November 2021

Decisions and How Doctors Make Them: Modeling Multilevel Decision-Making within Diagnostic Medicine

Michelle S. Kaplan
University of South Florida

Decisions and How Doctors Make Them:

Modeling Multilevel Decision-Making within Diagnostic Medicine

by

Michelle S. Kaplan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology
College of Arts and Sciences
University of South Florida

Co-Major Professor: Michael T. Brannick, Ph.D.
Co-Major Professor: Sandra Schneider, Ph.D.
Michael D. Coovert, Ph.D.
Walter C. Borman, Ph.D.
Jason Beckstead, Ph.D.
Michael T. Braun, Ph.D.

Date of Approval:
October 28, 2021

Keywords: Choice, Medical, Diagnosis, Judgment

Table of Contents

List of Tables

List of Figures

Abstract

Effective decision-making is critical and necessary for organizational success across a wide range of occupations, situations, and industries. However, decision making, by its nature, is not always a direct process of a single decision leading to a direct outcome. Rather, it can often become a multilevel process whereby one decision's outcome leads to information that is used in subsequent larger or other types of decisions. The decision-making process then becomes progressively more complex and more difficult to navigate as these decisions compound within one another. Thus, decision-makers must find an appropriate way to approach such decisions. Understanding the multilevel nature of decision-making and how to optimize the final solution can have implications across a variety of areas. This dissertation aims to address those multilevel decisions in diagnostic medicine where the decision requires the assessment of multiple informational inputs. A psychometric approach was taken to look at different models pertaining to how these decisions can be made with the greatest degree of classification accuracy. Ultimately, tree-based models outperformed all other methods and were found to have the most applicability to diagnostic medicine. While some constraints related to tree-based modeling are noted, examples are shown to discuss how these models can be used to enrich current medical approaches. Possible implications for future research are examined.

Chapter 1: Introduction

Effective decision-making is critical and necessary for organizational success across a wide range of occupations, situations, and industries (Highhouse, Dalal, & Salas, 2013; Simon, 1987). Understanding the processes implicit to decision-making can impact professions and the professionals within them from investors to teachers to doctors. Across the spectrum of jobs, job types, and job tasks, individuals are constantly put into a position whereby they will have to make large and impactful decisions that can affect the work and lives of themselves and countless others. Individual, team, and organizational success relies on individuals making effective decisions. Such decisions cut across organizational domains including selection, training, and leadership to impact any number of individuals within those domains (Highhouse et al., 2013).

Decisions within organizational psychology have a number of common characteristics, including having elements of uncertainty, risk, and overall ambiguity (Kahneman & Tversly, 1980; Markowitz, 1952; Rode, Cosmides, Hell, & Tooby, 1999; Tversky & Kahneman, 1979; Tversky & Kahneman, 1980). In this manuscript, the terms "decisions" and "choices" will be used interchangeably. When individuals make choices, each choice is associated with some outcome or outcomes. However, given the stochastic nature of decision-making, there is a level of uncertainty that accompanies the decision process. This is evidenced in key areas of decision making research, including prospect theory (Kahneman & Tversky, 1979), cumulative prospect theory (Tversky & Kahneman, 1992), portfolio theory (Markowitz, 1952), signal detection theory (Green & Swets, 1966), and field theory (Busemeyer & Townsend, 1993).

Likewise, decisions carry a certain degree of risk when outcomes have varying possibilities of occurrence across the spectrum of good and bad, right and wrong, optimal and suboptimal. This spectrum of "correct" decision-making creates a risk for the decision-maker of potentially making an erroneous or otherwise flawed choice. For example, Rode and colleagues (1999) show how factors such as uncertainty and risk lead to decision ambiguity, whereby individuals' ability to correctly select an option is impacted. This too is evidenced across a number of common decision-making theories, including those listed above. Uncertainty and risk impact the decision-making process and lead to individuals making biased choices (Kahneman & Tversky, 1979; Markowitz, 1952; Rode et al., 1999).

Because individuals are not optimal when compared to analytical decision models in situations with high stakes (i.e., risk), uncertainty (i.e., probabilistic information; Atkinson, 1957; Keeney & Raiffa, 1976), or ambiguity regarding what is best or important (Kahneman & Tversky, 1980; Tversky & Kahneman, 1992), it is crucial to learn more about different models of decision-making to determine how to improve decision-making processes and ultimately yield aids to make the process more effective. Given that conditions of risk, uncertainty, and ambiguity are ones that people across disciplines frequently face, it is unlikely that individuals will make all the correct decisions necessary to consistently reach the optimal conclusion. These issues can then compound and exponentially increase upon one another when multiple decisions are involved in an overarching process.

This is because decision making, by its nature, is not always a direct process of a single decision leading to a direct outcome, but is rather often a multilevel process whereby one decision’s outcome leads to information that is used in subsequent larger or other types of decisions. Thus, the decision-making process becomes progressively more complex and more difficult to navigate as these decisions compound within one another. Subsequently, decision-makers facing such decisions must find an appropriate way to approach these choices. There are numerous ways to model these decision processes, some better than others. And it is important to try to determine which of these various models of the decision-making process provide optimal versus suboptimal decisions overall.

When selecting models for decision-making, it is important to optimize correct decisions under conditions of uncertainty (Shirangi, Mehrdad, & Durlofsky, 2016). Optimal models will predict the correct decision with the highest degree of frequency over numerous scenarios whereas suboptimal models will predict the correct decisions with lower levels of frequency. However, it is also important for models to be relatively usable and not overwhelming in their level of complexity. As such, the most favorable models will be able to predict optimally without being overly complicated, balancing goodness of fit relative to optimal decision-making with overall simplicity (Burnham & Anderson, 2002).

One way in which these large and critical decisions can be approached is by breaking them down into a number of constituent components (Simon, 1959), which range in type from situational cues to other, smaller decisions. In the latter case, a multitude of small decisions culminate in a final decision that dictates what action(s) the individual takes based on the available information from their prior choices. When this approach is taken, it is critical that the decisions made at each step of the process are ones that maximize the likelihood of reaching the correct conclusion.

Fortunately, recent work shows that it is possible to make large decisions correctly even when many of their constituent smaller decisions are made incorrectly (Braun & Kaplan, 2017). Thus, while it may seem more intuitive to take the approach of maximizing optimal decision-making across all of the smaller decisions, these results highlight the importance of alternative approaches. Most importantly, it raises the need to identify correct patterns of decision-making across the smaller decisions that will ultimately lead to the most effective final solution.

Understanding the multilevel nature of decision-making and how to optimize the final solution can have implications across a variety of areas. For example, individuals must choose what career they want to pursue. In this decision process, many smaller choices must be made regarding what career to select: which field or area is of long-term interest, what level of education is the individual willing to pursue, what level of compensation do they want to receive, etc. Alternatively, when organizations choose to hire a new employee, they must make decisions such as: what level of education do they want a potential recruit to have, what personality traits are important to them, what degree of importance do they place on experience in the field, and many other factors (Hammond, Keeney, & Raiffa, 2015; Keeney & Raiffa, 1976; Keeney & Raiffa, 1993). These combinatory decision processes, that include numerous small decisions comprising the overarching decision made, can be demonstrated across areas and avenues wherein individuals are constantly under the burden of making difficult, oftentimes life-altering decisions with little understanding of how best to make these decisions.

In this research, I examined various models of these decision processes in the context of medical diagnosis with the aim of providing a framework on which preliminary diagnosis can ultimately be improved. The medical practice of diagnosis is a prime example of how large-scale decisions can be broken down into smaller decision units. When a doctor examines a patient, there are a multitude of small but critical decisions regarding what questions to ask during the patient interview, as well as what actions to perform and tests to administer during the patient physical examination. The collection of decisions made during the physical exam and patient interview impact the type of information they receive from the patient. This, in turn, will affect the ultimate diagnosis and proposed treatment plan. As such, it is important to understand common decision-making patterns that doctors employ during this process and how those patterns relate to successful patient diagnosis and treatment. Doing so not only allows for the development of potential training interventions to improve doctor effectiveness but also provides a roadmap for understanding how other sets of decisions within organizations relate to differentially effective outcomes.

This dissertation aimed to address those multilevel decisions whose complexity and magnitude requires the assessment of multiple informational inputs stemming from smaller decisions subsumed within larger ones. First, I reviewed the current literature regarding decision-making, with an emphasis on how it is currently being researched from the multilevel perspective. Then, this study explored various methods for improving multilevel decision-making paradigms across the literature. For this research, archival data collected from the Morsani College of Medicine at USF was used to understand the multilevel decision-making process within the context of medicine. I utilized a psychometric approach to look at different models pertaining to how these macro-level decisions can be made with the greatest degree of classification accuracy. Throughout the manuscript, special attention is given to the implications of optimal multilevel decision-making processes on organizational outcomes, both in the medical application being utilized in this research and beyond to other possible applications.

Decision-Making in Organizational Sciences

Understanding how individuals make decisions has been a key area of cognitive psychology, business, philosophy, and economics for decades (Edwards, 1954; Kahneman, 2011; Kahneman & Tversky, 1980; Orasanu, Calderwood, & Zsambok, 1993; Orasanu & Connolly, 1993; Simon, 1959; Simon, 1987; Yates, 1990). These different areas have looked across the various avenues relating to what decisions individuals are making and how they were made. A plethora of research has been conducted studying various facets of decision-making from prospects to risky choices (Kahneman & Tversky, 1980; Markowitz, 1952; Rode et al., 1999). And yet, despite the research in decision-making that exists across so many disciplines, it is still incredibly sparse within organizational psychology. And as this literature continues to be relatively unknown to organizational psychologists, many have called for greater communication between disciplines both historically and more currently (Dalal, Bonaccio, Highhouse, Ilgen, Mohammed, & Slaughter, 2010; Edwards, 1954). The research that does exist on decision-making within the context of the organizational sciences often focuses on practices such as selection, which the current study will not explore in depth due to the lack of research into the dynamic and multilevel components of these decisions.

As a result of this limited knowledge base, many simple questions about decision-making in the organizational context have yet to be answered. Such questions include when and how people make decisions, in which contexts, utilizing which mechanisms, and relying on which information. When making decisions, individuals may need to sift through enormous amounts of information and use a wide variety of variables and tools to guide them toward the decision they ultimately make (Plous, 1993). And while some basic research has been done on singular decisions, very little is known about how individuals undergo this process of multilevel decision-making in organizational contexts. There is little information about which factors are used, when they are used, and how best to use them.

However, in other domains such as cognitive psychology, business, philosophy, and economics, research has examined how decisions are affected by constantly changing conditions, how individuals reactions to those changes, and what influences the environment can have (Brehmer, 1992; Orasanu, Calderwood, & Zsambok, 1993; Orasanu & Connolly, 1993). When individuals make decisions, they weigh their alternatives, the consequences of choosing each, and make a choice based on some set of goals, purposes, or values (Orasanu & Connolly, 1993). These other domains have conducted research to model the overall decision process and understand approaches for making optimal decisions. Research into decision-making often looks at ways that individuals pull available information together to utilize in making their choices (Orasanu & Connolly, 1993). As these types of dynamic conditions and information utilization are key in the organizational context, understanding this type of research from other domains is necessary to build a cross-disciplinary bridge into organizational research.

In the context of organizations, individuals are constantly placed in the position wherein they have to make these types of decisions. Individuals are often given tasks to complete at work that are inundated with numerous components. When this occurs, they will source information from a number of different places and people (Gray & Meister, 2004). How these choices are made often context dependent, and yet is overwhelmingly done in the purpose of achieving a larger goal (Gray & Meister, 2004; Phelps, Heidl, & Wadhwa, 2012). However, because these processes have had so little study in the organizational context, we do not yet know how individuals approach these processes or what the optimal method of making these choices is. But, by gaining a greater understanding of how individuals choose to approach varying situations, and the smaller decisions they make en route to these larger choices, we can better learn how to help individuals optimize their decision processes so that they more often make the correct large-scale decision through the overarching multilevel process.

Multilevel Decision-Making

Multilevel theory looks to understand those phenomena that share variables with complex relationships across different levels of understanding and analysis (Klein & Kozlowski, 2000; Klein, Tosi, & Cannella Jr., 1999; Kozlowski & Klein, 2000; Xu, 1989). In multilevel research, variables exist in a hierarchy, where variables exist within the context of other variables. In the common organizational example, the organization is split into the macro level (i.e. the organization), the meso level (i.e. the group), and the micro level (i.e. the individual) with variables that can be studied across each level (Kozlowski & Klein, 2000). Understanding multilevel phenomena is crucial across various disciplines and subject matters to gain a deeper knowledge of how processes occur (Klein & Kozlowski, 2000; Kozlowski & Klein, 2000). Looking at organizational issues from the multilevel perspective not only adds nuance to the overarching understanding of those issues, but also lends itself to a better understanding of overarching patterns and behaviors that underlie those issues.

In the organizational context, the idea of multilevel decision-making generally refers to the more common ideas in multilevel theory: studying how the decisions of a few people culminate in larger decisions for a group or organization. However, this form of inter-individual multilevel decision-making is beyond the scope of the current research. Rather, in this study, I focused on the phenomenon of intra-individual multilevel decision-making. Specifically, I looked at how the decisions that an individual makes impact other decisions that they themselves will be making, or intra-individual decision-making. The question examined in this research examines how the information collected from micro-level decisions contributes to the overall efficacy of the intra-individual multilevel decision-making process.

Although research into multilevel theory has shown that organizational behavioral patterns integrate into various processes and effects that can transcend singular levels of study (going beyond only individuals to also include the groups and organization they make up), it is not often applied to the context of individuals. And yet, the individual's actions can also be separated into multiple levels of analysis wherein some variables surrounding a singular individual's actions can share complex multi-layered relationships with one another. Multilevel theory specifies that organizational phenomena exist at multiple levels of analysis and unfold over time (Kozlowski & Klein, 2000), and the decisions facing organizational workers also encounter these same phenomena.

In the context of decision making, multilevel theory can be applied to the study of how complex, large-scale (or macro-level) decisions are made through a process of compiling information from constituent simple, small-scale (or micro-level) decisions into an ultimate choice. In such macro-level decisions, there are key micro-level decisions that yield relevant information. These micro-level decisions focus on what kind of information should be sought and what lines of inquiry should be pursued. Throughout the multilevel decision-making process, these micro-level decisions are consistently made producing information that is ultimately used in the macro-level decision. This sequence of micro-level decisions represents the first level of the decision model, culminating in the second level, macro-level decision.

Within organizations, key macro-level decisions (e.g., who to hire; what technology to adopt) can be decomposed into a multitude of micro-level decisions which inform one or more of the choices from the macro-level decision. For example, when deciding amongst job candidates for an open position (i.e., the macro-level decision), human resource employees must make decisions about each candidate (i.e., micro-level decisions) such as which, if any, references to contact and how to weigh previous work experience. Moreover, there is a question of how much information is needed from the micro-level decisions to ultimately make the macro-level decisions, such as how many references to contact, how many questions to ask during the interview, and to what extent social media should be utilized. Additionally, there may be uncertainty in the information received, even though the micro-level choice to get the information was the correct one. For example, the applicant may have been fired for theft, and the correct decision is made to ask a reference why the applicant left their last job. However, the reference lies, saying that there was no problem. Similarly, a patient may complain about pain that has nothing to do with the diagnosis, or may misremember the sequence and timing of events important for the diagnosis. In situations such as this, the decision characteristics of uncertainty, risk, and ambiguity originate in the micro-level decisions and combine to manifest at the macro-level. As such, the ability to make the best macro-level decision is inextricably linked to making correct micro-level decisions.

Multilevel decision-making is a complex process with a multitude of potential ways in which micro-level decisions can make up macro-level decisions. There are nearly limitless possibilities for the number of macro-level choices, how many micro-level decisions are related to each macro-level choice, how the micro-level decisions are distributed across choices, and how the micro-level decisions are nested under macro-level choices. Likewise, there are a plethora of ways in which uncertainty, risk, and ambiguity can color the micro-level decisions. Completely crossing the variety of macro-level decision structures with the multitude of micro-level decision properties creates macro-level decision types too numerous to describe or hope to study with traditional means. With the wide proliferation of methods for understanding decisions and how they are made, it is important to use techniques that can incorporate numerous criteria across levels to increase overall decision utility (Triantaphyllou, 2000).

Applications of Multilevel Decision-Making Across Disciplines

While inside the organizational sciences, research into multilevel decision-making has been scarce– outside of them, research has been extensive. Everything from portfolio theory in economics to signal detection theory in biology to field theory in cognitive psychology has looked into ways of cataloging and understanding multilevel decision-making structures (Busemeyer & Townsend, 1993; Fernholz, 2002; Green & Swets, 1966; Markowitz, 1952; Shefrin & Statman, 2000). These various archetypes of multilevel decision-making processes and how to understand them can build a basis from which to expand into the organizational sciences. From early research in decision making (Edwards, 1954) to more modern work on the topic (Dalal et al., 2010), there have been calls for greater communication between disciplines to facilitate this greater understanding of the ubiquitous processes that span across them. But despite this wealth of literature on multilevel decision-making in other areas, the organizational sciences have yet to embrace and study these concepts.

