2015

# The Value of Diagnostic Software and Doctors' Decision Making

Babatunde Ayodele Alaofin
*Walden University*

# Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

Babatunde Alaofin

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

Review Committee
Dr. Robert Kilmer, Committee Chairperson,
Applied Management and Decision Sciences Faculty

Dr. Godwin Igein, Committee Member,
Applied Management and Decision Sciences Faculty

Dr. Raghu Korrapati, University Reviewer
Applied Management and Decision Sciences Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2015

Abstract

The Value of Diagnostic Software and Doctors' Decision Making

by

Babatunde A. Alaofin

MS, Georgia Southwestern State University, 1994

BA, Allen University, 1991

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Sciences

February 2015

Abstract

The prevalence of medical misdiagnosis has remained high despite the adoption of diagnostic software. This ongoing controversy about the role of technology in mitigating the problem of misdiagnosis centers on the question of whether diagnostic software does reduce the incidence of misdiagnosis if properly relied upon by physicians. The purpose of this quantitative, cross-sectional study based on planned behavior theory was to measure doctors' opinions of diagnostic technology's medical utility. Recruitment e-mails were sent to 3,100 AMA-accredited physicians through their database that yielded a sample of 99 physicians for the study. One-sample *t* tests and, where appropriate because of non-normal data, one-sample Wilcoxon signed-rank tests were conducted on the data to address the following key research questions on whether diagnostic software decreases misdiagnosis in healthcare versus unassisted human diagnostic method, if physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare, and if liability concerns prevent physicians from using diagnostic software. It was found that in the opinion of those surveyed (a) diagnostic software was likely to result in fewer misdiagnoses in healthcare than unassisted human diagnostic methods, (b) when speaking for themselves, physicians thought they used diagnostic software frequently enough to decrease misdiagnoses, and (c) physicians agreed they were not prevented from using diagnostic software because of liability concerns. The study's social significance is the affirmation of diagnostic software's usefulness: Policy and technology stakeholders can use this finding to speed the adoption of diagnostic software, leading to a reduction in the socially costly problem of misdiagnosis.

# The Value of Diagnostic Software and Doctors' Decision Making

by

Babatunde A. Alaofin

MS, Georgia Southwestern State University, 1994

BA, Allen University, 1991

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Sciences

Walden University

February 2015

Dedication

To my children, Tobi, Tola and Tawalade, for understanding the importance of education early in life. To my lovely wife, Tope, for her unconditional love and support throughout this journey. To my friends, Seye, Seyi, Jide, Bamidele and Pastor Bello–-yes, the journey is over.

Acknowledgments

I would like to give the utmost thanks and gratitude to God through Jesus Christ, my personal savior. With Him, all things are possible. To Dr. Robert Kilmer, I would like to express my sincere gratitude and appreciation for his advice, guidance, and support. Without his support and dedication to helping complete this dissertation after a tumultuous start, I am certain it would have been a much more difficult process. To Dr. Godwin Igein and Dr. Rauhu Korrapati, members of my dissertation committee, thank you very much for your assistance and valuable inputs. Finally, to my lovely wife, Tope. Thank you honey for a job well done in supporting me and raising our children during this odyssey of pursuing this doctorate degree. I certainly would not have gotten through this without her support.

Table of Contents

List of Tables

List of Figures

# Chapter 1: Introduction to the Study

## Background

Medical misdiagnosis is an immensely costly problem. Globally, misdiagnosis is responsible for millions of patient deaths every year; in the United States, about 100,000 people die every year because of misdiagnosis (Leavitt & Leavitt, 2011). Misdiagnosis leads to economic costs as well, by raising the already high cost of healthcare delivery. The costs of an inaccurate or slow diagnosis are high and include costs of delayed treatment, litigation, malpractice insurance payouts, and the lost economic productivity of the patient (Schweitzer, 2007). The dynamics of medical decision-making are changing in response to increased pressures on the global healthcare system. In developed countries, the amount of money spent on healthcare is typically the largest single component of gross domestic product (GDP; Krugman & Wells, 2009). Given the human and economic problems created by misdiagnosis, there is added pressure to bring new efficiencies to the delivery of healthcare (Cleverly, Cleverly, & Song, 2010). These pressures affect the practice of diagnosis, specifically in creating an imperative for diagnoses to be made more quickly and accurately (Goldsmith, 2011).

The combination of these pressures and the availability of increasingly sophisticated medical technology have resulted in the popularity of diagnostic medical software in most developed countries (Scott & Rundall, 2007). Diagnostic software has been widely available since the 1990s, but advancements in the underlying artificial intelligence (AI) of such software and its integration with other aspects of healthcare information technology have resulted in an increase in the use of diagnostic software in the United Kingdom (Graham, 2010; Hawe, 2010) and the United States in particular (Cleverly et al., 2010). This development has been praised as well as critiqued (Spekowius & Wendler, 2006). Supporters of diagnostic software emphasize its accuracy and speed; detractors of diagnostic software have suggested that the use of such software predisposes physicians to be lazy (Arora, 2010), and that diagnostic software can make inaccurate recommendations when dealing with complex or nuanced medical problems (Bligh, 2009). Thus, the immediate background for this study was the ongoing controversy (Ahlers, Jaeger, & Jakstat, 2010) about the role of technology in mitigating the problem of misdiagnosis. The first part of the controversy centered on the question of whether diagnostic software can indeed reduce the incidence of misdiagnosis but is not properly relied upon by physicians for this end. The other part questioned whether diagnostic software could not reduce the incidence of misdiagnosis because misdiagnosis emerges from factors that are beyond the ability of diagnostic software to address. What was already known is that, despite an adoption rate that has been estimated between 55 and 70% (Chernick, 2011; Felder & Mayrhofer, 2011), the implementation rate of diagnostic software has not coincided with a reduction in the prevalence of misdiagnosis; what was not known is why the use of diagnostic software has not concomitantly reduced diagnostic error rates.

In this chapter, I describe the problem statement, identify a meaningful gap in the current research literature, and present evidence that the problem is relevant and demanded attention. I present theoretical framework that is associated with the foundation of the study and frame the research questions, hypotheses, research design, and methodology in a manner that built upon the existing research. I discuss the purpose, nature, and the significance of the study. I describe the potential contributions of the study to the advancement of the discipline as well as the social change implication in accordance to the scope of the study.

## Problem Statement

The problem is the prevalence of high levels of misdiagnosis (Leavitt & Leavitt, 2011) despite the widespread adoption of diagnostic software and the improvements in such software over time. Throughout the literature, authors have suggested several possibilities for why this problem might exist. There have been suggestions that the inherent possibility of misdiagnosis is high because of the interaction of several complex factors that cannot be addressed by software (Sokolowski & Banks, 2011). There were also suggestions that diagnostic software is capable of reducing misdiagnosis but that physicians insufficiently or infrequently rely upon it. The academic dimension of this problem was the absence of more definitive knowledge why misdiagnosis has persisted well into the age of diagnostic software (Sokolowski & Banks, 2011).

## Purpose of the Study

The purpose of this quantitative study was to draw upon physician-provided data to determine why, at least in physicians' opinions, the prevalence of misdiagnosis has remained high despite the widespread adoption of diagnostic software. For the first two research questions of the study, the independent variable was whether diagnostic software was used, and the dependent variable was reduction of misdiagnosis. For the third research question of the study, the independent variable was knowledge of diagnostic software, and the dependent variable was reduction of misdiagnosis. For the fourth research question of the study, the independent variable was liability concern, and the dependent variable was use of diagnostic software.

## Research Questions and Hypotheses

The research questions that guided the study were as follows:

Research Question 1: Does use of diagnostic software decrease misdiagnosis in healthcare versus unassisted human diagnostic methods?

$H_0$1: Diagnostic software use has more misdiagnoses in healthcare than unassisted human diagnostic methods.

$H_A$1: Diagnostic software use has less misdiagnoses in healthcare than unassisted human diagnostic methods.

Research Question 2: Do physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare?

$H_0$2: Physicians do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare.

$H_A$2: Physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare.

Research Question 3: Is physicians' knowledge of diagnostic software extensive enough to decrease misdiagnosis in healthcare?

$H_0$3: Physicians' knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare.

$H_A3$: Physicians' knowledge of diagnostic software is extensive enough to decrease misdiagnosis in healthcare.

Research Question 4: Do liability concerns prevent physicians from using diagnostic software?

$H_04$: Liability concerns do not prevent physicians from using diagnostic software.

$H_A4$: Liability concerns prevent physicians from using diagnostic software.

## Nature of the Study

Addressing the question of why the widespread adoption of diagnostic software has not coincided with a decrease in the prevalence of misdiagnosis could be achieved with both a quantitative and a qualitative approach. A quantitative approach could determine which of the possible answers to this question—diagnostic software insufficiency, insufficient/improper use by physicians, liability, or other reasons—is more popular with physicians, and to determine whether answers to this question vary significantly depending on the demographic and professional characteristics of physicians. A qualitative approach could provide a narrative explanation of results; for example, if the quantitative analysis reveals that physicians think that diagnostic software is diagnostically useful but under used, then qualitative analysis could be an appropriate means of determining why physicians do not use diagnostic software more frequently, despite its utility. A quantitative approach was used in this study in order to obtain necessary empirical insight into physicians' attitudes towards diagnostic software, insight that can be used to inform future quantitative as well as qualitative research. I used a quantitative, cross-sectional approach in this study to examine whether physicians think there is a connection between diagnostic software used and misdiagnosis in a manner that addressed some of the gaps in the literature noted in Chapter 2. More detail about the method is provided in Chapter 3.

## Theoretical Base

In order to explore the question of diagnostic software's utility as a means of reducing misdiagnosis, some theory capable of explaining the interface between physicians and software is required. Accordingly, the first theoretical base for the study centered on the theory of distributed cognition (Ajzen, 2005), which suggested that the combination of humans and technology results in higher quality and more quickly rendered decisions, as long as humans use technology with sufficient frequency and skill. The theory of distributed cognition predicts that human decision-makers will employ software or other technology to assist them when the benefits of doing so (in terms of the quantity and quality of decisions) outweigh the costs (such as the emotional burdens of delegating some aspects of decision-making to machines or feeling a loss of control or expertise). These aspects of the theory of distributed cognition underlie the approach to answering the research questions of this study. The second theoretical foundation for the study is the theory of planned behavior, which was a specific model for explaining the human component of a human-software system of distributed cognition (Ajzen, 2005; Herath, 2010). The theory of planned behavior suggests that attitudinal perceptions of usefulness or other kinds of benefit drive behavioral decisions such as software adoption; the theory thus provides support for including diagnostic software adoption rates and attitudes in the same model.

## Definition of Terms

*Adoption of diagnostic software:* As one of the independent variables of the study, adoption of diagnostic software is a dichotomous variable with two levels: adoption and non adoption.

*Cognitive bias*: A cognitive bias "is a generic defect in human reasoning based on flawed methods of collecting, processing, or analyzing information" (Schwab, 2008, p. 23).

*Computing*: According to Woods and Woods (2000), "computing involves using numbers to count, solve problems, and gather information" (p. 7). Computing is a method of diagnosis or decision-making that is highly dependent on numerical analysis, and in which decisions are reached only if quantifiably justified.

*Diagnosis*. Diagnosis refers to "the process of evaluating a patient's medical condition with the aim of choosing an appropriate treatment" (McPhee, Papadakis, & Rabow, 2011. p. 56).

*Diagnostic outcomes*: As one of the dependent variables of the study, diagnostic outcomes was a categorical variable  measured on a Likert scale assessing the degree of doctors' agreement with the proposition that diagnostic software reduced misdiagnosis.

*Distributed cognition*: Distributed cognition is computation that is "part of the larger system of decision-making" (Hazlehurst, Gorman, & McMullen, 2008, p. 11). Thus, a doctor working to make a diagnosis with the help of his or her ratiocination, a medical manual, and a software interface would be part of a three-component system of distributed cognition.

*Intuition*, sometimes referred to as experiential intuition, was defined by Duggan (2005) in the following way:

> [Decision-makers] study a situation (Step A), and the problem and solution come to them at the same time (Step B). They think through the implications to arrive at a course of action (Step C), and then commit to it, or reject it if they think it will not work (Step D). In all four steps, they look for patterns of similarity and difference with other situations they have lived or learned about. (p. 9)

There are many variant definitions of intuition, but, in this study, Duggan's definition will be used.

*Planned behavior*: Planned behavior is a model of human action in which, according to Herath (2010), it is possible to "explain human actions by understanding the following inputs and the interactions between them: Individual beliefs; collective beliefs; beliefs about beliefs; and motivation" (p. 317). Planned behavior is thus a construct to explain human actions that consists of elements of rationality, social determinism, and classic behaviorism (Skinner, 1938). The planned behavior model (Herath, 2010) will be explained further in Chapter 2.

*Representativeness heuristic*: The representativeness heuristic (or problem-solving method) is the form of "cognitive bias that bases decisions based on available data rather than on all data, or at least a larger body of data" (Zilberberg, 2011, p. 69). Thus, for example, a doctor who has treated four patients in a row who have the same extremely rare disease might overestimate the actual prevalence of this disease among the general population.

*Satisficing*: Satisficing is a form of decision-making, sometimes employed by doctors that was originated by Simon (1947) and defined by Garnham and Oakhill (1994) as follows:

> A satisficer recognizes that making the best decision is a time-consuming process, and that the difference between a satisfactory decision and the best one will probably not justify the effort of computing utilities. In satisficing, a criterion is set for a satisfactory decision and the first alternative that meets that criterion is accepted…Satisficing is a simpler procedure than computing and comparing utilities, since the decision maker has simply to compare alternatives with the criterion, as they are encountered (p. 186).

*Software*: Software consists of "lines of instruction, written in a computer language, that direct a machine (or other software) to take a particular action" (Madhavji, 2006, p. 11).

**Assumptions**

One of the key assumptions of the study was that the results could be generalized, applied, and would demonstrate significant value. Another key assumption in this study was that doctors are able to introspect validly on the nature of their interaction with technology. My third assumption of this study was that the participants in the survey voluntarily provided honest responses to the best of their knowledge and understanding. My final assumption of this study was that doctors who reply to the survey are representative of non busy doctors rather than busy doctors.

## Limitations

The main limitation of this study was the small sample size. Physicians work under intense time constraints and are difficult to recruit (Creswell, 2009). It is likely, therefore, that the results are not highly generalizable to the entire population of American doctors. Another limitation of the study was that only computer software would be part of the study, as distinct from other forms of medical technology. This limitation means that some aspects of the doctor-technology system of distributed cognition were not examined.

## Scope and Delimitations

This quantitative study involved the use of a web-based survey to collect data on the relationship between the unchanged rates of misdiagnosis in the United States and the use or nonuse of diagnostic software. I delimited the study to a randomly selected sub population of 3,100 AMA-accredited, licensed, and practicing American doctors.

## Significance of the Study

Given the human costs of misdiagnosis, the rising economic costs of healthcare, and society's increasing impatience with substandard medical treatment, the use of diagnostic software represents a possible solution to both cost and efficiency problems that are widely noted in the literature (Capps, Dranove, & Lindrooth, 2010; Skinner, 2011; Yong, Saunders, & Olsen, 2010). The fact that the prevalence of misdiagnosis has remained high despite the widespread adoption of diagnostic software requires further analysis. This information can be useful to multiple stakeholders interested in improving hospitals' diagnostic performance, including healthcare policy-makers, hospital administrators, physicians, and software engineers. The social change implications of such improvement are significant, as it can result in fewer patient deaths from misdiagnosis and in this sense serve all patients reliant on accurate diagnosis for good medical outcomes.

## Summary and Transition

It is unlikely that the pressures on healthcare economics will ease in the near future, especially as much of the developed world enters a period in which the majority of its population will be aged and ill or under immediate threat of illness (Mankiw, 2011). It is all the more necessary to understand why diagnostic accuracy has not substantially improved, especially given that diagnoses are such an important predictor of the quality and speed of subsequent healthcare (Simel & Rennie, 2008). In this study, I examined whether physicians think there is a connection between diagnostic software use and misdiagnosis.

Because healthcare costs are spiraling out of control all over the world (Krugman & Wells, 2009; Mankiw 2011), any cost-efficient and feasible improvement in diagnostic efficiency would be a welcome development, as such improvements lower the overall cost of healthcare (Cleverly et al., 2010).

In Chapter 2 of this study, I present not only the review of the literature but also the theoretical foundation for the study and link the literature to the key variables and concepts. I build upon the foundation of theories and empirical studies that apply to diagnostics in medicine, with a special emphasis on scholarly work about doctors' and diagnostic software’s decision-making processes. I establish the need to research the relationships between the independent and dependent variables in Chapter 2.

Chapter 2: Literature Review

## Overview of the Literature Review

The purpose of this study was to draw upon physician-provided data to determine why, at least in physicians' opinions, the prevalence of misdiagnosis has remained high despite the widespread adoption of diagnostic software. The problem addressed in the study was the prevalence of high levels of misdiagnosis (Leavitt & Leavitt, 2011) despite an adoption rate for diagnostic software that is between 55 and 70% (Chernick, 2011; Felder & Mayrhofer, 2011). Throughout the literature, authors have suggested that diagnostic software does indeed have the ability to both assist physicians with their diagnoses and to provide sound diagnoses in its own right in a manner that will be examined later in the chapter. The questions that do not appear to be addressed in the literature are whether physicians are using diagnostic software frequently and expertly enough to avail themselves of its benefits.

## Literature Search Strategy

In order to investigate what previous literature has stated about this topic, searches for *medical misdiagnosis, diagnostic software, diagnostic technology, physician opinions* and *diagnostic software,* and *distributed cognition in healthcare* were conducted on the EBSCO Host, Science Direct, Google Scholar, and ProQuest databases. Older literature was included in the literature review because there are seminal theories of diagnosis and diagnostic technology. The extensive literature on the utility, or lack of utility, of diagnostic software exemplifies how the search strings and associated review of studies disclosed the existence of only a few studies (for example, Dreiseitl, 2005) that drew on a sample of physicians to answer the question of why misdiagnosis remains so prevalent despite the sophistication and widespread adoption of diagnostic technology.

## The Science of Diagnosis

The Oxford English Dictionary (2011) defined *diagnosis* as the "determination of the nature of a diseased condition; identification of a disease by careful investigation of its symptoms and history; also, the opinion (formally stated) resulting from such investigation" (para. 1). *Diagnosis* has Greek roots; in ancient Greek, the roots are "*δια-* through, thoroughly, asunder + *γιγνώσκειν* to learn to know, perceive" (Oxford English Dictionary). Thus, *diagnosis* suggests the acquisition of thorough knowledge.

In the Western medical tradition, the first great study of medicine was the Hippocratic Corpus (Kelly, 2010), the name given to the collected writings of Hippocrates (an ancient Greek physician, known as the father of medicine, who lived from 460 B.C.E. to 370 B.C.E.). The Hippocratic Corpus, over 70 medical treatises written by Hippocrates or his students in the 5$^{th}$ and 4$^{th}$ centuries B.C.E., is the earliest surviving and scientific discussion of diagnostic science (Renouard, 2010), and is, therefore, an appropriate starting point for any discussion of diagnostics. According to Hippocrates, diagnosis has the following components: (a) gathering of evidence, including physical evidence and verbal evidence (gathered from speaking to the patient) pertaining to a patient's symptoms; (b) fitting knowledge about the symptoms to a specific disease, whether known or postulated; (c) determining the most appropriate treatment for the disease; and (d) fine-tuning the treatment based on ongoing observations of the interaction between the patient and the proposed treatment. This process is discussed at length in “On Regimen in Acute Diseases,” when Hippocrates (Adams, 1849), writing in the 5th century BCE, used the word *diagnosis* for the first time in writing (p. 282). Adams, who provided a translation of the Hippocratic Corpus in 1849, argued that Hippocrates’ model of diagnosis remained highly influential: “Hippocrates and his followers…in a great measure anticipated all the results of modern diagnosis” (p. 307).

One of the revolutions in diagnostic science in the age of contemporary medicine was the study, analysis, and categorization of a vast number of diseases, which scientists were often able to understand on a molecular and genetic level (Bynum, 2008; Cunha, 2011). The vast accumulation of knowledge of disease meant that, over time, the diagnostic process became oriented to fitting observed symptoms to already-known diseases; after all, by the end of the end of the 20th century, the variations of disease were comprehensively understood, and the task of diagnosticians focused on fitting symptoms to disease (Hodler, Schulthess, & Zolikofer, 2011).

