can assist in reducing the problems associated with health information systems implementation adoption (Ludwick & Doucette, 2009).

Healthcare Operations and Standards: Information Assurance

The factors that increase challenges for implementation—patient privacy, patient safety, and information security—also are factors in the area of *Information Assurance*, a subset of the *Healthcare Operations and Standards* theme. For the purpose of this study, *Information Assurance* is defined by discussions related to the risks associated with transmitting, storing, and processing of sensitive and non-sensitive healthcare data and information, and the management of said risks. Over the duration of a typical healthcare process, some hundreds of people, both direct patient care providers and non-care providers, may have access to a patient's personal digital medical information (Cannoy & Salam, 2010), thus, creating a terrible risk of breach of patient privacy. The Health Information Insurance and Portability Act (HIPAA) established some provisions and requirements for addressing information assurance risks (Mercuri, 2004) and Georgetown University Medical Center provides another example of the *Information Assurance* theme emerging in their use of a self-directed risk assessment method to comply with the information security provisions of HIPAA (Collmann, Alaoui, Nguyen, & Lindisch, 2005). An additional concern for information assurance experts comes from the pervasive communications tools available to both patients and providers. How do the information assurance specialists ensure patient data and privacy security when patients and providers establish communications through the myriad of means available in today's environment of cell

phones, instant messaging, e-mail, etc? To be certain, this challenging area will remain in the forefront of medical informatics literature for the duration.

Knowledge Transfer/Communication: Extending Beyond the Organization

As with the general *Healthcare Operations and Standards* theme, the *Knowledge Transfer/Communication* theme further subdivides. The inherent idea in the sub-theme *Extending Beyond the Organization*, is that of passing information between people, groups, or elements within a healthcare organization to those outside the organization. The information transfer is usually, but not necessarily, two-way and it may be synchronous or asynchronous (Wilson, 2003). The emergence of this theme is attributed, in part, to the spread of newer and cheaper communications technologies such as cellular/smart telephones, social media, and health/healthcare information exchanges (Shachak & Jadad, 2010). Indeed, the effect of social media alone is so great the venerable American Medical Association has issued a policy statement on healthcare provider professionalism when using social media (Chretien, Azar, & Kind, 2011). Although smartphones—cellular phones that have additional communications applications built in—and instant messaging are tools that provide greater access for patients to healthcare administrators and providers than has existed in the past, they also provide more complex challenges in managing the information risks (Bønes, Hasvold, Henriksen, & Strandenæs, 2007; Nguyen, Fuhrer, & Pasquier-Rocha, 2009). Despite the risks, the communications technologies are great tools for including the patient more in his/her own care and we can expect to see more discussion of them in the medical informatics literature. The capability to transfer patient care data electronically among various healthcare organizations is known as *health information exchange* (HIE); with greater governmental and commercial interest in sharing data, HIE is a sought after technology in the healthcare community and there

are ample indications that HIE will continue to fuel the research in this area (Korst, Aydin, Signer, and Fink, 2011; Patel, Abramson, Edwards, Malhotra, and Kaushal, 2011; Sengupta, Calman, and Hripcsak, 2008; Wright, et al., 2010).

Knowledge Transfer/Communication: Internal to the Organization

The second subdivision in the major theme *Knowledge Transfer/Communication* lies in the conveyance of information among people, groups, and other elements within an organization. While there is some overlap with *Knowledge Transfer/Communication: Extending Beyond the Organization*—information that has traversed the boundaries of an organization often has made the rounds within the organization—*Knowledge Transfer/Communication: Internal to the Organization* is a theme evident in the literature. This *Internal to the Organization* theme can, like the previous theme, include instant messaging, social media, and smartphones. Nevertheless, it can also include messaging designed into the electronic medical records and other healthcare systems and other similar technologies. The effect that communications technologies has on continuity of patient care is the subject of some concern, in that while one might expect the improved technology to improve continuity of care, that is not necessarily the case (Horwitz & Detsky, 2011). Research about medical knowledge centers, repositories of medical knowledge available to assist with providers' educational and research needs, is another area that resides in the *Internal to the Organization* theme (Haux, Ammenwerth, Herzog, & Knaup, 2002). With continued rapid advances in communication technology, we can expect medical informatics scholars to carry on research in intra-organizational knowledge transfer and communication.

Knowledge Transfer/Communication: Patient-Provider

While *Patient-Provider* communication as a sub-theme of *Knowledge Transfer/Communication* could nestle within the *Knowledge Transfer/Communication: Extending Beyond the Organization* theme, there was adequate literature to support the separation. As mentioned in the *Information Assurance* discussion above, communications tools abound. With this abundance, research opportunities are plentiful for the medical informatics scholar. Many research questions arise such as: What are the professionally acceptable means of communication between patient-provider? How does communication method affect the patient-provider relationship? What medical knowledge can/should the provider share with the patient and how? Obviously, some topics have already been addressed—hence, the emergence of the *Patient-Provider* theme in the first place—such as whether a secure internet-based electronic messaging system is effective for augmenting patient care in general practices (Bergmo, Kummervold, Gammon, & Dahl, 2005), or investigations of patient controlled health records, which afford patients direct access—usually via the internet—to their health data (Bourgeois, Taylor, Emans, Nigrin, & Mandl, 2008). The use of video conferencing between patients in their homes and healthcare providers has been documented and investigated for effectiveness in patient care—providing another example of the extent of the *Knowledge Transfer/Communication: Patient-Provider* theme (Bakken, et al., 2006).