Signal Detection Theory

In biology, signal detection theory looks at how a number of pieces of information can be utilized within a system in order to detect a signal (DeCarlo, 1998; Green & Swets, 1966; McNicol, 2005; Pastore & Sheirer, 1974; Wickens, 2002). Signal detection looks at how smaller decisions about the system, in its constituent components and as a whole, can be utilized to make the larger decision regarding the presence or absence of a signal. In this traditional sense of signal detection, receiver operating characteristic (ROC) curves can be used to separate variables that can and cannot be sensed in the visual field (Green & Swets, 1966; Spackman, 1989; Swets, 1996). The curves illustrate the ability of an individual to correctly classify a signal as present or absent based on various underlying criteria. These curves are utilized to find the point at which optimal classification decisions are made. While signal detection is not in itself a traditional multi-level decision process, it utilizes a similar procedure wherein multiple pieces of information are gathered at one level and ultimately used to make a larger judgment at another level.

When thinking about applying these ideas to people, cognitive psychologists look at individuals as active decision-makers who use complex analysis of perceptual information to make decisions and choices under conditions of uncertainty. They utilize characteristics such as experience, expectations, physiological state and a number of other factors that can affect the signal detection threshold (Banks, 1970; Lockhart & Murdock, 1970; Stanislaw & Todorov, 1999; Swets, 1996). And while commonly applied to the areas of visual perception, this can also be applied to areas of cognition in how people execute complex choices. It has also been applied to the realm of medical decision-making, wherein doctors often have to determine whether or not an issue exists within a patient and what that issue is (Lusted, 1971; Swets, 1996). For example, radiologists must view images and correctly sense and further utilize that information to make a diagnosis (Lusted, 1971).

Understanding these applications of signal detection theory provides a background from which we can build a conceptualization of intra-individual multilevel decision-making and its ability to be applied to medical decisions. When doctors are confronted with a new patient, they need to make micro-level decisions about what portions of the patient's symptoms to attend to, how much to attend to them, and to what degree each symptom constitutes a pertinent issue.

They then must use the subsequent information, weighing the relative importance of each component piece of information gathered from the micro-level decisions, to make the macro-level decision regarding what preliminary diagnosis to give the patient.

Portfolio Theory

Another exemplification of a multilevel decision-making process that can give insight into the decision-making that individuals perform, portfolio theory looks at the process of selecting stocks into overall stock portfolios (Constantinides & Malliaris, 1995; Elton & Gruber, 1997; Markowitz, 1952; Markowitz, 1991; Markowitz, 2010). In portfolio theory, a mathematical framework is created to assemble stocks into a larger stock portfolio, attempting to maximize expected return on investment (ROI) while minimizing potential risks (Markowitz, 1952). When individuals choose to invest in a stock portfolio, micro-level decisions must be made within the context of this mathematical framework to determine which stocks to select into which portfolios, and the information gathered regarding risk vs. ROI from each of these stocks is utilized when appraising which portfolio an individual should select in the overall macro-level decision. Similar to the ROC in signal detection theory, portfolio theory utilizes the capital allocation line (CAL) to determine the optimal portfolio for an individual to select (Arnold, 2002; Sharpe, 1970). This line illustrates the amount of risk of choosing a given portfolio compared to the potential ROI. CALs can thus be used to determine optimal investment decisions.

When applying this theory to individuals and their ability to make decisions, it is important to underscore how many different aspects of various factors can account for how individuals make the micro-level decisions regarding the stocks placed within portfolios and later macro-level decisions regarding the portfolios chosen to invest in. These decisions are laced with uncertainty given a constant and shifting stock market, changes in behavioral attitudes over time, and the potential for unanticipated external events affecting both the risk and ROI of any given portfolio or stock (Elton & Gruber, 1997; Fernholz, 2002; Shefrin & Statman, 2000). These decisions are dynamic and constantly evolving to suit the needs of the individuals making the choices as well as the overall environment in which the choices are made (Elton & Gruber, 1997; Fernholz, 2002). As such, individuals need to be able to take into account each aspect of information and make appropriate decisions to optimize not only how they build a portfolio but also which portfolios they may choose to invest in. Any given stock exists in uncertainty and individuals need to be able to make decisions that support the best portfolio, rather than focusing on any given stock (Markowitz, 1952; Shefrin & Statman, 2000).

Knowing that the macro-level decision should be the overwhelming priority in multilevel decision-making provides a basis of how information should be sorted in the process. While micro-level decisions have value and are important to the overall effectiveness of the process, the focus should be on making the best possible macro-level decision. When applied to the medical decision-making context then, it is important to recognize that making the correct diagnosis is the most crucial element and should supersede the need to make every correct micro-level decision regarding which tests to run and which questions to ask.

Decision Field Theory

And in yet another discipline that has looked at multilevel decision-making, cognitive psychologists have studied decision field theory (Busemeyer & Townsend, 1993) wherein decisions are made in dynamic and uncertain environments looking at the overall deliberative process that individuals undergo while making an ultimate choice (Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993; Roe, Busemeyer, & Townsend, 2001). Decision field theory assumes that decision makers evaluate any given option relative to the other available alternatives. They then deliberate about their options and make micro-level (or attribute-wise) comparisons between the similarities and differences of their overall options. In this theory, deliberation is seen as a process that includes understanding and sorting through large amounts of information while weighing various consequences of the different attributes and components involved with each option. (Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993; Roe et al., 2001). The probability of any given choice being made is mapped against the amount of time and information needed to make the decision where probabilities of a given event occurring are built from an individual's past experience with similar events (Busemeyer & Townsend, 1992; Busemeyer & Townsend, 1993). Ultimately, decision field theory attempts to make predictions about the cognitive processes and various underlying components of a given choice.

Decision field theory looks at decisions in the context of a connected network comprised of multialternative preferential choices (Busemeyer & Diederich, 2002; Roe et al., 2001). It looks at the choice principles that exist across dynamic decision process, including consideration for the possibility of irrelevant and unnecessary alternatives (Busemeyer & Diederich, 2002). Specifically, decision field theory looks at how an individual's preferences change and evolve over time through a stochastic process of diffusion until they eventually reach a decision. The information that goes into a decision constantly shifts depending on time pressures, choice contexts, and relative uncertainty (Busemeyer & Townsend, 1992; Busemeyer & Townsend, 1993). While this theory primarily focuses on how decisions can fluctuate with time (with preferences and probabilities shifting across time) many of its core concepts can be applied beyond this domain.

Multilevel decision-making exists in a stochastic framework with micro-level decisions constantly altering information flow and created new and different environments within which the macro-level decisions are made. With a situation that changes rapidly depending on how micro-level decisions are made, it is crucial to understand how this dynamic flow of information can influence overall choice. Within medical decisions, doctors must constantly evolve their approach based on information as it is gathered from patients, accounting for ever-changing variables influencing their ultimate diagnosis.

Medical Decisions

There is much contention about the best way to make medical diagnostic decisions (Hunink, Weinstein, Wittenberg, Drummond, Pliskin, Wong, & Glasziou, 2014). However, understanding these overarching theories of multilevel decision-making across other domains lends a perspective to how they can be studied in a medical context. It provides a background regarding which components are involved and how the process operates. This helps create a more theory-oriented approach for the diagnostic decision process, accounting for its inherent dynamic and multilevel nature. However, the approaches that currently exist within medicine do not take these cross-disciplinary ideas into account and thus utilize an approach that is unlikely to be optimal (Braun & Kaplan, 2017; Mamede, Schmidt, & Rikers, 2007).

Medical Decision Aids in Practice

There are over 1,000 medical practice guidelines and tools for doctors to use when making judgments in their clinical practice (Weingarten, Riedinger, Conner, Johnson, & Ellrodt, 1994). However, the complexity of implementing any existing guideline creates difficulties in practice for doctors to actually use them (Cohen & Kataoka-Yahiro, 2009; Henriksen & Brady, 2013; Weingarten et al., 1994). The diagnostic performance of such tools often is dependent on the contexts in which they are utilized and how appropriately they are handled (Boussadi, Caruba, Karras, Berdot, Degoulet, Duriex, & Sabatier, 2011; Henriksen & Brady, 2013; Weingarten et al., 1994).

When these tools are properly utilized, the overall diagnostic performance is improved and individuals make fewer errors than the trained medical professionals who do not utilize them (Boussadi et al., 2011; Henriksen & Brady, 2013; Hero, Gerhards, Thiart, Hellhammer, & Linden, 2012; Novis, Zarbo, & Valenstein, 1999). Unfortunately, due to the complexity of implementing these types of tools, they are commonly left unused or underutilized (Boussadi, et al., 2011; Henriksen & Brady, 2013). Additionally, most diagnostic aids currently in practice are specifically targeted toward one specialty within medicine rather than being made for general diagnostic use (Boussadi et al., 2011; Henriksen & Brady, 2013; Hero et al, 2012; Novis, Zarbo, & Valenstein, 1999; O'Connor, Tugwell, Wells, Elmslie, Jolly, Hollingworth, … & Mackenzie, 1998; Schroy, Emmons, Peters, Glick, Robinson, Lydotes, … & Prout, 2011). Moreover, the majority of these aids are aimed primarily for patient use rather than doctor use, focusing on how to involve patients in the diagnostic process (Dolan & Frisina, 2002; Elwyn, Laitner, Coulter, Walker, Watson, & Thomson, 2010; Hamann, Leucht, & Kissling, 2003; Moulton & King, 2010; Peele, Siminoff, & Ravdin, 2005).

Even those aids that do exist for doctors often rely on either solely theoretical models of decision-making or otherwise use statistical models that make it difficult to extrapolate preliminary diagnoses (Elwyn, Frosch, Thomson, Joseph-Williams, Lloyd, Kinnersley, & Edwards, 2012; Heald, Kim, Sischo, Cooper, & Wolfgang, 2002; Smith, Doctor, Meyer, Kalet, & Phillips, 2009). Oftentimes, these tools simply provide a step-by-step theoretical guide for how to address patients and ask them questions to best elucidate answers from them (Elwyn et al., 2012; Weingarten et al., 1994). However, the theoretical guides suffer from the issue of focusing on helping doctors make the correct micro-level decisions (e.g. knowing which questions to ask, understanding which tests to run) rather than helping doctors make the correct macro-level decision (i.e. the diagnosis).

In other cases, techniques such as computational models, neural networks, and Bayesian networks are used (Heald et al., 2000; Smith et al., 2009). However, the computational modeling techniques that are currently utilized fail to provide more general support for diagnoses. Rather, these types of models focus primarily on one specific type or class of illness. In the case of Bayesian network modeling, which could be utilized in different contexts, the preliminary diagnosis still suffers. This is because in Bayesian techniques, the function of the model is to continually update and change as more information becomes available to it. And while this would help the diagnostic process overall, it would not aid doctors in the process of developing a preliminary diagnosis.

As such, when new guidelines and tools are made to assist doctors with diagnostics as well as other areas of their practice, it is important to consider if and how they can actually be utilized. The current research, unlike much of the other literature in the medical decision aid realm, hopes to ultimately improve preliminary diagnostic decision-making that doctors undergo. In this study, the emphasis was on modeling the decision process and attempting to extrapolate which models provide greater classification accuracy of diagnostic decisions in the overarching medical context. By modeling this decision process and gaining an understanding of how these diagnostic decisions can be made with greater classification accuracy, the current research aims to take the first step toward ultimately improving preliminary diagnostic accuracy.

Potential Models of Medical Decision Processes

In practice, individuals favor the dominance model which states that: the more correct micro-level decisions you make, the more likely you are to make a correct macro-level decision based on the information gathered from the micro-level decisions (Elwyn et al., 2012; Weingarten et al., 1994). In the medical context, these micro-level decisions would be made up of each choice to ask a particular question of a patient or to run a particular test. The USF Medical School currently trains students to gain the most information possible during a patient exchange. This translates to medical students being instructed to ask more questions and run more tests under the assumptions that more information at the micro-level will lead to better decision-making at the macro-level. Students are then graded on their ability not only to make correct macro-level decisions, but also correct micro-level decisions.

However, the literature and various computational models that have been previously built show us that a whole is not necessarily equal to the sum of its parts (Braun & Kaplan, 2017; Dillemuth, 2009; Kubovy & Van Den Berg, 2008). Macro-level decisions made based on information collected from correct micro-level decisions will not necessarily be correct. Each micro-level decision is made under conditions of uncertainty where the potential outcomes and information gathered cannot be known in advance. As previously discussed, this dynamic interchange of information gained from each micro-level decision changes the nature of how the macro-level decision can be made and from which information it is made. This dynamic nature of the multilevel decision-making process means that more information is not necessarily better. Rather, quality of information is likely to be more important than overall quantity of information gained.

Increasing the amount of information gained can have many outcomes, not necessarily only positive ones. By increasing the amount of information gained, the likelihood that decision-makers will gain incorrect information increases as well. This then, in turn, increases the likelihood of an incorrect macro-level decision being made. This applies even in the case where all micro-level decisions are made correctly (e.g. all pertinent questions are asked during patient history, all appropriate examinations are conducted) because simply making all of the correct micro-level decisions will not necessarily lead to correct macro-level choices. With the addition of more correct micro-level decisions, there becomes an issue of too much information being proliferated throughout the decision-making process, making it more difficult to locate the most important facts. Thus, it is unlikely and implausible to presume that sheer quantity of correct micro-level decisions would be able to accurately predict whether or not an individual will make a correct macro-level decision. Rather than predicting that the relations between quantity of information and the decision quality (outcome) is zero, and then attempting to affirm the null hypothesis, I assessed whether the relations are small enough to be considered inferior to other approaches.

A more appropriate method of approaching these multilevel decisions would be to look at an approach that values quality of information over quantity of information. The key is in trying to understand which information is the most helpful and which micro-level decisions are the most important to elucidating this information. If we can understand which specific pieces of information are the most crucial to the given macro-level decision and which micro-level decisions result in the discovery of that information, then we should be able to better predict how these decisions can best be made. Thus, approaches based on quality of information are likely to perform better than dominance model approaches.

When focusing on quality of information, it is important to understand which micro-level decisions are the most important and how they can be most optimally grouped together to yield correct macro-level decisions. These micro-level decisions can thus be considered subgroups that represent stable patterns of small yes/no decisions made by participants during the patient interview and physical examination. Their final medical diagnosis will serve as the ultimate judge of success or failure.

One possible method of locating the key micro-level decisions that feed into making the correct macro-level decisions is k-means cluster analysis. This allows for the ability to separate individuals into groups on the basis of which micro-level decisions they made. Because all decision-makers follow different pathways toward making decisions and can utilize a variety of types of information, it is likely that creating these clusters of differing types of decision makers can show which types of decision makers are the most likely to make correct macro-level decisions.

With k-means cluster analysis, data points are brought together into groups based on relative similarity to one another with the goal being to minimize the distance from each data point to the overall cluster. However, there are other techniques that cluster data in more refined ways. Latent Class Analysis (Muthén, 2004) is an approach that can be used to identify distinct patterns of decisions in a sample. LCA is an approach that utilizes categorical responses across a number of items/stimuli to infer underlying subgroups based on the observed patterns (Muthén, 2004). Applied to the current situation, these subgroups that LCA can create represent stable patterns of micro-level decisions made by participants during the patient interview and physical examination with their final medical diagnosis serving as the ultimate judge of success or failure.

Similar to k-means cluster analysis– observed data are analyzed, connections between data points are found, and the data are grouped into clusters. And while cluster analysis is generally quicker to perform, LCA is a statistically superior model with much more theoretical support (Heinen, 1996; Muthén, 2004; Vermunt & Magidson, 2002). LCA is able to accommodate a much larger variety of data, while accounting for potential missing data within a set. Furthermore, it can accommodate different relative weights of data points, thus providing a stronger method of grouping micro-level decisions wherein it is likely that some micro-level decisions are relatively more important than others for the purposes of making the correct macro-level choice.

By creating these clusters and then seeing how predictive they ultimately are of the macro-level decision (i.e. the diagnosis), an understanding can be gained regarding which groups are most likely to make the correct diagnosis. And knowing which groups make the correct diagnosis can lead to understanding which micro-level decisions are the most important in the overarching decision-making process. However, both of the above approaches suffer from the same core issue. K-means cluster analysis and LCA infer the structure of the data without regard to potential classification issues. Rather, they group individuals into latent clusters based on the dataset and infer their grouping structure based on similarities in the independent variables within that dataset. As such, these approaches still have issues in how groupings are determined and under what circumstances.