Hippocrates (as cited in Adams, 1849), for his part, placed equal emphasis on observed symptoms and grand etiological theories that were intended to explain the ultimate origins of disease. This emphasis steadily fell away by the Middle Ages. Hersen and Thomas (2006) described the key post-Hippocratic developments in diagnostic science as follows:

> Throughout the classical era, diagnoses were made based on presumed etiology, as when Hippocrates rooted the illnesses he diagnosed (mania, melancholia, and paranoia) in various imbalances of black bile, yellow bile, blood, and

> phlegm…Basing diagnostic assessments on such etiologic conceits changed only when the Swiss physician and natural philosopher Paracelsus (1490-1541) developed the concept of *syndromal diagnosis*. Paracelsus defined the syndrome as a group of signs and symptoms that co-occur in a common pattern and thereby, presumably, characterize a particular abnormality or disease state, but for which etiology is unknown, perhaps unknowable. Syndromal diagnosis is epitomized today in the *DSM*, which continues its focus on the signs and symptoms of diseases, rather than presumed etiologies, which are unnecessary for diagnostic purposes (p. 4).

As Hensen and Thomas (2006) wrote from the perspective of psychology and psychiatry, their reference to the DSM—the abbreviation for the Diagnostic and Statistical Manual of Mental Disorders—does not apply to all of medicine. However, the concept of syndromal diagnosis and the accompanying importance of evidence-based, empirical pattern recognition and fitting symptoms to diseases do indeed characterize the entire tradition of modern Western medicine, also known as biomedicine (Hughes, 2011). Of course, it is not necessary to agree, along with Hensen and Thomas, that Hippocrates's emphasis on etiology, or the study of the causes of diseases, lessened the importance of the Hippocratic practice of diagnosis and symptom fitting. Robson and Baek (2009) argued that Hippocrates's belief in no etiological concepts (such as the belief that the color and volume of bodily fluids determined aspects of personality) should not distract attention from the remarkably modern paradigm of Hippocratic diagnosis. Hess, MacIntyre, and Mishoe (2011) pointed out that Hippocrates's diagnosis of digital clubbing could stand alongside modern examples of sound diagnosis. Regardless of the role of etiology, the next steps in diagnostic science are straightforward. After fitting symptoms to a disease, the diagnostic authority either administers or recommends the administration of some treatment (Carpenito-Moyet, 2008).

There are many kinds of treatments, including pharmacological treatments, surgery, and other interventions (such as radiation; Foster, 2010). Whatever the precise composition of a treatment, the next and final stage in the diagnostic model is to monitor what happens to the patient so that a treatment can be modified if necessary or carried through to the termination of a patient's symptoms (Monahan, Neighbors, & Green, 2010).

For example, the post traumatic stress disorder (PTSD) diagnostic guidelines contain reference material on which a human doctor or diagnostic software could draw to make a diagnosis of PTSD. In this case, for a diagnosis of PTSD to be reached, the patient has to present with a specific set of symptoms spread across criteria A-F as shown in Table 1. However, even though PTSD is a highly studied disease with a known etiology and biological mechanisms (American Psychiatric Association, 2000), there is still room for ambiguity and discretion in making this diagnosis based on the diagnostic guidelines for PTSD (American Psychiatric Association, 2000, p. 256) in Table 1.

Table 1

*DSM Criteria for PTSD*

| Criterion | Description |
|---|---|
| A: Stressor | The person has been exposed to a traumatic event in which both of the following have been present: (1)The person has experienced, witnessed, or been confronted with an event or events that involve actual or threatened death or serious injury, or a threat to the physical integrity of oneself or others. (2) The person's response involved intense fear, helplessness, or horror. Note: in children, it may be expressed instead by disorganized or agitated behavior. |
| B:Intrusive Recollection | The traumatic event is persistently re-experienced in at least one of the following ways: (1) Recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions. Note: in young children, repetitive play may occur in which themes or aspects of the trauma are expressed. (2) Recurrent distressing dreams of the event. Note: in children, there may be frightening dreams without recognizable content. (3) Acting or feeling as if the traumatic event were recurring (includes a sense of reliving the experience, illusions, hallucinations, and dissociative flashback episodes, including those that occur upon awakening or when intoxicated). Note: in children, trauma-specific reenactment may occur. (4) Intense psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event. (5) Physiologic reactivity upon exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event |
| C:Avoidant / Numbing | Persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness (not present before the trauma), as indicated by at least three of the following: (1) Efforts to avoid thoughts, feelings, or conversations associated with the trauma. (2) Efforts to avoid activities, places, or people that arouse recollections of the trauma. (3) Inability to recall an important aspect of the trauma. (4) Markedly diminished interest or participation in significant activities. (5) Feeling of detachment or estrangement from others. (6) Restricted range of affect (e.g., unable to have loving feelings). (7) Sense of foreshortened future (e.g., does not expect to have a career, marriage, children, or a normal life span). |
| D:Hyper-Arousal | Persistent symptoms of increasing arousal (not present before the trauma), indicated by at least two of the following: (1) Difficulty falling or staying asleep. (2) Irritability or outbursts of anger. (3) Difficulty concentrating. (4) Hyper-vigilance. (5) Exaggerated startle response. |
| E: Duration | Duration of the disturbance (symptoms in B, C, and D) is more than one month. |
| F: Functional Significance | The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning. Acute: if duration of symptoms is less than three months; Chronic: if duration of symptoms is three months or more. |
| Specify if: | With or Without delay onset: Onset of symptoms at least six months after the stressor. |

*Note*. Adapted from American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fourth edition, text revision. Washington, D.C.: American Psychiatric Association, (2000, p. 256).

According to the American Psychiatric Association, at least two of the following five symptoms should be present in a PTSD-diagnosed patient: “(a) Difficulty falling or staying asleep, (b) Irritability or outbursts of anger, (c) Difficulty concentrating, (d) Hyper-vigilance, (e) Exaggerated startle response" (p. 256). However, there are cases of patients with PTSD who have exhibited only one of these symptoms (American Psychiatric Association, p. 256). This example shows the potential inexactitude of diagnostic science, regardless of the revolutionary advances that have occurred in medicine since the time of Hippocrates.

One of the problems of diagnosis is that patients can be radically different from each other, and therefore diseases can manifest themselves in different ways in different patients (Winkelman, 2008). However, diagnostic science is not designed to accommodate variation but rather to look for generalities and laws (Winkelman, 2008). Thus, as in the American Psychiatric Association's (2000) discussion of the symptoms of PTSD, statistical generalities have to be used in order to construct a portrait of the most common type of PTSD patient. In real life, PTSD patients might not conveniently display the symptoms that other patients have had, but they might suffer from PTSD all the same. Thus, in diagnostic science, there is a constant pull between the academic need to reach general conclusions about disease contrasted with the practical necessity of remembering that patients and their symptoms can be highly idiosyncratic (Winkelman, 2008).

The way of diagnostic science presented thus far in the literature review is what Lock and Nguyen (2011) called the consensus view of biomedicine. However, there are other views as well. Some scholars, for instance, have tried to argue that the diagnostic process is not merely a kind of science but also a cultural practice that is laden with inherent social values. Byrne (2012) offered a powerful recent articulation of this point of view: “Social construction recognizes that disease is not merely a biological fact but is an artifact of social interpretation. Diseases have meanings. Homosexuality used to be considered a disease; catching a cold and catching herpes are somehow different” (pp. 2-3). According to Byrne 2012; Freidson 1970; Schneider and Conrad 1981 who emphasize the social construction of disease, there is temptation to think of diagnostic science as somehow distinct from the society in which it takes place. However, there is a strong argument to be made that, because the concept of disease is itself fluid and culturally relative, so too diagnostic science should be fluid. Fadiman’s (1998) book-length account of the diagnosis of a child of Hmong ethnicity in California chronicles the value-laden nature of Western medicine, or biomedicine, when it comes into conflict with other traditions of belief and medicine. While issues of social construction will not be considered in this study, it is nonetheless important to be aware of the limits of diagnostic science.

The science of diagnosis as it is understood in contemporary times can, relying on the authorities whose work was discussed above, be summarized as follows. Diagnosis begins with some form of evidence collection, typically relying on a combination of physical evidence (such as a patient’s blood) and the patient’s own subjective and phenomenological accounts of illness, some of which can be overlaid with social and cultural values. This evidence is then synthesized into a set of symptoms, that is, observable problems and abnormalities. Next, the diagnostician moves toward fitting the symptoms to a known disease and consults some established authority—including personal experience of past patients’ symptoms, current analysis of an individual patient's symptoms, or a reference guide (such as the Diagnostic and Statistical Manual of Mental Disorders or diagnostic textbooks that address specific fields such as autoimmune disease, musculoskeletal disease, or other areas)—to decide upon a suitable form of treatment. Finally, the patient is monitored to determine whether the chosen treatment is proving to be effective, or whether further modifications to the treatment plan are needed. With this nutshell definition of the scientific process in diagnosis serving as a foundation, it is possible to try to theoretically model what is known of diagnostics.

## A Brief Overview of American Healthcare: Diagnostic Issues

Diagnosis is a process that is driven by the kinds of diseases with which a population is faced. In the United States, the most common terminal diseases are listed in Table 2 (Centers for Disease Control, 2012, p. 3); these diseases are more commonly diagnosed than, for example, the kinds of infectious diseases that are more prevalent in the Global South.

Table 2

*Main Leading Causes of Death in the U.S.*

| Causes of death | Annual victims |
|---|---|
| Heart disease | 616,067 |
| Cancer | 562,875 |
| Stroke or cerebrovascular disease | 135,952 |
| Chronic lower respiratory illness | 127,924 |
| Accidents or unintentional injuries | 123,706 |
| Alzheimer's disease | 74,632 |
| Diabetes | 71,382 |
| Influenza and pneumonia | 52,717 |
| Nephritis or nephrosis | 46,448 |
| Septicemia | 41,144 |

*Note*. Adapted from Centers for Disease Control and Prevention. Death and Mortality. NCHS Web site.http://www.cdc.gov/nchs/fastats/deaths.htm. Accessed May 20, 2012

Thus, as can be seen from Table 2, the most common diagnoses in the United States are for heart disease and cancer. Overall, medical scholars have argued that the United States is a classic example of a developed country in which there has been a shift in the burden of disease, away from infectious diseases to so-called lifestyle diseases (Caperchione, Kolt, & Mummery, 2009). Whereas 19th-century Americans were routinely killed by infections, Americans now tend to die because of health problems that emerge from a highly sedentary and inactive national lifestyle (Edlin & Golanty, 2009). Thus, the vast majority of cases that present to American doctors have to do with heart disease, cancer, and other diseases that have come to predominate in the developed world (Edlin & Golanty, 2009).

There is currently no national clearinghouse of data for American diagnostic statistics. However, different sources in the literature offer insight into the state of American diagnostics. The outlook is decidedly mixed in terms of the quality and timeliness of diagnostic decisions. The New York City Comptroller, Liu (2011) released a report indicating that many New York City hospitals had what the comptroller called “dangerously long waiting times” (para. 1) for diagnostic mammograms. In one New York City hospital, the average wait for a diagnostic mammogram in 2011 was 50 working days (Liu, 2011). Given that there are tens of thousands of discrete diagnostic procedures and many thousands of hospitals in the United States, it is not possible to offer an overview of the national healthcare system’s diagnostic efficiency; however, some general conclusions can still be reached. In his report on mammogram waiting times, Liu noted that the healthcare system of New York City was characterized by wide disparities in the speed of diagnosis; while one hospital in the Queens borough took 50 working days to order diagnostic mammograms, there were other hospitals and clinics in New York City where this procedure could be ordered and completed in a single day (Liu, 2011).

The conclusion is that the American healthcare system is highly variable with respect to diagnostic efficiency. One trend is that publicly-funded hospitals, especially hospitals in urban centers, are perpetually under budget pressure and have developed dysfunctional operational cultures, resulting in the long-delayed diagnoses noted by Liu (2011) and other observers (Trautman, 2011). Another trend is that wealthier Americans are able to opt out of bottom-tier care by hiring concierge doctors, purchasing better health insurance policies, and patronizing private clinics (Stillman, 2010). For this reason, there are wide disparities in the quality and timeliness of diagnostic procedures depending on the hospital, the precise diagnostic procedure, and the city in which services are provided. These disparities are far less pronounced in countries with robust public medicine programs; in the United Kingdom, for example, wait times for medical diagnostic procedures are essentially standardized so that anyone who attends a healthcare facility associated with the National Health Service (NHS) can expect to receive diagnoses in the same timeframe, and of the same quality, regardless of which healthcare facility is chosen (Dimakou, Parkin, Devlin, & Appleby, 2010).

The main conclusion to be drawn is that in the United States, diagnosis of disease is not only a scientific process but also a market phenomenon that is affected by the

American healthcare financing system and various market pressures. For example, recent years have seen the rise of the phenomenon of diagnostic outsourcing, in which physicians in another country—India is a popular destination, given the rigor of medical education and language commonalities—are paid to diagnose diseases by looking at x-rays and other forms of patient data (Schneirdjans, Schneirdjans, & Schneirdjans, 2007).

The purpose of this brief overview of some diagnostic trends and issues in American healthcare has been to offer a context for some of the implications that will be drawn, and recommendations that will be advanced in Chapter 5. Clearly, diagnostic processes are part of a larger market system, and recommendations about how diagnostic processes can be improved ought to be grounded in an acknowledgement of that reality.

## A Theoretical Model of Diagnosis

In order to think more deeply about what diagnosis is, how it functions, and how it might vary depending on whether the diagnostician is a doctor or software, it is necessary to suggest and defend a more general theory of diagnosis. First, diagnosis appears to be a high-level cognitive skill. Additionally, because diagnosis is the basis for some form of intervention in the physical world (such as the administration of a drug or the initiation of a surgery), it also has a component of action. Formally speaking, diagnosis is what Linehan (1993) called cognitive-verbal behavior, which "includes such activities as thinking, problem solving, perceiving, imaging, speaking, writing, and gestural communication" (p. 17), all of which are activities that have been shown to be inalienable parts of the diagnostic process (Foster, 2010; Hess, MacIntyre, & Mishoe, 2011; Hughes, 2011). It makes sense, therefore, that a theoretical model of cognitive-verbal behavior could also serve as a theoretical model of diagnosis itself. One promising model of behavior is the planned behavior model of Herath (2010). The planned behavior model as a five-stage model of what goes on between the formation of a diagnostic belief and the transformation of that belief into actual diagnostic behavior, such as the issuance of a prescription recommendation as shown in Table 3.

Table 3

*Five Levels of Herath's (2010) Planned Behavior Model*

| Level | Description | Relation to other levels |
|---|---|---|
| 1 | 1a Behavioral beliefs: Beliefs formed by the individual out of a personal process of rational, purposive thinking.<br><br>1b Normative beliefs: Beliefs considered normal; highly accepted beliefs within a culture or sub-culture.<br><br>1c Control beliefs: Beliefs about beliefs (for example, judgments about the soundness of a belief) | 1a leads to 2a;<br>1b leads to 2b;<br>1c leads to 2c |
| 2 | 2a Attitudes: Attitudes are more concrete beliefs; they imply some intention to act in a certain way.<br><br>2b Subjective norms: Subjective norms are the collective knowledge of authority (for example, a medical manual).<br><br>2c Perceived behavioral control: An individual's belief about his or her ability to control and direct her beliefs. | 2a, 2b, and 2c all lead to 3a. |
| 3 | 3a Behavioral intention: A disposition to act in a particular way. | 3a interacts with 4a and 4b and leads to 5 |
| 4 | 4a Intrinsic motivation: A desire to act in a particular way based on personal reasons.<br><br>4b Extrinsic motivation: A desire to act in a particular way based on external, non-personal reasons. | 4a and 4b are variables that mediate or moderate between 3a and 5a, and also between 2a, 2b, 2c and 3a |
| 5 | 5a Actual behavior | Outcome of previous levels |

*Note*. Synthesized from Eliciting Salient Beliefs are Critical to Predict Behavioural Change in Theory of Planned Behavior Herath (2010).

Table 3 offers more detail on what these five levels are and how they interact; it should be noted that the descriptions of each level are broad paraphrases of Herath (2010) that have been modified to account specifically for diagnosis (for example, medical manuals were given as an example of subjective norms, an example that does not appear in Herath's own discussion of the model. The Planned Behavior Model has the advantages of strength, flexibility, and alignment with the existing literature on diagnosis. It also addresses a question that the model of distributed cognition does not, which is exactly how the human components of decision-making work within a broader welter of influences from society, the individual mind, and the environment (which includes diagnostic software). The only potentially complex aspect of the model is the role of motivation. Herath argued that, in previous versions of the Planned Behavior Model (such as the seminal version of the model created by Ajzen, 2005), researchers had failed to take adequate account of the role of motivation. In order words, according to Herath, researchers assumed that behavioral intentions—formed by the inputs in levels 1 and 2 of Table3—led straightforwardly to actual behavior (as was the case in Ajzen's original model). However, Herath argued that motivation was an important intermediate variable. This point can be illustrated by means of an example. Even if a doctor were to arrive at a behavioral intention to diagnose a particular drug as a result of beliefs and attitudes that strongly supported the prescription of that drug, a powerful kind of motivation—for example, the doctor's knowledge of, and guilt about, the fact that a previous patient on the same drug died—could cause the doctor to revise and reject the rational process of attitude formation. Thus, by making an accommodation for the role of motivation, Herath's Planned Behavior Model can accommodate both rational and irrational behavior. Herath's Planned Behavioral Model also provides an underpinning for one of the independent variables of the study, the use of diagnostic software, by grounding the decision to use such software in a rationalistic framework of choice.

Before examining how the Planned Behavior Model accommodates existing features of diagnostic science, a few further points about the model ought to be made. To begin with, variations of this model are often used to analyze patients' intentions to adopt health behaviors (Reneman, Geertzen, Groothoff, & Brouwer, 2008). However, as Herath (2010) pointed out, the Planned Behavior Model is population-agnostic, and can be applied to any human decision-making process. Second, it is also often the case that the Planned Behavior Model is used to model behavior that will take place weeks, months, or even years in the future (Li, Frieze, & Tang, 2010). However, the Planned Behavior Model can be applied to any decision that takes place more than a few seconds after the initial stimulus (Herath, 2010). Thus, the ways in which the Planned Behavior Model has historically been used in medical research should not be taken as limitations of the model itself. Having explained why the model might be useful to researchers interested in diagnostic behavior, it is natural to offer more detail on how and why the Planned Behavior Model fits with diagnostic science. Such a discussion, while being useful in its own right, will also serve as a foundation for a discussion of differences between human and machine diagnosis.

In order to understand how the Planned Behavior Model is a good description of what takes place in diagnosis, and thereby to set the stage for understanding how human and machine diagnosis are different, consider that diagnosis itself can be broken into three components: computation, satisficing, and intuition. In this section of the literature review, each of these components will be discussed on their own, after which the components' relevance to the Planned Behavior Model will be discussed in a separate section.

**Computation**

Computation at its most basic level can be understood as the use of mathematical processes to arrive at an output from an input (Berstein, 2011). There are two widely recognized forms of computation. In the mathematical model of computation, an input is transformed via a function; once the initial input is given, there are no additional steps, and the input and output are distinct from each other. In the engineering model of computation, inputs and outputs can be entangled, such that they interact with each other, and the computation is revised accordingly (Meyers, 2011). In terms of outputs themselves, computations have two forms: They can be closed and exhaustive, or they can have a confidence level. For example, the computation that 2 + 2 = 4 is a closed and exhaustive computation; the degree of certainty that the output follows from the inputs is absolute (Adam, 2011). An example of a kind of computation that is based on a confidence level is a forecast based on a Poisson distribution (Babu, 2011). For example, if one tries to forecast how many red cars will arrive at a stop sign based on previous observations of the sign, the forecast will always be an estimate; statistical methods can be used to indicate how much confidence researchers should have in the forecast.

As it can be imagined, computation in medical diagnosis tends toward confidence levels rather than absolutely certainty (Chernick, 2011). It is hardly ever the case that doctors think of a particular set of symptoms as absolutely indicative of a certain disease;

because of the complexity of disease, the fallibility of the process of matching symptoms to diseases, and other unforeseen variables—such as patients who lie about their symptoms—doctors prefer not to think of diagnoses as being absolutely certain or uncertain (First & Tasman, 2011). Additionally, because diagnosis contains a built-in process of monitoring and, if necessary, adjusting the treatment, diagnosis can be thought as a kind of engineering computation rather than as a kind of mathematical computation (McGann & Hutson, 2011).

Thus, having set aside mathematical computation, closed calculations, and absolute certainty as concepts that apply infrequently to medical diagnosis (Chernick, 2011; First & Tasman, 2011; McGann & Hutson, 2011), it would be useful to spend more time understanding how confidence levels and engineering computation function in diagnostic science. How, then, does computation function in diagnosis science? To begin with, the treatment (alongside the symptoms) should be thought of as the input, and the result as the output. Doctors wish to be as certain as they can be that a particular input will lead to the output of wellness (Nuttall & Rutt-Howard, 2011), however it is defined (e.g., cessation of symptoms, patient's self-reported health, etc.). In the process of computation, then, the goal of the doctor is to be as sure as possible that the input of treatment leads to the output of health (Bath-Hextall, Lymn, & Knaggs, 2011). The problem is: How can the doctor find the most appropriate input?