Perceptions and Managing Expectations of Information Technology

Addressing user *Perceptions and Managing Expectations of Information Technology* is not a new theme in medical informatics or management information systems. However, as technology continues to change and improve at a rapid pace, managing user perceptions and

expectations becomes increasingly complex. Healthcare workers expect more from the technology they use and expect systems implementations to be performed without affecting other work. Other aspects play into the increased challenges for information managers in managing expectations of information technology. For example, the aforementioned internet-based personally patient controlled healthcare records and patient-provider messaging add a new facet for medical informaticians to factor: the expectations of the patients. In the past, the focus was on those within the healthcare organization—providers and staff. With more complex systems and a greater level of technology integration and communication, information systems experts have to adapt to the myriad customer base. No longer is it acceptable to approach all customers with the same information technology product (Bakker & Hammond, 2003). Although this theme is established, more research is required in determining how information systems managers balance the perceived needs of healthcare employees *and* patients themselves.

Analytics

The *Analytics* theme has at its foundation, a focus on the analysis and management of medical health knowledge ideas, insights, and experiences. Aspects of the theme include healthcare applications of data mining, managing clinical information, and business intelligence approaches. With the advent of increasingly cheap data storage and the push toward electronic health data, the amount of electronic medical data available is greater than our ability to use it to maximum effectiveness in improving clinical care and operations (Ferranti, Langman, Tanaka, McCall, & Ahmad, 2010). Continued research reflection and growth is necessary in the *Analytic* area to overcome the overwhelming volume of data available and put it to good use. The medical informatics community has been researching methods of information retrieval for several years and we can expect that research to grow in the future (Baud & Ruch, 2002). One

such research study gave perspective on the Mayo Clinic's "Enterprise Data Trust," a collection of data to support business intelligence (Chute, Beck, Fisk, & Mohr, 2010). Another study provided a view of a hospital's transition to a new health information system with a focus on, among other things, data mining and reporting (Haug, Rocha, & Evans, 2003). The *Analytic* theme includes web-based and open source tools, as well (Liu, Marengo, & Miller, 2006). As long as the effect of Moore's law continues to push down the prices for data storage and governmental and organizational entities continue the drive toward electronic health data, it is likely the *Analytic* theme will continue its prevalence in the literature.

Software as a Service

The migration of software applications from residing on the desktop computer toward residing on an in-house server or an internet-based server is an integral aspect of the *Software as a Service* theme. The computing evolution cycle is reverting back toward its initial stages, when mainframes housed the applications and users used "dumb" terminals, terminals with little to no computing power, to access the mainframe. Now, the term *thin client* has replaced dumb terminal and, with the improvement of data transfer speeds, it is becoming more efficient to manage software update and security requirements hosted on an organization's server that provides virtual software desktops to the thin clients. An extension of the server owned and managed by an organization is found in current approaches toward internet-based hosting—the organization leases server computing resources from another organization using the internet for access to the resources. While centralized health data and applications are not new in health organizations, there has been a recent emergence in internet-based health-related software such as Microsoft's *HealthVault* and Google's soon to be discontinued (i.e., January 1, 2012) *Google Health* (Bergmann, Bott, Pretschner, & Haux, 2007; Haas, Wohlgemuth, Echizen, Sonehara, &

Müller; Mandl & Kohane, 2008; Simborg, 2009). While Microsoft's and Google's offerings are primarily patient-focused, Practice Fusion, a free web-based electronic health record focuses on providing software as a service to healthcare providers (accessed August 18, 2011 <http://www.practicefusion.com/>). The growing emphasis of government and commercial entities toward electronic health data and applications, combined with improvements in computing and communications technology suggest continued relevance of *Software as a Service*.

Aspects of Healthcare Research

Introspective *Healthcare Research* in the field of medical informatics will continue as long as research is performed by humans and is therefore, subject to error. Medical informatics scholars will perform analyses such as those represented by the reviews in Chapter Two and the more specific look at the Telehealth medical informatics literature in JAMIA (W.R. Hersh, Patterson, & Kraemer, 2002). Meta-analytic and content analytic literature reviews will continue to be performed with scholars assessing the quality of research being performed and the results therein. The government influence on healthcare, through regulations such as HIPAA and laws such as the recently enacted American Recovery and Reinvestment Act of 2009 (ARRA), with its specific focus on health information technology, provides ample stimulation for *Healthcare Research* to continue to flourish in the medical informatics literature (Blumenthal, 2009). Indeed, the \$17 billion from the government to incentivize healthcare providers and organizations to adopt and use electronic health records will most certainly trickle down to the research community—if not through direct academic grants, assuredly in the form of consultant fees to academics with side businesses (Blumenthal, 2009). We can expect consortia of researchers to continue to provide *Healthcare Research* contributions such as those provided by members of the State Networks of Colorado Ambulatory Practices and Partners (SNOCAP), in

their investigation of missing clinical information (Smith, et al., 2005). Scholars have a thirst for knowledge; that thirst will continue to be slaked through continual investigation of newer statistical methods such as hierarchical linear modeling or structural equation modeling and the application of these methods toward medical informatics hypotheses and problems (Gagnon, et al., 2003; Ko & Dennis, 2011). Overall, *Healthcare Research* as a medical informatics theme will continue to remain strong for years to come.

Diffusion of Innovations Perceived Characteristics of Innovations

As discussed previously, there is an identified need in the medical informatics literature for researchers to strengthen research rigor and value through increased application of theory (Brennan, 2008). Through theory, scholars can provide a common perspective from which others can conceptualize a phenomenon and answer the question, “Why should colleagues give credence to this particular representation of the phenomena” (Whetten, 1989, p. 491). Theory provides metaphorical pictures and can incite greater insight for readers than that from only an explanation of what occurred. *Diffusion of Innovations* is one theoretical model that has been widely applied across research disciplines and time (Rogers, 2003), and it proves useful as a representative theoretical model for medical informatics. Diffusion of Innovations is a large model, offering many facets that can assist understanding of why certain phenomena we are seeing in the medical informatics literature occurs. For this analysis, the five perceived characteristics of innovations—*relative advantage*, *compatibility*, *complexity*, *trialability*, and *observability*—were drawn from the diffusion model as concepts to identify in the medical informatics literature. During the manual coding phase, the study identified the perceived characteristics that met reliability minimums—*relative advantage* and *compatibility*—in a combined 12 percent (276) of the articles.