Thus, an approach that allows for micro-level choices to lead people down multiple pathways (rather than one grouping structure) will be optimal over these approaches. Techniques that can allow for more nuanced classification of individuals are likely to perform better than these more simplistic clustering approaches when it comes to predicting macro-level decision-making. Given the sheer volume of potential predictor variables that exist (every micro-level decision comprised of every question asked during patient history, every test conducted during physical examination), few techniques can wholly account for ways in which to classify the predictor variables in an appropriate fashion. However, one family of methods that is designed to deal with this issue is tree-based methods (Breiman, Friedman, Olshen, & Stone, 1984; Cutler, Cutler, & Stevens, 2009).

Tree-based methods can be used to classify problems across a number of predictor variables (Cutler et al., 2009). Similar to the previous methods, each data point is placed into any number of distinct groups with the goals of using the predictor variables to classify observations. However, this technique classifies data points in ways that include variable importance, associations between variables, and how the variables relate to and interact with one another in predicting the response. As such, these techniques provide the groundwork for creating groupings of micro-level decisions based on their relation to the macro-level decision and are thus more likely to yield correct prediction of macro-level decision-making.

However, to better enhance this technique, a method known as bagging (Breiman, 1996) can be used. This method utilizes bootstrapping to substantially increase the predictive abilities of the trees. While it is much more difficult to use and to interpret, it can help account for issues such as unstable predictors. For example, if one micro-level decision plays a much stronger role the overall process, such that any grouping without that decision made correctly ultimately fails in making the correct macro-level decision, bagged trees will be better able to account for it.

To improve on this technique, random forests can be used to increase randomness into the samples during the tree-building procedure. This increases both the speed and accuracy of the models by reducing bias and correlation between the trees (Breiman, 2001; Cutler et al., 2009).

This is because multiple trees are grown and are given a much larger degree of specification for making them as dissimilar as possible, reducing the relationship between them and thus improving overall predictive power.

This type of model comparison has been utilized on data with known structures across a few different areas, including age prediction and car crash severity prediction (Iranitalab & Khattak, 2017; Rendall, Pereira, & Reis, 2017). These studies compare across regression analyses, cluster analyses, and machine learning tree-based methods (Iranitalab & Khattak, 2017; Rendall, Pereira, & Reis, 2017). Across these areas, tree-based methods have better predictive abilities and generally outperform alternate methods.

> *Hypothesis 1:* Utilizing k-means cluster analysis will yield groupings of individuals that will better predict correct macro-level decisions compared to dominance models.
>
> *Hypothesis 2:* Utilizing latent class analysis will yield groupings of individuals that will better predict correct macro-level decisions compared to k-means cluster analysis.
>
> *Hypothesis 3*: Utilizing traditional regression trees will better predict correct macro-level decisions compared to clustering techniques.
>
> *Hypothesis 4:* Utilizing bagged regression trees will better predict correct macro-level decisions compared to traditional classification trees.
>
> *Hypothesis 5:* Random forests of micro-level decisions will better predict correct macro-level decisions compared to bagged classification trees.

Chapter 2: Method

Study Design

This study aimed to provide models with greater classification accuracy for decision-making processes of multilevel decisions in the medical context. More specifically, the study aimed to model diagnostic decisions made during medical simulations. Archival data provided by the Morsani College of Medicine at USF were utilized to describe how these decisions are made. Data were initially gathered for instructional/training purposes. Students in this program undergo a multitude of training and assessment simulations that are intended to hone their clinical decision-making abilities. In these simulations, they interact with a confederate patient, who presents the symptoms of a specified set of medical problems.

It was the student's goal to correctly assess the patient's condition and identify which ailment they suffer from, given information obtained during the interview and physical examination. To make their final diagnosis (i.e., the overarching choice or macro-level decision), participants needed to make many small but critical decisions regarding: which questions to ask in the interview, which aspects of the patient to take note of during the physical examination, and what tests to conduct. A variety of psychometric techniques were utilized to model diagnostic decisions to learn which sets of micro-level decisions are the most likely to lead to better macro-level decisions.

Participants

This study used an archival dataset provided by the Morsani College of Medicine at USF. Approximately 130 individuals were included in the data sample. These individuals were all third-year medical students at the USF Morsani College of Medicine. Demographic data were not included in the data received for the project, and thus no demographics were used in the data analysis.

Procedure

The simulation utilized for this study is the Comprehensive Clinical Performance Exam (CCPX). The CCPX is a training exercise that medical students undergo as preparation for their certification exams. The goal of the simulation is to provide students with realistic medical situations in which assessors know the correct portions of history, physical examination, and other areas that students should attend to as well as the correct diagnoses.

When students begin the simulation, they enter an examination room where a confederate (referred to as a standardized patient) is waiting for them. The standardized patient then describes a specific medical concern that has brought them in for a doctor's visit. The medical school trains standardized patients on a number of specific medical scenarios. These scenarios vary widely, with issues ranging from alcohol dependence to extreme fatigue. For the purposes of this study, I looked at seven such scenarios. Specifically, these scenarios include patients with key symptoms of:

- Alcohol Dependency
- Dizziness
- Fatigue
- Hematuria
- Hoarseness
- Low Back Pain
- Night Sweats

Once the standardized patient explains the key symptom for their given scenario, the medical student begins taking the patient's history and conducting a physical examination. The standardized patients are trained to respond the same way to any questions asked of them for all students, and the responses are intended to be consistent with the correct diagnosis (i.e., errors by standardized patients are not deliberately injected into the encounters). During the history, students are expected to ask questions regarding general facets of the individual's medical history, specific symptoms they may or may not be exhibiting that are related to the key symptom, past medical history, and social factors history. During the physical examination, students are expected to take note of general appearance, vital signs, and examine any region of the body that may be related to the key symptom and/or other symptoms mentioned during the patient history (invasive bodily inspections are simulated by the standardized patient handing the medical student a card indicating the result of the inspection). After the students have completed both the history and the physical examination, they must then come up with two to four preliminary diagnoses and order appropriate diagnostic tests to confirm or refute these diagnoses.

Throughout this process, students make notes of each segment of the patient interaction in an online chart. For each scenario, reviewers at the medical school have a rubric that outlines which questions the students should be asking during the patient interview, what they should be attending to during the physical examination, which preliminary diagnoses are appropriate given the patient symptoms, and which diagnostic tests should be ordered given the information they are provided with assuming that the student asked the right questions and made the proper preliminary diagnosis. Students are then rated by reviewers based on the quantity of correct choices they made within each category. For example, if they were supposed to take note of four specific items regarding the patient's social factors history, then examiners will rate them higher or lower based on how many of these items the student correctly marked. The total number of items that should be marked is dependent on the scenario, but ranges between forty-five and sixty-eight items. The total number of possible correct items per scenario were forty-five for Alcohol Dependency, sixty-two for Dizziness, sixty-four for Fatigue, fifty-eight for Hematuria, fifty-two for Hoarseness, sixty-eight for Low Back Pain, and sixty-eight for Night Sweats.

Unlike real-world situations, the simulated nature of these scenarios enables a known correct diagnosis for decision-making patterns to be compared against the medical students' choices. The questions students asked in the patient history and physical examinations were recorded in the patient chart. Similarly, their preliminary diagnoses were also available in this online chart. For any given scenario, the rubric included anywhere between two and six potentially correct diagnoses and students were able to list anywhere between two and four possible options in the online chart. The total number of possible correct diagnoses per scenario were two for Alcohol Dependency, four for Dizziness, four for Fatigue, six for Hematuria, four for Hoarseness, five for Low Back Pain, and four for Night Sweats. Reviewers then rated students based on the quantity of correct preliminary diagnoses they listed (up to four, depending on the scenario). Thus, the dependent variable for this study was computed using the overall quantity of correct potential diagnoses made by a given participant within their scenario.

For the study, medical students' online charts were coded by research assistants utilizing the reviewer rubrics as a coding key. Each item on the rubric represented a piece of information that was relevant to the scenario. As such, the items were coded into a dataset of dichotomous variables with a 0 indicating that the student did not elicit a relevant piece of information and a 1 indicating that they did. Any instances where the students performed an action that was not on the rubric, a note was made to indicate the action and a separate coding key was made for these notes. There were two coders, rating all individuals independently, for the scenario, and all data were tested for interrater reliability. All variables had greater than 90% agreement between raters. Any disputes were brought to conference with a third rater and were resolved by discussion.

Given the nature of the dataset and the question being asked, the macro-level decisions (i.e. diagnoses) were conceptualized as the dependent variables while the micro-level decisions (i.e. questions asked during patient interview and tests run during physical examination) were conceptualized as the independent variables. Diagnostic tests ordered after the completion of the scenario were not included as data points in this study because the results of these tests were not made available to the students when determining preliminary diagnoses. For all hypotheses across scenarios, quantity of correct preliminary diagnoses was the operationalization for the dependent variable. For the dominance models, the micro-level decisions were aggregated within scenario to provide a single number for each individual denoting the quantity of overall micro-level decisions made as well as a number for each individual denoting the quantity of correct (keyed as 'should be asked' by instructor's rubric) micro-level decisions made. For the clustering approaches, the micro-level decisions were utilized to form groupings dependent on technique used (i.e. k-means cluster analysis or LCA). For the tree-based methods, the micro-level decisions were utilized completely independently within the regression trees, that is, each micro-level decision was considered as a separate independent variable.

Chapter 3: Results

In this study, I looked at each of the seven scenarios separately conducting analyses in the programming software R (see code in Appendix A). To ensure that models were predicting significantly better than one another, I looked at the difference in *r* between models. In order to calculate these *r* values, the data were subset into two parts: one for deriving the model and one for cross-validation. In the cross-validation, 70% of the data was used to fit the models and 30% was used to test the models. The two subsets were created using sampling with replacement. The derivation models were then used to predict the number of correct diagnoses in the cross-validation sample. The cross-validation *r* value was computed between the model prediction and total number of correct potential diagnoses. This process was completed independently 2000 times for each scenario to produce a distribution of cross-validation correlations for each model for each scenario. The resulting empirical sampling distributions of cross-validation *r* follow the spirit of bootstrap sampling (Efron, 1987; 1988) but did not specifically employ the R packages for bootstrapping due to sampling issues[1]. Findings from the various analyses conducted were then compared by looking at the differences in *r* value for the models, where the models with higher *r* values indicated better prediction. The results looked at both the means of the empirical

[1] The bootstrap packages in R utilize sampling with replacement. However, the method used when creating data subsets for cross-validation also utilized sampling with replacement. As such, using the bootstrap packages in R would have caused data to be resampled twice, leading to inconsistencies in results found.

sampling distribution as well as at the 95% confidence intervals. The 95% confidence intervals were created analytically using the empirical distributions to calculate upper and lower bounds of *r* across 2000 iterations for each of the seven analysis methods and in each scenario using empirical mean and standard deviation.

Regression Models

Analysis 1 (Overall Quantity Micro-Decisions)

I began with testing whether the quantity of micro-level decisions significantly predicts correct macro-level decisions. Being the quantity (sum of) of micro-level decisions each medical student made, the independent variable was continuous. The dependent variable was also continuous, as medical students were able to give multiple possible correct preliminary diagnoses. As such, the data were analyzed using linear regression, allowing me to look at the effect that the quantity of micro-level decisions has on correct macro-level choices. The following *r* values were found for the seven scenarios:

- Alcohol Dependency, *r*(107) = .011, 95% CI [-.335, .357]
- Dizziness, *r*(111) = -.004, 95% CI [-.314, .307]
- Fatigue, *r*(112) = .138, 95% CI [-.199, .475]
- Hematuria, *r*(108) = .095, 95% CI [-.313, .503]
- Hoarseness, *r*(108) = .018, 95% CI [-.354, .390]
- Low Back Pain, *r*(112) = -.002, 95% CI [-.310, .306]
- Night Sweats, *r*(113) = .099, 95% CI [-.291, .490] (*see Table 1*)

With all *r* values approaching 0 across all scenarios, the findings indicate that overall quantity of micro-level decisions made did not predict correct macro-level decisions.

Analysis 2 (Correct Micro-Decisions)

Likewise, I tested whether quantity of correct micro-level decisions impacts macro-level decision-making utilizing the same technique of linear regression by substituting the independent variable from quantity of overall micro-level decisions made to quantity of correct micro-level decisions made. The following *r* values were found for the seven scenarios:

- Alcohol Dependency, $r(107) = .002$, 95% CI [-.101, .105]
- Dizziness, $r(111) = -.003$, 95% CI [-.329, .323]
- Fatigue, $r(112) = .196$, 95% CI [-.148, .540]
- Hematuria, $r(108) = .045$, 95% CI [-.372, .461]
- Hoarseness, $r(108) = .033$, 95% CI [-.352, .418]
- Low Back Pain, $r(112) = -.001$, 95% CI [-.268, .265]
- Night Sweats, $r(113) = .033$, 95% CI [-.326, .393] (*see Table 1*)

Similar to the previous regression, all *r* values across scenarios were extremely low. This indicates that quantity of correct micro-level decisions made (similar to overall quantity) did not ultimately predict correct macro-level decisions.

Classification Models

For the second set of analyses, I utilized two different techniques to classify individuals into different groups dependent on their micro-level decisions: k-means cluster analysis and LCA. Then, for each of these techniques, I used the resultant groupings as the categorical independent variable and once again utilized the macro-level decisions as the dependent variable. The number of groupings was dependent on what each technique found for common latent decision patterns among the students. For each set of analyses, I then conducted an Analysis of Variance (ANOVA) on each scenario to see if and when groupings significantly predicted correct macro-level decisions.

Analysis 3 (k-means Clustering)

When conducting the k-means clustering, the gap statistic was calculated to determine the number of clusters present in each scenario. However, after utilizing this technique on all seven scenarios, only two scenarios presented with clusters: Alcohol Dependency and Hematuria. In the other five scenarios, all participants fell within one unified cluster. As such, ANOVAs were only run on the two scenarios that presented with clusters. In both of these scenarios, two clusters were found across participants. The results of the ANOVAs were then used to create correlation matrices with ANOVA values for the clusters being compared with the number of correct diagnoses in the cross-validation samples. The following *r* values were found for those scenarios:

- Alcohol Dependency, *r*(107) = .109, 95% CI [-.283, .502]
- Hematuria, *r*(108) = .002, 95% CI [-.345, .349] (*see Table 1*)

While the majority of scenarios did not present with latent clusters, the two scenarios that did present with them had very low *r* values. These findings indicate that Hypothesis 1 was not supported. This indicates that there are likely no consistent latent clusters that exist within the micro-level decisions made across various medical scenarios in the data from this study.

Analysis 4 (Latent Class Analysis)

When conducting the LCA, the latent classes were calculated using a variable selection method based on creating the optimal latent class model using BIC (Dean & Raftery, 2010; Fop, Smart, & Murphy, 2017). However, this method encountered a similar issue to that found when conducting Analysis 3. Specifically, after conducting the LCA on all seven scenarios, only two presented with latent classes: Alcohol Dependency and Hematuria. In the other five scenarios, all participants fell within one unified class. As such, ANOVAs were only run on the two scenarios that presented with clusters. Similar to the k-means analysis, the results of the ANOVAs were then used to create correlation matrices with ANOVA values for the clusters being compared with the number of correct diagnoses in the cross-validation samples. The following *r* values were found for those scenarios:

- Alcohol Dependency, *r*(107) = .010, 95% CI [-.339, .359]
- Hematuria, *r*(108) = .011, 95% CI [-.357, .379] (*see Table 1*)

Possessing the same issue as Analysis 3, the majority of scenarios did not present with latent classes and the two that did present with them had very low *r* values. These findings indicate that Hypothesis 2 was also not supported. Similar to the cluster analysis, this indicates that there are likely no consistent latent classes that exist within the micro-level decisions made across various medical scenarios in the data from this study. While the findings from these two methods are discouraging, they were not entirely unanticipated, given the variety of paths that doctors can choose in finding a diagnosis and the possibility that these different paths can yield similar outcomes. With more sophisticated methods of analysis (such as those that follow), these different paths can be better taken into account and utilized.

Tree-Based Models

Similar to Analyses 3 and 4, Analyses 5-7 also used techniques to classify individuals into groups. However, in this set of analyses, the use of regression trees categorized individuals into groups dependent on how the macro-level decision was made (i.e., the diagnostic outcome variable). The regression model then fit the key micro-level decisions to the diagnostic decision. Because these techniques fit models to the correct diagnoses, cross-validation was used where 70% of the data was used to fit the models and 30% was used to test the models. Once again, this was done separately for all seven scenarios. Three different techniques of developing regression trees were utilized: the traditional approach (simple regression trees), bagged, and random forests. While the more complex approaches (i.e. bagged and random forests) don't require a validation subset to accurately create trees, the traditional approach does. As such, in the traditional approach, the data were split so that 70% was used to fit the models, 15% to validate the models, and 15% to test the models.