In computational terms, one solution to the problem is what is known as a decision tree. In a decision tree, the computing system—whether it is a human, a computer, or a human using a computer—works through all of the available forking paths on an if-then basis: That is, calculating if a particular decision is taken, what will its utility be? (Bekkerman, Bilenko, & Langford, 2011). Chess computers provide a simple example of decision trees. In deciding how to make a chess move, chess computers employ decision trees to calculate the respective costs and benefits of all available moves, or as many moves as the computer's central processing unit (CPU) can model. Current supercomputers can calculate the costs and benefits of trillions of moves in just a few seconds (Nielsen, 2011). Of course, calculations of this sort are based on a confidence level; there is no one right move, but rather a move that is rated higher than other moves (Lefrancois, 2011).

In the world of the physician, one use of the computational decision tree is to match observed symptoms to any number of diseases that could fit the symptoms. In some cases, such a decision tree might be small; for example, some patients have symptoms that are highly typical of a handful of diseases, prompting the doctor to take a closer look at those diseases and their possible connection to the symptoms (Clavien & Trotter, 2011). In other cases, a symptom could be typical of hundreds, or even thousands, of known diseases (Reiss, Shadomy, & Lyon, 2011). In such a case, a so-called brute force computational approach would be to examine every one of the possible diseases for further evidence of matching symptoms; in the actual practice of medicine, however, it is more common for doctors to collect more data that can narrow down the number of diseases with which a symptom might be associated (Gifford, 2011).

One interesting computational study was that of Martin, Perez, and Muller's (2009), which analyzed the role of Bayesian statistics in medical decision-making. As

Martin et al. pointed out, diagnosing a disease can be highly dependent on when a patient is examined, and when particular data is collected from the patient. According to Martin et al., some diseases progress in a more or less random fashion, meaning that more regular sampling of a patient's condition might be necessary to reach a proper diagnosis. On the other hand, some diseases proceed in an almost linear fashion, creating urgency for the doctor to move from the process of gathering evidence to the process of administering a treatment to effect a cure. Martin et al. argued that there is an obvious computational role for diagnostic software that can accurately estimate, on behalf of a doctor, when data should be gathered from a patient, and calculate the significance of gathered data. Martin et al.'s discussion emphasized the rising importance of statistical analysis in modern medicine, and argued that software performs more accurately than doctors in determining when patients should be monitored for particular diseases. As Martin et al., argued, even if physicians perform this kind of statistical analysis themselves, the act of doing so might deplete valuable time and energy that the physician needs for more cure-oriented actions. Thus, Martin et al. concluded that there appears to be a clear role for diagnostic software in making statistical calculations about when to gather patient data, and deciding the statistical significance of the gathered data.

Whereas Martin et al. (2009) discussed the specific computational utility of diagnostic software at a specific stage of diagnosis (evidence collection), Schwab (2008) made a more general point about the advantages of computation that has to do with heuristics, which is defined as "a method of solving problems that puts aside belief in things like causality and argumentation from the known to the unknown" (Bowman &

Frega, 2012, p. 348). In an article about the limits of medical decision-making, Schwab pointed out that “human judgment is governed by generally expedient heuristics (a flipped coin will come up heads half the time) that lead to predictable biases (people expect a flipped coin to land heads, then tails, then heads, then tails, etc.)” (p. 1861). Doctors are not immune to what Schwab called the heuristics and biases theory of decision-making. The heart of the problem is that, as Schwab put it, human psychology is committed to “sense-making processes” (p. 1865). All humans have some bias toward the need for events, actions, and behaviors to make sense. Computation does not have this bias; it is agnostic to the various decision-making fallacies enumerated by Schwab. Thus, in addition to serving the kind of positive utility described by Martin et al., it is also possible that computation as it applies to diagnostic decision-making has a negative utility: By rooting diagnosis in statistics, computation prevents doctors from committing cognitive errors related to the innate human desire for sense and meaning.

Another discussion of the fallibility of physician decision-making and the superiority of the computational approach appeared in Gorini and Pravettoni’s (2011) recent article on cognitive bias in the diagnostic process. Goritni and Pravettoni identified two important flaws in physician decision-making; heuristic bias (which was also discussed by Schwab, 2008) and under-reliance on the statistical likelihood of disease:

> …physicians often use [representative heuristics] to match symptoms of the patient against prototypes or mental templates of diagnoses. However, relying on the representativeness heuristic can lead a diagnostician to only look at and search for the prototypical manifestations of a disease. This can lead to an incorrect or

> delayed diagnosis when aspects of a patient's presentation are atypical. In some instances, the reliance on the representativeness heuristic leads to a ‘base-rate neglect’. Base-rate neglect includes the failure to adequately take into account the prevalence of a particular disease. When the true prevalence of a disease is ignored, it may lead to the overestimation of improbable diagnoses, which, in turn, is disadvantageous for the patient and can result in an over-utilization of resources (p. 548).

Because diagnostic calculation is probabilistic, diagnostic software programs do not succumb to the representativeness heuristic; as will be discussed later in the literature review, software programs offer likelihoods, not certainties, and computation cannot be misled by the motivation to fit symptoms to diseases in a particular and biased way. Additionally, because computation is non-emotional reasoning, relying on diagnostic computations is warranted when either physicians are trying to get an idea of the likelihood of a disease, quite apart from how the patient or the physician would feel about seeing a diagnosis revealed as accurate or inaccurate. For these reasons, the software engineers Gorini and Pravettoni (2011) defended the use of diagnostic software and other forms of computational decision-making in the diagnostic context.

On the other hand, as Bucknall (2010) argued in a recent article on the nature of medical error in emergency diagnoses, computation also has its limits. According to Bucknall, “80% of medical error results from system flaws” (p. 152), with the system defined as the sum of human, machine, and process inputs. Thus, for example, it is of little use to employ diagnostic software to indicate the best blood sampling times to test for a certain disease when the patient cannot be reached by the hospital, or has not been directed to follow up. One of the implications of Bucnkall's argument is that, while diagnostic software is excellent at computational tasks, these tasks in themselves do not significantly reduce the innate risk of medical error, because they take place within an existing system. If that system is flawed, then so is the utility of diagnostic software. At the same time, Bucknall argued that the value of doctors lies in their ability to step outside a flawed system, a quality that Bucknall considered an extremely important component of diagnostic success in emergency medicine in particular. Another implication of Bucknall's work is that a physician's knowledge of certain aspects of diagnostic software, one of the independent variables of the study, can affect the ultimate efficacy of diagnostic software.

Another limitation of computation is that computation becomes more complicated once a decision tree has already been generated and followed to a conclusion (Segal & Shahar, 2009). For example, if a doctor has used computational methods to identify six diseases with which a symptom might be compatible, there are diminishing returns to further computation. At some point, doctors might not be able to reduce the possible number of symptom-matching diseases; in addition to the obvious limitations on time that exist in many diagnostic environments (such as an emergency room in a busy urban environment), evidence itself is limited (Croskerry, Cosby, & Schenkel, & Wears, 2008). When a doctor runs out of tests and other diagnostic aids, and is still left with a handful of diseases that match the patient's symptoms, other diagnostic skills have to be called upon. One of these skills is known as satisficing.

**Satisficing**

To return to the previous example, imagine that a doctor has identified six diseases with which a given patient's symptoms are consistent, and that there is no further use for computational roles in determining precisely what disease the patient might be suffering from. What can a doctor do? Garnham and Oakhill (1994) provided one possible answer as follows: The doctor, following the principle of satisficing, could simply choose the first diagnostic alternative that met a specified criterion, such as accounting for a highly important symptom.

A doctor starting with a list of six diseases that are equally likely to be responsible for a given set of symptoms could thus resort to satisficing by administering a treatment for a single disease and watching the result (Gigerenzer & Gray, 2011). The initial treatment would be guided by the use of bounded rationality, meaning that the physician would apply his or her existing knowledge to narrow down the possible choices for treatment (Gigerenzer & Gray, 2011). If the treatment for that disease resulted in a cessation of the symptoms, then the doctor might assume that the diagnostic process had been successful (Groopman, 2007). The choice of which disease to try to treat first is driven by any number of considerations. For example, a doctor could try to treat the disease that was the most life threatening or that required the earliest intervention. If all of the diseases compatible with a patient's symptoms were equal in their danger, then the doctor might randomly choose one of the diseases to try to treat (Montgomery, 2006).

Satisficing is, in its way, a kind of experimentation. If the experimentation is successful, the diagnostic process is over—especially in a medical atmosphere in which there is extreme pressure on doctors to achieve results and move on to the next patient (Schwartz & Bergus, 2008). If the experimentation is unsuccessful, however, then satisficing can lead back into computation. It was noted earlier in the literature review that the engineering model of computation assumes an ongoing interaction between inputs and outputs. If a satisficing doctor noticed that a particular treatment failed to result in a cessation of symptoms, but gave rise to a new piece of medical evidence, then the computation process could begin again, as the doctor tried to fit new symptoms to a disease set (Shaw, Ramachandra, Lucas, & Robinson, 2011). What is more common, however, is for doctors to try treatments in turn until they observe cessation of symptoms (Felder & Mayrhofer, 2011). Thus, it can be argued that the goal of medical computation is to lower the number of possible fits between symptom and disease—ideally, to 1 (that is, a unique fit), but sometimes to 2-6 fits, in which the doctor works through different possibilities in an experimental manner (Rao, 2007).

Satisficing has its weaknesses, among them the lack of diagnostic precision (Zilberberg, 2011), but also its strengths. One of the strengths of satisficing is the self-correcting nature of the practice (Simon, 1947). When doctors observe that a diagnosis reached through satisficing is incorrect, they move on pragmatically, factoring other knowledge into their diagnostic decisions and subjecting their earlier decisions to more critical scrutiny, achieving what Ryan (2010) called reflective inquiry, and what other scholars call, more generically, metacognition (Mamede, Rikers, & Schmidt, 2012).

Satisficing, while proven to be a common tool in the diagnostic arsenal, has limitations. Sometimes, especially in emergencies, a doctor might lack the time or the means to try several treatments; at other times, a patient might be too fragile to endure several treatments in succession. In cases of this kind, doctors often rely on a third kind of diagnostic skill: Intuition.

**Intuition**

Duggan (2005) defined intuition, as it functions for expert and scientifically minded decision makers, as a form of decision-making based on "patterns of similarity and difference with other situations" (p. 9) experienced by the decision-maker. Intuition can be understood as a specific form of physician knowledge not easily rendered into diagnostic rules.

The popular understanding of intuition is that of a sixth sense or some other pseudo-mystical capability. However, as Duggan pointed out, intuition as it functions among scientific decision-makers are not mystical or frivolous, but rather a diagnostic skill rooted in experience. Such experiences can be highly idiosyncratic and resistant to computational analysis (Kattan & Cowen, 2009). For example, doctors might have noted that many past patients in a particular situation reacted to a particular drug with toxic shock, and might thus assume that a current patient in the same situation would respond in the same way. The medical literature is replete with examples of intuition as a diagnostic skill (Chapman & Sonnenberg, 2003; Flynn & Van Schaik, 2003; Plessner, Betsch, & Betsch, 2008). Intuition is often the last diagnostic computation, one that comes after computation has been exhausted and satisficing is impossible or deemed too risky (Flynn & Van Schaik, 2003).

Intuition has another important role in diagnostic science. It also serves as the form of introspection that affords doctors a stronger belief in their beliefs, and that leads doctors to rely on the received wisdom of medical manuals and the pooled knowledge of their profession (Groopman, 2007). In this way, intuition—in combination with computation and satisficing—can be directly related to the Planned Behavior Model of Herath (2010).

Recent literature on intuition in medical decision-making has called attention to some unique strengths of the human touch. For example, McDermott (2008) pointed out that there were many variables in whether or not a patient would accept and comply with a specific treatment plan for a diagnosed disease, including “the way treatment options are framed and presented” (p. 665) by the doctor and “denial mechanisms” (p. 665). As McDermott argued, diagnostic decisions require patient compliance to be successful. Take the case of a patient whom the diagnosing doctor knows to have a particular bias: For example, a reluctance to take a medicine that is injected into the bloodstream. In this kind of case, diagnostic software would not know which of the available treatments to, as it were, pitch to the patient; the doctor’s knowledge of patient context, and skill at framing, are required to achieve compliance. Sometimes this aspect of the doctor’s skill is not based on a direct judgment about the patient’s compliance, but becomes an innate quality built out of thousands of clinical encounters (McDermott, 2008). Software designers are striving towards the ability to build a kind of intuition into software using expert systems, although, in the context of this study, it remains to be seen what doctors think of such systems.

**Computation, Satisficing, Intuition, and the Planned Behavior Model**

The planned behavior model (Ajzen, 2005; Herath, 2010) was chosen as one of the theoretical and conceptual bases for this study because it appeared to be a useful account of the entire diagnostic cycle, from evidence collection to administration of treatment, as a human would proceed (but not as software would proceed, because software does not take therapeutic actions). This model complements the analysis of distributed cognition explored earlier in the literature review, via an examination of how diagnostic software adds knowledge on which a doctor-software system can act. In addition, the three classic components of diagnosis (that is, computing, satisficing, and intuition) can fit within the Planned Behavior Model. In reference to Table 3, computing is a source of behavioral beliefs (element 1a of the model); intuition is a source of beliefs about beliefs (element 1c) and motivation (elements 4a and 4c), as is also the force that promotes belief in normativity (1b). Finally, computation, satisficing, and intuition work together to determine doctors' attitudes, behavioral intentions, and actual behavior.

It is important to note that the relationship between the three classic elements of diagnosis (computation, satisficing, and intuition) and the Planned Behavior Model is a conceptual relationship. The relationship between the model and the elements of diagnosis has not been explored in the literature, and is better thought of as a means of interweaving behavioral theory with diagnostic science rather than as a precise description of what takes place during diagnostic decisions.

It is possible to think of the diagnostic process as a search to reduce possibilities (Sox & Higgins, 1988) in the match between a set of symptoms and diseases that match those symptoms. In this process, the role of computation is to eliminate as many inappropriate or poor matches as possible. The role of satisficing is to treat the possible disease matches and make empirical observations to test whether the diagnosis is working. The role of intuition is to bypass both computation and satisficing in those cases in which the physician makes what is essentially an educated, experience-based diagnostic guess. Diagnostic software can contribute in each of these domains of diagnosis.

For example, if fed the input that a patient has elevated alpha fetoprotein, diagnostic software could either return a single suggestion based on statistical likelihoods of the fit between symptom and disease—for example, *Patient has liver cancer*—or else a list of all of the diseases with which an elevation of alpha fetoprotein is consistent, a list that can be sorted based on likelihood. The default setting on various medical software packages, including DiagnosisPro and Connectance, is to return a list of diseases with which a symptom is consistent, sometimes accompanied by a percentage quantification of the fit between symptom and disease (e.g., 5% of patients with hemothorax have atypical mycobacteria). Thus, the current medical software packages can engage in computation, satisficing, and a form of intuition (Newborn, 2003). Once again, however, it remains to be discovered in the qualitative portion of this study what doctors think of diagnostic software's intuitive or pseudo-intuitive capabilities.

## The Role of Software Technology in Diagnoses

Medical technology has existed from the beginning of the practice of medicine, given what is known about ancient human attempts at surgery (Cockburn, Cockburn, &

Reyman, 1998). In this study, however, only computer software will be discussed under the rubric of medical technology. It is appropriate, then, to delimit the discussion to computer software and to tie its development to medical decision-making. With that limitation in mind, the remainder of this final section of the literature review will dwell on recent research on diagnostic software. The purpose of this discussion is to gather as much knowledge as possible about what scholars think of the characteristics, strength, and weaknesses of diagnostic medical software. This discussion will be integrated with the earlier discussions of theory, and of the three frames—computation, satisficing, and intuition—of diagnostic decision-making.

To begin with, diagnostic medical software can be divided into two broad categories. One kind of diagnostic software is embedded into diagnostic medical machines or databases; this kind of software does not make a diagnosis, per se, but rather provides information that is extremely important for the doctor in making a diagnosis (Eadie, Taylor, &Gibson, 2012). Another kind of software is specifically designed to be diagnostic in nature; fed data about a set of symptoms, it returns a diagnosis or list of diagnoses, accompanied by relevant information (Eadie et al., 2012). Both of these kinds of software will be discussed in turn.

Many different kinds of diagnostic software accompany many different kinds of medical devices. One common and instructive example is that of software that accompanies a computerized tomography (CT) machine (De Palma, 2011). CT machines, which have a characteristic tube-like shape, are designed to take a three-dimensional image of human subjects. For example, CT scanning is often performed in order to determine whether tumors, calcification, infarction, or other conditions are present in a patient (DePalma, 2011). A patient who comes to a doctor complaining of an unaccountable pain in the jaw might turn out to be suffering from a tumor that is pressing down on a nerve, and that might show up in a CT scan (DePalma, 2011).

On its own, a CT scan conveys purely visual information. This information has to be interpreted (DePalma, 2011). A great deal of data is interpreted directly in the CT machine interface by diagnostic software that is part of the CT package (Fujiyoshi, Kadowaki, Kadowaki, Sekikawa, Ohkubo, & Miura, 2011). For example, the medical equipment company Siemens (2012) sells a workstation to accompany its CT unit, and one of the functions of the workstation is to be able to perform calcium analysis. For example, Siemens' CT software can calculate what is known as a calcium score for the patient, on a vessel-by-vessel basis. Typically, if a patient were to have a CT scan for calcification, it would be the software itself—for example, Siemens' Crealife CT Calcium Score Analysis Function—that would make the diagnosis of calcification, as in the case of hardened arteries (Baumuller, Leschka, Desbiolles, Stolzmann, Scheffel, & Seifert et al., 2009). Of course, that is not the only case in which calcification of the arteries could be diagnosed. A doctor could reach the same conclusion by means of a traditional differential diagnosis. The advantage of the CT software is that it can make a diagnosis whose accuracy cannot be replicated by a doctor. For example, in order to determine a calcium score for different blood vessels in a patient's body, the doctor could theoretically probe into the vessels with a scalpel, take samples, and make manual calculations, but such a procedure would be unnecessary, invasive, and dangerous in comparison with simply obtaining a calcium score from CT software (Watanabe, Nakazawa, Higashi, Itoh, & Naito, 2011).

The use of computer-assisted diagnosis for radiology is well supported. According to Eadie et al., (2012), the significant error rate in radiology is between 2-20%; in other words, radiologists working without the assistance of any form of diagnostic software tend to make clinically-significant errors anywhere from 2 to 20% of the time. Eadie et al.'s meta-review of 147 empirical studies on computer-assisted diagnosis within the field of radiology revealed that software assistance is associated with between half and a fourth of the rate of error as compared to unassisted human diagnosis. As Eadie et al. noted, however, a number of factors make it difficult to quantify the difference between the accuracy of unassisted human doctors and the accuracy of computer-assisted doctors; for example, computer-associated diagnostics vary significantly in purpose, design, and characteristics, so that making precise comparisons is methodologically difficult. Nonetheless, Eadie et al. suggested that, at worst, computer-assisted diagnostic systems could improve a radiologist's diagnostic accuracy by 25% and at best by 50%. As a result, computer-assisted diagnosis in the field of radiology has become nearly ubiquitous in the United States, and is spreading in many other countries (Eadie et al., 2012).

Researchers in other areas of medicine have replicated results of the kind obtained in Eadie et al.'s (2012) study. Renz, Bottcher, Diekmann, Poellinger, Maurer, and Pfeil et al. (2012) discovered that computer-assisted diagnostic software embedded within a breast magnetic resonance imaging (MRI) machine was able to achieve diagnostic accuracy of 93.5%, sensitivity of 96.5%, and specificity of 75.5%. The accuracy rate obtained by computer-assisted diagnosis in this case is between 10-20% greater than historic accuracy rates achieved without software (Renz et al., 2012).

It is interesting to observe that, in studies in which doctors register their disapproval of diagnostic software, they fail to take CT and similar software into account. Doctors utilize diagnostic software on a routine basis because of the superior computing speed of such software. Indeed, much of the work that is done by this software is now an indispensable part of medicine; many tasks that doctors once did by hand, involving mental calculation and manual measurement, now take place in an electronic environment (Reece, 2009, p. 15).

However, when doctors protest about diagnostic software, they are typically referring not to the kind of diagnostic software that is embedded in CT machines, but rather to software, that, fed a particular set of symptoms, returns a diagnosis. Such software is typically designed to run on a hand-held device that accompanies the doctor, although it can also run on personal computer workstations or laptops (Randeree, 2007). Differential diagnosis generation software is, in some ways, an extension of printed diagnostic manuals, which also serve as a reference guide to physicians trying to make a diagnosis. One of the differences between diagnostic manuals and diagnostic software is that doctors control the pace and quality of their interaction with manuals. In other words, a doctor is the one who makes the decision to consult a manual. On the other hand, diagnostic software is designed to accompany doctors into consultations with patients, which some doctors have found to be intrusive (O’Malley, Grossman, Cohen, Kemper, &

Pham, 2010). One of the classic complaints that doctors have about diagnostic software is that it represents some kind of curb on their autonomy; thus, it is not necessarily the mere fact of diagnostic software that some doctors have protested, but rather a pattern of administrative decisions that is seen as foisting particular tools and practices on doctors (Queenan, Angst, & Devaraj, 2011).