Although it was not the primary goal of this analysis, identifying perceived characteristics of innovation from the diffusion of innovation model proved useful and enlightening. The benefit of this aspect of this dissertation is in complementing Brennan's argument for greater theoretical development in medical informatics literature (2008). It is arguably a very subjective analysis to review the research of others *post hoc* and suggest that the authors could have used a particular theoretical method in their research. However, it is the study's intent to proffer the idea and support for continued theoretical development and to argue in support of the *Diffusion of Innovations* as one such theoretical model. There is value gained in identifying areas in which specific theoretical concepts can be applied to previous research. For instance, Goth (2010), provides an example in which aspects of both compatibility and relative advantage could be appropriately used, in this case, in the discussion of strengthening biologists' educational background through the addition of computer science coursework. For example, Goth identifies and expresses the opinions of several academics, both from biology and computer science; the academics' quotes illustrate their stances based on their experiences and needs of their opinions of their representative fields, prime material for discourse of the *compatibility* of injecting computer science into biology education. Likewise, the degree to which adding more computer science study to biologist training was perceived to be better than the *status quo* was exemplified in the academics' quotes Goth related.

In truth, diffusion theory is a very generalizable theory; it is understandable to accept that authors of papers discussing a new innovation, whether a physical system or a new idea, could reasonably apply aspects of diffusion theory. As one aspect, extrapolating the perceived characteristics of innovations could guide an investigation. Establishing how the different characteristics are related could further strengthen the study. Obviously, *Diffusion of*

Innovations is not the only theoretical model that could effectively be used in the medical informatics research. There are several established models that would provide a good basis for theory development in the field of medical informatics—the *Information Systems Success Model*, *General Systems Theory*, and *Expectation Disconfirmation Theory*—to name a few. The continuing importance the medical informatics community should garner is to give attention to strong theory development. While empirical medical informatics studies abound, the field will benefit from more devotion to theoretical grounding in studies.

Limitations

While this study is thorough, it is not without room for improvement. For example, the automated method used to identify themes and topics arising from a large collection of articles published about the theme is acceptable; however, it does not delineate specific calls by editors for a special edition of the journal focused on a specific topic. The last edition of JAMIA for the year 2002—volume 6 (Supplement 1)—is a good illustration of that phenomenon. Lenert, Burstin, Connell, Gosbee, and Phillips (2002), referring to Kohn, Corrigan, and Donaldson's *To Err is Human: Building a Safer Health System* (2000), introduce the supplemental issue for the discussion of patient safety. Do the 31 articles in the supplement affect the results of the various analyses used? The answer is an obvious “yes”. However, that leads to the next question, though. Does it affect the outcome of the study? Assuming the editors of JAMIA have some insight to the medical informatics field, an argument can be made that the supplemental does affect the outcome of the study, but not in a negative way; the call for and subsequent publication of a supplemental issue with a singular focus is another indicator of the importance of said focus in the field. To look at it from another perspective, the editors did not choose to publish a supplement about calibration failures of pulse oximeters—devices that measure the oxygen

saturation of a patient's blood—because it is unlikely that topic holds great interest for medical informaticists.

Another limitation of this inquiry lies in the sampling. While texts from each sampled discipline (i.e., medical informatics, medicine, and management information systems) were included, the three fields were unequally represented in the dataset. There were substantially fewer articles available in the non-medical informatics mainstream publications (i.e., CACM, ISR, JAMA, MISQ, and NEJM) than from the medical informatics mainstream publications (i.e., IJMI, JAMIA), and therefore, more article texts from the medical informatics publications were collected. Further, there were far fewer management information systems published articles than those found in the medical discipline. This may be an indicator of a greater interest about medical informatics in the field of medicine than that of management information systems, it may be a factor of the editor interest and direction for each of the non-medical informatics publications, or a combination of both. Regardless of the underlying causes, there is an imbalance in the dataset. Interestingly, this limitation suggests another.

In a similar way, the scope of the disciplines represented in the study was limited. As in all studies, there has to be a defining line drawn to determine the extent of the data collected. The fields chosen to be included were medical informatics, medicine, and management information systems. Because the study focuses on medical informatics, the inclusion of texts from medical informatics is apparent. Based on calls for extending medical informatics research beyond the boundaries of medical informatics, the closely related fields of management information systems and medicine were included. However, the two fields are not the only disciplines that are closely associated with medical informatics. This study could have benefited from the inclusion of publications in other closely related fields such as computer science and

nursing. Likewise, inclusion of disciplines that are less closely related such as electrical engineering and medical law may have shed additional light on the field of medical informatics.

Another concern for the study is with the data sampling. Editorials, abstracts, and introductory texts of articles were used to produce a collection of 2,686 pages of text for analysis. While a seemingly large volume, collecting only abstracts and abbreviated texts of articles without abstracts does provide limited data from which themes can be drawn. Full texts would substantially increase the volume of data to analyze, perhaps ten to twenty-fold, and therefore, should provide richer results in the analysis. While the greater volume of text may require a small increase in computing resources and time, the effect on manual coding would be detrimental. Coding fatigue was a concern in the manual coding phase using only the abstracts and abbreviated texts during this study; including full texts would certainly increase coder fatigue, likely to a level that would substantially affect the results of the study. In other words, there is an inverse relationship between the data analysis methods and the volume of text to analyze—the greater the volume, the better the results of the automated analysis will likely be, but the poorer the results of the manual coding will likely be. There is no heuristic that states at what point the perfect balance of tradeoffs occurs. There are variables that can affect each method. To factor for these variables, steps were employed in the manual coding procedures to minimize coding fatigue and to reiterate the operationalized definitions. For the automated analysis, procedures were repeated and data entry methods were varied to ensure consistency of results.