Analysis 5 (Simple Regression Trees)

The following *r* values were found for the seven scenarios:

- Alcohol Dependency, $r(107) = .268$, 95% CI [-.298, .835]
- Dizziness, $r(111) = .335$, 95% CI [-.245, .914]
- Fatigue, $r(112) = .341$, 95% CI [-.254, .938]
- Hematuria, $r(108) = .307$, 95% CI [-.276, .890]
- Hoarseness, $r(108) = .310$, 95% CI [-.269, .889]
- Low Back Pain, $r(112) = .268$, 95% CI [-.302, .837]
- Night Sweats, $r(113) = .311$, 95% CI [-.251, .872] (*see Table 1*)

With relatively high average *r* values found consistently across scenarios, Hypothesis 3 was supported. This suggests that utilizing traditional regression trees will better predict correct macro-level decisions on average compared to clustering techniques (as well as traditional regression techniques). However, it should be noted that the empirical 95% confidence intervals across the 2000 iterations for this method were rather large, which also suggests a high degree of variability in successful prediction across different cases of tree development (*see Figures 1-7 for full r distribution for each scenario*).

Looking across the figures, it is clear that when the simple regression trees were able to construct successful models, they predicted quite well. But similarly, it can be seen that there was a large number of cases across the 2000 iterations whereby the simple regression trees failed to construct models. This high level of variability in successful model building is likely due to constraints from the size of the dataset, further exacerbated by the need to split the data for validation and testing purposes (see Discussion for further consideration of this issue).

Furthermore, in the context of the current study, when the simple trees were built, they ultimately used fairly different inputs depending on the scenario. To understand this issue, we can look at two contrasting example cases from different scenarios. The first example we can look at is a sample model (one possible iteration out of 2000 total) from the Alcohol Dependency scenario. In this exemplar model from the Alcohol Dependency scenario, the variables utilized exclusively came from the list identified as important by the medical school (*see Table 2)*.

And from these qualitative results, we can see the relative importance and weight of specific items that the medical school thought important for medical students to consider. Additionally, we can note that not all variables considered important by the medical school were considered important by the model given that only 12/45 variables were flagged by the model. Moreover, not all of the scenarios solely relied on medical school criteria for model building. For example, in an example iteration of the Hematuria scenario, the model relied first on variables outside of those identified by the medical school before even beginning to use those that were identified by the medical school (*see Table 3*). Thus, while the first example helps to illustrate the importance of honing in on and catching key relevant diagnostic information, this example helps to illustrate the relative importance of ruling out superfluous information as well.

Analysis 6 (Bagged Trees)

The following r values were found for the seven scenarios:

- Alcohol Dependency, $r(107) < .001$, 95% CI [-.279, .280]
- Dizziness, $r(111) = .105$, 95% CI [-.177, .387]
- Fatigue, $r(112) = .248$, 95% CI [-.023, .519]
- Hematuria, $r(108) = .028$, 95% CI [-.277, .334]
- Hoarseness, $r(108) = .069$, 95% CI [-.206, .344]
- Low Back Pain, $r(112) = -.037$, 95% CI [-.378, .303]
- Night Sweats, $r(113) = .110$, 95% CI [-.164, .384] (*see Table 1*)

While some of the scenarios (Dizziness, Fatigue, Night Sweats) had medium r values, this trend was not consistent across all seven. Moreover, all of the values found were lower than those in Analysis 5 with traditional regression trees outperformed bagged regression trees across scenarios. Thus, Hypothesis 4 was not supported. However, it should be noted that the 95% confidence intervals across the 2000 scenarios for this method were generally smaller than those found in Analysis 5, displaying less variability across iterations of this method (*see Figures 1-7 for full r distribution for each scenario*).

Analysis 7 (Random Forests)

The following *r* values were found for the seven scenarios:

- Alcohol Dependency, $r(107) = -.069$, 95% CI [-.338, .200]
- Dizziness, $r(111) = .203$, 95% CI [-.074, .480]
- Fatigue, $r(112) = .199$, 95% CI [-.090, .489]
- Hematuria, $r(108) = .082$, 95% CI [-.219, .383]
- Hoarseness, $r(108) = .070$, 95% CI [-.198, .337]
- Low Back Pain, $r(112) = -.078$, 95% CI [-.381, .224]
- Night Sweats, $r(113) = .138$, 95% CI [-.118, .395] (*see Table 1*)

As with Analysis 6, only some of the scenarios had medium *r* values (Dizziness, Fatigue, Night Sweats), however the trend was not consistent across scenarios and all values were lower than Analysis 5. These findings indicate that traditional regression trees also outperformed random forests. Additionally, the random forests did not consistently outperform the bagged trees with relatively similar values across scenarios, thus not supporting Hypothesis 5. Once again, it should be noted that the 95% confidence intervals across the 2000 scenarios for this method were generally smaller than those found in Analysis 5, displaying less variability across iterations of this method (*see Figures 1-7 for full r distribution for each scenario*).

Table 1.
r value and Bootstrap Confidence Interval comparing Analysis by Scenario

	Alcohol Dependency	Dizziness	Fatigue	Hematuria	Hoarseness	Low Back Pain	Night Sweats
Analysis 1 (Regression with All)	.011 [-.335, .357]	-.004 [-.314, .307]	.138 [-.199, .475]	.095 [-.313, .503]	.018 [-.354, .390]	-.002 [-.310, .306]	.099 [-.291, .490]
Analysis 2 (Regression with Correct)	.002 [-.101, .105]	-.003 [-.329, .323]	.196 [-.148, .540]	.045 [-.372, .461]	.033 [-.352, .418]	-.001 [-.268, .265]	.033 [-.326, .393]
Analysis 3 (K-means Cluster Analysis)	.109 [-.283, .502]	NA	NA	.002 [-.345, .349]	NA	NA	NA
Analysis 4 (Latent Class Analysis)	.010 [-.339, .359]	NA	NA	.011 [-.357, .379]	NA	NA	NA
Analysis 5 (Simple Regression Tree)	.268 [-.298, .835]	.335 [-.245, .914]	.341 [-.254, .938]	.307 [-.276, .890]	.310 [-.269, .889]	.268 [-.302, .837]	.311 [-.251, .872]
Analysis 6 (Bagged Regression Tree)	.000 [-.279, .280]	.105 [-.177, .387]	.248 [-.023, .519]	.028 [-.277, .334]	.069 [-.206, .344]	-.037 [-.378, .303]	.110 [-.164, .384]
Analysis 7 (Random Forest)	-.069 [-.338, .200]	.203 [-.074, .480]	.199 [-.090, .489]	.082 [-.219, .383]	.070 [-.198, .337]	-.078 [-.381, .224]	.138 [-.118, .395]

Table 2.
Order of Importance in Exemplar Alcohol Dependency Scenario

1	Patient is annoyed that his/her children want him/her to cut down on drinking
2	Intermittent epistaxis
3	Hospitalization for abdominal pain 2 years ago
4	Told liver and pancreas inflammation
5	Allupurinol 300 mg daily
6	No Allergies
7	Drinks 3-4 scotches daily past few years
8	Vital Signs (VS): Temperature (T): 98.0
9	Scattered bruises on the forearms
10	Started drinking in his/her teens
11	Working in security over past two years
12	CAGE[2] screening: ¾

[2] CAGE is an alcoholism screening that includes four questions

Table 3.
Order of Importance in Exemplar Hematuria Scenario

Not identified by medical school	
1	Patient displays no symptoms of chest infections
2	No history of these abnormalities of illnesses
3	Prior prostate exams were normal
4	Patient is in a committed relationship
Identified by medical school	
5	Progressive difficulty initiating urinary stream
6	Dribbling
7	The hematuria was noted throughout without the voiding
8	No prior episode

Figure 1.
Cross-Validated *r* for Analyses 5-7 across 2000 Iterations in Alcohol Dependency Scenario

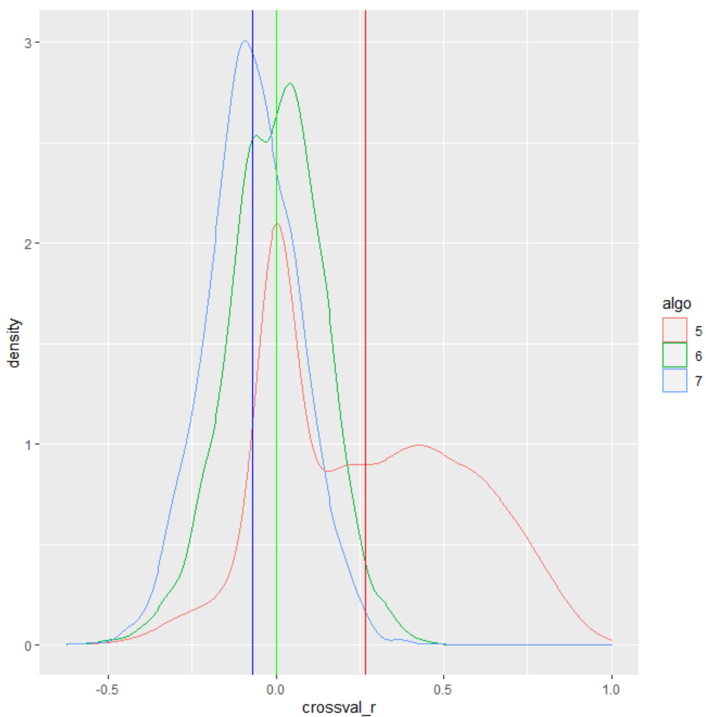


Figure 2.
Cross-Validated *r* for Analyses 5-7 across 2000 Iterations in Dizziness Scenario

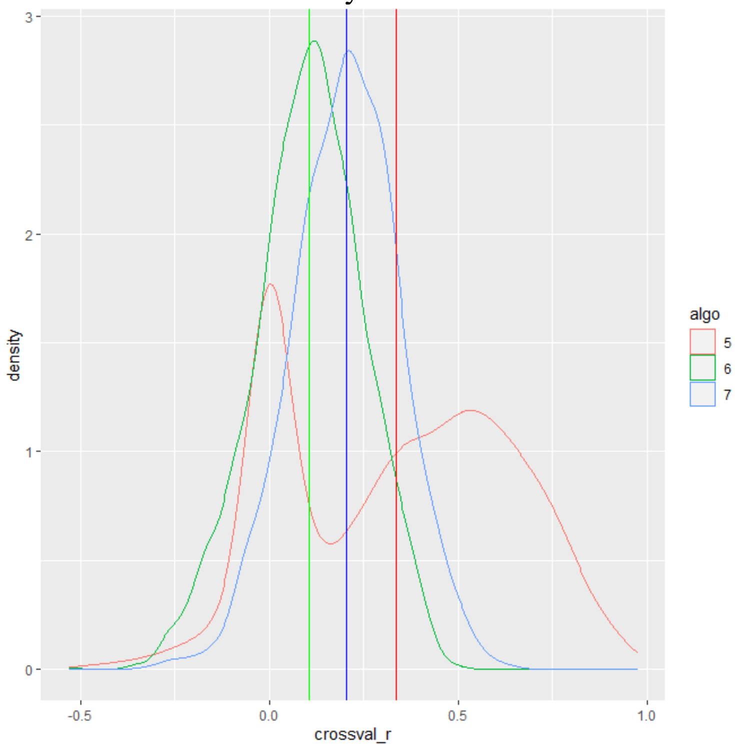


Figure 3.
Cross-Validated *r* for Analyses 5-7 across 2000 Iterations in Fatigue Scenario

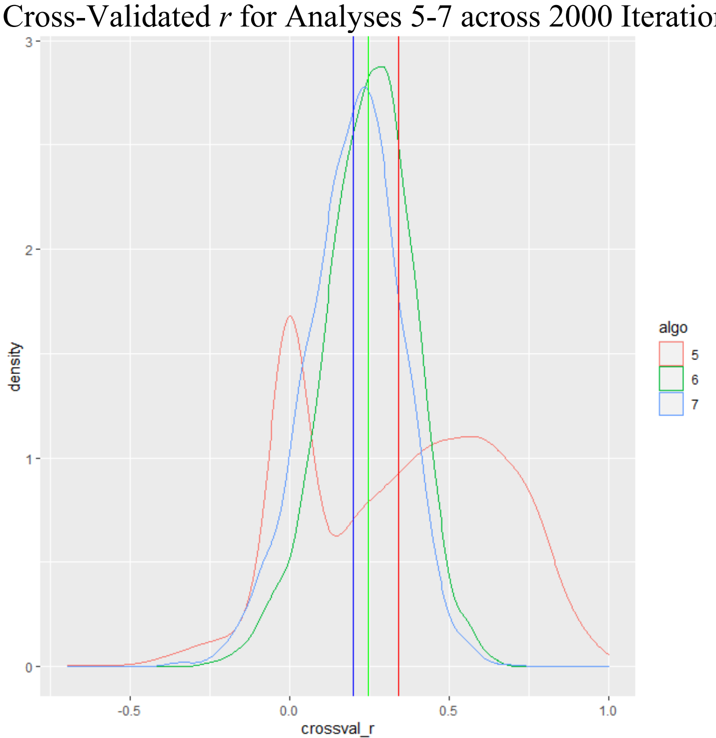


Figure 4.
Cross-Validated *r* for Analyses 5-7 across 2000 Iterations in Hematuria Scenario

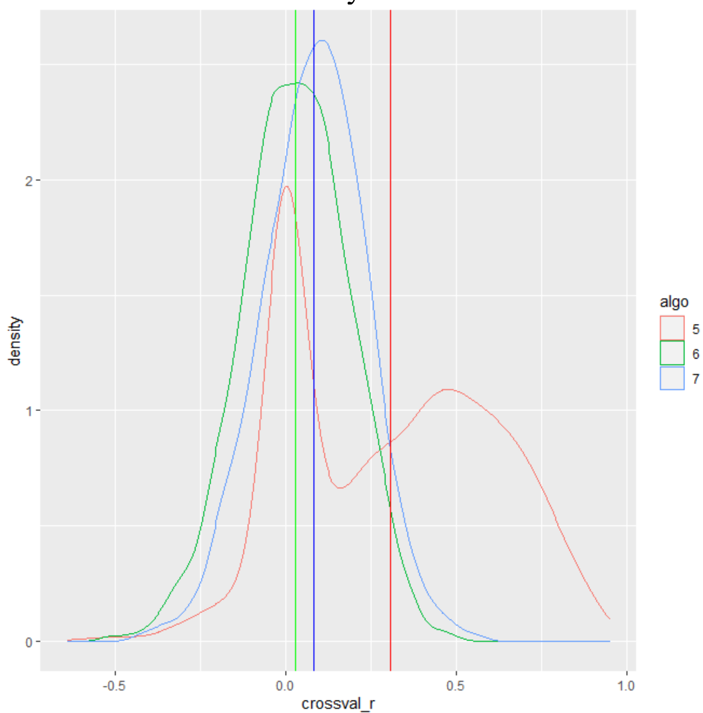


Figure 5.
Cross-Validated *r* for Analyses 5-7 across 2000 Iterations in Hoarseness Scenario

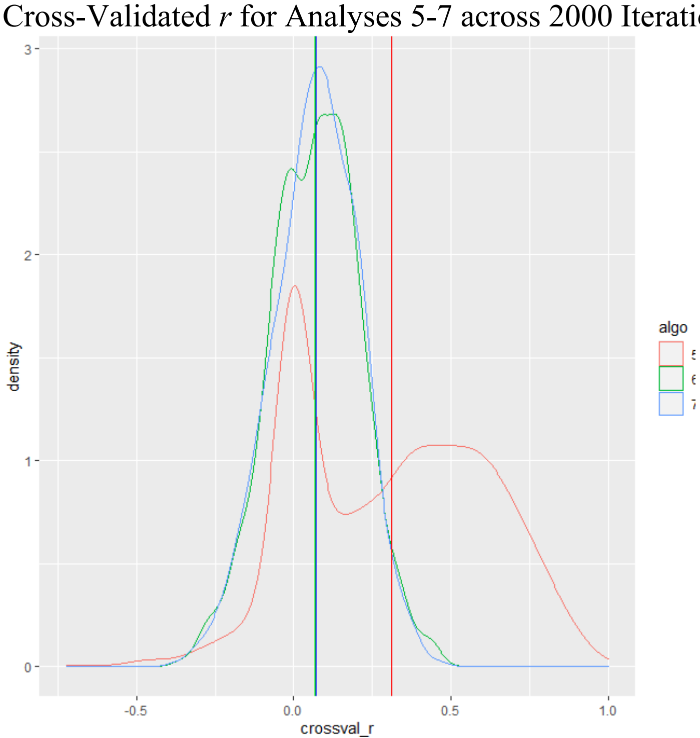


Figure 6.
Cross-Validated *r* for Analyses 5-7 across 2000 Iterations in Low Back Pain Scenario