Bond, Schwartz, Weaver, Levick, Giulianio, and Graber (2011) evaluated differential diagnosis generators with performance testing. The findings indicated that, in the field of differential diagnosis in particular, software is of varying strength; it is not yet the case that, as in the field of radiology, computer-assisted diagnosis has risen to the level of a must-have tool in differential diagnosis. Bond et al. (2011) discovered that only two differential diagnosis generators, Isabel ™ and DxPlain ™, performed well in testing. Ranked on a 5-point scale based on performance in achieving accurate diagnosis in 20 test cases, both Isabel ™ and DxPlain ™ were able to achieve a mean rating of 3.45. There are thus two major differences between differential diagnosis generators and computer-assisted diagnosis in the fields of radiology and image analysis in general. First, differential diagnosis generators are less accurate than computer-assisted diagnosis in image analysis (Bond et al., 2011). Second, there is a wide variance in the performance level of commercially available differential diagnosis generators (Bond et al., 2011), which can be interpreted in a number of ways. First, it could be the case that image analysis is innately simpler than solving primary medicine cases. Second, image analysis is more advanced than differential diagnosis software. Since there do not appear to be empirical studies that compare the sophistication of differential diagnosis software with computer-assisted diagnosis in image analysis, these questions have not been resolved; nonetheless, they ought to be kept in mind by future researchers interested in obtaining a more detailed understanding of the differences between the two major kinds of diagnostic software in the medical marketplace.

There is clearly empirical support for the proposition that both differential diagnosis software and computer-assisted diagnostic software for image analysis work, although at differing levels of accuracy. There is also support (bolstered by Eadie et al.'s, 2012 meta-review of 147 studies of computer-assisted image analysis in radiology) that radiologists and other doctors who analyze images work extensively with computer-assisted diagnostic systems, especially in the United States. The main open question in the literature, and the one that is most germane to this study, is the question of how doctors engage with differential diagnostic software in real-world settings.

Some empirical studies have added important insights to what is known about this topic. For example, Ramnarayan, Winrow, Coren, Nanduri, Buchdahl, and Jacobs et al. (2006) conducted a study of how pediatricians used differential diagnosis generators while on the job. Ramnarayan et al. (2006) discovered that, when given freedom of choice in resorting to the use of differential diagnosis generators in a pediatric hospital, physicians chose to access the system only 8.6% of the time, and to examine actual diagnostic advice only 2, 55% of the time. The mean usage time of the diagnostic system was only 1 minute, 38 seconds. Ramnarayan et al. reported that the main obstacle to diagnostic software use cited by the physicians in the study was technical; many physicians reported difficulty using the interface of the system, and some physicians noted that they were working from locations in which access to the differential diagnosis system was difficult to obtain.

Ramnarayan et al.'s (2006) study has a number of important implications for the topic of physician use of diagnostic software. To begin with, the study raises the possibility that, when use of differential diagnosis generators is not mandated, numerous physicians might simply be refusing diagnostic software because of issues of accessibility or perceived difficulty. The study provides support for the point made by Umscheid and Hanson (2011) and Bond et al. (2011) in different contexts, which is that differential diagnostic generation software has some quality gaps. Umscheid and Hanson summarized a number of non-technical and non-access-related reasons that physicians have historically given in order to justify the avoidance of diagnostic software. One such reason is that, in many medical contexts, the presenting cases are of what Umscheid and Hanson (2011) described as a "bread and butter" (p. 6) character. Another reason is that many doctors practice surgery and other practices that Umscheid and Hanson distinguished from so-called cognitive medicine. A third reason is that, when a presenting case is complex and a physician has a lack of knowledge, a more likely outcome than consulting a differential diagnosis generator is to consult a senior colleague or a colleague with more experience in treating the presenting set of symptoms. Based on these reasons, Umscheid and Hanson concluded that differential diagnosis software is useful in a number of contexts that might seldom manifest themselves in a doctor's career. However, Umscheid and Hanson reached this conclusion based on a meta-review of only a few empirical studies, and other views about the nature of the relationship between physicians and differential diagnosis generation software should be sought out and analyzed.

While Umscheid and Hanson (2011) argued that differential diagnosis generators were not, overall, superior in accuracy or specificity to unaided physician diagnosis except in limited circumstances, a pseudo-experiment conducted by David, Chira, Eells, Ladrigan, Papier, and Miller et al. (2011) reached markedly different conclusions. David et al. (2011) worked with a sample of patients whose cellulitis had been misdiagnosed by the admitting team. Interestingly, David et al. found that, in 64% of the cases, a differential diagnostic software package known as Visual Dx had included the correct diagnosis, which had been ignored or overridden by the physicians in the admitting team. David et al. discovered that unassisted physicians were only 14% accurate in diagnosing stasis dermatitis that presented with some of the characteristics of cellulitis. Thus, at least in their limited field of misdiagnosed cellulitis and based on what was after all a small sub-sample (N=28) of misdiagnosed patients, the differential diagnostic system studied by David et al. appeared to be 50% more accurate than an unassisted admitting team. Results of this kind indicate that differential diagnostic software might have its own pockets of excellence, especially in contrast to human diagnosticians; one strength of VisualDX, for example, was its integration with visual data (David et al., 2011). This result implies that differential diagnostic generators that are able to benefit from the more advanced forms of image analysis outshine generic differential diagnostic systems that rely on text and mote limited kinds of input. This implication is supported by the empirical literature, since there are many studies in which the accuracy and specificity of differential diagnostic software embedded in image databases appears to be high (see for example the meta-review of studies in David et al., 2011's literature review). There are fewer studies on the utility of differential diagnosis generators that are disconnected from visual data.

Another empirical study confirming the superiority of visually based computer-assisted diagnosis software over unassisted physicians was that of Puech, Betrouni, Makni, Dewalle, Villers, and Lemaitre (2009). Puech et al. (2009) discovered that a computer-assisted diagnostic tool was able to successfully diagnose 77% of instances of prostate cancer appearing in magnetic resonance imaging (MRI) files, whereas expert radiologists examining the same MRI data only achieved a 70% rate of success. The findings of this sort indicate why, in radiology, diagnostic software is not an afterthought, but rather directly integrated into the ordinary diagnostic processes of radiologists.

Of course, many physicians have also indicated their satisfaction with non-visually-based diagnostic software, explaining that such software narrows down potential diagnoses, saves time by taking the place of a medical manual, and also serves as a convenient resource for patients, since diagnoses can be printed or emailed to patients from within the software interface (Chowdhury, Roy, & Saha, 2011, p. 221). Thus, the reaction to diagnostic software—at least in the United States—can best be described as mixed. Many doctors admire the computational robustness and convenience of diagnostic software, but resent such software's potential to cut into the autonomy of their practices (Menachemi, Matthews, Ford, Hikmet, & Brooks, 2009). In terms of the Planned Behavior Model (Herath, 2010), doctors' beliefs about their own decisions is thus a determinant of how they feel about diagnostic software. Doctors who prize the autonomy of individual and independent beliefs and attitudes as a part of the diagnostic process might wish to distance themselves from software (Seeley, 2009). Older doctors might have a resistance to learning and re-learning new technology (Seeley, 2009). Other doctors take a different approach, seeing diagnostic software not as a form of competition but rather as an extension of the existing infrastructure of medical informatics (Shield, Goldman, Anthony, Wang, Doyle, & Borkan, 2010). Doctors who dislike learning new technology, who have physical difficulties with reading digital data, or who dislike having to install new software might also be at odds with using more software in their practices (Seeley, 2009).

It should be emphasized that the purpose of diagnostic software is not, and has seldom been described as, an attempt to replace the doctor. The role of diagnostic software can be better understood through the concept of distributed cognition. According to Hazlehurst, Gorman, and McMullen (2008), distributed cognition is the act of cognition diffused across a system that is larger than the individual doctor is. For example, a doctor sitting and thinking about a diagnosis would be engaging in pure cognition, but a doctor consulting a reference manual or interacting with software would become part of what Hazlehurst et al. called a model of distributed cognition. The act of cognition cannot be said to reside in any one component of the system, but is distributed across it. In contemporary times, the spread of computer technology has brought distributed cognition to many fields; yet, as Hazlehurst et al. pointed out, distributed cognition has been around at least as long as humans have created machines and tools (from the abacus to the personal computer) that can assist them in cognition. According to Hazlehurst et al. (2008), the goal of studying the diagnostic process as it manifests itself in software versus the doctor is not to argue on behalf of one or the other component of the system. Doctors and software will continue to work together as part of a system of distributed cognition that is deeply embedded in medicine. While accepting the reality of distributed cognition, however, it is still necessary to be able to better understand the division of labor between humans and software tools. Hsiung (2012) has pointed out that the pace of software adoption in American hospitals has been far slower than the adoption of other forms of medical technology. As Hsiung has stated, many doctors have been perfectly willing to take advantage of tools to assist them in arriving at their diagnostic decisions, but have resented the incursion of software that can take the input of a set of symptoms and return a diagnosis as an output. Whether this attitude is rational or not, it has been observed in many different medical contexts, leading to the conclusion that doctors take special pride in diagnosis and are wary of relinquishing their role to software. Thus, all of the forms of computational utility enabled by diagnostic software are best thought of parts of a distributed cognition system in which the key diagnostic role is still played by the doctor.

The key problem acknowledged in the literature, especially literature pertaining to the American healthcare system, is the rate of misdiagnosis. Umscheid and Hanson (2011) pointed out that deaths from misdiagnosis in the United States have remained study somewhere between 40,000 and 80,000 a year, depending on the methodology by which such deaths are ascribed to misdiagnosis. There is no simple way in which to illuminate the possible relationship between the unchanged rates of misdiagnosis in the United States and the use or non-use of diagnostic software. However, given that software of this kind has been available for over 30 years and was specifically designed in order to lower the rates of misdiagnosis, it is worth continuing to gather and analyze data on physicians' usage of diagnostic software, and to understand what usage patterns might have to do with accurate diagnosis, misdiagnosis, and other medical outcomes.

The empirical literature offers some insights into the relationship between diagnostic software and diagnostic outcomes. First, there is strong evidence that computer-assisted diagnostic systems that are tied into visual databases are both highly accurate and highly specific in their diagnoses. Eadie et al.'s (2012) meta-review of 147 studies in the field of radiology found that such diagnostic software was nearly ubiquitous in the United States, and routinely achieved diagnostic accuracy rates over 90%. Second, there is some research suggesting that physician adoption of differential diagnostic generation software—as opposed to diagnostic software embedded in machines or associated with image databases, as in radiology—is low, whether because of technical problems and perceived inconvenience (Ramnarayan et al., 2006) or whether because physicians seldom encounter a genuine need to use such software (Umscheid & Hanson, 2011). Nonetheless, there is other research (David et al., 2011) suggesting that differential diagnostic software can perform better than unassisted physicians in some fields, such as dermatology can.

In general, the sheer number of symptom permutations, types of medicine, and differences between diagnostic software packages makes it difficult to reach conclusions about the relationship between such software and diagnostic success. Nonetheless, there is compelling evidence that diagnostic software is, at best, superior to human diagnostic procedures and, at worst, able to function as a helpful adjunct to human diagnostic processes.

## Summary and Conclusion

In general, the literature is in agreement (Ahlers et al., 2010; Bath-Hextall et al., 2011; Carpenito-Moyet, 2008; First & Tasman, 2011; Gifford, 2011) that diagnosis is a process characterized by the following steps: (a) Gathering of evidence, including physical evidence and verbal evidence (gathered from speaking to the patient) pertaining to a patient’s symptoms; (b) fitting knowledge about the symptoms to a specific disease, whether known or postulated; (c) determining the most appropriate treatment for the disease; and (d) fine-tuning the treatment based on ongoing observations of the interaction between the patient and the proposed treatment. However, there is some debate on the question of how culture- and value-laden the process of diagnosis is; with certain scholars (Byrne, 2012; Fadiman, 1988), having argued that diagnosis is heavily influenced by culture. On balance, however, there is stronger support in the literature for the idea that diagnosis is a repeatable, rigorous scientific process with the discrete steps enumerated above.

Diagnosis was shown to have three components: Computation, satisficing, and intuition. The main computing concepts (Lefrancois, 2011; Nielsen, 2011) that apply in medicine are those of (a) engineering computing, in which the input and output can interact several times over the course of diagnosis; (b) the decision tree, in which the utility of each possibility is calculated independently; and (c) confidence levels, in which relative rather than absolute recommendations are made based on statistical likelihoods that an observed symptom or set of symptoms corresponds with a disease. Computation can prevent certain cognitive fallacies, such as the representativeness heuristic, from manifesting themselves during the diagnostic process (Schwab, 2009) and serve as the basis for a self-correcting form of satisficing. Satisficing can be a comparison of several alternative solutions followed by a choice of the solution that seems more likely to succeed, or else an experimental means of working through possibilities (for example, six diagnoses of six different diseases that are all equally likely matches of an observed symptom set) in which the uniquely human characteristic of on-the-fly learning is important (Garnham & Oakhill, 1994). Finally, doctors often use intuition to shorten the decision process, make difficult diagnostic decisions, or raise the chances that patients will comply with a treatment implied by a specific diagnosis (Kattan & Cowen, 2009). Overall, the utility of software seems to be limited to computation; however, the literature review did not contain any studies that tried to quantify human advantages in satisficing and intuition versus the software advantage in computation.

# Chapter 3: Research Method

## Introduction to Research Method

The purpose of this quantitative study was to draw upon physician-provided data to determine why, at least in physicians' opinions, the prevalence of misdiagnosis has remained high despite the widespread adoption of diagnostic software. Knowledge that is more definitive is needed about why misdiagnosis has persisted well into the age of diagnostic software. The research method described and defended in this chapter was intended to generate such knowledge.

In addressing the question of why the widespread adoption of diagnostic software has not coincided with a decrease in the prevalence of misdiagnosis, a quantitative approach was necessary to determine which of some of the possible answers to this question—diagnostic software insufficiency, insufficient/improper use by physicians, or liability—is more popular with physicians and to determine whether answers to this question vary significantly depending on the demographic and professional characteristics of physicians, which also served as control variables in the study design in a manner described later in the chapter.

## Restatement of Research Questions and Hypotheses

The research questions associated with the study were as follows:

Research Question 1: Does use of diagnostic software decrease misdiagnosis in healthcare versus unassisted human diagnostic methods?

$H_0$1: Diagnostic software use has more misdiagnoses in healthcare than unassisted human diagnostic methods.

$H_A1$: Diagnostic software use has less misdiagnoses in healthcare than unassisted human diagnostic methods.

Research Question 2: Do physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare?

$H_02$: Physicians do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare.

$H_A2$: Physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare.

Research Question 3: Is physicians’ knowledge of diagnostic software extensive enough to decrease misdiagnosis in healthcare?

$H_03$: Physicians’ knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare.

$H_A3$: Physicians’ knowledge of diagnostic software is extensive enough to decrease misdiagnosis in healthcare.

Research Question 4: Do liability concerns prevent physicians from using diagnostic software?

$H_04$: Liability concerns do not prevent physicians from using diagnostic software.

$H_A4$: Liability concerns prevent physicians from using diagnostic software.

## Research Design and Approach

According to Windelband (1913), nomothesis—the hallmark of quantitative research—is the search for laws and generalizations; nomothetic methods try to examine a research phenomena in the most general way possible. On the other hand, idiography is the study of unique phenomena. In Windelband's example, an idiographic researcher might spend time and effort on trying to understand a single painting, whereas a nomothetic researcher might look for general differences between two classes of paintings.

Diagnostic practices can, and indeed should, be studied from both nomothetic and idiographic perspectives because these two perspectives complement and enrich each other. In most qualitative methodologies, idiography requires rich narrative data (Lapan, Quartaroli, & Riemer, 2011) whereas nomothesis can be carried out with basic survey or numeric data (Mehl, Conner, & Csikzentmihalyi, 2011). In this study, nomothetic data were gathered using an original survey designed to answer the four research questions of the study.

There are numerous quantitative designs available. Experimental, pre-experimental, and quasi-experimental designs are all reliant on the experimenter's control or partial control of variables (Creswell, 2009). In this study, I could not control or manipulate factors related to the use of diagnostic software, so these three approaches were not appropriate. A cross-sectional design is based on collecting data from participants at a single point at time; such designs can be used either to measure change in response to some manipulation of variables or else the prevalence of some attitude or behavior in a test population (Creswell, 2009). A cross-sectional design was, therefore, an appropriate design for the study, given the focus on measuring physicians' adoption of and attitudes to diagnostic software at a single point in time.

## Setting and Sample

Conceptually—not physically, because the study was not reliant on direct observation or on-site analysis—the setting for this study was the world of medical diagnostics. According to Shealy (2011), diagnostic skill is taught and is required to be learned in all medical schools that offer an accredited degree in biomedicine; thus, it was assumed that all doctors in the sample were, in their own ways, experts in the process of diagnosis. The study qualification criteria for doctors were, simply, (a) being currently qualified and practicing as a doctor in the United States, (b) using English comfortably enough to participate in the survey, (c) having access to diagnostic software, and (d) giving consent. Given that the study is in English, it made the most sense to sample doctors from English-speaking countries. As the United States has by far the largest population of native English speakers in the world (Yoshihara, Sylva, & Eberstadt, 2011), and is home to most of the major diagnostic software providers currently in operation (Kramme, Hoffmanm, & Pozos, 2011), sampling from the United States was logical.

In the United States, the AMA offers a master list of every U.S.-licensed physician; according to the AMA (2012), there were over 814,000 licensed physicians in the United States. I sent recruitment emails to 3,100 AMA-accredited physicians through their professional database licensees' distribution list with the use of Survey Monkey platform that yielded a sample of 99 physicians for the study.  According to Kennett and

Salini (2011), the average response rate for a marketing campaign in which the target message recipient has some innate interest is between 2 and 3%. Assuming that this result is achieved, reaching out to 3,100 doctors (by a combination of e-mail, where available, and mass mailing) yielded a likelihood of 62 doctor respondents.

A sample size over 60 is in line with the sample sizes reported by some previous scholars working with medical populations (Keeney et al., 2011) and should therefore be considered acceptable, even though it was not anticipated whether this sample size was sufficiently large or qualified enough to yield rich data for the study, or whether a sample over 60 was sufficient for all statistical procedures. The power analysis in Figure 1 revealed that, with an effect size of 0.5 and an $\alpha$ of .05, a sample size over 45 is sufficient for a one sample *t* test (that is, a *t* test comparing a group mean versus a hypothesized value) at a power of 0.95.

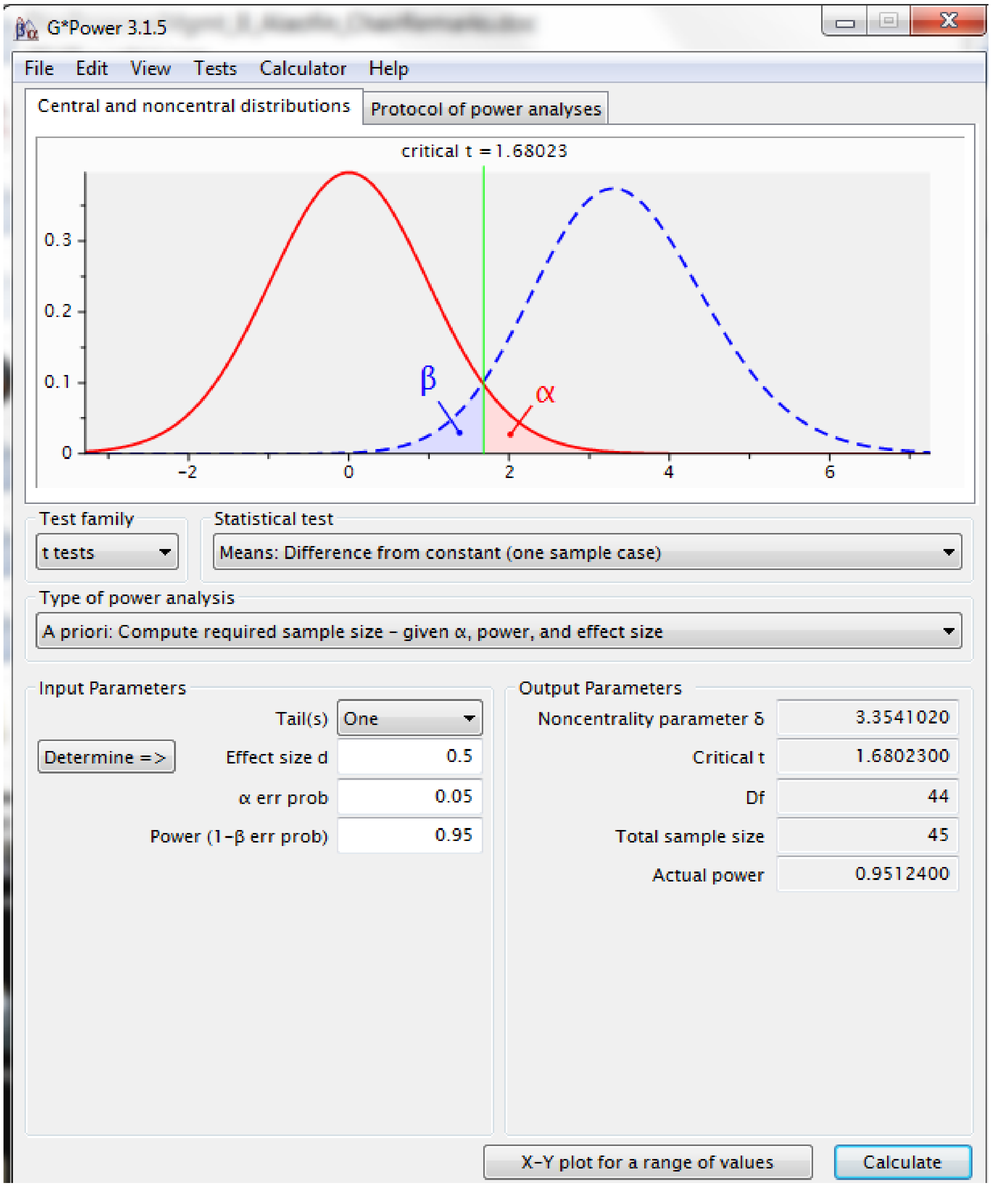


*Figure 1.* Power analysis.