Recommendations for Future Research

As mentioned, expanding this study to include fields outside of mainstream medical informatics will yield interesting insights and has been a call relatively unheeded in prior research (Andrews, 2003; Haux, 2010; Morris & McCain, 1998). Although the related fields of

medicine and management information systems were included, future research should include other closely related disciplines *and* disciplines further removed from medical informatics. Nursing and medical law are fields closely related to medical informatics and provide good examples for fields to include. Both computer science and electrical engineering have focuses that are generally closer to the specific characteristics of technology than to the focus of medical informatics. A future study including an investigation into the literature of the computer science and electrical engineering fields could provide an understanding of newer software and technologies, those that are on the “leading edge.” A view from a strong technical perspective—a perspective down at the bits and bytes, the transistors and resistors—could enlighten the medical informatics field with a richer understanding of capabilities and potential capabilities of information systems. Perhaps, this enlightenment could provide medical informatics scholars with material for expanding medical informatics research.

In this mixed methods analysis, two methods were balanced to ascertain the themes emerging from the medical informatics discipline. The first method was the semi-automated, centering resonance analysis. The second was the manual coding. Each provides strengths and weaknesses. Together, they provide strong results. One of the weaknesses that emerged was the relatively small sample of article texts. While 2,686 pages of text may appear large for manual coding—it is—it is small for automated coding. To continue and extend this study, full texts of the articles should be analyzed using centering resonance analysis. It can be reasonably assumed that the volume of texts would increase by at least ten-fold. The benefits of the automated approach using full texts are several. By eliminating the manual coding element, we eliminate coding fatigue and reliability concerns. The full text analysis will provide a more in-depth view of each article.

Another future research direction is found in the corollary to the previous suggestion. A manual coding analysis could be performed using full texts of articles similar to those collected for the current study, but with a limited stratified randomized sample of articles from each publication. Limiting the number of articles included would minimize coding fatigue without severely affecting the outcome of the investigation. Performing manual coding would contribute the insights and experiences of the coders in a way unavailable through the automated coding of centering resonance analysis. Each of these studies would provide confirmation or contradiction to the current study. Either result advances the field.

The results of this analysis are based in the literature and suggest a direction for future research in the realm of practice. An interesting perspective on the themes would be found in an action-based or grounded theory approach. Direct observation of aspects of medical informatics in practice is essential in expanding the understanding of the discipline and in maximizing the benefits that medical informatics practitioners offer. Using the themes found in this study as guiding principles, subsequent analysis in a healthcare facility would enrich the discipline.

Toward the end of this research, the U.S. Congress passed the Budget Control Act of 2011 (S.365) and President Obama signed it into law on August 2, 2011 (Public Law Number 112-25). The Act provides a multi-part process to reduce the federal deficit and directs the establishment of a special bipartisan *Joint Select Committee on Deficit Reduction*. Further, the Act establishes strict spending limits and debt reduction measures, to include automatic across-the-board federal spending reductions by up to \$1.2 trillion if the Joint Committee fails to agree on and accomplish savings of \$1.2 trillion. Although Medicaid and a few other large budget categories (e.g., Social Security, military pay, and veterans' benefits) are exempted from the across-the-board cuts, Medicaid is not. Also, because the aforementioned large budget

categories are exempted, the remaining areas of the budget will be hit with a proportionally larger across-the-board cut. While the ramifications of the law for healthcare in general, and medical informatics in specific, are unclear, it is clear that the continued emphasis on improved fiduciary responsibility will have a dramatic affect on healthcare in the United States. The Law provides ample research opportunity for studies related to the financial benefit and cost-cutting measures associated with health information systems. The magnitude of this legislation is enormous – the call for future research in this area will likely be drowned out by the calls for analysis from the leaders in government, public, and private entities of healthcare.

Concluding Remarks

While it may be cliché in stating that the future is bright for the field of medical informatics, it is not an understatement. With continuing legislation emphasizing digital health records, dramatic and rapid improvements in technology, and the ever-pressing need to reduce healthcare costs, the demand for medical informatics is great. The discipline provides a synthesis of information systems, healthcare, operations, and research—and it does so in a manner unlike any other field. Although medical informatics is relatively young, the field has established deep roots and a strong foundation. We can expect to see persistent growth and maturity in the field as scholars, practitioners, and researchers continue to provide value to the healthcare of the ever-increasing population.

References

- American Recovery and Reinvestment Act/Health Information Technology for Economic and Clinical Health Act, House of Congress, 111th Sess.
- Anderson, J., Gremy, F., & Pages, J. (Eds.). (1974). *Education in informatics of health personnel*. Amsterdam: North-Holland Publ. Co.
- Andrews, J. E. (2003). An author co-citation analysis of medical informatics. *Journal of the Medical Library Association, 91*(1), 47.
- Aronsky, D., Jones, I., Lanaghan, K., & Slovis, C. (2008). Supporting patient care in the emergency department with a computerized whiteboard system. *Journal of the American Medical Informatics Association, 15*(2), 184-194.
- Bakken, S., Grullon-Figueroa, L., Izquierdo, R., Lee, N. J., Morin, P., Palmas, W., et al. (2006). Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. *Journal of the American Medical Informatics Association, 13*(6), 660.
- Bakker, A. R., & Hammond, W. (2003). Report of conference track 1: basic bottlenecks. *International Journal of Medical Informatics, 69*(2), 295-296.
- Baud, R., & Ruch, P. (2002). The future of natural language processing for biomedical applications. *International Journal of Medical Informatics, 67*(1), 1-5.