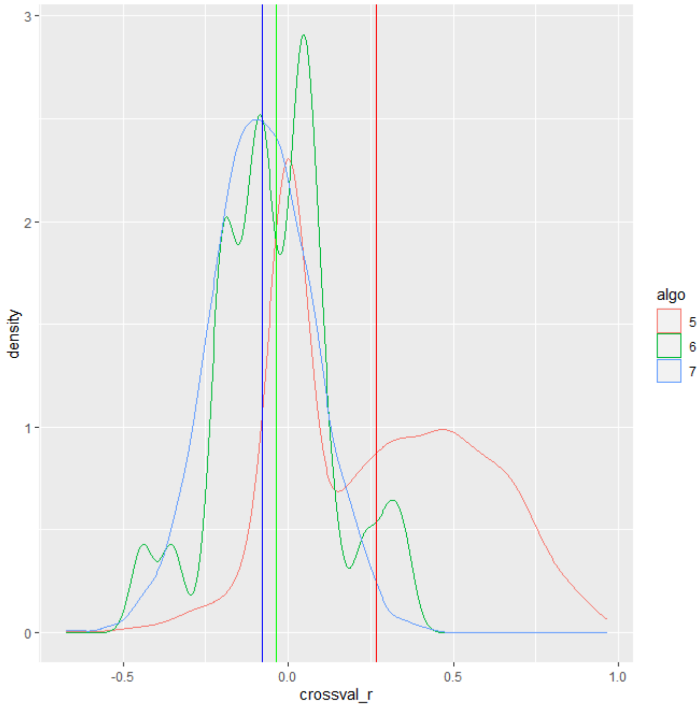
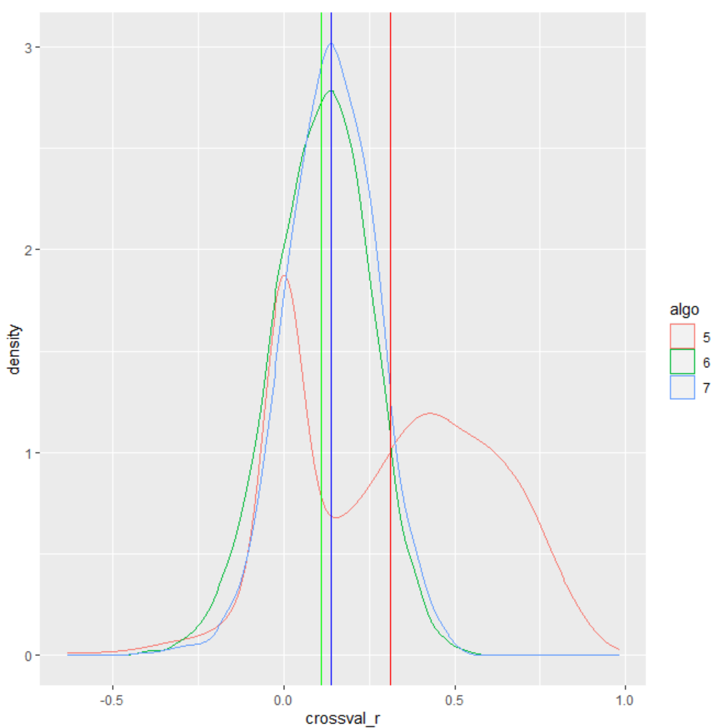


Figure 7.
Cross-Validated *r* for Analyses 5-7 across 2000 Iterations in Night Sweats Scenario



Chapter 4: Discussion

The ability to make effective decisions is a necessary and important role across situations (Highhouse et al., 2013; Simon, 1987). However, decision making is a complex process, wherein a multitude of factors ultimately play a role. The more complex and difficult decisions become, the more decision-makers need help in approaching and making these decisions. By gaining a greater understanding of the processes involved in decision making, we can work toward enhancing how these decisions are ultimately made.

This dissertation looked at the particular issues posed by intra-individual multilevel decisions. In these cases, decisions become a multilevel process when one decision's outcome leads to information that is used in subsequent larger or other types of decisions. In this particular area of decision making, a lot of the complexity arises from individuals needing to assess multiple informational inputs stemming from those smaller decisions in order to correctly make the larger ones. This study looked at how this particular process is conducted within the medical domain. Medicine was a particularly illustrative example of this phenomenon due to the way in which diagnostic decisions are made. Doctors must constantly make decisions while working with patients (deciding on which questions to ask and which tests to run), and then use the information gathered from these decisions to ultimately develop a diagnosis for the patient's issues. However, there is a great deal of contention regarding how best to make these diagnostic decisions (Hunink et al., 2014), and still minimal research regarding the multilevel factors involved within them.

In this study, I looked to gain a greater understanding of the multilevel process involved in making the diagnostic decision using a psychometric approach. At the time that data were collected, the medical school utilized a dominance approach for instructing students (whereby students are graded as more effective diagnosticians dependent on the sheer quantity of micro-level information they gathered during patient examination). The medical school likely employs this approach as a means by which to build foundational medical knowledge and aid students in gathering information in situations where the diagnostic outcomes are ultimately unknown. That said, the present research found that this approach likely does not lead to doctors synthesizing correct preliminary diagnoses in the cases studied. Quantity of micro-level decisions was not found to impact the quality of the macro-level decisions made. Regardless of which medical scenario individuals were placed within, the number of questions asked and tests run did not ultimately impact whether or not they came to the correct conclusions regarding diagnosis. Even when only accounting for micro-level decisions that were considered correct (determined by a panel of medical experts as being the appropriate questions to ask a patient coming in with a given complaint), the results were unchanged. Quantity of correct micro-level decisions does not impact the quality of the overall macro-level decisions made. This much was aligned with expectations and previous research (Braun & Kaplan, 2017; Mamede et al., 2007).

Beyond looking at the extant methods utilized by the medical school, this study also aimed to look at various other methods of understanding the interrelationship between micro- and macro-level decisions involved in the diagnostic process that did not solely rely on quantity. Specifically, the aim was to try and understand which information gathered at the micro-level would be the most helpful and important at the macro-level. To this end, two different types of approaches were used to better look at quality, rather than quantity, of information gathered. The first approach attempted to look for consistent patterns of micro-level decision making that individuals might be using to come up with a diagnoses. In this approach, patterns were assessed psychometrically using methods of k-means clustering and LCA. However, neither of these approaches consistently found underlying patterns in the decision making. In fact, only two of the medical scenarios were able to produce patterns and even in those scenarios, the predictive ability of those patterns was statistically negligible.

This finding, while inconsistent with the hypotheses, was not entirely unexpected. The data used in this study had an enormous number of predictor variables captured within each dataset. Every question asked during patient history and every test conducted during physical examination were considered micro-level decisions and thus counted as possible predictors leading to datasets with anywhere from fifty-three to 107 total predictor variables. Given the sheer volume of potential predictor variables that existed within the data, few techniques are fully capable of accounting for this level of classification of predictor variables. Based on the results found in this study, k-means cluster analysis and latent class analysis were not sufficient for datasets this vast in complexity.

The second alternative approach attempted to look at the quality of the micro-level decisions in a different way that overcomes many of the shortcomings that the clustering methods had. In the second method, tree-based methods were used to classify micro-level decisions with each data point being placed into groups using a variety of factors including variable importance, associations between variables, and how the variables relate to and interact with one another in predicting the diagnoses (Breiman et al., 1984; Cutler, et al., 2009). In this approach, micro-level decisions were evaluated using simple regression trees, bagged regression trees, and random forests.

The results indicated that simple regression trees did a reasonably effective job at predicting correct diagnoses across all scenarios. This finding corresponded with initial predictions by being both predictive of correct diagnosis and outperforming other techniques, including those currently in use by the medical school. By effectively predicting correct diagnosis regardless of medical scenario, this method of using regression trees shows that it is possible to predict correct macro-level decision-making in the diagnostic process across various cases and possibilities. It also provided evidence for the importance of quality in micro-level decision making being more influential than quantity. This implies that there are a great deal of future possibilities for modeling and predicting correct diagnoses and provides a stepping stone for future endeavors that look to better understand the multi-level nature of the diagnostic decision process.

Initially, I had also predicted that each layer of added complexity involved in the tree-based approaches would also add predictive ability, but these hypotheses were not supported by the findings of this study. Rather, the more complex bagged trees and random forests were vastly outperformed by the simple regression trees. However, this finding does not necessarily invalidate these two methods as there were other important limiting factors that could have impacted the results found.

The Medical Context

Although the current research suggests that the dominance approach utilized by the medical school likely does not lead to correct preliminary diagnoses in the cases seen here, there are still other reasons why it is employed from the perspective of medical training. From the current study, we can know that some questions and tests likely matter more than others, but these questions inevitably differ between different medical situations. The key, for doctors, is finding which questions/tests are needed for a correct diagnosis, and which are not. In this process of finding which questions need to be asked and which tests need to be run, individuals simply cannot avoid collecting some degree of unnecessary or extraneous information. And ideally, as doctors collect and sort through information, they can slowly narrow down from the less useful information to the more useful information.

While in this study, the diagnostic outcome (the correct macro-level decision) was known for each medical situation; in practice, this is not the case. By instructing medical students to make more micro-level decisions (i.e. collect more information), they are able to learn how to engage in this process of narrowing on their own. That said, the findings here seem to indicate that there are differences in how different individuals move throughout this process, wherein some are better able to make the micro-level decisions needed to enable them to come to the correct diagnosis.

The regression trees utilized in this study, as a type of machine learning, are able to pull useful bits of information from large amounts of data and categorize them simply. This is likely how they were able to identify the most predictive micro-level decisions across the various datasets utilized and create a high level of prediction for preliminary diagnosis. When taking the results of these trees and looking at how various individuals performed within any given scenario, we can see that there are differences in correct diagnosis between those individuals who were better able to collect and sort through the information they needed and those who were less able. This indicates that there are different ways for individuals to go through this decision process more or less optimally, and the current research provides the first stepping stones to understanding this process.

Given the findings in the current study related to the dominance models, it is important to consider how possible additions to the medical context from machine learning approaches may be able to improve diagnostic accuracy. While the present work is only the beginning of understanding the statistical possibilities of applying machine learning approaches to this context, additional work can help to bring about ways in which these models can be used to supplement the processes that doctors are already going through. By narrowing down the possibilities, these types of techniques could potentially help doctors more quickly and more accurately hone in on correct diagnoses.

Limitations and Future Directions

Due to the sample sizes involved, this study was limited in psychometric approaches that could be used. All of the datasets used in this study were relatively small (particularly for the machine learning approaches). Samples were drawn from a pool of approximately 130 individuals, creating datasets ranging between 107-113 individuals. These small sample sizes, by nature, impact both the psychometric approaches that could be used, and the results of the psychometric approaches that were used. For example, in a more robust dataset, more complex machine learning algorithms could have potentially been tried (and were initially considered), but due to small sample size, there would not have been sufficient data to appropriately train such models. And even in the models utilized in this study, there were still many issues that arose and which could likely be attributed to the sample sizes.

In the case of the simple regression trees, while the results found were encouraging, there are still some issues. First, the modal value for the simple regression tree cross-validation (across

2000 iterations) is zero across all scenarios. This could potentially indicate a similar issue to that found in the cluster analyses, i.e. all individuals had the same predicted value and thus their data "clustered" together. Additionally, the simple regression tree also contained some rather high cross-validation values of *r* as well as high variability across cases, likely due to the combination of factors caused by small datasets (further fragmented by splitting data for validation), utilizing sampling with replacement, and effects from pruning the trees.

With the more complex tree-based methods, the results were likely even more impacted by the issue of sample size. Generally, tree-based methods utilize sample sizes larger than those found in this study (Cutler et al., 2009; Kim, 2008) by orders of magnitude. And as the level of complexity increases, so too should the sample size. With bagging and random forests, numerous trees are being constructed across the data and these numerous trees are then consolidated to form the final prediction. Using these approaches with ample data generally allows for these trees to be built with little issue. But due to limitations in the data used, such examinations were not fully possible. As such, the findings concerning the more complex tree-based methods could be a statistical artifact dependent on the samples used for this study rather than a true finding regarding the efficacy of more complex tree-based approaches.

Another issue to consider is the matter of what information the tree-based models actually used for prediction. Tree-based models statistically select the most informative variable that produces the clearest split between groups through a method of recursive binary splitting. Specifically, they start at all possible observations, and select the predictor that is the best at successfully splitting the data into two branches. Then, they continue to split down the branches in a similar pattern until ultimately reaching a predefined stopping point (in this study, that point was defined as the place where additional splits were no longer meaningful). The problem of using a method that relies solely on the statistics, and less on the context, is that it can be difficult to hone in on what the relevant parts of the dataset ultimately are.

In other areas of study that utilize this type of information searching (i.e. Information Theory, Shannon, 1948), entropy is used to define and extract the most efficient searching strategies. Where entropy is concerned, the larger the number of possible options, the more uncertainty individuals need to deal with throughout. In larger datasets, the items that end up being the most predictive often end up being those items that eliminate incorrect possibilities more so than those items that confirm the correct solution. This is because eliminating incorrect options also reduces the total possibilities and leads to more refined searching over time. Thus, it is important to consider that a key possibility is the matter of the doctors asking the necessary wrong questions they need to root out all incorrect alternatives.

Given the differences that doctors would be confronted with in various medical scenarios, it is important to consider what the relative meaning of the information gathered is. In practice, doctors largely do not know what specific diagnosis any specific patient may have. Understanding which information is useful (both for ruling out diagnoses as well as possibly including them) is the key to helping patients and finding the correct diagnosis more expeditiously. While there are some differences in which types of questions and examinations were the most useful across scenarios (as demonstrated by the findings of this study), the chief similarity among them is their ability to help the doctors narrow down the total number of possible diagnoses. Without knowing the true diagnosis in advance, the most important thing for doctors, regardless of patient or scenario, will be to know how to narrow down the options and to learn which questions will best help them to do that.

Tree-based models, along with other similar statistical approaches, can aid doctors in this process, but should also be paired with subject matter expertise. Because tree-based models are a purely statistical approach, they do not take greatly from the context in which the data were created. And because of their machine learning nature, they become increasingly more difficult to work with and interpret with each additional level of complexity. The goal then for future work in this area should be to marry the insights we can gain from these statistical approaches with the contextual understanding that already exists in the field from which the data originates (e.g. medicine). Future research should also attempt to look at these tree-based approaches, along with other machine learning techniques, on larger datasets that can be more accurately tested for their predictive abilities. And specifically, they should try to understand which criteria are the most important for reducing the diagnostic possibilities and thus hone in on the correct diagnosis.

By gaining a better understanding of how, when, and why these models are able to predict for correct diagnoses, we can better learn and understand how these decision processes are developed and properly applied in clinical contexts. We can, hopefully, gain a greater understanding of what differentiates those individuals who are better at narrowing down key information from those who are worse. And ultimately, we can begin to develop tools that will aid doctors in their diagnostic reasoning and improve on their overall accuracy when treating patients for various illnesses.

Moreover, the current research has implications that reach beyond the context of medicine. By gaining a greater understanding of this type of decision-making process at a more general level, there can be widespread repercussions across a wide variety of individuals, groups, and organizations spanning throughout a range of fields and decisions being made, from employee selection to organizational business decisions. Given the interplay between statistical models and potential contexts seen in this study, the possibilities for future research grow beyond the medical field to any field in which similar decisions are made.

Ultimately, this endeavor looked to psychometrically assess and understand multilevel decisions that require the assessment of multiple informational inputs. The findings of the current study indicate the importance of not only gaining correct information through micro-level decisions, but also the importance of ruling out incorrect information. This notion of funneling from all available options to the most likely correct options by taking away the least likely options first is one that both organizations and individuals can use while making complex and decisions and one that should continue to be explored in future research.

Conclusion

Making decisions in an effective manner is an important feature of individual and organizational success. However, due to the overwhelming complexity of the decision-making process, decisions are not always made in an effective manner. And when decisions increasingly include more layer and nuance, it often become increasingly difficult to make them. By looking at large scale decisions through the lens of multilevel reasoning, we can begin to understand the overall process whereby outcomes of one or more smaller scale decisions can feed into the outcomes of larger scale decisions. The dissertation looked at analyzing these multilevel decision processes within the field of medical diagnostics in the pursuit of better understanding the statistical underpinnings of how these decisions can be made. While the statistical findings were insightful in understanding parts of this decision process, it was clear that they were not all-encompassing. And though statistics can be a useful tool to aid in understanding this process, it is likely something that should be paired with existing expertise, as illustrated by the examples provided throughout this study. As shown throughout this manuscript, continuing work in this area can help us better understand where and when statistics can and should meet with context. Ultimately, by gaining a greater understanding of this process, we can work toward building better tools to consistently make correct decisions moving forward.