## Data Collection and Analysis

The first question asked was *What kind of diagnostic software do you use?* While none of the research questions pertained to what kind of diagnostic software did physicians use, it was still important to gather information on this point for descriptive purposes. A brief definition and some examples of diagnostic software were provided as part of the survey.

The first research question that guided the study were as follows: Does use of diagnostic software decrease misdiagnosis in healthcare versus unassisted human diagnostic methods?

$H_0 1$: Diagnostic software use has more misdiagnoses in healthcare than unassisted human diagnostic methods. Data for this question were collected by presenting the following two survey prompts:

1. *My use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods.*
2. *In general, physicians' use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods.*

All of the survey questions differentiate between doctors' own use of diagnostic software and their general perceptions about diagnostic software. Responses for all of the survey questions were conducted on a 7-point Likert scale, with 1 = *completely disagree*, 2 = *disagree*, 3 = *somewhat disagree*, 4 = *neither agree nor disagree*, 5 = *somewhat agree*, 6 = *agree*, and 7 = *completely agree*.

The operationalization of the null hypothesis was *H*o: The value of the mean survey response was greater than or equal to 4. I would have rejected the null hypothesis if the observed mean response to the survey question were sufficiently small so that the *p* value was less than the alpha of 0.05. Two one-tailed one-sample *t* tests were done, one for each survey prompt.

The second research question that guided the study were as follows: Do physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare?

$H_0$2: Physicians do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare. Data for this question were collected by presenting the sample the following two surveys prompts:

1. *I do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare,*
2. *In general, physicians do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare.*

The operationalization of the null hypothesis was *H*o: The value of the mean survey response was greater than or equal to 4. I would have rejected the null hypothesis if the observed mean response to the survey question were sufficiently small so that the *p*-value was less than the alpha of 0.05. Two one-tailed one-sample *t* tests were done, one for each survey prompt.

The third research question that guided the study were as follows: Is physicians' knowledge of diagnostic software extensive enough to decrease misdiagnosis in healthcare?

$H_0$3: Physicians' knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare. Data for this question were collected by presenting the following two survey prompts:

1. *My knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare.*
2. *In general, physicians' knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare.*

The operationalization of the null hypothesis was *H*o: The value of the mean survey response was greater than or equal to 4. I would have rejected the null hypothesis if the observed mean response to the survey question were sufficiently small so that the *p*-value was less than the alpha of 0.05.  Two one-tailed one-sample *t* tests were done, one for each survey prompt.

The fourth research question that guided the study were as follows: Do liability concerns prevent physicians from using diagnostic software?

$H_0$4: Liability concerns do not prevent physicians from using diagnostic software. Data for this question were collected by presenting the following two questions:

1. *Liability concerns do not prevent me from using diagnostic software.*
2. *In general, liability concerns do not prevent physicians from using diagnostic software.*

The operationalization of the null hypothesis was *H*o: The value of the mean survey response was greater than or equal to 4. I would have rejected the null hypothesis if the observed mean response to the survey question were sufficiently small so that the *p*-value was less than the alpha of 0.05.

Hypothesis testing for the research questions were carried out with the assistance of standard statistical software (i.e., SPSS™). According to Dewberry (2004), Likert scales are continuous: “a common way of obtaining continuous data in organizational research is with a Likert scale” (p. 9). A *t* test was, therefore, appropriate to use with a Likert scale.

No instruments other than the Likert-type scales discussed above were used in the study. Not using an instrument allowed doctors to define the concepts of expertise, frequency of use, liability, and the capability of diagnostic software in their own personal and professional contexts instead of requiring them to align their responses with others’ operationalizations of these concepts. Additionally, since physicians are among the busiest of all professionals, it was unlikely that they would submit to the administration of several scales; it was therefore a research advantage to design an instrument that could be answered in only a few minutes.

## Instrumentation and Materials

The instrument for data collection was an online survey, hosted on the Survey Monkey™ platform that presented the following prompts, in addition to demographic questions about the doctor’s age, practice area, and gender:

1. *What kind of diagnostic software do you use?*
2. *My use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods.*

3. *In general, physicians' use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods.*
4. *I do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare.*
5. *In general, physicians do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare.*
6. *My knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare.*
7. *In general, physicians' knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare.*
8. *Liability concerns do not prevent me from using diagnostic software.*
9. *In general, liability concerns do not prevent physicians from using diagnostic software.*

Because this was a new questionnaire, a reliability analysis was conducted on the first 15 to 20 responses. Cronbach's alpha was used to determine whether the survey instrument had sufficient reliability. Changes were made to the survey items based upon the results of this analysis. The proposed sampling strategy to select 3,100 doctors randomly from the master list of American Medical Association yielded a likelihood of 62 doctor respondents. This was theoretical and increased validity. Physicians who completed the survey may have different characteristics such as, different areas of practice and specializations. It is likely, therefore, that the results were not highly generalizable to the entire population of American doctors.

## Protection of Human Participants

The four commonly-accepted (Creswell, 2009) categories of ethical assurances in research are offering protection from harm, using informed consent, defending the privacy of study subjects, and being honest with professional colleagues and study subjects. This study posed no innate harms to subjects and will use best practices (Creswell, 2009) for informed consent. IRB approval (# 07-16-13-0036160) was obtained prior to the collection of any data. Respondent privacy was defended in the following ways: First, the surveys did not collect respondents' names or other data that could be used by outside parties to identify respondents; second, all electronic forms were encrypted; third, all electronic forms received by the researcher from Survey Monkey™ were placed on a password-protected laptop to which only the researcher had access, and were furthermore backed up by online storage at Box.com ™ in case of laptop theft or loss; finally, all data will be destroyed no earlier than seven years after the study has been accepted by the researcher's institution. Honesty with professional colleagues and study subjects were ensured by means of following the ethical precepts of research, and by means of publishing the raw data of the study for open scrutiny at the end of the study.

## Conclusion

The quantitative approach described and defended in this chapter was designed to address the question of why the widespread adoption of diagnostic software has not coincided with a decrease in the prevalence of misdiagnosis. The quantitative approach determined which of the possible answers to this question—diagnostic software insufficiency, insufficient / improper use by physicians, or liability—was more plausible to physicians, in a manner that casted more light on why the problem of misdiagnosis has persisted despite the proliferation of technology.

# Chapter 4: Results

## Introduction to Results

The purpose of this study is to draw upon physician-provided data to determine why, at least in physicians' opinions, the prevalence of misdiagnosis has remained high despite the widespread adoption of diagnostic software. This chapter contains three sections for reporting and interpreting the results of the study. The first section is an overview of descriptive statistics. The second section consists of the inferential statistics associated with the research hypotheses. The third section consists of findings that are not related to the research questions but that still cast light on the question of physicians' relationships with diagnostic software.

The research questions of the study were as follows:

Research Question 1: Does use of diagnostic software use decrease misdiagnosis in healthcare versus unassisted human diagnostic methods?

Research Question 2: Do physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare?

Research Question 3: Is physicians' knowledge of diagnostic software extensive enough to decrease misdiagnosis in healthcare?

Research Question 4: Do liability concerns prevent physicians from using diagnostic software?

I collected data from July 8, 2013 to October 16, 2013. There were 99 surveys distributed, of which 97 were completed, indicating a completion rate of roughly 98%. One of the 97 respondents did not answer the questions pertaining to Hypotheses 5 to 8.

Recruitment emails were sent to 3,100 physicians; the response rate was roughly 3%, close to the rate anticipated in Chapter 3. There was a deviation from the data collection plan presented in Chapter 3. The sample of 99 was likely to be representative of the physician population of 814,000, given that the a priori sample size analysis in Figure 1 suggested a sample of 45 in order to achieve a confidence interval of 95% in the context of the chosen statistical procedure and effect size.  The Cronbach's Alpha of the survey was .761, indicating a high level of internal consistency.

## Descriptive Statistics of Demographic Variables

The specialty areas of physician's bar chart in Figure 2 indicated that there were 97 respondents in the sample from a wide variety of specialization areas of medicine.

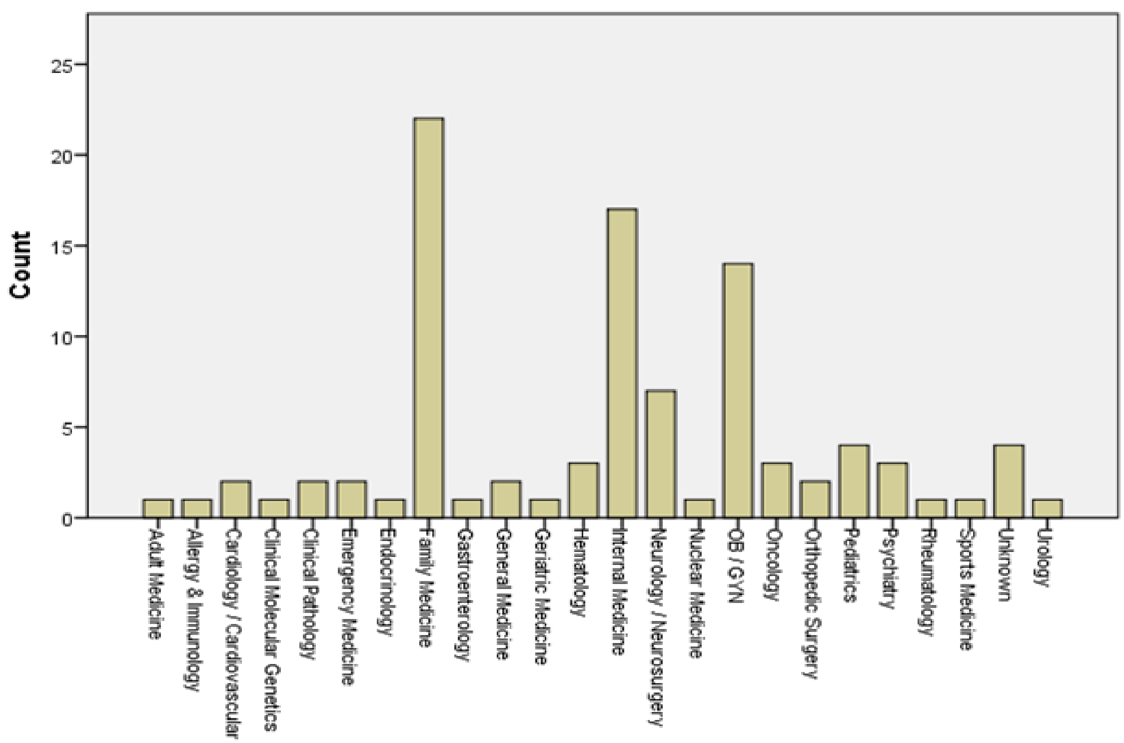


*Figure 2.* Specialty areas of physicians in sample.

According to data collected on specialty areas of physicians in the sample presented in Figure 2, the most well-represented categories of medicine in the survey were family medicine (22 out of 97 respondents), internal medicine (17 out of 97 respondents), and OB/GYN (14 out of 97 respondents).

Further analysis of descriptive statistics revealed that the majority of the sample (59.8%) had practiced medicine between 0 and 10 years (Table 4), that the sample was geographically well-distributed (Table 5), the sample tended to be fairly young (Table 6), and that men and women were equal in number (Table 7).

Table 4

*Experience of Physicians in Sample*

| | | Frequency | Percent |
|---|---|---|---|
| | Less than 5 years | 36 | 37.1 |
| | 5-10 years | 22 | 22.7 |
| | 11-20 years | 18 | 18.6 |
| | 21-30 years | 11 | 11.3 |
| | More than 30 years | 9 | 9.3 |
| | Total | 96 | 99.0 |
| Missing | | 1 | 1.0 |
| Total | | 97 | 100.0 |

Table 5

*Geographic Location of Physicians in Sample*

| | | Frequency | Percent |
|---|---|---|---|
| | New England | 17 | 17.5 |
| | Mid-Atlantic | 16 | 16.5 |
| | South Atlantic | 18 | 18.6 |
| | South | 11 | 11.3 |
| | Midwest | 9 | 9.3 |
| | Mountain | 9 | 9.3 |
| | Pacific | 16 | 16.5 |
| | Total | 96 | 99.0 |
| Missing | System | 1 | 1.0 |
| Total | | 97 | 100.0 |

Table 6

*Age of Physicians in Sample*

| | Frequency | Percent |
|---|---|---|
| Less than 30 years old | 23 | 23.7 |
| 30-40 years old | 26 | 26.8 |
| 41-50 years old | 20 | 20.6 |
| 51-60 years old | 16 | 16.5 |
| 61-65 years old | 7 | 7.2 |
| More than 65 years | 5 | 5.2 |
| Total | 97 | 100.0 |

Table 7

*Gender of Physicians in Sample*

| | Frequency | Percent |
|---|---|---|
| Male | 51 | 52.6 |
| Female | 46 | 47.4 |
| Total | 97 | 100.0 |

Descriptive statistics pertaining to participants' usage of diagnostic software were also collected and are shown in Tables 8 to 14. I presented the results of the data collected on use and non use of different diagnostic software packages among the sampled physicians in Tables 8 to 14.

Table 8

| *Access to Various Diagnostic Medical Packages* | Frequency | Percent |
|---|---|---|
| None | 42 | 43.3 |
| Isabel | 17 | 17.5 |
| DXplain | 7 | 7.2 |
| Your rapid diagnosis | 8 | 8.2 |
| Diagnosis pro | 6 | 6.2 |
| Connectance | 9 | 9.3 |
| Search engines | 1 | 1.0 |
| Use more than 1 software | 7 | 7.2 |
| Total | 97 | 100.0 |

Table 9

| *Diagnostic Medical Packages Used* | Frequency | Percent |
|---|---|---|
| None | 42 | 43.3 |
| Isabel | 17 | 17.5 |
| DXplain | 7 | 7.2 |
| Your rapid diagnosis | 8 | 8.2 |
| Diagnosis pro | 6 | 6.2 |
| Connectance | 9 | 9.3 |
| Search engines | 1 | 1.0 |
| Use more than 1 software | 7 | 7.2 |
| Total | 97 | 100.0 |

Table 10

| *Length of Access to Diagnostic Software* | Frequency | Percent |
|---|---|---|
| Not applicable | 42 | 43.3 |
| Less than 6 months | 8 | 8.2 |
| 6 months to 1 year | 10 | 10.3 |
| 1-3 years | 14 | 14.4 |
| 3-5 years | 17 | 17.5 |
| Use more than 5 years | 6 | 6.2 |
| Total | 97 | 100.0 |

Table 11

| *Length of Using Diagnostic Software* | Frequency | Percent |
|---|---|---|
| Not applicable | 42 | 43.3 |
| Less than 6 months | 8 | 8.2 |
| 6 months to 1 year | 12 | 12.4 |
| 1-3 years | 12 | 12.4 |
| 3-5 years | 17 | 17.5 |
| Use more than 5 years | 6 | 6.2 |
| Total | 97 | 100.0 |

Table 12

| *Access to Types of Diagnostic Software* | Frequency | Percent |
|---|---|---|
| None | 42 | 43.3 |
| Isabel | 11 | 11.3 |
| DXplain | 8 | 8.2 |
| Your rapid diagnosis | 6 | 6.2 |
| Diagnosis pro | 5 | 5.2 |
| Connectance | 8 | 8.2 |
| Search engines | 5 | 5.2 |
| EasyDiagnosis | 2 | 2.1 |
| NxOpinion | 1 | 1.0 |
| Use more than 1 software | 9 | 9.3 |
| Total | 97 | 100.0 |

Table 13

*Diagnostic Software Currently Used*

| | Frequency | Percent |
|---|---|---|
| None | 42 | 43.3 |
| Isabel | 11 | 11.3 |
| DXplain | 8 | 8.2 |
| Your rapid diagnosis | 6 | 6.2 |
| Diagnosis pro | 5 | 5.2 |
| Connectance | 8 | 8.2 |
| Search engines | 5 | 5.2 |
| EasyDiagnosis | 2 | 2.1 |
| NxOpinion | 1 | 1.0 |
| Use more than 1 software | 9 | 9.3 |
| Total | 97 | 100.0 |

Table 14

*Length of Time Using Current Software*

| | Frequency | Percent |
|---|---|---|
| Not applicable | 47 | 48.5 |
| Less than 6 months | 7 | 7.2 |
| 6 months to 1 year | 14 | 14.4 |
| 1-3 years | 9 | 9.3 |
| 3-5 years | 14 | 14.4 |
| Use more than 5 years | 6 | 6.2 |
| Total | 97 | 100.0 |

One of the important insights that emerged from Tables 8 to 14 was that slightly over 43% of the sample reported not using, and not having ever used, diagnostic software. In the Other Findings section of this chapter, some of the differences between diagnostic software adopters and non adopters were explored in greater depth. In the next section of the chapter, inferential statistics were calculated in order to perform the hypothesis tests to provide information for answering the research questions of the study.

## Inferential Statistics

Responses for all of the survey questions followed a 7-point Likert scale, with 1 = *completely disagree*, 2 = *disagree*, 3 = *somewhat disagree*, 4 = *neither agree nor disagree*, 5 = *somewhat agree*, 6 = *agree*, and 7 = *completely agree*.

The research questions and hypotheses of the study were as follows:

Research Question 1: Does use of diagnostic software decrease misdiagnosis in healthcare versus unassisted human diagnostic methods?

$H_0$*1*: Diagnostic software use has more misdiagnoses in healthcare than unassisted human diagnostic methods.

$H_A$1: Diagnostic software use has less misdiagnoses in healthcare than unassisted human diagnostic methods.

Research Question 2: Do physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare?

$H_0$2: Physicians do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare.

$H_A$2: Physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare.

Research Question 3: Is physicians' knowledge of diagnostic software extensive enough to decrease misdiagnosis in healthcare?

$H_0$3: Physicians’ knowledge of diagnostic software is not extensive enough to decrease misdiagnosis in healthcare.

$H_A$3: Physicians’ knowledge of diagnostic software is extensive enough to decrease misdiagnosis in healthcare.

Research Question 4: Do liability concerns prevent physicians from using diagnostic software?

$H_0$4: Liability concerns do not prevent physicians from using diagnostic software.

$H_A$4: Liability concerns prevent physicians from using diagnostic software.

One-sample *t* tests were carried out on each of the null hypotheses. Before the *t* tests, descriptive statistics were collected for each of the survey questions. These descriptive statistics included four measurements: (a) *N* (the number of respondents who answered the prompt), (b) mean (the mean response score, on a Likert scale of 1-7), (c) standard deviation, and (d) standard error of the mean. These descriptive statistics, as well as the inferential statistic of the 95% confidence interval, are presented in Table 15 and Table 16. Note that one of the 97 respondents did not respond to any of the Table 16 questions, and that another respondent did not respond to the question for H7.