- Bergmann, J., Bott, O. J., Pretschner, D. P., & Haux, R. (2007). An e-consent-based shared EHR system architecture for integrated healthcare networks. *International Journal of Medical Informatics*, 76(2-3), 130-136.
- Bergmo, T. S., Kummervold, P. E., Gammon, D., & Dahl, L. B. (2005). Electronic patient-provider communication: will it offset office visits and telephone consultations in primary care? *International Journal of Medical Informatics*, 74(9), 705-710.
- Blumenthal, D. (2009). Stimulating the adoption of health information technology. *New England journal of medicine*, 360(15), 1477-1479.
- Bønes, E., Hasvold, P., Henriksen, E., & Strandenæs, T. (2007). Risk analysis of information security in a mobile instant messaging and presence system for healthcare. *International Journal of Medical Informatics*, 76(9), 677-687.
- Bourgeois, F. C., Taylor, P. L., Emans, S. J., Nigrin, D. J., & Mandl, K. D. (2008). Whose personal control? Creating private, personally controlled health records for pediatric and adolescent patients. *Journal of the American Medical Informatics Association*, 15(6), 737-743.
- Brannen, J. (2005). Mixing methods: The entry of qualitative and quantitative approaches into the research process. *International Journal of Social Research Methodology*, 8(3), 173-184.
- Brennan, P. (2008). Standing in the Shadows of Theory. *Journal of the American Medical Informatics Association*, 15(2), 263.

- Brown, L. (1981). *Innovation diffusion: A new perspective*: Routledge.
- Cajori, F. (1991). *A history of mathematics* (Fifth ed.): Chelsea Pub Co.
- Cannoy, S. D., & Salam, A. (2010). A framework for health care information assurance policy and compliance. *Communications of the ACM*, 53(3), 126-131.
- Chretien, K. C., Azar, J., & Kind, T. (2011). Physicians on Twitter. *JAMA: the journal of the American Medical Association*, 305(6), 566.
- Chute, C. G., Beck, S. A., Fisk, T. B., & Mohr, D. N. (2010). The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *Journal of the American Medical Informatics Association*, 17(2), 131.
- Cimino, J. (1998). The concepts of language and the language of concepts. *Methods of information in medicine*, 37, 311-311.
- Collen, M. (1986). Origins of medical informatics. *Western Journal of Medicine*, 145(6), 778.
- Collmann, J., Alaoui, A., Nguyen, D., & Lindisch, D. (2005). Safe teleradiology: Information assurance as project planning methodology. *Journal of the American Medical Informatics Association*, 12(1), 84.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach*: Sage Publications, Inc.
- Corman, S. R., & Dooley, K. (2006). *Crawdad Text Analysis System (Version 1.2)*. Chandler, Arizona: Crawdad Technologies, LLC.

- Corman, S. R., Kuhn, T., McPhee, R. D., & Dooley, K. J. (2002). Studying Complex Discursive Systems. Centering Resonance Analysis of Communication. *Human Communication Research*, 28(2), 157-206.
- Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*: Sage Publications, Inc.
- Creswell, J., & Clark, V. (2007). *Designing and conducting mixed methods research*: Sage Publications, Inc.
- Cruz, P., Neto, L., Muñoz-Gallego, P., & Laukkanen, T. (2010). Mobile banking rollout in emerging markets: evidence from Brazil. *Marketing*, 28(5), 342-371.
- Damman, O. (2010). An International Comparison of Web-based Reporting About Health Care Quality: Content Analysis. *J Med Internet Res*, 12(2), e8.
- Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- DeLone, W., & McLean, E. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3(1), 60-95.
- DeShazo, J. P., LaVallie, D. L., & Wolf, F. M. (2009). Publication trends in the medical informatics literature: 20 years of " Medical Informatics" in MeSH. *BMC medical informatics and decision making*, 9(1), 7.
- Dunn, R. (2007). *Haimann's healthcare management*: Health Administration Press.

- Eggers, S., Huang, Z., Chen, H., Yan, L., Larson, C., Rashid, A., et al. (2005). Mapping medical informatics research. In H. Chen, S. Fuller, C. Friedman & W. Hersh (Eds.), *Medical Informatics: Knowledge management and data mining in biomedicine* (pp. 35-62). New York, NY: Springer Science + Business Media, Inc.
- Ferranti, J. M., Langman, M. K., Tanaka, D., McCall, J., & Ahmad, A. (2010). Bridging the gap: leveraging business intelligence tools in support of patient safety and financial effectiveness. *Journal of the American Medical Informatics Association*, 17(2), 136.
- Gagnon, M. P., Godin, G., Gagné, C., Fortin, J. P., Lamothe, L., Reinharz, D., et al. (2003). An adaptation of the theory of interpersonal behaviour to the study of telemedicine adoption by physicians. *International Journal of Medical Informatics*, 71(2-3), 103-115.
- Ginn, R. V. N. (1997). *The History of the US Army Medical Service Corps*: Office of the Surgeon General and Center of Military History, United States Army.
- Goth, G. (2010). CS and biology's growing pains. *Communications of the ACM*, 53(3), 13-15.
- Greenes, R. A., & Siegel, E. R. (1987). *Characterization of an emerging field: approaches to defining the literature and disciplinary boundaries of medical informatics*.
- Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 30(3), 611.
- Haas, S., Wohlgemuth, S., Echizen, I., Sonehara, N., & Müller, G. Aspects of privacy for electronic health records. *International Journal of Medical Informatics*, 80(2), e26-e31.
- Halamka, J., Mandl, K., & Tang, P. (2008). Early Experiences with Personal Health Records. *Journal of the American Medical Informatics Association*, 15(1), 1-7.

Hasman, A., & Haux, R. (2006). Modeling in Biomedical Informatics—An Exploratory Analysis (Part 1). *Methods Inf Med*, 45, 638-642.