References

Acker, F. (2008). New findings on unconscious versus conscious thought in decision making: Additional empirical data and meta-analysis. *Judgment and Decision Making*, *3*(4), 292.

Aguirre-Rodriguez, A., Bosnjak, M., & Sirgy, M. J. (2012). Moderators of the self-congruity effect on consumer decision-making: A meta-analysis. *Journal of Business Research*, *65*(8), 1179-1188.

Arnold, T. (2002). Advanced portfolio theory: why understanding the math matters. *Journal of Financial Education*, 79-96.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological review*, *64*(6p1), 359.

Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. S. (2002). Computer-mediated communication and group decision making: A metaanalysis. Organizational behavior and human decision processes, 87(1), 156-179.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological bulletin*, *74*(2), 81.

Boussadi, A., Caruba, T., Karras, A., Berdot, S., Degoulet, P., Durieux, P., & Sabatier, B. (2013). Validity of a clinical decision rule-based alert system for drug dose adjustment in patients with renal failure intended to improve pharmacists' analysis of medication orders in hospitals. *International journal of medical informatics*, *82*(10), 964-972.

Braun, M. T. & Kaplan, M. S. (2017, April). Missing the forest for the trees; Perils in multilevel decision-making. Presentation at the 32nd annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta psychologica*, *81*(3), 211-241.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Belmont, CA: Wadsworth. *International Group*, *432*, 151-166.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Burnham, K.P.; Anderson, D.R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.), Springer-Verlag

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, *43*(3), 345-370.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, *23*(3), 255-282.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, *100*(3), 432.

Cohen, S. M., & Kataoka-Yahiro, M. (2009). Themes in the literature related to cardiovascular disease risk reduction. *Journal of Cardiovascular Nursing*, *24*(4), 268-276.

Constantinides, G. M., & Malliaris, A. G. (1995). Portfolio theory. *Handbooks in operations research and management science*, *9*, 1-30.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2009). Tree-based methods. In *High-Dimensional*

Data Analysis in Cancer Research (pp. 1-19). Springer, New York, NY.

Dalal, R. S., Bonaccio, S., Highhouse, S., Ilgen, D. R., Mohammed, S., & Slaughter, J. E. (2010). What if industrial–organizational psychology decided to take workplace decisions seriously?. *Industrial and Organizational Psychology*, *3*(4), 386-405.

Dean, N., & Raftery, A. E. (2010). Latent class analysis variable selection. Annals of the Institute of Statistical Mathematics, 62(1), 11.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological methods*, *3*(2), 186.

Dillemuth, J. A. (2009). Navigation tasks with small-display maps: The sum of the parts does not equal the whole. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *44*(3), 187-200.

Dolan, J. G., & Frisina, S. (2002). Randomized controlled trial of a patient decision aid for colorectal cancer screening. *Medical Decision Making*, *22*(2), 125-139.

Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, *51*(4), 380.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association*, *82*(397), 171-185.

Efron, B. (1988). Bootstrap confidence intervals: good or bad?. *Psychological bulletin*, *104*(2), 293.

Elton, E. J., & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, *21*(11-12), 1743-1759.

Elwyn, G., Frosch, D., Thomson, R., Joseph-Williams, N., Lloyd, A., Kinnersley, P., ... & Edwards, A. (2012). Shared decision making: a model for clinical practice. *Journal of general internal medicine*, *27*(10), 1361-1367.

Elwyn, G., Laitner, S., Coulter, A., Walker, E., Watson, P., & Thomson, R. (2010). Implementing shared decision making in the NHS. *Bmj*, *341*, c5146.

Fernholz, E. R. (2002). Stochastic portfolio theory. In *Stochastic Portfolio Theory* (pp. 1-24). Springer, New York, NY.

Fop, M., Smart, K. M., & Murphy, T. B. (2017). Variable selection for latent class analysis with application to low back pain diagnosis. The Annals of Applied Statistics, 2080-2110.

Gray, P. H., & Meister, D. B. (2004). Knowledge sourcing effectiveness. *Management Science*, *50*(6), 821-834.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Hamann, J., Leucht, S., & Kissling, W. (2003). Shared decision making in psychiatry. *Acta Psychiatrica Scandinavica*, *107*(6), 403-409.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (2015). *Smart choices: A practical guide to making better decisions*. Harvard Business Review Press.

Heald, C. W., Kim, T., Sischo, W. M., Cooper, J. B., & Wolfgang, D. R. (2000). A computerized mastitis decision aid using farm-based records: An artificial neural network approach. *Journal of Dairy Science*, *83*(4), 711-720.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Sage Publications, Inc.

Henriksen, K., & Brady, J. (2013). The pursuit of better diagnostic performance: a human factors perspective. *BMJ quality & safety*, *22*(Suppl 2), ii1-ii5.

Hero, T., Gerhards, F., Thiart, H., Hellhammer, D. H., & Linden, M. (2012). Neuropattern: a new translational tool to detect and treat stress pathology. II. The Teltow study. *Stress*, *15*(5), 488-494.

Highhouse, S., Dalal, R. S., & Salas, E. (Eds.). (2013). *Judgment and decision making at work*. Routledge.

Hunink, M. M., Weinstein, M. C., Wittenberg, E., Drummond, M. F., Pliskin, J. S., Wong, J. B., & Glasziou, P. P. (2014). *Decision making in health and medicine: integrating evidence and values*. Cambridge University Press.

Iranitalab, A., & Khattak, A. (2017). Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis & Prevention*, *108*, 27-36.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1980). Prospect theory. *Econometrica*, *12*.

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (Vol. 2, p. 528). New York: Wiley.

Keeney, R. L., & Raiffa, H. (1976). Decision analysis with multiple conflicting objectives. *Wiley& Sons, New York*.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.

Kim, Y. S. (2008). Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size. Expert Systems with Applications, 34(2), 1227-1234.

Klein, K. J., & Kozlowski, S. W. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. Jossey-Bass.

Klein, K. J., Tosi, H., & Cannella Jr, A. A. (1999). Multilevel theory building: Benefits, barriers, and new developments. *Academy of Management review*, *24*(2), 248-253.

Kozlowski, S. W., Chao, G. T., Grand, J. A., Braun, M. T., & Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods*, *16*(4), 581-615.

Kozlowski, S. W., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes.

Kubovy, M., & Van Den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological review*, *115*(1), 131.

Lockhart, R. S., & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, *74*(2), 100.

Lusted, L. B. (1971). Signal detectability and medical decision-making. *Science*, *171*(3977), 1217-1219.

Mamede, S., Schmidt, H. G., & Rikers, R. (2007). Diagnostic errors and reflective practice in medicine. *Journal of evaluation in clinical practice*, *13*(1), 138-145.

March, J. G., & Simon, H. A. (1958). Organizations.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, *7*(1), 77-91.

Markowitz, H. M. (1991). Foundations of portfolio theory. *The journal of finance*, *46*(2), 469-477.

Markowitz, H. M. (2010). Portfolio theory: as I still see it. *Annu. Rev. Financ. Econ.*, *2*(1), 1-23.

McNicol, D. (2005). *A primer of signal detection theory*. Psychology Press.

Moulton, B., & King, J. S. (2010). Aligning ethics with medical decision-making: the quest for informed patient choice. *The Journal of Law, Medicine & Ethics*, *38*(1), 85-97.

Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for* *the social sciences*, *345*(368), 106-109.

Novis, D. A., Zarbo, R. J., & Valenstein, P. A. (1999). Diagnostic uncertainty expressed in prostate needle biopsies: a College of American Pathologists Q-Probes study of 15 753 prostate needle biopsies in 332 institutions. *Archives of Pathology and Laboratory Medicine*, *123*(8), 687-692.

O'Connor, A. M., Tugwell, P., Wells, G. A., Elmslie, T., Jolly, E., Hollingworth, G., ... & Mackenzie, T. (1998). Randomized trial of a portable, self-administered decision aid for postmenopausal women considering long-term preventive hormone therapy. *Medical Decision Making*, *18*(3), 295-303.

Orasanu, J., Calderwood, R., & Zsambok, C. E. (1993). *Decision making in action: Models and methods* (pp. 138-47). G. A. Klein (Ed.). Norwood, NJ: Ablex.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. *Decision making in action: Models and methods*, *1*, 3-20.

Pastore, R. E., & Scheirer, C. J. (1974). Signal detection theory: Considerations for general application. *Psychological Bulletin*, *81*(12), 945.

Peele, P. B., Siminoff, L. A., Xu, Y., & Ravdin, P. M. (2005). Decreased use of adjuvant breast cancer therapy in a randomized controlled trial of a decision aid with individualized risk information. *Medical Decision Making*, *25*(3), 301-307.

Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of management*, *38*(4), 1115-1166.

Plous, S. (1993). *The psychology of judgment and decision making*. Mcgraw-Hill Book Company.

Rendall, R., Pereira, A. C., & Reis, M. S. (2017). Advanced predictive methods for wine age prediction: Part I–A comparison study of single-block regression approaches based on variable selection, penalized regression, latent variables and tree-based ensemble methods. *Talanta*, *171*, 341-350.

Rode, C., Cosmides, L., Hell, W., & Tooby, J. (1999). When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition*, *72*(3), 269-304.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review*, *108*(2), 370.

Schroy III, P. C., Emmons, K., Peters, E., Glick, J. T., Robinson, P. A., Lydotes, M. A., ... & Prout, M. (2011). The impact of a novel computer-based decision aid on shared decision making for colorectal cancer screening: a randomized trial. *Medical Decision Making*, *31*(1), 93-107.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 623–656.

Sharpe, W. F. (1970). *Portfolio theory and capital markets*. McGraw-Hill College.

Shefrin, H., & Statman, M. (2000). Behavioral portfolio theory. *Journal of financial and quantitative analysis*, *35*(2), 127-151.

Shirangi, Mehrdad G.; Durlofsky, Louis J. (2016). "A general method to select representative models for decision making and optimization under uncertainty". *Computers & Geosciences*. **96**: 109–123.

Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American economic review*, *49*(3), 253-283.

Simon, H. A. (1987). Making management decisions: The role of intuition and emotion. *Academy of Management Perspectives*, *1*(1), 57-64.

Smith, W. P., Doctor, J., Meyer, J., Kalet, I. J., & Phillips, M. H. (2009). A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic Bayesian network and a Markov model. *Artificial intelligence in medicine*, *46*(2), 119-130.

Spackman, K. A. (1989, January). Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the sixth international workshop on Machine learning* (pp. 160-163). Morgan Kaufmann.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.

Swets, J. A. (1996). Signal Detection Theory in Psychology and Diagnostics: Collected Papers.

Tversky, A., & Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Tversky, A., & Kahneman, D. (1980). *The framing of decisions and the rationality of choice* (No. TR-2). Stanford University, CA, Department of Psychology.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied latent class analysis*, *11*, 89-106.

Weingarten, S., Riedinger, M., Conner, L., Johnson, B., & Ellrodt, A. G. (1994). Reducing lengths of stay in the coronary care unit with a practice guideline for patients with congestive heart failure: insights from a controlled clinical trial. *Medical care*, 1232-1243.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press, USA.

Xu, J. (1989). *Theory of multilevel methods* (Vol. 8924558). Ithaca, NY: Cornell University.

Yates, L. (1993). Feminism and Australian state policy: some questions for the 1990s. *Feminism and social justice in education: International perspectives*, 167-185.

Appendix A: R code for Analyses

```
#####################################################################################
#1- FIRST ANALYSIS
#Linear Regression
#IV = Continuous (Total number decisions)
#Variable Name: IVA (IV ALL)
#DV = Continuous (Total number correct diagnoses)
#####################################################################################

#Build function for multiple iterations
Reg1.f <- function(d, i){
 d2 <- d[i,]
 n1 = round(nrow(d2)*.7)
 n2 = round(nrow(d2)*.3)
 index1 <- sample(1:nrow(d2), n1, replace = TRUE)
 index2 <- sample(1:nrow(d2), n2, replace = TRUE)
 # Create a train and tests from the original data frame
 Training_dataset <- d2[index1, ] # subset data to training indices only
 Testing_dataset <-  d2[index2, ]# subset data to test indices only

 Dataframe <- data.frame(Training_dataset $DV, Training_dataset $IVA)
 reg1 <- lm(Dataframe$ Training_dataset.DV~Dataframe$ Training_dataset.IVA)
 Testing_dataset $DV <- as.numeric(Testing_dataset $DV)
 predictedvalue <- Testing_dataset $IVA*reg1$coefficients[2]+reg1$coefficients[1]

 dat.one <- data.frame(predictedvalue, Testing_dataset$DV)  # creating a matrix with test values
 rmat1 <- cor(dat.one)                    # find the correlation matrix
 if(sd(predictedvalue)==0) {rmat1[2, 1] <- 0}
 if(sd(Testing_dataset$DV)==0){rmat1[2, 1] <-0}
 r1 <-  rmat1[2, 1]                       # Pearson's r
 rsq1 <- r1^2                             # Sample r-squared value
 return(c(r1))
}
output1 <- matrix(0, 2000, 1)
for (i in 1:2000){
 output1[i, ] <- Reg1.f(dataset, )
}

stem(output1)
out1 <- data.frame(output1)
str(out1)
empmean1 <- mean(output1)
empsd1 <- sd(output1)
empLB1 <- empmean1 - 1.96*empsd1
empUB1 <- empmean1 + 1.96*empsd1
```

```
##################################################################################
# 2- SECOND ANALYSIS
#Linear Regression
#IV = Continuous (Total number Correct decisions)
#Variable Name: IVC (IV CORRECT)
#DV = Continuous (Total number correct diagnoses)
##################################################################################

#Build function for multiple iterations
Reg2.f <- function(d, i){
  d2 <- d[i,]
  n1 = round(nrow(d2)*.7)
  n2 = round(nrow(d2)*.3)
  index1 <- sample(1:nrow(d2), n1, replace = TRUE)
  index2 <- sample(1:nrow(d2), n2, replace = TRUE)
  # Create a train and tests from the original data frame
  Training_dataset  <- d2[index1, ] # subset data to training indices only
  Testing_dataset  <-  d2[index2, ]# subset data to test indices only

  Dataframe <- data.frame(Training_dataset $DV, Training_dataset $IVC)
  reg2 <- lm(Dataframe$Training_dataset .DV~Dataframe$Training_dataset .IVC)
  Testing_dataset $DV <- as.numeric(Testing_dataset $DV)
  predictedvalue <- Testing_dataset $IVC*reg2$coefficients[2]+reg2$coefficients[1]

 dat.two <- data.frame(predictedvalue, Testing_dataset $DV)  # creating a matrix with test values
  rmat2 <- cor(dat.two)                    # find the correlation matrix
  if(sd(predictedvalue)==0) {rmat5[2, 1] <- 0}
  if(sd(Testing_dataset $DV)==0){rmat2[2, 1] <-0}
  r2 <-  rmat2[2, 1]                       # Pearson's r
  rsq2 <- r2^2                             # Sample r-squared value
  return(c(r2))
}
output2 <- matrix(0, 2000, 1)
for (i in 1:2000){
  output2[i, ] <- Reg2.f(Dataset, )
}

stem(output2)
out2 <- data.frame(output2)
str(out2)
empmean2 <- mean(output2)
empsd2 <- sd(output2)
empLB2 <- empmean2 - 1.96*empsd2
empUB2 <- empmean2 + 1.96*empsd2
```

```
#####################################################################################
# 3- THIRD ANALYSIS
#Cluster analysis w/ ANOVA
#IV = Categorical (Grouping variable determined by k-means clustering)
#DV = Continuous (Total number correct diagnoses)
#####################################################################################

###Determine total number of clusters for analysis
#Determine number of clusters by computing gap statistic
clusterstat <- clusGap(Dataset, kmeans, 10, B = 2000, verbose = interactive())
gap <- clusterstat$Tab[, 3]
gap.se <- clusterstat$Tab[, 4]
gapstat <- maxSE(gap, gap.se)

#Run k-means cluster analysis
kmodel <- kmeans(Dataset, centers = gapstat)
kclust <- kmodel$cluster
kmeansresults <- mutate(Dataset, cluster = kclust)
DatasetClust <- mutate(kmeansresults, DV = Dataset$DV)

#Conduct ANOVA with K-means results as IV (GROUP/grouping variable)
kmeansANOVA <- aov(Dataset$DV~kmeansresults$cluster)
summary(kmeansANOVA)

#Build function for multiple iterations
AOV3.f <- function(d, i){
  d2 <- d[i,]
  n1 = round(nrow(d2)*.7)
  n2 = round(nrow(d2)*.3)
  index1 <- sample(1:nrow(d2), n1, replace = TRUE)
  index2 <- sample(1:nrow(d2), n2, replace = TRUE)
  # Create a train and tests from the original data frame
  Training_dataset  <- d2[index1, ] # subset data to training indices only
  Testing_dataset  <-  d2[index2, ]# subset data to test indices only

  Dataframe <- data.frame(Training_dataset $DV, Training_dataset $cluster)
  aov3 <- aov(Dataframe$Training_dataset .DV~Dataframe$Training_dataset .cluster)
  Testing_dataset $DV <- as.numeric(Testing_dataset $DV)
  predictedvalue <- Testing_dataset $cluster*aov3$coefficients[2]+aov3$coefficients[1]

dat.three <- data.frame(predictedvalue, Testing_dataset $DV)  # creating a matrix with test
values
  rmat3 <- cor(dat.three)                    # find the correlation matrix
  if(sd(predictedvalue)==0) {rmat3[2, 1] <- 0}
  if(sd(Testing_dataset $DV)==0){rmat3[2, 1] <-0}
  r3 <-  rmat3[2, 1]                         # Pearson's r
```

```
  rsq3 <- r3^2                          # Sample r-squared value
  return(c(r3))
}
output3 <- matrix(0, 2000, 1)
for (i in 1:2000){
  output3[i, ] <- AOV3.f(NightsweatsClust, )
}

stem(output3)
out3 <- data.frame(output3)
str(out3)
empmean3 <- mean(output3)
empsd3 <- sd(output3)
empLB3 <- empmean3 - 1.96*empsd3
empUB3 <- empmean3 + 1.96*empsd3
```

```
####################################################################################
# 4- FOURTH ANALYSIS
#Latent Class Analysis (LCA)
#IV = Categorical (Grouping variable determined by LCA)
#DV = Continuous (Total number correct diagnoses)
####################################################################################

#Determine number of clusters by computing fit using BIC
listnumbergroups <- fitLCA(Dataset, G = 1:10, X = NULL, ctrlLCA = controlLCA())
numbergroups <- as.numeric(as.character(unlist(listnumbergroups[[1]])))

#Define variables used
lcavaruse <- with(Dataset, cbind(list of variable names[3])~1)
lcamodel <- poLCA(lcavaruse, data=Dataset, nclass=numbergroups)
lcaclust <- lcamodel$predclass
lcaresults <- mutate(Dataset, predclass = lcaclust)
Dataset <- mutate(lcaresults, DV = Nightsweats$DV)

#Conduct ANOVA with LCA results as IV (GROUP/grouping variable)
lcaANOVA <- aov(Nightsweats$DV~lcaresults$predclass)
summary(lcaANOVA)

#Build function for multiple iterations
AOV4.f <- function(d, i){
  d2 <- d[i,]
  n1 = round(nrow(d2)*.7)
  n2 = round(nrow(d2)*.3)
  index1 <- sample(1:nrow(d2), n1, replace = TRUE)
  index2 <- sample(1:nrow(d2), n2, replace = TRUE)
  # Create a train and tests from the original data frame
  Training_dataset  <- d2[index1, ] # subset data to training indices only
  Testing_dataset  <-  d2[index2, ]# subset data to test indices only

  Dataframe <- data.frame(Training_dataset $DV, Training_dataset $predclass)
  aov4 <- aov(Dataframe$Training_dataset .DV~Dataframe$Training_dataset .predclass)
  Testing_dataset $DV <- as.numeric(Testing_dataset $DV)
  predictedvalue <- Testing_dataset $predclass*aov4$coefficients[2]+aov4$coefficients[1]

  dat.four <- data.frame(predictedvalue, Testing_dataset $DV)  # creating a matrix with test
values
  rmat4 <- cor(dat.four)                    # find the correlation matrix
  if(sd(predictedvalue)==0) {rmat4[2, 1] <- 0}
  if(sd(Testing_dataset $DV)==0){rmat4[2, 1] <-0}
  r4 <-  rmat4[2, 1]                        # Pearson's r
  rsq4 <- r4^2                              # Sample r-squared value
```