Table 15

*Descriptive Statistics of the Sample (Personal Use)*

| *H* | Survey Questions | *N* | Mean | *SD* | *SE* Mean | 95% Confidence Interval of the Mean | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower | Upper |
| *H*1 | My use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | 55 | 2.47 | 2.01 | 0.27 | 1.92 | 3.01 |
| *H*2 | I do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | 55 | 3.24 | 1.93 | 0.26 | 2.71 | 3.75 |
| *H*3 | My knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | 55 | 3.60 | 2.06 | 0.28 | 3.04 | 4.15 |
| *H*4 | Liability concerns do not prevent me from using diagnostic software. | 55 | 5.31 | 1.99 | 0.27 | 4.77 | 5.84 |

Table 16

*Descriptive Statistics of the collected* Sample (Physicians in General)

| *H* | Survey Questions | *N* | Mean | *SD* | *SE* Mean | 95% Confidence Interval of the Mean | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower | Upper |
| *H*5 | In general, physicians' use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | 96 | 3.89 | 2.09 | 0.21 | 3.46 | 4.30 |
| *H*6 | In general, physicians do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | 96 | 3.74 | 1.90 | 0.19 | 3.35 | 4.13 |
| *H*7 | In general, physicians' knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | 95 | 3.99 | 1.92 | 0.20 | 3.60 | 4.38 |
| *H*8 | In general, liability concerns do not prevent physicians from using diagnostic software. | 96 | 5.22 | 1.64 | 0.17 | 4.88 | 5.55 |

After the descriptive statistics were collected, an attempt was made to measure whether there was a significant difference between the mean of the response and the value of 4; 4 was chosen as the cutoff value because it was the mean value between 1 and 7 on the Likert scale. The *p* values are presented in Table 17.

Table 17

*One-Sample* T *Test Results*

| | Survey question | *t* | *p* (two-tailed) | *p* (one-tailed) | Effect Size |
|---|---|---|---|---|---|
| H1 | My use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | -5.641 | <.001 | <.001 | .76 |
| H2 | I do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | -2.929 | .005 | .0025 | .40 |
| H3 | My knowledge of diagnostic Software is not extensive enough to result in a decrease in misdiagnoses. | -1.440 | .156 | .078 | .29 |
| H4 | Liability concerns do not prevent me from using diagnostic software. | 4.880 | <.001 | >.999 | .66 |
| H5 | In general, physicians' use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | -.514 | .183 | .0915 | .05 |
| H6 | In general, physicians do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | -1.340 | .593 | .2965 | .13 |
| H7 | In general, physicians' knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | -.053 | .958 | .479 | .005 |
| H8 | In general, liability concerns do not prevent physicians from using diagnostic software. | 7.297 | <.001 | >.999 | .74 |

Since a one-tailed approach was used, the directionality of the one-sample *t*-tests was ≥ 4 for each of the hypotheses. With this point in mind, the results of the one-sample *t*-tests were as follows.

A one-sample *t*-test was conducted to evaluate H1, whether doctors' personal use of diagnostic software was likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. The null hypothesis was that personal diagnostic software use resulted in more misdiagnoses in healthcare than unassisted human diagnostic methods. The results indicated that doctors' mean level of agreement with this prompt (M = 2.47, $SD$ = 2.008) was significantly lower than 4, $t(96) = -5.641$, $p < .001$, so the null hypothesis was rejected. Some support was therefore found for the alternative hypothesis, which was the claim that personal diagnostic software use resulted in fewer misdiagnoses in healthcare than unassisted human diagnostic methods. Table 18 and Table 19 contain the test results for each of the eight survey prompts:

Table 18

*Hypothesis Testing Results (Personal Use)*

| Null hypothesis | | *p*(one tailed) | Result |
|---|---|---|---|
| H1 | My use of diagnostic software diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | <.001 | Rejected |
| H2 | I do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | .0025 | Rejected |
| H3 | My knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | .078 | Retained |
| H4 | Liability concerns do not prevent physicians from using diagnostic software. | >.999 | Retained |

Table 19

*Hypothesis Testing Results (Physicians in General Use)*

| Null hypothesis | | *p*(one tailed) | Result |
|---|---|---|---|
| H5 | In general, physicians' use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | .0915 | Retained |
| H6 | In general, physicians do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | .2965 | Retained |
| H7 | In general, physicians' knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | .479 | Retained |
| H8 | In general, liability concerns do not prevent physicians from using diagnostic software. | >.999 | Retained |

A one-sample *t*-test was conducted to evaluate H2, whether doctors thought that they used diagnostic software frequently enough to decrease misdiagnoses in healthcare. The null hypothesis was that physicians thought they did not use diagnostic software frequently enough to decrease misdiagnosis in healthcare. The results indicated that doctors' mean level of agreement with this prompt (M = 3.24, *SD* = 1.934) was significantly lower than 4, $t(96) = -2.929$, $p = .0025$, so the null hypothesis was rejected. Some support was therefore found for the alternative hypothesis, which was that physicians thought they used diagnostic software frequently enough to decrease misdiagnosis in healthcare.

A one-sample *t*-test was conducted to evaluate H3, whether doctors thought that their knowledge of diagnosis software was extensive enough to result in a decrease in misdiagnoses. The null hypothesis was that doctors would agree with this prompt. The results indicated that doctors' mean level of agreement with this prompt (M = 3.60, *SD* = 2.060) was not significantly lower than 4, $t(96) = -1.440$, $p = .078$, I have not found evidence to reject this statement, so the null hypothesis was retained.

A one-sample *t*-test was conducted to evaluate H4, whether doctors thought that liability concerns did not prevent them from using diagnostic software. The null hypothesis was that liability concerns did not prevent physicians from using diagnostic software. The results indicated that doctors' mean level of agreement with this prompt (M = 5.31, *SD* = 1.990) was significantly higher than 4, $t(96) = 4.880$, $p = 1.000$, I have not found evidence to reject this statement, so the null hypothesis was retained.

A one-sample *t*-test was conducted to evaluate H5, whether doctors thought that, in general, physician use of diagnostic software was likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. The null hypothesis was that, in general, diagnostic software use resulted in more misdiagnoses in healthcare than unassisted human diagnostic methods. The results indicated that doctors' mean level of agreement with this prompt (M = 3.89, *SD* = 2.092) was not significantly lower than 4, $t(96) = -.514$, $p = .0915$, I have not found evidence to reject this statement, so the null hypothesis was retained.

A one-sample *t*-test was conducted to evaluate H6, whether doctors thought that, in general, physicians do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. The null hypothesis was that in general, physicians did not use diagnostic software frequently enough to decrease misdiagnosis in healthcare. The results indicated that doctors' mean level of agreement with this prompt (M = 3.74, *SD* = 1.904) was not significantly lower than 4, *t*(96) = -.1.340, *p* = .2965, I have not found evidence to reject this statement, so the null hypothesis was retained.

A one-sample *t*-test was conducted to evaluate H7, whether doctors thought that, in general, physician knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnosis. The null hypothesis was that in general, physicians' knowledge of diagnostic software was not extensive enough to decrease misdiagnosis in healthcare. The results indicated that doctors' mean level of agreement with this prompt (M = 3.99, *SD* = 1.921) was not significantly lower than 4, *t*(95) = -.053, *p* = .479, I have not found evidence to reject this statement, so the null hypothesis was retained.

A one-sample *t*-test was conducted to evaluate H8, whether doctors thought that, in general, liability concerns prevented physicians from using diagnostic software. The null hypothesis was that in general, liability concerns did not prevent physicians from using diagnostic software. The results indicated that doctors' mean level of agreement with this prompt (M = 5.22, *SD* = 1.636) was not significantly lower than 4, *t*(96) = 7.297, *p* = 1.000, I have not found evidence to reject this statement, so the null hypothesis was retained.

Physicians thus indicated that they thought their personal use of diagnostic software to be associated with error reduction, and in particular that their personal use of diagnostic software was frequent enough for the purpose of error reduction. Interestingly, doctors did not agree with the proposition that their knowledge of diagnostic software was extensive enough for misdiagnosis reduction purposes. Additionally, doctors were not personally deterred from using diagnostic software by liability concerns. When asked to speak for their profession, physicians did not agree with the claims that (a) diagnostic software use is associated with diagnostic error reduction, (b) diagnostic software is used frequently enough to make a difference, (c) physician knowledge of diagnostic software is extensive, and (d) physicians are deterred by liability concerns. As a result, I retained the null hypothesis.

Interestingly, the results of hypothesis testing for the fifth through the eight null hypotheses were the same when the one-sample *t* tests were run separately for those physicians who had never used diagnostic software and those physicians who had used diagnostic software. Additionally, all eight hypothesis tests were run again on separate sub-samples. First, the sample was divided into men and women. Second, the sample was divided into two groups of physicians, namely (a) those in family medicine, OB/GYN, and internal medicine; and (b) those in every other specialty area. Third, the sample was divided into physicians who were 40 or younger versus physicians who were over 40. The results of hypothesis testing did not differ significantly on these subgroups; in no case did a *p* value that was below .05 in Tables 18 and 19 change to .05 or over, and in no case did a *p* value that was .05 or over in Tables 18 and 19 fall under 05. There is therefore some support for the claim that the findings of the study hold across gender, specialty area, and physician age.

The implications of these findings relevant to the literature will be discussed further in Chapter 5. Before proceeding to the discussion of findings, I will present information about some of the assumptions of the one-sample *t* tests and procedures related to non-parametric techniques.

## Other Findings

According to Sheskin (2003), there are a number of assumptions that must be met by the one-sample *t* test. One assumption is that the sample has been randomly drawn. In this study, a randomly selected sub-population of 3,100 AMA-accredited physicians through Survey Monkey yielded a sample of 99 physicians using the techniques described in Chapter 3, so the assumption of randomness was met. The independence assumption of the one-sample *t* test is that observations are independent. In this study, each physician recorded one answer to every question independently from the other participants, so the observations are independent from each other. Another assumption is the use of interval or ratio data. Likert data is often treated as being interval data (Sheskin, 2003), since the range of Likert scale in this study is 7 and the test statistic comes from the sum of all of the responses from the respondents, so this assumption was also met. A final statistical assumption of the one-sample *t* test is that the data pass a test of normality such as the Shapiro-Wilk W test or the Kolmogorov-Smirnov test. If normality is not observed, then a non-parametric option to the one-sample *t* test, such as the one-sample Wilcoxon signed-rank test, can be used. I ran tests of normality on each of the eight survey prompts associated with the hypotheses of the study as seen in Table 20.

Table 20

*Tests of Normality*

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| H1 My use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | .339 | 55 | .000 | .716 | 55 | .000 |
| H2 I do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | .248 | 55 | .000 | .858 | 55 | .000 |
| H3 My knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | .236 | 55 | .000 | .850 | 55 | .000 |

Table 20

| | | | | | | |
|---|---|---|---|---|---|---|
| H4 Liability concerns do not prevent me from using diagnostic software. | .327 | 55 | .000 | .795 | 55 | .000 |
| H5 In general, physicians do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | .182 | 55 | .000 | .887 | 55 | .000 |
| H6 In general, physicians' use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | .216 | 55 | .000 | .884 | 55 | .000 |
| H7 In general, physicians' knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | .199 | 55 | .000 | .912 | 55 | .001 |
| H8 In general, liability concerns do not prevent physicians from using diagnostic software? | .273 | 55 | .000 | .808 | 55 | .000 |

*Note.* a. Lilliefors Significance Correction

table continues

Because the Shapiro-Wilk W and the Kolmogorov-Smirnov tests of normality both disclosed that the data for the eight prompts tested in the study were not distributed normally, the one-sample Wilcoxon signed-rank test was also conducted on the data. In conducting the Wilcoxon signed-rank tests, the null hypothesis in each instance was that the median response $\geq 4$. Because the interval data range for the Wilcoxon signed-rank test are the same as for the one-sample *t* test, no further assumption testing was conducted. Since a one-tailed approach was used, the directionality of the one-sample Wilcoxon signed-rank test was $\geq 4$ for each of the hypotheses.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H1, whether doctors' personal use of diagnostic software was likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. The null hypothesis was that personal diagnostic software use resulted in more misdiagnoses in healthcare than unassisted human diagnostic methods. The results indicated that doctors' mean level of agreement with this prompt (M = 2.47, $SD = 2.008$) was significantly lower than 4, $t(96) = 138.0$, $p < 0.001$, so the null hypothesis was rejected. Some support was therefore found for the alternative hypothesis, which was the claim that personal diagnostic software use resulted in fewer misdiagnoses in healthcare than unassisted human diagnostic methods.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H2, whether doctors thought that they used diagnostic software frequently enough to decrease misdiagnoses in healthcare. The null hypothesis was that physicians thought they did not use diagnostic software frequently enough to decrease misdiagnosis in healthcare. The results indicated that doctors' mean level of agreement with this prompt (M = 3.24, *SD* = 1.934) was significantly lower than 4, $t(96) = 273.0$, $p < 0.001$, so the null hypothesis was rejected. Some support was therefore found for the alternative hypothesis, which was that physicians thought they used diagnostic software frequently enough to decrease misdiagnosis in healthcare.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H3, whether doctors thought that their knowledge of diagnosis software was extensive enough to result in a decrease in misdiagnoses. The null hypothesis was that doctors would agree with this prompt. The results indicated that doctors' mean level of agreement with this prompt (M = 3.60, *SD* = 2.060) was significantly lower than 4, $t(96) = 389.0$, $p < 0.001$, so the null hypothesis was rejected. Some support was therefore found for the alternative hypothesis that doctors thought that their knowledge of diagnosis software was extensive enough to result in a decrease in misdiagnoses.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H4, whether doctors thought that liability concerns prevent them from using diagnostic software. The null hypothesis was that liability concerns did not prevent physicians from using diagnostic software. The results indicated that doctors' mean level of agreement with this prompt (M = 5.31, *SD* = 1.990) was significantly higher than 4, $t(96) = 1081.0$, $p = 0.99$, I have not found evidence to reject this statement, so the null hypothesis was retained.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H5, whether doctors thought that, in general, physician use of diagnostic software was likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. The null hypothesis was that, in general, diagnostic software use resulted in more misdiagnoses in healthcare than unassisted human diagnostic methods. The results indicated that doctors' mean level of agreement with this prompt (M = 3.89, *SD* = 2.092) was significantly lower than 4, $t(96) = 1337.5$, $p <0 .001$, so the null hypothesis was rejected. Some support was therefore found for the alternative hypothesis, which was the claim that general diagnostic software use resulted in fewer misdiagnoses in healthcare than unassisted human diagnostic methods.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H6, whether doctors thought that, in general, physicians do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. The null hypothesis was that in general, physicians thought they did not use diagnostic software frequently enough to decrease misdiagnosis in healthcare. The results indicated that doctors' mean level of agreement with this prompt (M = 3.74, *SD* = 1.904) was significantly lower than 4, $t(96) = 1557.5$, $p = 0.001$, so the null hypothesis was rejected.  Some support was therefore found for the alternative hypothesis, which was that physicians, in general, thought they used diagnostic software frequently enough to decrease misdiagnosis in healthcare.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H7, whether doctors thought that, in general, physician knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnosis. The null hypothesis was that in general, physicians' knowledge of diagnostic software was not extensive enough to decrease misdiagnosis in healthcare. The results indicated that doctors' mean level of agreement with this prompt (M = 3.99, *SD* = 1.921) was significantly lower than 4, $t(96)$

= 1642.5, $p = 0.008$, so the null hypothesis was rejected. Some support was therefore found for the alternative hypothesis that doctors thought that their knowledge of diagnosis software was extensive enough to result in a decrease in misdiagnoses.

A one-sample Wilcoxon signed-rank test was conducted to evaluate H8, whether doctors thought that, in general, liability concerns prevented physicians from using diagnostic software. The null hypothesis was that in general, liability concerns did not prevent physicians from using diagnostic software. The results indicated that doctors' mean level of agreement with this prompt (M = 5.22, *SD* = 1.636) was not significantly lower than 4, $t(96) = 3458.5$, $p < 0.099$, I have not found evidence to reject this statement, so the null hypothesis was retained.

Table 21

*Hypothesis Testing Based on Wilcoxon Signed-Rank Test (Personal Use)*

| **Null Hypothesis** | | **Wilcoxon Test Statistic** | ***p* (two-tailed)** | ***p* (one-tailed)** | **Result** |
|---|---|---|---|---|---|
| H1 | My use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic method | 138.0 | <0.001 | <0.001 | Rejected |
| H2 | I do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | 273.0 | <0.001 | <0.001 | Rejected |
| H3 | My knowledge of Diagnostic Software is not extensive enough to result in a decrease in misdiagnoses. | 389.0 | <0.001 | <0.001 | Rejected |
| H4 | Liability concerns do not prevent me from using diagnostic software. | 1081.0 | <0.001 | 0.99 | Retained |

Table 22

*Hypothesis Testing Based on Wilcoxon Signed-Rank Test (Physicians in General)*

| **Null Hypothesis** | | **Wilcoxon Test Statistic** | ***p* (two-tailed)** | ***p* (one-tailed)** | **Result** |
|---|---|---|---|---|---|
| H5 | In general, physicians' use Of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. | 1337.5 | <0.001 | <0.001 | Rejected |
| H6 | In general, physicians do not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. | 1557.5 | 0.001 | 0.001 | Rejected |
| H7 | In general, physicians' knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnoses. | 1642.5 | 0.020 | 0.008 | Rejected |
| H8 | In general, liability concerns do not prevent physicians from using diagnostic software. | 3458.5 | 0.01 | <0.099 | Retained |

One-sample Wilcoxon signed-rank tests were conducted to evaluate H1 and H5, whether doctors use of diagnostic software were likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. Speaking for themselves and the profession, the results indicated that doctors' mean levels of agreement with these prompts were significantly lower than 4, so the null hypotheses were rejected. Some support were therefore found for the alternative hypotheses, which were the claims that diagnostic software use resulted in fewer misdiagnoses in healthcare than unassisted human diagnostic methods.

One-sample Wilcoxon signed-rank tests were conducted to evaluate H2 and H6, whether doctors thought that they used diagnostic software frequently enough to decrease misdiagnoses in healthcare.  Speaking for themselves and the profession, the results indicated that doctors' mean levels of agreement with these prompts were significantly lower than 4, so the null hypotheses were rejected. Some support were therefore found for the alternative hypotheses, which were the claims that physicians thought they used diagnostic software frequently enough to decrease misdiagnoses in healthcare.

One-sample Wilcoxon signed-rank tests were conducted to evaluate H3 and H7, whether doctors thought that their knowledge of diagnostic software are not extensive enough to result in decrease in misdiagnoses. Speaking for themselves and the profession, the results indicated that doctors' mean levels of agreement with these prompts were significantly lower than 4, so the null hypotheses were rejected. Some supports were therefore found for the alternative hypotheses, which were the claims that physicians knowledge of diagnostic software are not extensive enough to result in a decrease in misdiagnoses.

One-sample Wilcoxon signed-rank tests were conducted to evaluate H4 and H8, whether doctors thought that liability concerns did not prevent them from using diagnostic software. Speaking for themselves and the profession, the results indicated that doctors' mean levels of agreement with this prompt were significantly higher than 4, I have not found evidence to reject this statement, so the null hypotheses were retained.

## Summary

The answers to the research questions of the study were as follows:

Research Question 1: *Does use of diagnostic software decrease misdiagnosis in healthcare versus unassisted human diagnostic methods?* Speaking for themselves, physicians did not agree with the null hypothesis that personal diagnostic software use resulted in more misdiagnoses in healthcare than unassisted human diagnostic methods. Therefore, I can conclude that the physicians believe that personal diagnostic software results in fewer misdiagnoses than unassisted human diagnostic methods. Speaking for the profession, I would say that there is insufficient evidence to refute the null hypothesis that general diagnostic software use resulted in more misdiagnoses in healthcare than unassisted human diagnostic methods.

Research Question 2: *Do physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare?* Speaking for themselves, but not for the profession in general, physicians did not agree with the null hypothesis that they did not use diagnostic software frequently enough to decrease misdiagnosis in healthcare. Therefore, I can conclude that the physicians believe frequent use that personal diagnostic software results in fewer misdiagnoses than unassisted human diagnostic methods. Speaking for the profession in general, I would conclude that there is insufficient evidence to refute the null hypothesis that physicians did not use diagnostic software frequently enough to decrease misdiagnosis in healthcare.

Research Question 3: *Is physicians' knowledge of diagnostic software extensive enough to decrease misdiagnosis in healthcare?* Speaking for themselves and the profession, I would say that there is insufficient evidence to refute the null hypothesis that physicians' knowledge of diagnostic software was not extensive enough to decrease misdiagnosis in healthcare.

Research Question 4: *Do liability concerns prevent physicians from using diagnostic software?* Speaking for both themselves and the profession, there is statistically significant evidence that the doctors in the sample have a different opinion than "neither agree or disagree". Majority of the physicians in this sample thought that liability concerns were not preventing physicians from using diagnostic software.

Having arrived at these results, the focus of the fifth and concluding chapter of the study will be on relating these findings to previous empirical findings and theoretical models, acknowledging limitations, and providing recommendations for scholars and practitioners.