Hasman, A., & Haux, R. (2007). Modeling in biomedical informatics—An exploratory analysis: Part 2. *International Journal of Medical Informatics*, 76(2-3), 96-102.

Haug, P. J., Rocha, B. H. S. C., & Evans, R. S. (2003). Decision support in medicine: lessons from the system. *International Journal of Medical Informatics*, 69(2-3), 273-284.

Haux, R. (2010). Medical informatics: past, present, future. *International Journal of Medical Informatics*, 79(9), 599-610.

Haux, R., Ammenwerth, E., Herzog, W., & Knaup, P. (2002). Health care in the information society. A prognosis for the year 2013. *International Journal of Medical Informatics*, 66(1-3), 3-21.

Hayes, A., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89.

Health Care and Education Reconciliation Act 111th Congress(2010).

Hersh, W. R. (2002). Medical Informatics: Improving Health Care Through Information. *JAMA: Journal of the American Medical Association*, 288(16), 1955-1955.

Hersh, W. R., Patterson, P. K., & Kraemer, D. F. (2002). Telehealth: the need for evaluation redux. *Journal of the American Medical Informatics Association: JAMIA*, 9(1), 89.

Horwitz, L. I., & Detsky, A. S. (2011). Physician Communication in the 21st Century. *JAMA: the journal of the American Medical Association*, 305(11), 1128.

Howe, K. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard.

Educational researcher, 17(8), 10.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*: Sage Pubns.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.

Keen, P. G. W. (1980). *MIS Research: Reference Disciplines and a Cumulative Tradition*. Paper presented at the International Conference on Information Systems, Philadelphia, Pennsylvania.

Ko, D. G., & Dennis, A. R. (2011). Profiting from Knowledge Management: The Impact of Time and Experience. *Information Systems Research*, 22(1), 134-152.

Kohn, L. T., Corrigan, J., & Donaldson, M. S. (2000). *To err is human: building a safer health system* (Vol. 6): Natl Academy Pr.

Krippendorff, K. (2004). *Content Analysis An Introduction to Its Methodology* (2nd ed.). London: Sage Publications.

Krippendorff, K. (2007). Computing Krippendorff's alpha reliability. *Departmental Papers (ASC)*, 43.

Leech, N., Barrett, K., & Morgan, G. (2008). *SPSS for intermediate statistics: Use and interpretation* (Third ed.). New York: Lawrence Erlbaum Assoc Inc.

- Lenert, L., Burstin, H., Connell, L., Gosbee, J., & Phillips, G. (2002). Federal patient safety initiatives panel summary. *Journal of the American Medical Informatics Association*, 9(Suppl 6), S8.
- Liu, N., Marenco, L., & Miller, P. L. (2006). ResourceLog: an embeddable tool for dynamically monitoring the usage of web-based bioscience resources. *Journal of the American Medical Informatics Association*, 13(4), 432.
- Ludwick, D., & Doucette, J. (2009). Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. *International Journal of Medical Informatics*, 78(1), 22-31.
- Mandl, K. D., & Kohane, I. S. (2008). Tectonic shifts in the health information economy. *New England journal of medicine*, 358(16), 1732-1737.
- Manns, M. (2002). An investigation into factors affecting the adoption and diffusion of software patterns in industry. *De Montfort University, United Kingdom, Leicester*.
- Marias, J. (1967). *History of philosophy*: Dover Pubns.
- Mcphee, R. D., Corman, S. R., & Dooley, K. (2002). Organizational Knowledge Expression and Management: Centering Resonance Analysis of Organizational Discourse. *Management Communication Quarterly*, 16(2), 274-281.
- Mercuri, R. T. (2004). The HIPAA-potamus in Health Care Data Security. *Communications of the ACM*, 47(7), 25-28.

- Morris, T. A., & McCain, K. W. (1998). The structure of medical informatics journal literature. *Journal of the American Medical Informatics Association*, 5(5), 448.
- Neuendorf, K. A. (2002). *The content analysis guidebook*: Sage Publications, Inc.
- Nguyen, M. T., Fuhrer, P., & Pasquier-Rocha, J. (2009). Enhancing E-health information systems with agent technology. *International Journal of Telemedicine and Applications*, 2009, 1-13.
- Onwuegbuzie, A., & Leech, N. (2005). Taking the “Q” out of research: Teaching research methodology courses without the divide between quantitative and qualitative paradigms. *Quality and Quantity*, 39(3), 267-295.
- Patient Protection and Affordable Care Act, 111th Congress, 2nd Sess.(2010).
- Pawar, B. (2009). *Theory Building for Hypothesis Specification in Organizational Studies*. Thousand Oaks, CA: Sage Publications.
- Rainer, R., & Miller, M. (2005). Examining differences across journal rankings. *Communications of the ACM*, 48(2), 91-94.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free press.
- Schuemie, M., Talmon, J., Moorman, P., & Kors, J. (2009). Mapping the domain of medical informatics. *Methods Inf Med*, 48(1), 76-83.
- Shachak, A., & Jadad, A. R. (2010). Electronic health records in the age of social networks and global telecommunications. *JAMA: the journal of the American Medical Association*, 303(5), 452.