[3] All variables used for LCA should be listed here

```
  return(c(r4))
}
output4 <- matrix(0, 2000, 1)
for (i in 1:2000){
  output4[i, ] <- AOV4.f(Dataset, )
}

stem(output4)
out4 <- data.frame(output4)
str(out4)
empmean4 <- mean(output4)
empsd4 <- sd(output4)
empLB4 <- empmean4 - 1.96*empsd4
empUB4 <- empmean4 + 1.96*empsd4
```

```
##################################################################################
# 5- FIFTH ANALYSIS
#Simple Regression Trees
#Root node = DV
#DV = Continuous (Total number correct diagnoses)
##################################################################################

#Split data so that it is partitioned into 3 groups:
# 1. training set (70% split off from main sample to be used to train model)
# 2. validation set (15% split off from main sample to be used to validate model)
# 3. testing set (15% split off from main sample to be used to test model)

# Look at the data and build function
str(DatasetTrees)
Lotsa.trees.f <- function(d, i){
  d2 <- d[i,]
  #  d2 <- DatasetTrees
  # Set seed and create assignment
  # set.seed(1)
  n1 = round(nrow(d2)*.7)
  n2 = round(nrow(d2)*.15)
  n3 = round(nrow(d2)*.15)
  index1 <- sample(1:nrow(d2), n1, replace = TRUE)
  index2 <- sample(1:nrow(d2), n2, replace = TRUE)
  index3 <- sample(1:nrow(d2), n3, replace = TRUE)
  # Create a train, validation and tests from the original data frame
  Training_dataset <- d2[index1, ] # subset data to training indices only
  Validation_dataset <- d2[index2, ]  # subset data to validation indices only
  Testing_dataset <-  d2[index3, ]# subset data to test indices only


  ### Train the model
  simplemodel <- rpart(formula = DV ~ .,
                    data = Training_dataset,
                    method = "anova")

 ### Hypertuning and validating model
  #Hypertuning with calculation of optimal minsplit and maxdepth
  #minsplit = minimum number of datapoints needed to create leaf
  #maxdepth = maximum number of branches in tree
  #Start these calculations by creating base hypergrid to grid data onto

  # Establish a list of possible values for minsplit and maxdepth
  minsplit <- seq(1, 4, 1)
  maxdepth <- seq(1, 6, 1)
```

```
# Create a data frame containing all combinations
hyper_grid <- expand.grid(minsplit = minsplit, maxdepth = maxdepth)
#Continue calculations for minsplit/maxdepth- grid training data onto base hypergrid
# Number of potential models in the grid
num_models <- nrow(hyper_grid)
# Create an empty list to store models
Dataset_models <- list()

# Write a loop over the rows of hyper_grid to train the grid of models
for (i in 1:num_models) {
  # Get minsplit, maxdepth values at row i
  minsplit <- hyper_grid$minsplit[i]
  maxdepth <- hyper_grid$maxdepth[i]
    # Train a model and store in the list
  Dataset_models[[i]] <- rpart(formula = DV ~ .,
                              data = Training_dataset,
                              method = "anova",
                              minsplit = minsplit,
                              maxdepth = maxdepth)
}


###Validate model using newly tuned trained model
# Number of potential models in the grid
num_models <- length(Dataset_models)
# Create an empty vector to store RMSE values
rmse_values <- c()

# Write a loop over the models to compute validation RMSE
for (i in 1:num_models) {
  # Retrieve the i^th model from the list
  model <- Dataset_models[[i]]
  # Generate predictions on grade_valid
  pred <- predict(object = model,
                 newdata = Validation_dataset)
  # Compute validation RMSE and add to the
  rmse_values[i] <- rmse(actual = Validation_dataset$DV,
                         predicted = pred)
}

# Identify the model with smallest validation set RMSE
best_model <- Dataset_models[[which.min(rmse_values)]]
# Compute test set RMSE on best_model
pred <- predict(object = best_model,
               newdata = Testing_dataset)
simpleRMSE <- rmse(actual = Testing_dataset$DV,
```

```
                    predicted = pred)

  # Calculate r
  dat.five <- cbind(pred, Testing_dataset$DV)  # creating a matrix with test values
  rmat5 <- cor(dat.five)                    # find the correlation matrix
  if(sd(pred)==0) {rmat5[2, 1] <- 0}
  if(sd(Testing_dataset$DV)==0){rmat5[2, 1] <-0}
  r5 <-  rmat5[2, 1]                        # Pearson's r
  rsq5 <- r5^2                              # Sample r-squared value
  return(c(r5))
}
output5 <- matrix(0, 2000, 1)
for (i in 1:2000){
  output5[i, ] <- Lotsa.trees.f(DatasetTrees, )
}
empmean5 <- mean(output5)
empsd5 <- sd(output5)
empLB5 <- empmean5 - 1.96*empsd5
empUB5 <- empmean5 + 1.96*empsd5
```

```
###################################################################################
# 6- SIXTH ANALYSIS
#Bagged Regression Trees
#Root node = DV
#DV = Continuous (Total number correct diagnoses)
###################################################################################

#Build function
Trees.six.f <- function(d, i){
  d3 <- d[i,]
  baggedassignment <- sample(1:2, size = nrow(d3), prob = c(0.7, 0.3), replace = TRUE)

  # Create a train and test datasets from the original data frame
  Training_dataset <- d3[baggedassignment == 1, ]  # subset data to training indices only
  Testing_dataset <- d3[baggedassignment == 2, ]  # subset data to test indices only

  # Train bagged model
  # Default set to 25 iterations, change using nbagg if needed at later time
  trainbaggedmodel <- train(DV ~ .,
                         data = Training_dataset,
                         method = "treebag",
                         trControl = trainControl(method = "cv", number = 10),
                         nbagg = 200)


  # Generate predicted classes using the model object
  baggedclasspredict <- predict(trainbaggedmodel, Testing_dataset)
  dat.six <- cbind(baggedclasspredict, Testing_dataset$DV)  # creating a matrix with test values
  rmat6 <- cor(dat.six)    # find the correlation matrix
  r6 <-  rmat6[2, 1]                        # Pearson's r
  rsq6 <- r6^2
  # Calculate final RMSE
  baggedRMSE <- RMSE(baggedclasspredict, Testing_dataset$DV)
  # print(baggedRMSE)
  return(r6)
}
output6 <- matrix(0, 2000, 1)
for (i in 1:2000){
  output6[i, ] <- Trees.six.f(DatasetTrees, )
}
empmean6 <- mean(output6)
empsd6 <- sd(output6)
empLB6 <- empmean6 - 1.96*empsd6
empUB6 <- empmean6 + 1.96*empsd6
```

```
################################################################################
# 7- SEVENTH ANALYSIS
#Random Forests
#Root node = DV
#DV = Continuous (Total number correct diagnoses)
################################################################################

library(randomForest)
# Train a Random Forest
#set.seed(1)  # for reproducibility
Trees.seven.f <- function(d, i){
  d4 <- d[i,]
  baggedassignment <- sample(1:2, size = nrow(d4), prob = c(0.7, 0.3), replace = TRUE)
  Training_dataset <- d4[baggedassignment == 1, ]  # subset data to training indices only
  Testing_dataset <- d4[baggedassignment == 2, ]  # subset data to test indices only

  randommodel <- randomForest(formula = DV ~ .,
                              data = Training_dataset)

 #Calculate RMSE
  randomMSE <- which.min(randommodel$mse)
  randomRMSE <- sqrt(randommodel$mse[which.min(randommodel$mse)])

  #Tune random forest
   res <- tuneRF(x = subset(Training_dataset, select = -DV),
               y = Training_dataset$DV,
               ntreeTry = 2000,
               doBest = TRUE)

  #Calculate final RMSE
  rmsenew <- sqrt(res$mse[which.min(res$mse)])

  randompredict <- predict(res, Testing_dataset)

  finalrandomRMSE <- RMSE(randompredict, Testing_dataset$DV)

  #Calculate r
  dat.seven <- cbind(randompredict, Testing_dataset$DV)  # creating a matrix with test values
  rmat7 <- cor(dat.seven)    # find the correlation matrix
  r7 <-  rmat7[2, 1]                    # Pearson's r
  rsq7 <- r7^2
  return(r7)
}
output7 <- matrix(0, 2000, 1)
for (i in 1:2000){
  output7[i, ] <- Trees.seven.f(DatasetTrees, )
```

```
}
#
empmean7 <- mean(output7)
empsd7 <- sd(output7)
empLB7 <- empmean7 - 1.96*empsd7
empUB7 <- empmean7 + 1.96*empsd7
```

Appendix B: Medical School Rubrics

Alcohol Dependency

General Description	Percentage[4]
Core Symptom: discuss alcohol consumption, and the possibility of quitting	97%
Age	97%
Sex	94%
Brought in by son/daughter	91%

History	Percentage
Drinks 3-4 scotches daily past few years	78%
Started drinking in his/her teens	32%
Recently had license revoked for driving under the influence	60%
Living with son/daughter	60%
Held 3 different jobs	1%
Working in security over past two years	23%
Problems getting to work on time	16%
CAGE screening: 3/4 (CAGE is an alcoholism screening that has 4 Q's in it)	10%
Patient has tried to cut down on drinking in the past when hospitalized for liver and pancreas problems	50%
Patient is annoyed that his/her children want him/her to cut down on drinking	34%
Occasionally has a drink first thing in the morning	29%

Positive Symptom History	Percentage
Bruising with minimal trauma	40%
Intermittent epistaxis	6%

[4] Percentage of medical students who noted each item during history/physical examination

Negative Symptom History	Percentage
No jaundice	1%
No increasing abdominal girth	4%
No anxiety	1%
No depression	3%

Past Medical History	Percentage
No prior diagnosis (dx) liver disease	2%
No prior diagnosis (dx) of alcohol use disorder	1%
No prior diagnosis (dx) bleeding disorder	1%
Hospitalization for abdominal pain 2 years ago	55%
Told liver and pancreas inflammation	81%
Medication: • Omeprazole 20 mg daily	28%
• Allupurinol 300 mg daily	41%
Other illness: • Gout for 10 years	51%
• Peptic ulcer disease diagnosed (dx) 5 years ago	37%
No Allergies	34%
Family history: • Father: died of liver failure and alcohol use at age 55	34%

Social-Factors History	Percentage
Alcohol: 3-4 scotches per day for past few years	80%
Tobacco: 35 pack year	34%
No illicit drugs intake (cocaine)	34%

Physical Examination	Percentage
General Appearance: • The patient is unkempt, with disheveled clothing	14%
• Alert and oriented	1%
• Patient is somewhat angry about questioning, but cooperative	3%
Vital Signs (VS): • BP: 120/80	15%
• P: 80	13%
• T: 98.0	16%
• RR: 14/min	14%
Skin: • Yellowish	66%
• Scattered bruises on the forearms	50%
• Telangiectasias on the neck, palms, and palmar erythema	61%

Possible Diagnoses[5]	Percentage
Alcohol Use Disorder	79%
Liver Disease	81%

[5] All diagnoses were considered equally viable options. The dependent variable for the study was computed using the overall quantity of correct potential diagnoses made by a given participant within their scenario.