# Chapter 5: Discussion, Conclusions, and Recommendations

## Introduction

The purpose of this quantitative study was to draw upon physician-provided data to determine why, at least in physicians' opinions, the prevalence of misdiagnosis has remained high despite the widespread adoption of diagnostic software. The study was carried out in a quantitative, cross-sectional manner that relied primarily on the inferential technique of the one-sample *t* test, Wilcoxon Signed Ranked test, and the administration of a diagnostic software attitudes and usage survey of 3,100 AMA-accredited physicians of whom 97 completed the survey. The purpose of this study was to measure physicians' attitudes to diagnostic software in a manner that could identify physician-perceived hindrances to and benefits of the use of diagnostic software.

In addressing the question of why the widespread adoption of diagnostic software has not coincided with a decrease in the prevalence of misdiagnosis, I used a quantitative approach to determine which of some of the possible answers to this question—diagnostic software insufficiency, insufficient/improper use by physicians, or liability—is more popular with physicians and to determine whether answers to this question vary significantly depending on the demographic and professional characteristics of physicians, with an emphasis on gender, specialty area, and physician age.

AMA offers a master list of U.S.-licensed physicians through their professional database licensees. I contacted one the licensees that provided a list of 3,100 physicians. I sent recruitment emails to these AMA-accredited physicians and invited them to complete my questionnaire on Survey Monkey platform. A total of 99 physicians responded from where I collected the data for the analysis.

All of the survey questions differentiated between doctors' own use of diagnostic software and their general perceptions about diagnostic software. Responses for all of the survey questions were conducted on a 7-point Likert scale, with 1 = *completely disagree*, 2 = *disagree*, 3 = *somewhat disagree*, 4 = *neither agree nor disagree*, 5 = *somewhat agree*, 6 = *agree*, and 7 = *completely agree*.

To determine the relationship between the variables, for the first two research questions of the study, the independent variable was whether diagnostic software was used and the dependent variable was reduction of misdiagnosis. For the third research question of the study, the independent variable was knowledge of diagnostic software and the dependent variable was reduction of misdiagnosis. For the fourth research question of the study, the independent variable was liability concern and the dependent variable was use of diagnostic software. I used the online survey tool, SurveyMonkey, to collect research data and the SPSS computing application to analyze the data. I conducted two one-tailed one-sample *t* tests on each survey question and used tables to report the result of the survey data.

**Research Question 1**

For Research Question 1, in the survey I asked if physicians thought the use diagnostic software helped to decrease misdiagnosis in healthcare versus unassisted human diagnostic methods.  The result of my analysis led to rejection of $H_0 1$. Speaking for themselves, but not for the profession in general, physicians indicated their thought that diagnostic software did in fact decrease misdiagnosis versus unassisted human diagnostic methods.

**Research Question 2**

For Research Question 2, I asked in the survey if physicians use diagnostic software frequently enough to decrease misdiagnosis in healthcare. The result of my analysis led to rejection of $H_0$2. Speaking for themselves, but not for the profession in general, physicians indicated their belief that they used diagnostic software frequently enough to detect misdiagnoses in healthcare.

**Research Question 3**

For Research Question 3, I asked in the survey if physicians' knowledge of diagnostic software was extensive enough to decrease misdiagnosis in healthcare. Speaking for both themselves and the profession in general, physicians did not indicate that there was enough knowledge of diagnostic software to decrease misdiagnoses in healthcare. I have not found evidence to reject this statement, so my analysis resulted in retaining $H_0$3.

**Research Question 4**

For Research Question 4, I asked if liability concerns prevent physicians from using diagnostic software. Speaking for both themselves and the profession in general, there is statistically significant evidence that the doctors in the sample have a different opinion than "neither agree or disagree". If I had done a two-tailed test, I would have rejected the null hypothesis that the doctors would neither agree or disagree that liability concerns prevent physicians from using diagnostic software. The majority of the physicians in this sample clearly indicated that liability concerns would not prevented physicians from using diagnostic software.

The results of hypothesis testing for the first and second null hypotheses were rejected using both one-sample *t* tests and one-sample Wilcoxon signed-rank tests. Therefore, one-sample *t* test and one-sample Wilcoxon signed-rank test agreed on this result. The results of hypothesis testing for the fourth and eighth null hypotheses were the same when one-sample *t* tests and one-sample Wilcoxon signed-rank test were run separately for those physicians who thought liability concerns did not prevent them from using diagnostic software.

Finally, the results of hypothesis testing for the Null Hypotheses 3, 5, 6, and 7 that were rejected using one-sample Wilcoxon signed-rank tests were not rejected using one sample *t* test. These results indicated that there were statistically significant difference between one-sample *t* test and one-sample Wilcoxon signed-rank test. Since the findings of the one-sample Wilcoxon signed-rank tests in this instance were not in accordance with the findings of the one-sample *t* tests, the limitation of non normality did appear to compromise the results of the study. The purpose of the previous chapter of the study was to present and comment on the statistical characteristics of the results associated with the study. The purpose of the present chapter of the study is to (a) summarize the relevant findings, (b) explore the implications of the findings with respect to the literature reviewed in Chapter 2, (c) generate recommendations for scholars and physicians, (d) discuss the limitations of the study, and (e) discuss the significance of the study. Each of these purposes will be addressed in a separate section of the chapter.

**Summary of Findings**

The research question-based findings of the study were as follows. First, physicians disagreed that their personal use of diagnostic software was likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods. Thus, there is statistical support for the conclusion that physicians in the sample thought that use of diagnostic software was likely to result in less misdiagnoses than unassisted human diagnostic methods. This is an important facet of the findings in that it exhibits the perception of value of diagnostic software on behalf of physicians. Second, I asked in the survey if physicians use diagnostic software frequently enough to decrease misdiagnoses in healthcare. Speaking for themselves, but not for the profession in general, physicians disagreed that they did not personally use diagnostic software frequently enough to decrease misdiagnoses in healthcare. On the other hand, speaking for the profession in general, there is insufficient statistical evidence to refute the null hypothesis that physicians did not use diagnostic software frequently enough to decrease misdiagnoses in healthcare. Finally, there was evidence that suggested fairly strongly that physicians agreed that liability concerns did not prevent them from using diagnostic software and that physicians agreed that physicians in general were not prevented from using diagnostic software because of liability concerns.

In the next section of this chapter, these findings will be discussed in relation to the existing literature on diagnostic software. Particular attention will be paid to the ways in which the findings of the study complement, contradict, or add context to the existing empirical findings as well as to theories of diagnostic software use.

**Relation of Findings to Literature**

Perhaps the most important finding in the current study with respect to the existing literature was the finding that physicians did not think that the use of diagnostic software would result in more misdiagnoses than unassisted human diagnostic methods. This finding can be interpreted as an endorsement of diagnostic software. If, in physicians' opinions, the use of diagnostic software is likely to result in fewer diagnostic errors than unaided human diagnostics, then diagnostic software is serving its intended function of improving clinical outcomes. If so, then the appropriate follow-up question is to ask why, if physicians endorse the clinical usefulness of diagnostic software, misdiagnosis continues to be such a pressing problem in the American healthcare system. While this question was not directly posed to the participants in the study, an answer to it can nonetheless be inferred from the fact that 42 of the 97 participants in the study, or just over 43% of the participants, had never used any form of diagnostic software as of the time of being surveyed. However, physicians in the sample seemed to think they used the diagnostic software enough to decrease misdiagnoses than unassisted human diagnostic methods. They just thought other doctors have used it enough. To address such a great discrepancy, I would recommend that future researchers seek to determine not only the usage rates and confidence of doctors in relation to diagnostic software but also the cost.

It might be true that diagnostic software is clinically effective but not widespread enough to lower misdiagnosis in a significant way. If so, then it is natural to ask why the adoption rate of diagnostic software is not higher. Technology adoption theories typically share the assumption that technology is adopted to the extent that it is found useful by adopters, barring mitigating factors such as high expense, the absence of organizational support, and high effort of use (Shah & Gardner, 2008; Spekowius & Wendler, 2006; Van Grembergern & De Haes, 2009). While these factors were not measured in the current study, the existing literature suggests that diagnostic software is relatively easy to use and is not expensive relative to other items of medical technology (Menachemi et al., 2009; Renz et al., 2012; Seeley, 2009; Shield et al., 2010). Additionally, I found that liability was not a concern for physicians' vis-à-vis diagnostic software. The main uninvestigated variable is likely institutional support. It is not clear, at least based on the existing research, whether the use of diagnostic software is institutionally supported in the American healthcare establishment. The absence of such support would help to explain the otherwise paradoxical finding that physicians find diagnostic software to be useful in reducing medical error while overall adoption of this technology remains relatively low.

## Relation of Findings to Theory

The questions in the study were not designed to explore the roles or characteristics of the three theoretical roots of diagnosis—that is, computing, satisficing, and intuition. However, the findings did not generally support the planned behavior model (Ajzen, 2005). Physicians who had not actually used diagnostic software had the same kinds of opinions about its usefulness and capabilities as physicians who had used diagnostic software. As discussed in Chapter 4, the one-sample *t* test results for non adopters and adopters of diagnostic software yielded the same hypothesis testing results.

Clearly, then, there are many nonadopters of diagnostic software that nonetheless have a positive view of such software and who might not have adopted the software because of institutional pressures or other structural reasons. One of the shortcomings of the planned behavior model in this regard is that it is not as adept at measuring the impact of external forces, such as organizational mandates, in the formation of individual behavior. Future scholars who study the adoption or non adoption of diagnostic software among physicians might be better served with a theoretical model more capable of measuring environmental and institutional pressures.

## Limitations of the Study

One of the statistical limitations of the study that became apparent during the process of data analysis was that three medical specialties—family medicine, OB/GYN, and internal medicine—represented more than 56% of the entire sample. The analysis in Chapter 4 revealed that segmenting the sample into the three medical specialties versus all other specialties did not result in any non significant *p* values becoming significant or significant *p* values becoming non significant. Future researchers who draw a sample of physicians that is more balanced with respect to medical specialty can likely overcome this limitation. In more general terms, the small size of the study ($N$ = 97) is likely to constitute an innate limit to the validity and reliability of results (note that the measured Cronbach's Alpha was .761); studies with significantly larger samples might find that the results of the present study might not be replicated. Another limitation of the study was that the data for the eight survey prompts relating to diagnostic software use were not normally distributed, leading to the supplemental use of the one-sample Wilcoxon signed-rank test.

## Recommendations for Scholars and Physicians

One appropriate recommendation for future researchers is to draw a sample of physicians that is more balanced with respect to the representation of different medical specialties. The fact that over half of the respondents came from three of the 24 occupations surveyed did not significantly impact the results of hypothesis testing, but drawing a larger and more balanced collection of specialty areas might nonetheless improve future results. It would also be appropriate for scholars to attempt to understand the apparent paradox of physicians' thoughts in the clinical utility of diagnostic software combined with a low adoption rate. This may be due to the economic factors associated with diagnostic software. It is recommended that future researchers seek to determine not only the usage rates and confidence of doctors in relation to diagnostic software but also the cost. Given that 43% of respondents have not adopted diagnostic software, yet a substantial proportion of respondents find such software to be valuable, there must be additional variables that this study failed to address.

The current study contained no findings that can explain this apparent paradox. It might be the case that physicians find diagnostic software useful in personal practice but they are not allowed to use it in certain practice settings. Alternatively, the costs associated with establishing a diagnostic software-capable system may be too great for smaller organizations. Whatever the case, scholars ought to investigate this paradox further; a study with a larger sample might find that the paradox does not exist at all and that physicians' opinions of the usefulness of diagnostic software are aligned with the actual adoption rate of diagnostic software.

A recommendation for physicians is to speak more frequently to their peers about both the benefits and disadvantages of diagnostic software. Roughly, 43% of participants in the current study were non adopters of diagnostic software; the 57% of physicians who were adopters to initiate a dialogue with non adopting peers and to share information could be instrumental in promoting diagnostic software adoption. Communication amongst medical professionals related to diagnostic software may also help to identify the most effective products available to suit their specific needs. In addition, the existence of such dialogue could help physicians who have already adopted diagnostic software to learn and adopt best practices in the use of such software from more advanced peers.

## Significance of the Study

Diagnostic software is a popular, cost-efficient, and clinically powerful healthcare tool (O'Malley et al., 2010; Puech et al., 2009; Randeree, 2007). The purpose of this quantitative, survey-based study was to draw upon physician-provided data to determine why, at least in physicians' opinions, the prevalence of misdiagnosis has remained high despite the widespread adoption of diagnostic software. It was found that physicians thought diagnostic software to be more capable of reducing misdiagnosis than unassisted human diagnosis, which can be interpreted as a general endorsement of diagnostic software. However, overall adoption of diagnostic software remained low. The study was significant because the results provided empirical insight to this apparent paradox, which might be explained through the variable of institutional support for diagnostic software. Additionally, I presented findings that called attention to the absence of gender-, specialty-, and age-related differences, which might suggest that physicians' attitudes to diagnostic software are related to core concerns of the profession rather than to demographic differences.

## Implications for Social Change

The social significance of the study lies in its affirmation of the usefulness of diagnostic software, at least in the opinion of physicians. Medical misdiagnosis is an enormously costly social problem in terms of lives lost and health compromised; according to Leavitt and Leavitt (2011), approximately 100,000 people die every year in the United States because of misdiagnosis. If physicians themselves think that diagnostic software is superior to unassisted human diagnosis, as was found in the current study, then there is additional support for the use of diagnostic software to reduce the incidence of misdiagnosis. If this study contributes, however modestly, to the increased adoption of diagnostic software, then it will have helped to address the various social problems—including problems of lost life, health, and productivity—caused by misdiagnosis. Thus, the main implication for social change in the study is its support for the broader use of diagnostic software by physicians, which might result in the lowering of misdiagnosis.

# References

AMA. (2012). *Doctor finder*. Retrieved from https://extapps.ama-assn.org/doctorfinder/

Adam, J. A. (2011). *Mathematics in nature*: *Modeling patterns in the natural world*. Princeton, NJ: Princeton University Press.

Adler, M. & Ziglio, E. (1996). *Gazing into the oracle*. New York, NY: Jessica Kingsley.

Agnihotri, V. K. (1995). *Public policy analysis and design*. New Delhi, India: Concept Publishing Company.

Ahlers, M. O., Jaeger, D., & Jakstat, H. A. (2010). Concept of computer-assisted clinical diagnostic documentation systems for the practice with the option of later scientific evaluations. *International Journal of Computers and Dentistry*, *13*(3), 233-250. http://ijcd.quintessenz.de/index.php

Ajzen, I. (2005). *Attitudes, personality, and behavior*. New York, NY: McGraw-Hill.

Alaofin, T. (2014). *Life Savers: How the use of medical diagnostic software, as a second opinion to your doctor's diagnosis, can save your life.* Morgan Hill, CA: Bookstand Publishing.

Arora, N., Weaver, K. E., Clayman, M. L., Oakley-Girvan, I., & Potosky, A. L. (2009). Physicians' decision-making style and psychosocial outcomes among cancer survivors. *Patient Education and Counseling, 77*(3), 404-412. doi:http://dx.doi.org/10.1016/j.pec.2009.10.004

Arora, P. (2010). Digital gods: The making of a medical fact for rural diagnostic software. *Information Society, 26*(1), 70-79.

Babu, R. (2011). *Discrete mathematics*. New Delhi, India: Pearson Education India.

Bath-Hextall, F., Lymn, J., & Knaggs, R. (2011). *The new prescriber: An integrated approach to medical and non-medical prescribing*. New York, NY: John Wiley & Sons.

Baumuller, S., Desbiolles, L., Leschka, S., Scheffel, H., Seifert, B., & Stolzmann, P. et al. (2009). Dual-source versus 64-section CT coronary angiography at lower heart rates: Comparison of accuracy and radiation dose. *Radiology, 253*, 56-64. doi:http://dx.doi.org/10.1148/radiol.2531090065

Bekkerman, R., Bilenko, M., & Langford, J. (2011). *Setting up machine learning: Parallel and distributed approaches*. Cambridge, England: Cambridge University Press.

Bernstein, D. S. (2011). *Matrix mathematics*: *Theory, fact, and formulas*. Princeton, NJ: Princeton University Press.

Betrouni, N., Dewalle, A. S., Lemaitre, L. Makni, N., Puech, P., & Villers, A. (2009). Computer-assisted diagnosis of prostate cancer using DCE-MIR data: Design, implementation, and preliminary results. *International Journal of Computer-Assisted Radiology and Surgery, 4*(1), 1-10. doi: 10.1007/s11548-008-0261-2

Bligh, J. (2009). Who can resist Foucault? *Journal of Medicine and Philosophy, 34*(4), 368-383. doi:10.1093/jmp/jhp028

Bond, W. F., Schwarz, L. M., Weaver, K. R., Levick, D., Giuliano, M., & Graber, M. L. (2011). Differential diagnosis generators: An evaluation of currently available computer programs. *Journal of General Internal Medicine, 27*(2), 213-219.

doi:10.1007/s11606-011-1804-8

Bowman, W. & Frega, A. L. (2012). *The Oxford handbook of philosophy in music education*. Oxford, England: Oxford University Press.

Bucknall, T. K. (2010). Medical error and decision making: Learning from the past and present in intensive care. *Australian Critical Care*, *23*(3), 150-156. doi:http://dx.doi.org/10.1016/j.aucc.2010.06.001

Bynum, W. F. (2008). *History of medicine: A very short introduction*. Oxford, England: Oxford University Press.

Byrne, J. P. (2012). *Encyclopedia of the Black Death*. New York, NY: ABC-CLIO.

Caperchione, C. M., Kolt, G. S., & Mummery, W. K. (2009). Physical activity in culturally and linguistically diverse migrant groups to Western society: A review of barriers, enablers, and experiences. *Sports Medicine, 39*(3), 167-177. doi:10.2165/00007256-200939030-00001

Capps, C., Dranove, D., & Lindrooth, R. C. (2010). Hospital closure and economic efficiency. *Journal of Health Economics, 29*(1), 87-109. doi:10.1016/j.jhealeco.2009.10.006

Carpenito-Moyet, L. J. (2008). *Nursing diagnosis: Application to clinical practice*. Amsterdam, Netherlands: Wolters Kluwer Health.

Centers for Disease Control. (2012). *Deaths and mortality*. Retrieved from *http://www.cdc.gov/nchs/fastats/deaths.htm*

Chapman, G. B. & Sonnenberg, F. A. (2003). *Decision making in healthcare: Theory, psychology, and applications*. Cambridge, England: Cambridge University Press.

Chernick, M. R. (2011). *The essentials of biostatistics for physicians, nurses, and clinicians*. New York, NY: John Wiley & Sons.

Chira, S.,David, C. V., Eells, S. J., Ladrigan, M., Miller, L. G., & Papier, A. et al. (2011). Diagnostic accuracy in patients admitted to hospitals with cellulitis. *Dermatology Online Journal*, *17*(3), 1. doi:http://dx.doi.org/10.1017/s0950268810001408

Chowdhury, S. R., Roy, A., & Saha, H. (2010). ASIC design of a digital fuzzy system on chip for medical diagnostic applications. *Journal of Medical Systems, 35*(2), 221-235. doi:10.1007/s10916-009-9359-5

Clavien, P. A. & Trotter, J.F. (2011). *Medical care of the liver transplant patient*. New York, NY: John Wiley & Sons.

Cleverly, W., Cleverly, J. O., & Song, P. H. (2010). *Essentials of healthcare finance*. New York, NY: Jones & Bartlett.

Cockburn, A., Cockburn, E., & Reyman, T. A. (1998). *Mummies, disease, and ancient cultures*. Cambridge, England: Cambridge University Press.

Cohen, B. H. (2007). *Explaining psychological statistics*. New York, NY: Wiley.

Creswell, J. W. (2009). *Research methods*. Thousand Oaks, CA: Sage.

Croskerry, P., Cosby, K. S., Schenkel, S. M., & Wears, R. L. (2008). *Patient safety in emergency medicine*. Philadelphia, PA: Lippincott, Williams, & Wilkins.

Cunha, B. A. (2011). *Antibiotic essentials*. Atlanta, GA: Jones & Bartlett.

DePalma, M. (2011). *Evidence-based interventional spine care*. New York, NY: Demos.

Dewberry, C. (2004). *Statistical methods for organizational research: Theory and practice*. New York, NY: Psychology Press.

Dimakou, S., Parkin, D., Devlin, & Appleby, J. (2009). Identifying the impact of government targets on waiting times in the NHS. *Healthcare Management Science, 12*(1), 1-10. doi:10.1007/s10729-008-9069-4

Dreiseitl, S. (2005). Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artificial Intelligence in Medicine, 33*(1), 25-30. doi:http://dx.doi.org/10.1016/j.artmed.2004.07.007

Duggan, W. R. (2005). *Strategic intuition in army planning*. Washington, D.C.: DIANE.