- Shortliffe, E., Perreault, L., Wiederhold, G., & Fagan, L. (Eds.). (2001). *Medical Informatics-Computer Applications in Health Care and Biomedicine* (2nd ed.). New York: Springer.
- Shortliffe, E. H., & Cimino, J. J. (2006). *Biomedical informatics: computer applications in health care and biomedicine*: Springer Verlag.
- Simborg, D. W. (2009). The limits of free speech: the PHR problem. *Journal of the American Medical Informatics Association*, 16(3), 282.
- Sittig, D., & Kaalaas-Sittig, J. (1995). A quantitative ranking of the biomedical informatics serials. *Methods of information in medicine*, 34, 397-397.
- Smith, P. C., Araya-Guerra, R., Bublitz, C., Parnes, B., Dickinson, L. M., Van Vorst, R., et al. (2005). Missing clinical information during primary care visits. *JAMA: the journal of the American Medical Association*, 293(5), 565.
- Sterling, T. D., Pollack, S. V., & Center, C. U. M. C. (1964). *Medcomp: Handbook of computer applications in biology and medicine*: Medical Computing Center, College of Medicine, Univ. of Cincinnati.
- Sutton, R., & Staw, B. (1995). What theory is not. *Administrative Science Quarterly*, 40(3).
- Tan, J. K. H., & Global, I. (2009). *Medical informatics: concepts, methodologies, tools, and applications*: Medical Information Science Reference.
- Tate, W. L., Ellram, L. M., & Kirchoff, J. O. N. F. (2010). Corporate social responsibility reports: A thematic analysis related to supply chain management. *Journal of Supply Chain Management*, 46(1), 19-44.

- Weber, R. P. (1990). *Basic content analysis (Quantitative Applications in the Social Sciences No. 49)*. Newbury Park, CA: Sage.
- Weigel, F. K., Landrum, W. H., & Hall, D. J. (2009). *Human-Technology Adaptation Fit Theory For Healthcare*. Paper presented at the Twelfth Annual Conference of the Southern Association for Information Systems (SAIS), Charleston, SC.
- Whetten, D. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, 490-495.
- Wilson, E. V. (2003). Asynchronous health care communication. *Communications of the ACM*, 46(6), 79-84.

Appendix A: Sample SQLite Query Code

Below is the code for extracting the JAMIA publication information. The code for the other publications was identical except for a change in the “**JAMIA**” listed in the fourth to the last line of the code and displayed in bold text:

```
SELECT
items.itemID,
titleValue.value AS title,
creatorData0.LastName AS lastname1,
creatorData0.FirstName AS firstname1,
creatorData1.LastName AS lastname2,
creatorData1.FirstName AS firstname2,
creatorData2.LastName AS lastname3,
creatorData2.FirstName AS firstname3,
creatorData3.LastName AS lastname4,
creatorData3.FirstName AS firstname4,
creatorData4.LastName AS lastname5,
creatorData4.FirstName AS firstname5,
creatorData5.LastName AS lastname6,
creatorData5.FirstName AS firstname6,
creatorData6.LastName AS lastname7,
```

creatorData6.FirstName AS firstname7,
creatorData7.LastName AS lastname8,
creatorData7.FirstName AS firstname8,
creatorData8.LastName AS lastname9,
creatorData8.FirstName AS firstname9,
creatorData9.LastName AS lastname10,
creatorData9.FirstName AS firstname10,
abstractNoteValue.value AS abstract,
pubValue.value AS pubname,
dateValue.value AS date
FROM items

LEFT JOIN itemData AS abstractNoteData ON items.itemID = abstractNoteData.itemID
AND abstractNoteData.fieldID = 90

LEFT JOIN itemDataValues AS abstractNoteValue on abstractNoteValue.valueID =
abstractNoteData.valueID

LEFT JOIN itemData AS titleData ON items.itemID = titleData.itemID AND
titleData.fieldID = 110

LEFT JOIN itemDataValues AS titleValue on titleValue.valueID = titleData.valueID

LEFT JOIN itemData AS pubData ON items.itemID = pubData.itemID AND
pubData.fieldID = 12

LEFT JOIN itemDataValues AS pubValue on pubValue.valueID = pubData.valueID

LEFT JOIN itemData AS dateData ON items.itemID = dateData.itemID AND
dateData.fieldID = 14
LEFT JOIN itemDataValues AS dateValue on dateValue.valueID = dateData.valueID
LEFT JOIN itemTypes ON items.itemTypeID = itemTypes.itemTypeID
LEFT JOIN itemCreators AS itemCreators0 ON items.itemID = itemCreators0.itemID
AND itemCreators0.orderIndex = 0
LEFT JOIN creators AS creators0 on creators0.creatorID = itemCreators0.creatorID
LEFT JOIN creatorData AS creatorData0 on creators0.creatorDataID =
creatorData0.creatorDataID
LEFT JOIN itemCreators AS itemCreators1 ON items.itemID = itemCreators1.itemID
AND itemCreators1.orderIndex = 1
LEFT JOIN creators AS creators1 on creators1.creatorID = itemCreators1.creatorID
LEFT JOIN creatorData AS creatorData1 on creators1.creatorDataID =
creatorData1.creatorDataID
LEFT JOIN itemCreators AS itemCreators2 ON items.itemID = itemCreators2.itemID
AND itemCreators2.orderIndex = 2
LEFT JOIN creators AS creators2 on creators2.creatorID = itemCreators2.creatorID
LEFT JOIN creatorData AS creatorData2 on creators2.creatorDataID =
creatorData2.creatorDataID
LEFT JOIN itemCreators AS itemCreators3 ON items.itemID = itemCreators3.itemID
AND itemCreators3.orderIndex = 3
LEFT JOIN creators AS creators3 on creators3.creatorID = itemCreators3.creatorID

LEFT JOIN creatorData AS creatorData3 on creators3.creatorDataID =
creatorData3.creatorDataID

LEFT JOIN itemCreators AS itemCreators4 ON items.itemID = itemCreators4.itemID
AND itemCreators4.orderIndex = 4

LEFT JOIN creators AS creators4 on creators4.creatorID = itemCreators4.creatorID

LEFT JOIN creatorData AS creatorData4 on creators4.creatorDataID =
creatorData4.creatorDataID