Dizziness

General Description	Percentage
Core Symptom: intermittent dizziness	89%
Age	90%
Sex	90%

History	Percentage
Duration of Symptoms: 2 days	94%
Onset/precipitating factors: • Seems to occur with getting up from a lying position	68%
• Could occur at any time	50%
• No head trauma	23%
Character/quality: • Sensation of room spinning around	68%
Aggravating/relieving factors: • Seems to occur with getting up from a lying position	68%
• No specific relieving factors	31%
Duration: • Episodes last 20-30 min	44%
• Symptoms are progressive	36%
• No previous episodes	23%
Functional: • Causes patient to feel unsteady	28%

Positive Symptom History	Percentage
Decreased hearing left ear x1 day	94%
Feels unsteady, however no falls	27%
Nausea	76%
Vomited several times since yesterday	79%

Negative Symptom History	Percentage
No fever	34%
No weight loss	11%
No pain, fullness, or discharge from the ears	23%
No sore throat, sinus drainage, cough, upper respiratory infection symptoms	39%
No headache	49%
No numbness, tingling, weakness, tremor, visual disturbance, or other neurological deficits	59%
No heat or cold intolerance	1%

Past Medical History	Percentage
Hypertension for three years	90%
No malignancy (i.e. worsening of symptoms)	0%
No strokes	3%
No previous ear problems	4%
Diuretic medications taken previously	84%
No allergies	53%
Family history: • Father (age 76), mother (age 78) and brother alive and well	13%
• No Meniere's disease, no hearing, or vertigo problems in family history	9%

Social-Factors History	Percentage
No alcohol problems	72%
No tobacco intake	70%
No illicit drugs intake (cocaine)	57%
No exposure to ill persons or persons with similar symptoms	9%

Physical Examination	Percentage
General Appearance: Anxious appearing male in no acute distress	53%
Vital Signs (VS): • P 80/min	10%
• BP 135/90 Lying	6%
• P 85/min	38%
• BP 130/90 Sitting	53%
• T 98.6	77%
• RR 16/min	71%
• No orthostasis	1%
Throat/Ear Examination: • HEENT: NC/AT, PERRLA, EOMI without nystagmus.	53%
• Normal external ear canal bilat.	18%
• Normal TM with normal light reflex.	59%
• Mouth and oropharynx normal	14%
Lungs • Normal breathing sounds	74%
Cardiac: • Normal s1 s2, regular rhythm	86%
• No murmurs, rubs, gallops	67%
Neuro: • Cranial nerves II-IX grossly intact except decreased hearing in left ear	58%

• Rinne - normal: air conduction>bone conduction in left ear	21%
• Weber lateralizes to right ear	25%
Motor strength • 5/5 D	30%
Sensory Testing • Normal to light touch	21%
Reflexes: • DTRs 2+ symmetric	17%
• Babinski negative bilaterally	3%
Cerebellar: • finger to nose	9%
• heel to shin normal	6%
Gait: • normal • Romberg normal	32%

Possible Diagnoses	Percentage
Meniere's Disease	68%
Acoustic neuroma or other intracranial lesion	17%
Vestibular neuritis	29%
Benign Paroxysmal vertigo	72%

Fatigue

General Description	Percentage
Core Symptom: chronic fatigue	98%
Age	96%
Sex	96%

History	Percentage
Duration of Symptoms: 3 months	95%
Onset/precipitating factors: • very gradual onset	17%
• no precipitating event	27%
Frequency: Daily	3%
Aggravating factors • No, not specifically worse any time of the day	10%
Relieving factors • None, it is constant	15%
Severity: Severe	1%
Functional: • no longer want to go on my 30 minute daily walks with dog	29%
• it is a struggle to get through work	2%
• decreased sexual activity with spouse since so fatigued	5%

Positive Symptom History	Percentage
Unintentional weight loss 10 lb over several months	69%
Poor sleep, wakes frequently at night, snores	76%
SIGECAPS (i.e. major depression): • depressed mood for 3 mos	54%
• some loss of interest in activities	55%
• poor concentration	50%
• poor appetite	63%
Endocrine: • cold intolerance	37%
• dry skin	38%
• hair thinning	38%
GI: some occasional abdominal cramping x 1 month	20%

Negative Symptom History	Percentage
No fever	38%
No chills	22%
No night sweats	18%
SIGECAPS: • no feeling of guilt or worthlessness	22%
• not suicidal	33%
Endocrine: • No hoarseness	0%
Cardiopulmonary: • no SOB (shortness of breath)	47%
• no chest pain or chest pressure	39%
Gastrointestinal: • no brbpr (bright red blood per rectum)	45%

• no diarrhea	49%
• no melena	44%
• no nausea	34%
• no vomiting	26%

Past Medical History	Percentage
Colon polyp at age 60 years	40%
Overdue for colonoscopy	3%
Health maintenance: • if female - mammogram up to date • if male - DRE up to date	9%
No malignancy	1%
No past surgical history	51%
No allergies	59%
Family History: • Father died in a car accident at age 60	14%
• Mother died at age 80 after a hip fracture	15%
• No malignancy, colon cancer, lung cancer	15%

Social-Factors History	Percentage
No alcohol	72%
Tobacco: 1 pack per day (PPDx) for 40 years	58%
• Quit 1 year ago	47%
No illicit drugs intake (cocaine)	67%
Above has caused functional limitations	0%

Physical Examination	Percentage
General appearance: • Tired appearing	11%
• Patient is in discomfort	3%
Vital Signs (VS): • P 80/min	64%
• BP 110/70	63%
• T 98.6	79%
• RR 13/min	61%
HEENT (head, ears, eyes, nose, throat): • Normal oral mucosa	27%
Thyroid: • Non palpable	45%
Lymph nodes: • No cervical, supraclavicular, periumbilical, or inguinal adenopathy	44%
Abdomen: • normal BS, soft, nontender	63%
• no hepatosplenomegaly	10%
• no rebound tenderness	3%
Extremities: • no cyanosis	8%
• clubbing edema	28%

Possible Diagnoses	Percentage
Anemia (due to colorectal cancer, or other)	64%
Depression	66%
Hypothyroidism	81%
Lung cancer	19%

Hematuria

General Description	Percentage
Core Symptom: Painless hematuria with clots	99%
Age	98%
Sex	97%
Duration of Symptoms: this morning	77%

History	Percentage
Onset/precipitating factors: • No trauma	27%
• No vigorous physical activity	6%
Duration: one episode this morning	4%
Character/quality: • Painless	33%
• Grossly visible red blood with clots	72%
Frequency: once	5%
Aggravating factors: • None	3%
Relieving factors: • None	3%
Timing • The hematuria was noted throughout without the voiding	6%
• No prior episode	51%

Positive Symptom History	Percentage
Genitourinary (GU): • Progressive difficulty initiating urinary stream	42%
• Decreased strength of stream	21%
• Increased frequency of urination	49%
• Less volume during urination	28%
• Nocturia past several months	28%
Genitourinary (GU) cont: • Dribbling	34%
• Urgency	35%

Negative Symptom History	Percentage
General: • No fever	73%
• No chills	48%
• No fatigue	16%
• No weight loss or change in appetite	57%
Genitourinary (GU): • No dysuria, or foul smelling urine	66%
• No impotence or hematospermia	7%
Genitourinary (GU) cont: • No penile discharge	12%
• No flank pain	20%
Gastrointestinal (GI): • No abdominal pain	41%
• No nausea or vomiting or diarrhea	34%

Past Medical History	Percentage
Benign prostatic hyperplasia (BPH) for 3 years	98%
No history (hx) of prostate cancer	6%
No history (hx) of bladder cancer	6%
No history (hx) of renal disease	21%
No UTIs (urinary tract infections) or STI (sexually transmitted illness/disease)	12%
No hypertension (HTN)	2%
Medication: • Terazosin 20 mg po (taken orally) qhs (every night)	74%
No allergies	64%
Family history: • Mother with kidney disease died age 75	67%
• Dad died age 65 of lung cancer	64%

Social-Factors History	Percentage
No alcohol problems	80%
Tobacco: cigarettes 1 pack per day (PPDx) for 30 years	92%
No illicit drugs intake (cocaine)	70%
Works as a painter for 35 years	41%

Physical Examination	Percentage
General Appearance: patient appears well in no acute distress	66%
Vital Signs (VS): • BP: 140/80	69%
• P: 80	61%
• T: 98.0	61%
• RR: 14/min	59%
Cardiac: • RRR (regular rate and rhythm) normal S1, S2	92%
Abdomen: • Normal BS soft nontender	84%
• No HS megaly	21%
• No suprapubic tenderness	15%
• Mild right CVA (costovertebral angle) tenderness	35%
• No CVA (costovertebral angle) tenderness on left	20%
Extremities: • No clubbing	6%
• No edema	27%

Possible Diagnoses	Percentage
Bladder cancer	89%
Urolithiasis	32%
Urinary tract infection (UTI)	33%
Prostate cancer	38%
Renal cancer	45%
Polycystic kidney disease	3%

Hoarseness

General Description	Percentage
Core Symptom: progressive hoarseness	100%
Age	89%
Sex	85%
Duration of Symptoms: 4 months	97%
Notation whether patient is Smoker	94%

History	Percentage
Onset/precipitating factors: • Teacher overuses voice	26%
• No recent upper respiratory tract infections (URI or URTI)	7%
Character/quality: • Prior intermittent hoarseness she attributed to overused voice	16%
Aggravating/relieving factors: • None	13%
Duration: • 4 months	35%
• Progressive worsening over 4 months	51%

Positive Symptom History	Percentage
Weight loss 10 lbs over 3 months	74%
Sensation of lump in throat	63%
Progressive nonproductive cough over the past month	40%
Poor appetite	29%

Negative Symptom History	Percentage
No fever	65%
No fatigue	21%
No sore throat, sinus drainage, no upper respiratory tract infections (URI or URTI) symptoms (sx)	23%
No hemoptysis	17%
No wheezing, chest pain, shortness of breath (SOB) stridor	60%
No dysphagia, odonophagia, reflux	65%
No numbness, tingling, weakness, tremor, visual disturbance	10%
No heat or cold intolerance	19%

Past Medical History	Percentage
No malignancy	0%
No thyroid disease	2%
Positive gastroesophageal reflux disease (GERD)	12%
Medication: Omeprazole 20 mg daily	9%
Gastroesophageal reflux disease (GERD) controlled with Omeprazole	2%
No allergies	63%
Family history: • Father died of lung cancer at age 55	75%

Social-Factors History	Percentage
No alcohol problems	69%
Tobacco intake: 1-2 pack per day (PPDx) for 40 years	95%
No illicit drugs intake (cocaine)	60%

Physical Examination	Percentage
Vital Signs (VS): • P: 85/min	70%
• BP: 130/80	71%
• T: 98.6	76%
Throat/Ear Examination: • HEENT: Normal oral mucosa	20%
• Pharynx without erythema, or exudate	62%
• No sinus tenderness	5%
• Thyroid nonpalpable	62%
• No cervical axillary, or supraclavicular lymphadenopathy	60%
Lungs: • Clear to auscultation	86%
Cardiac: • normal s1 s2, regular rhythm	83%
• No murmurs, rubs, gallops (MRG)	69%
Abdomen: • Normal breath sound (BS)	20%
• Soft, nontender	33%
• No hepatosplenomegaly	7%
• No masses	6%
Neuro: • Speech normal	1%
• Voice hoarse	33%
Reflexes: • 2+ symmetric	9%
• Muscle 5/5 throughout	0%

Possible Diagnoses	Percentage
Laryngeal cancer	44%
Lung cancer with recurrent laryngeal nerve involvement	44%
Chronic Laryngitis due to irritant (smoking or voice overuse)	14%
Vocal cord nodule	7%

Low Back Pain

General Description	Percentage
Core Symptom: low back pain	100%
Age	97%
Sex	98%
Duration of Symptoms: 1 week	96%

History	Percentage
Onset/precipitating factors: • Lifting boxes	96%
Location/radiation: • Lumbar, occasional radiation to posterior thighs	80%
Duration: 1 week	94%
Character or quality: • Achy in quality	35%
• At its worse, will be a sharp pain	56%
Frequency: pain is constantly present	45%
Aggravating factors: • Bending	23%
• Lifting	12%
• Leaning forward	33%
• Walking and standing	37%
Relieving factors: • Lying down with hips and knees bent with a pillow under the knees = some relief	49%
• Aleve = some relief	83%
Severity: • 8/10 at max, 3-4/10 min	60%

• Red flag sign: Sometimes has pain at night, but does not waken patient from sleep	6%
Prior back pain: • Past year, back pain radiating into thighs	36%
• Relieved with rest	21%
• Relieved with leaning forward	2%
• This occurs with prolonged walking	13%
Functions: • Back pain limiting sexual activity	2%

Positive Symptom History	Percentage
Weight loss 5 lb in the last few months, no change in appetite	31%

Negative Symptom History	Percentage
No fever	39%
No chills	25%
No numbness of the LE	32%
No saddle anesthesia	21%
No weakness of the LE	34%
No tingling LE	29%
No urinary incontinence, no dysuria, no hematuria	62%
No fecal incontinence	56%

Past Medical History	Percentage
Enlarged prostate diagnosis (dx) a few years ago Does not regularly get evaluated	79%
Health maintenance: • Last colonoscopy was at age 50	1%
No malignancy	21%
No prior back injuries, degenerative joint disease (DJD), or other back conditions	6%
Medication: • OTC (over the counter) Aleve 2 or 3 per day	35%
No allergies	59%
Family history: • Father: died at age 70 of colon cancer	66%
• Mother: died at age 80 of heart problems	39%
• Son: age 40, alive and well	2%
• Daughter: age 45, alive and well	2%
• No siblings	0%

Social-Factors History	Percentage
Alcohol: 3 martinis per day	69%
No tobacco use	71%
No illicit drugs intake (cocaine)	60%
Works in construction, involves a lot of physical exertion, including heavy lifting	52%
CAGE: one positive (CAGE is an alcoholism screening that has 4 Q's in it)	7%

Physical Examination	Percentage
General Appearance: well-nourished male, appears uncomfortable due to the back pain	31%

Vital Signs (VS): • BP: 136/84	79%
• P: 86	79%
• T: 98.6	79%
• RR: 12/min	79%
Abdomen: • Normal BS soft nontender	25%
• No suprapubic tenderness	0%
• No CVA (costovertebral angle) tenderness	5%
• No hepatosplenomegaly	2%
Musculoskeletal: • Lumbar spine: o Tenderness to palpation over the lower lumbar midline, and paraspinal bilaterally	84%
o Pain elicited in the low back, and posterior thighs with SLR (straight leg raise) bilaterally	49%
o Pain does not radiate past knee	1%
o No tenderness SIJ (Sacroiliac joint)	0%
Neuro: • Normal sensory exam LE	42%
• LE strength 5/5 hip, knee, ankle flexion/extension	31%
• Reflexes 2+ knee, ankle	29%
Extremities: • No cyanosis, clubbing edema	7%
• Hips no tenderness trochanter, FROM	0%
• Knees nontender	0%
• No effusion, FROM	1%

Possible Diagnoses	Percentage
Lumbar strain	55%
Metastatic prostate cancer to lumbar spine	42%
Lumbar spinal stenosis with exacerbation	20%
Lumbar osteoporotic compression fracture	39%
Lumbar herniated disc	68%

Night Sweats

General Description	Percentage
Core Symptom: night sweats and 10 lb weight loss over past month	100%
Age	96%
Gender: M	95%
Key modifier: homosexual	66%

History	Percentage
Duration of Symptoms: 1 month	96%
Onset/precipitating factors: • no known exposure to ill persons	29%
• previous travel to Dominican Republic	65%
Frequency: every night	39%
Aggravating factors • Only occur at night	18%
Relieving factors • Nothing	11%
Timing: never previous similar sx (symptoms)	21%
Severity: clothes, and the bed are wet with sweat	55%

Positive Symptom History	Percentage
Fatigue	50%
Possible fever, feels warm, did not take temp at home, no chills	58%
Unintentional weight loss 10 lb over 1 month	75%
Endocrine: unsure if heat intolerance is present	3%

Negative Symptom History	Percentage
No change appetite	19%
Skin: • No rashes	22%
Cardiopulmonary: • no cough	72%
• no chest pain	39%
• no hemoptysis	21%
Cardiopulmonary: • no SOB (shortness of breath)	55%
• no chest pain or chest pressure	1%
GI (Gastrointestinal): • No abdominal pain, N,V, diarrhea, melena, or hematochezia	59%
GU (Genitourinary): • No dysuria, hematuria	33%
MSK (musculoskeletal): • No myalgias, arthralgias	7%
Endocrine: • no cold intolerance	4%
• no increased thirst	1%
• no excess urination	3%
Heme: no increased bleeding, bruising	9%

Past Medical History	Percentage
No history of hepatitis	0%
Vaccinated for hepatitis B	2%
Never tested for HIV	10%
No history of STIs	20%
No history of TB or known TB exposure	23%
Past medication use: Tylenol prn	36%
No allergies	65%
Family history: • Mom (50 years) alive and well	7%
• Dad (52 years) alive and well	7%
• No malignancy	16%

Social-Factors History	Percentage
No alcohol	64%
No tobacco	77%
No illicit drugs intake (cocaine)	73%
No exposure to ill persons or persons with similar symptoms	28%
Medical student	50%
Sexually active - 1 male partner who uses protection	73%
No known HIV exposure	4%
Travel to Dominican Republic on a medical mission trip 2 months ago	52%

Physical Examination	Percentage
General Appearance: well kempt male in NAD (no abnormality detected)	61%
Vital Signs (VS): • P: 90	82%
• T: 100.8	90%
• RR: 14/min	80%
Skin: • No rashes	21%
• No lesions	8%
HEENT: • normal sclera	13%
• no exophthalmos	8%
• oral mucosa without thrush or oral ulcers	43%
• thyroid not palpable	18%
No cervical, supraclavicular, axillary adenopathy	67%
Lungs: clear to A/P	93%
Cardiac: normal S1, S2	91%
Abdomen: • normal BS soft nontender	65%
• no HS megaly	31%
• no masses	11%
Extremities: • no clubbing	6%
• no edema	20%
Neuro: • biceps and patellar reflexes 2+ bilaterally	5%
• no tremor	1%

Possible Diagnoses	Percentage
Lymphoma	42%
TB	72%
HIV	63%
Hyperthyroidism	13%