Eadie, L. H., Taylor, P., & Gibson, A. P. (2012). Recommendations for research design and reporting in computer-assisted diagnosis to facilitate meta-analysis. *Journal of Biomedical Informatics, 45*(2), 390-397. doi:http://dx.doi.org/10.1016/j.jbi.2011.07.009

Edlin, G. & Golanty, E. (2009). *Health and wellness*. Atlanta, GA: Jones & Bartlett Learning.

El Khatib, M. F. (2007). A diagnostic software tool for determination of complexity in respiratory pattern parameters. *Computers in Biology and Medicine*, *37*(10), 1522-1527. doi:http://dx.doi.org/10.1016/j.compbiomed.2007.01.014

Fadiman, A. (1998). *The spirit catches you and you fall down: A Hmong child, her American doctors, and the collision of two cultures*. New York, NY: Macmillan.

Felder, S. & Mayrhofer, T. (2011). *Medical decision making*. New York, NY: Springer.

Field, A. (2009). *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage.

First, M. B. & Tasman, A. (2011). *Clinical guide to the diagnosis and treatment of mental disorders*. New York, NY: John Wiley & Sons.

Flynn, D. & Van Schaik, P. (2003). *Non-medical influences upon medical decision-making and referral behavior*. Westport, CT: Greenwood Publishing Group.

Foster, C. (2010). *The Washington manual of medical therapeutics*. New York, NY: Lippincott.

Fujiyoshi, A., Kadowaki, T., Kadowaki, S., Miura, K., Ohkubo, T., & Sekikawa, A. et al. (2011). Comparability in coronary artery calcium scores on CT can between two community-based cohort studies. *International Journal of Cardiology, 149*(2), 244-245. doi:http://dx.doi.org/10.1016/j.ijcard.2011.02.015

Garnham, A. & Oakhill, J. (1994) *Thinking and reasoning*. London: Wiley-Blackwell.

Gifford, F. (2011). *Philosophy of medicine*. New York, NY: Elsevier.

Gigerenzer, G. & Gray, J. A. (2011). *Better doctors, better patients*. Cambridge, MA: The MIT Press.

Goldsmith, S. (2011). *Principles of healthcare management*. New York, NY: Jones & Bartlett.

Gorini, A. & Pravettoni, G. (2011). An overview of cognitive aspects implicated in medical decisions. *European Journal of Internal Medicine, 22*(6), 547-553. doi: http://dx.doi.org/10.1016/j.ejim.2011.06.008

Goyzsche, P. C. & Wulff, H. R. (2008). *Rational diagnosis and treatment: Evidence-based clinical decision-making*. New York, NY: John Wiley & Sons.

Graham, I. (2010). *In the workplace*. London, England: Evans Brothers.

Groopman, J. (2007). *How doctors think*. Boston, MA: Houghton Mifflin.

Haeberli, F. B. (2005). *Developing expert consensus guidelines*. Madison, WI: University of Madison-Wisconsin.

Hawe, E. (2011). *Compendium of health statistics*. London, England: Radcliffe Publishing.

Hazlehurst, B., Gorman, P. N., & McMullen, C. K. (2008). Distributed cognition: An alternative model of cognition in medical informatics. *International Journal of Medical Informatics, 77*(4), 226-234. doi:http://dx.doi.org/10.1016/j.ijmedinf.2007.04.008

Helmer, O. & Rescher, N. (1959). On the epidemiology of the inexact sciences. *Management Science, 6*(11), 25-52. doi: http://dx.doi.org/10.1287/mnsc.6.1.25

Herath, C. S. (2010). Motivation as a potential variable to explain farmers' behavioral change in agricultural technology adoption decisions. *Economy and Management, 3*, 62-71.

Hersen, M. & Thomas, J. C. (2006). *Comprehensive handbook of personality and psychopathology*. New York, NY: Wiley.

Hess, D., MacIntyre, N., & Mishoe, S. (2011). *Respiratory care: Principles and practice*. Atlanta, GA: Jones and Bartlett.

Hippocrates. (1849). *The genuine works of Hippocrates* (Trans. F. Adams). London, England: The Sydenham Society. (Original work published between the fifth and fourth centuries B.C.E.).

Hodler, J., Schulthess, G. K., & Zolikofer, C. L. (2011). *Diseases of the heart, chest, and breast*. New York, NY: Springer.

Hsiung, R. C. (2012). Adoption of electronic health records by medical specialty societies. *Journal of the American Medical Informatics Association, 19*, 143. doi:http://dx.doi.org/10.1136/amiajnl-2011-000593

Hughes, J. C. (2011). *Thinking through dementia*. Oxford, England: Oxford University Press.

Kaplan, D. (2004). *The Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks, CA: Sage.

Kattan, M. W. & Cowen, M. E. (2009). *Encyclopedia of medical decision-making: Volume 1*. Thousand Oaks, CA: Sage.

Keeney, S., McKenna, H., & Hasson, F. (2011). *The Delphi technique in nursing and health research*. New York, NY: John Wiley & Sons.

Keith, K. D. (2011). *Cross-cultural psychology: Contemporary themes and perspectives*. New York, NY: John Wiley & Sons.

Kelly, K. (2010). *The history of medicine*. Washington, D.C.: Facts on File.

Keeney, S., McKenna, H., & Hasson, F. (2011). *The Delphi Technique in nursing and health research*. New York, NY: Wiley.

Kennett, R. S. & Salini, S. (2011). *Modern analysis of customer surveys: With applications using R*. New York, NY: John Wiley & Sons.

Kramme, R., Hoffmann, K. P., & Pozos, R. S. (2011). *Springer handbook of medical technology*. New York, NY: Springer.

Krugman, P. & Wells, R. (2009). *Economics*. New York, NY: Worth Publishers.

Lapan, S. D., Quartaroli, M. T., & Riemer, F. J. (2011). *Qualitative research: An introduction to methods and designs*. New York, NY: John Wiley & Sons.

Large, R. D. (1975). *Delphi methodology*. Chicago, IL: Northwestern University Press.

Leavitt, J. & Leavitt, F. (2011). *Improving medical outcomes*. New York, NY: Rowman & Littlefield.

Lefrancois, G. (2011). *Theories of human learning*. New York, NY: Cengage.

Li, M. Y., Frieze, I., & Tang, C. S. K. (2010). Understanding adolescent peer sexual harassment and abuse: Using the Theory of Planned Behavior. *Sexual Abuse, 22*(2), 157-171. doi:10.1177/1079063210363827

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. London, England: Guilford Press.

Liu, J. C. (2011). *Liu: Dangerous delays in women's healthcare at some city hospitals*. Retrieved from http://www.comptroller.nyc.gov/press/2011_releases/pr11-05-041.shtm

Lock, M. & Nguyen, V. K. (2011). *An anthropology of biomedicine*. New York, NY: John Wiley & Sons.

Madhavji, N. H. (2006). *Software evolution and feedback*. New York, NY: John Wiley & Sons.

Mamede, S., Rikers, R., & Schmidt, H. G. (2012). The role of reflection in medical practice: Continuing professional development in medicine. *Innovation and Change in Professional Education, 7*, 163-174.

doi:10.1007/978-94-007-1724-4_8

Mankiw, N. G. (2011). *Principles of economics*. New York, NY: Cengage.

Martin, J., Perez, C. J., & Muller, P. (2009). Bayesian robustness for decision making problems. *International Journal of Approximate Reasoning, 50*(2), 315-323. doi:10.1016/j.ijar.2008.03.017

Martin, J. W. (2011). *Unexpected consequences: Why the things we trust fail*. New York, NY: ABC-CLIO.

McDermott, R. (2008). Medical decision making: Lessons from psychology. *Urologic Oncology: Seminars and Original Investigations, 26*(6), 665-668. doi:http://dx.doi.org/10.1016/j.urolonc.2007.12.006

McGann, P. J. & Hutson, D. J. (2011). *Sociology of diagnosis*. New York, NY: Emerald Group.

McPhee, S., Papadakis, M., & Rabow, M. W. (2011). *Current medical diagnosis and treatment 2012*. New York, NY: McGraw-Hill.

Mehl, M. R., Conner, T. S., & Csikzentmihalyi, M. (2011). *Handbook of research methods for studying daily life*. London, England: Guilford Press.

Menachemi, N., Matthews, M., Ford, E. W., Hikmet, N., & Brooks, R. (2009). The relationship between local hospital IT capabilities and physician EMR. *Journal of Medical Systems*, *33*(5), 329-335. doi:10.1007/s10916-008-9194-0

Meyers, R. A. (2011). *Mathematics of complexity and dynamical systems*. New York, NY: Springer.

Monahan, F. D., Neighbors, M., & Green, C. (2010). *Manual of medical–surgical nursing care*. New York, NY: Elsevier.

Montgomery, K. (2006). *How doctors think: Clinical judgment and the practice of medicine*. Oxford, England: Oxford University Press.

Moustakas, C. E. (1994). *Phenomenological research methods.* Thousand Oaks, CA: Sage.

Newborn, M. (2003). *Deep Blue: An artificial intelligence milestone*. New York, NY: Springer.

Nielsen, M. (2011). *Reinventing disco very: The new era of networked science*. Princeton, NJ: Princeton University Press.

Nuttall, D. & Rutt-Howard, J. (2011). *The textbook of non-medical prescribing*. New York, NY: John Wiley & Sons.

O'Guinn, T. C., Allen, C. T., & Semenik, R. J. (2011). *Advertising and integrated brand promotion*. New York, NY: Cengage.

O'Malley, A. S., Grossman, J. M., Cohen, G. R., Kemper, N. M., & Pham, H. (2010). Are electronic medical records helpful for care coordination? Experiences of physician practices. *Journal of General Internal Medicine, 25*(3), 177-185. doi:10.1007/s11606-009-1195-2

*Oxford English Dictionary Online*. (2011). Diagnosis. Retrieved from *http://www.oed.com/Entry/51836*.

Plessner, H., Betsch, C., & Betsch, T. (2008). *Intuition in judgment and decision making.* Toronto, Canada: CRC Press.

Poole, D. L. & Mackworth, A. L. (2010). *Artificial intelligence: Foundations of computational agents*. Cambridge, England: Cambridge University Press.

Queenan, C. C., Angst, C. M., & Devaraj, S. (2011). Doctor's orders—if they're electronic, do they improve satisfaction? A complements/substitutes perspective. *Journal of Operations Management*, *29*(7-8), 639-649. doi:10.1016/j.jom.2011.03.001

Ramnarayan, P., Winrow, A., Coren, M., Nanduri, V., Buchdahl, R., & Jacobs, B. et al. (2006). Diagnostic omission errors in acute pediatric practice: Impact of a reminder system on decision-making. *Medical Informatics Decision Making, 6,* 37.

Randeree, E. (2007). Exploring physician adoption of EMRs: A multi-case analysis. *Journal of Medical Systems, 31*, 489-496. doi:10.1007/s10916-007-9089-5

Rao, G. (2007). *Rational medical decision-making: A case-based approach*. New York, NY: McGraw-Hill Medical.

Reece, R. L. (2009). *Obama, doctors, and health reform: A doctor assesses the odds for success*. New York, NY: iUniverse.

Reiss, E., Shadomy, H. J., & Lyon, M. (2011). *Fundamental medical mycology*. New York, NY: John Wiley & Sons.

Reneman, M. F., Geertzen, J. H., Groothoff, J. W., & Brouwer, S. (2008). General and specific self-efficacy reports of patients with chronic low back pain: Are they related to performances in a functional capacity evaluation? *Journal of Occupational Rehabilitation, 18*(2), 183-189. doi:10.1007/s10926-008-9129-0

Renouard, P. V. (2010). *History of medicine*. San Francisco, CA: Nabu Press.

Renz, D. M., Bottcher, J., Diekmann, F., Poellinger, A., Maurer, M. H., & Pfeil, A. (2012). Detection and classification of contrast-enhancing masses by a fully automatic computer-assisted diagnosis system for breast MRI. *Journal of Magnetic Resonance Imaging, 35*(5), 1077-1088. doi:10.1002/jmri.23516

Robson, B. & Baek, O. K. (2009). *The engines of Hippocrates*. New York, NY: Wiley.

Ryan, C. A. (2010). *Reflective inquiry in the medical profession*. New York, NY: Springer.

Schanll, R., Velez, O., John, R.M., & Bakken, S. (2011). Psychometric evaluation of the attitudes toward handheld decision support software scale in a sample of nursing students. *Compputers, Informatics, Nursing, 29*(4), 251-255. doi:10.1097/ncn.0b013e3181f9dc1d

Schneirdjans, M. J., Schneirdjans, A. M., & Schneirdjans, D. (2007). *Outsourcing management information systems*. San Francisco, CA: Idea Group.

Schwab, A. P. (2008). Putting cognitive psychology to work: Improving decision-making in the medical encounter. *Social Science and Medicine, 67*(11), 1861-1869. doi:10.1016/j.socscimed.2008.09.005

Schwartz, A. & Bergus, G. (2008). *Medical decision making: A physician's guide*. Cambridge, England: Cambridge University Press.

Schweitzer, S. O. (2007). *Pharmaceutical economics and policy*. Oxford, England: Oxford University Press.

Scott, T. & Rundall, T.G. (2007). *Implementing an electronic medical record system*. New York, NY: Radcliffe Publishing.

Seeley, B. E. (2009). Introducing a computer-based electronic record: Perceptions of clinicians. *Urologic Nursing*, *29*(5), 329-352. doi:10.1097/01.ncn.0000360477.08025.80

Segal, I. & Shahar, Y. (2009). A distributed system for support and explanation of shared decision-making in the prenatal testing domain. *Journal of Biomedical Informatics*, *42*(2), 272-286. doi:http://dx.doi.org/10.1016/j.jbi.2008.09.004

Shah, J. Y. & Gardner, W. L. (2008). *Handbook of motivation science*. London, England: Guilford Press.

Shaw, R., Ramachandra, V., Lucas, N., & Robinson, N. (2011). *Management essentials for doctors*. Cambridge, England: Cambridge University Press.

Shealy, N. (2011). *Medical intuition*. Berkeley, CA: ARE Press.

Shield, R. R., Goldman, R. E., Anthony, D. A., Wang, N., Doyle, R. J., & Borkan, J. (2010). Gradual electronic health record implementation: New insights on physician and patient adaptation. *Annals of Family Medicine, 8*(4), 316-326. doi:10.1370/afm.1136

Shih, T. K. & Hung, J. C. (2007). *Future directions in distance learning and communication technology*. San Francisco, CA: Idea Group.

Simon, H. (1947). *Administrative behavior: A study of decision-making processes in administrative organization.* New York, NY: Macmillan.

Siemens. (2012). *Accessory solutions*. Retrieved from http://www.medical.siemens.com/webapp/wcs/stores/servlet/ProductDisplay~q_catalogId~e_11~a_catTree~e_100010,1012315,1014512,1017118,1014401,1027872~a_langId~e_-11~a_productId~e_193054~a_storeId~e_10001.htm

Simel, D. L. & Rennie, D. (2008). *The rational clinical examination: Evidence-based clinical diagnosis*. New York, NY: McGraw-Hill Medical.

Skinner, B. F. (1938). *The behavior of organisms*. New York, NY: Macmillan.

Skinner, J. (2011). Understanding prices and quantities in the U.S. healthcare system. *Journal of Health Politics, Policy, and Law, 36*(4), 791-801. doi: 10.1215/03616878-1302939

Sokolowski, J. A. & Banks, C. M. (2011). *Modeling and simulation in the medical and health sciences*. New York, NY: John Wiley.

Sox, H. C. & Higgins, M. C. (1988). *Medical decision making*. New York, NY: ACP Press.

Spekowius, G. & Wendler, T. (2006). *Advances in healthcare technology*. New York, SNY: Springer.

Stillman, M. (2010). Concierge medicine: A “regular” physician’s perspective. *Annals of Internal Medicine, 152*(6), 391-392. doi:10.7326/0003-4819-152-6-201003160-00009

Szafran, R. (2011). *Answering questions with statistics*. Thousand Oaks, CA: Sage.

Tashakkori, A. & Teddlie, C. (2010). *Sage handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.

Trautman, D. E. (2011). Healthcare reform: 1 year later. *Nursing Management, 42*(4), 26-31. doi:10.1097/01.numa.0000394955.71466.ef

Umscheid, C. A. & Hanson, C. W. (2011). A follow-up report card on computer-assisted diagnosis—the grade: C+. *Journal of General Internal Medicine, 27*(2), 142-144. doi:10.1007/s11606-011-1944-x

U.S. Department of Veterans Affairs. (2012). *DSM-IV-TR criteria for PTSD*. Retrieved from http://www.ptsd.va.gov/professional/pages/dsm-iv-tr-ptsd.asp

Van Grembergen, W. & De Haes, S. (2009). *Enterprise governance of information technology*. New York, NY: Springer.

Watanabe, Y., Nakazawa, T., Higashi, M., Itoh, T., & Naito, H. (2011). Assessment of calcified carotid plaque volume: Comparison of contrast-enhanced dual-energy CT angiography and native single-energy CT. *American Journal of Roentgenology, 196*(6), 796-799.

Windelband, W. (1913). *Encyclopedia of the philosophical sciences*. New York, NY: Macmillan.

Winfield, R. D. (2011). *The living mind: From psyche to consciousness*. New York, NY: Rowman & Littlefield.

Winkelman, M. (2008). *Culture and health: Applying medical anthropology*. New York, NY: John Wiley & Sons.

Woods, M. & Woods, M. B. (2000). *Ancient computing: From counting to calendars*. New York, NY: Twenty-First Century Books.

Yong, P. L., Saunders, R. S., & Olsen, L. A. (2010). *The healthcare imperative: Lowering costs and improving outcomes*. Washington, D.C.: National Academies Press.

Yoshihara, S., Sylva, D. A., & Eberstadt, N. (2011). *Population decline and the remaking of great power politics*. Washington, D.C.: Potomac Books.

Zilberberg, M. D. (2011). Evidence under judgment: Can we oversee our own decision-making? *Archives of Internal Medicine*, *171*(16), 1496-1497. doi:10.1001/archinternmed.2011.355

Appendix A: Diagnostic Software Survey Form for Study

Management of Diagnostic Software Survey Form for Study

1. To which of the following types of diagnostic medical software have you ever had access?

   Isabel

   DXplain

   Your Rapid Diagnosis

   DiagnosisPro

   Connectance

   Other____________________

   None (Please skip Questions 2 – 12)

2. Which of the following types of diagnostic medical software have you ever used?

   Isabel

   DXplain

   Your Rapid Diagnosis

   DiagnosisPro

   Connectance

   Other____________________

   None (Please skip Questions 3 – 12)

3. How long have you had access to diagnostic medical software?

0 - 6 months

6 months – 1 year

1 - 3 years

3 - 5 years

More than 5 years

4. How long have you used diagnostic medical software?

0 - 6 months

6 months – 1 year

1 - 3 years

3 - 5 years

More than 5 years

5. To which of the following types of diagnostic medical software do you currently have access?

Isabel

DXplain

Your Rapid Diagnosis

DiagnosisPro

Connectance

Other__________________

None

6. Which of the following types of diagnostic medical software do you currently use?

Isabel

DXplain

Your Rapid Diagnosis

DiagnosisPro

Connectance

Other____________________

None

7. How long have you been using the software you are currently using?

0 - 6 months

6 months – 1 year

1 - 3 years

3 - 5 years

More than 5 years

**(Questions 8 – 12 are individual in nature)**

8. My use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

9. I do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

10. My knowledge of diagnostic software is not extensive enough that I am able to use it effectively?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

11. My knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnosis?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

12. Liability concerns do not prevent me from using diagnostic software?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

**(Questions 13 – 16 are general in nature)**

13. In general, physicians' use of diagnostic software is likely to result in more misdiagnoses in healthcare than unassisted human diagnostic methods?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

14. In general, physicians do not use diagnostic software frequently enough to decrease misdiagnosis in healthcare?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

15. In general, physicians' knowledge of diagnostic software is not extensive enough to result in a decrease in misdiagnosis?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

16. In general, liability concerns do not prevent physicians from using diagnostic software?

Strongly Disagree

Disagree

Somewhat Disagree

Neither Agree nor Disagree

Somewhat Agree

Agree

Strongly Agree

Not Applicable

17 What is your gender?

Male

Female

Transgender

18. How old are you?

Less than 30 years old

30-40 years old

41-50 years old

51-60 years old

61-65 years old

More than 65 years old

19. How long have you had a license to practice medicine?

Less than 5 years

5-10 years

11-20 years

21-30 years

More than 30 years

20. What is your specialty area? ______________________________________

21. In what area of the U.S. do you practice medicine?

New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont)

Mid-Atlantic (New Jersey, New York, Pennsylvania)

South Atlantic (Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia)

South (Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Oklahoma, Tennessee, Texas)

Midwest (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin)

Mountain (Arizona, Colorado, Idaho, Montana, New Mexico, Nevada, Utah, Wyoming)

Pacific (Alaska, California, Hawaii, Oregon, Washington)