LEFT JOIN itemCreators AS itemCreators5 ON items.itemID = itemCreators5.itemID
AND itemCreators5.orderIndex = 5

LEFT JOIN creators AS creators5 on creators5.creatorID = itemCreators5.creatorID

LEFT JOIN creatorData AS creatorData5 on creators5.creatorDataID =
creatorData5.creatorDataID

LEFT JOIN itemCreators AS itemCreators6 ON items.itemID = itemCreators6.itemID
AND itemCreators6.orderIndex = 6

LEFT JOIN creators AS creators6 on creators6.creatorID = itemCreators6.creatorID

LEFT JOIN creatorData AS creatorData6 on creators6.creatorDataID =
creatorData6.creatorDataID

LEFT JOIN itemCreators AS itemCreators7 ON items.itemID = itemCreators7.itemID
AND itemCreators7.orderIndex = 7

LEFT JOIN creators AS creators7 on creators7.creatorID = itemCreators7.creatorID

LEFT JOIN creatorData AS creatorData7 on creators7.creatorDataID =
creatorData7.creatorDataID

```

LEFT JOIN itemCreators AS itemCreators8 ON items.itemID = itemCreators8.itemID
AND itemCreators8.orderIndex = 8

LEFT JOIN creators AS creators8 on creators8.creatorID = itemCreators8.creatorID
LEFT JOIN creatorData AS creatorData8 on creators8.creatorDataID =
creatorData8.creatorDataID

LEFT JOIN itemCreators AS itemCreators9 ON items.itemID = itemCreators9.itemID
AND itemCreators9.orderIndex = 9

LEFT JOIN creators AS creators9 on creators9.creatorID = itemCreators9.creatorID
LEFT JOIN creatorData AS creatorData9 on creators9.creatorDataID =
creatorData9.creatorDataID

LEFT JOIN collectionItems ON items.itemID = collectionItems.itemID
INNER JOIN collections ON collectionItems.collectionID = collections.collectionID
AND collections.collectionName = "JAMIA"

LEFT JOIN deletedItems
ON items.itemID = deletedItems.itemID
WHERE deletedItems.itemID IS NULL

```

Appendix B: Coding Procedures Guide

General Instructions:

1. You are coding article abstracts, letters to the editor, or brief introductory sections (hereafter, “texts”), of journal articles searching for *Medical Informatics* themes. For the purpose of this study, Medical Informatics is defined as *the discipline dedicated to the systematic processing, analysis, and dissemination of health-related data through the application of digital information systems (computers) to various aspects of healthcare, research, and medicine.*

2. Following are the research questions and sub-questions for the study. It is important that you keep the research questions and sub-questions in the forefront of your mind while coding the texts:

Research Question 1: What themes have emerged in the medical informatics discipline since its inception in the 1960s?”

Sub-question 1: What themes emerge from the medical informatics literature?

Sub-question 2: What medical informatics themes emerge from the related fields of medicine and management information systems?

Research Question 2: What challenges exist for the medical informatics discipline?

Research Question 3: What future directions does the literature suggest for the field?

3. Familiarize yourself with the coding definitions below each time you begin a coding session and keep a copy of the definitions available during coding. Table A.1 and A.2 includes three general themes and most article texts should relate to at least one of them. Table Y includes

perceived characteristics of innovations as conceived in Rogers' *Diffusion of Innovations* theory. The themes and characteristics listed in Tables X and Y are preloaded in DiscoverText. The "DiscoverText keystroke" column lists the keys that can be used to quickly code a text with the respective theme or characteristic without having to use the mouse. If you code a text as "REMOVE," you should not code it with any other themes or characteristics. If you type in your code choices, separate multiple codes with a vertical line "|" (usually the SHIFT, BACKSLASH key).

4. Take at least a 10 minute break after coding 25 texts (i.e., abstracts or introductory sections) or after 1 hour, whichever occurs first. Even if you don't feel tired, take at least a 10 minute break before continuing to code and re-read the coding variable definitions below before starting.

Table B.1.
Themes

Theme	Description	DiscoverText keystroke
REMOVE	The article text is unrelated to medical informatics and should be removed from the dataset	R
Direct Patient Care	Information system/technology that contributes to good medicine and good health for the individual and aids organizational leaders in making decisions (e.g., electronic medical records)	D
Business Analytics	Exchange, creation, distribution, analysis, or adoption of medical and health knowledge ideas, insights, and experiences within and across organizations (e.g., data mining for health reporting)	B
Information Architecture	Contributes to well-organized, patient-centered health care and appropriate information management (e.g., intra-organizational health information systems architecture; health information data exchange standardization)	I


Table B.2
Perceived Characteristics of Innovations (Rogers, 2003, pg 170)


Characteristic	Description	DiscoverText keystroke
Relative Advantage	The degree to which an innovation is perceived as better than the idea it supersedes	A
Compatibility	The degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters	C
Complexity	The degree to which an innovation is perceived as difficult to understand and use	X
Trialability	The degree to which an innovation may be experimented with on a limited basis	T
Observability	The degree to which the results of an innovation are visible to others	O

5. In the coding section of DiscoverText, you will see the following information near the top of the screen:

(A) Relative Advantage (C) Compatibility (I) Information Architecture (O) Observability (T) Trialability (X) Complexity
 (B) Business Analytics (D) Direct Patient Care (N) Not Listed (R) REMOVE

Using the applicable DiscoverText keystroke from Table A.1/A.2, above, select the appropriate code(s) for each article text. You may use a mouse click on the appropriate entry box instead of the DiscoverText keystroke if you prefer.

6. Once you have selected all the appropriate codes, press the ENTER key on your keyboard or click the onscreen “Code” button  to move on to the next article text.

7. Once you are done with a coding session, click the red stop sign  to save your work and leave the coding session. You can do this when you take a break or when you have completed your coding assignment.

End of code book.