The Effects of Specific Support to Hypothesis Generation on the Diagnostic Performance of Medical Students

Carlos Nakamura

Department of Educational and Counselling Psychology

McGill University, Montreal

A thesis submitted to the

Faculty of Graduate Studies and Research

in partial fulfillment of the requirements of

the degree of Doctor of Philosophy

November 2006

Your file Votre référence
ISBN: 978-0-494-32223-9
Our file Notre référence
ISBN: 978-0-494-32223-9

Acknowledgements

I wish to thank my supervisor, Dr. Susanne P. Lajoie, for her unconditional support throughout my Ph.D. This thesis would not be possible without her constant guidance and encouragement. My thanks to the members of my Thesis Advisory Committee, Dr. Alenoush Saroyan and Dr. Robert J. Bracewell for their insightful feedback. My thanks to my colleagues, Andrew Chiarella and Dr. Gloria Berdugo, for all their academic and nonacademic support. Mr. Chiarella programmed the two computer applications that were used in this study. My thanks to Sandrine Turcotte for translating the abstract. I extend my gratitude to all members of the Applied Cognitive Science Research Group.

Abstract

The hypothetico-deductive method, which involves an iterative process of hypothesis generation and evaluation, has been used for decades by physicians to diagnose patients. This study focuses on the levels of support that medical information systems can provide during these stages of the diagnostic reasoning process. The physician initially generates a list of possible diagnoses (hypotheses) based on the patients' symptoms. Later, those hypotheses are examined to determine which ones best account for the signs, symptoms, physical examination findings, and laboratory test results. Hypothesis generation is especially challenging for medical students because the organization of knowledge in medical school curricula is disease-centered. Furthermore, the clinical reference tools that are regularly used by medical students (such as Harrison's Online, UpToDate, and eMedicine) are mostly organized by disease. To address this issue, *Abduction*, a hypothesis generation tool, was developed for this study. Sixteen medical students were asked to solve two patient cases in two different conditions: A (support of clinical reference tools chosen by the participant and *Abduction*) and B (support of clinical reference tools chosen by the participant). In Condition A, participants were able to generate the correct diagnosis in all 16 occasions (100%) and were able to confirm it in 13 occasions (81.25%). In Condition B, participants were able to generate the correct diagnosis in three out of 16 occasions (18.75%) and were able to confirm it once (6.25%). The implications of this study are discussed with respect to the cognitive support that *Abduction* can provide to medical students for clinical diagnosis.

Résumé

La méthode hypothetico-déductive, qui implique un processus itératif de génération et d'évaluation d'hypothèses, est employée depuis des décennies par des médecins pour diagnostiquer des patients. Cette étude se concentre sur les niveaux de l'aide que les systèmes d'information médicaux peuvent fournir pendant ces étapes du processus de raisonnement diagnostique. Initialement, le médecin produit une liste de diagnostiques possibles (hypothèses) basée sur les symptômes du patient. Plus tard, ces hypothèses sont examinées afin de déterminer lesquelles expliquent mieux les signes, les symptômes, les résultats d'évaluations physiques, et les résultats des tests en laboratoire. La génération d'hypothèses est particulièrement difficile pour les étudiants en médecine parce que l'organisation de la connaissance dans les programmes d'études médicales est centrée sur les maladies. De plus, les outils de référence clinique qui sont régulièrement employés par les étudiants en médecine (par exemple, *Harrison's Online*, *UpToDate*, et *eMedicine*) sont la plupart du temps organisés par maladie. Pour aborder cette question, *Abduction*, un outil de génération d'hypothèses, a été développé dans le cadre de cette étude. Seize étudiants en médecine ont été invités à résoudre deux cas cliniques sous deux conditions différentes: A (avec l'aide des outils de référence cliniques choisis par le participant et *Abduction*) et B (avec l'aide des outils de référence cliniques choisis par le participant). Pour la condition A, les participants ont pu produire un bon diagnostic à chacune des 16 occasions (100%) et ont pu le confirmer à 13 occasions (81.25%). Pour la condition B, les participants ont pu produire un bon diagnostic à trois occasions sur 16 (18.75%) et ont pu le confirmer une fois (6.25%). Les implications de cette étude sont discutées en ce qui

concerne l'aide cognitive que *Abduction* peut fournir aux étudiants en médecine pour le diagnostic clinique.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

> Medical students, house officers, and attending physicians encounter patients who complain of symptoms, who have signs discovered during physical examination, or have abnormal laboratory values identified by diagnostic tests. Patients with such complaints expect them to be explained and solved. Yet, the organization of knowledge in medical school curricula is better suited for patients who present with diagnoses and expect us to determine a set of complaints. We are better prepared to consider "I have a myocardial infarction, tell me my symptoms, signs, and enzyme levels" rather than "I have chest pain, what is wrong with me?" Realizing this, we spend our clinical years relearning and reorganizing medical knowledge into information packets which are more effective for the resolution of the patient problems we encounter (Mandin & DesCoteaux, 1998).

Patient diagnosis is a skill that is developed slowly by novice physicians through clinical practice. It is a challenging process because it requires the reorganization of knowledge acquired during medical school years into illness scripts[1] (Boshuizen & Schmidt, 1992; Boshuizen, Schmidt, Custers, & Van de Wiel, 1995; Charlin, Tardif, & Boshuizen, 2000; Rikers et al., 2002; Schmidt & Boshuizen, 1993a) or information packets (Mandin & DesCoteaux, 1998) that are better suited to patient diagnosis. Further, the application of medical knowledge to patient diagnosis is conditional to extensive practice. In the case of medical students, however, it is not always possible or desirable to practice clinical diagnosis on real patients. To bridge the gap between the kind of learning that happens in the classroom and the kind of learning that happens at hospital wards, many medical schools have adopted some form of problem-based learning (PBL)

[1] knowledge structures that "emerge from continuing exposure to patients and are, therefore, largely the result of extended practice." (Boshuizen & Schmidt, 1992, p. 207)

(Barrows & Tamblyn, 1980; Maudsley, 1999; Norman & Schmidt, 1992, 2000; Schmidt, 1993; Vernon & Blake, 1993). However, time-wise, PBL is more demanding of students and instructors alike when compared to more traditional approaches (Lillehaug & Lajoie, 1998; Stillman & Hanshaw, 1989). In order to optimize the development of medical students' clinical skills, a great deal of mentorship is required. In reality, medical schools can hardly afford to provide the ideal amount of mentorship required to maximize the learning curve of patient diagnosis. The recognition of this problem has led medical schools to encourage students to take a more proactive role in their clinical training and to place a greater emphasis on self-directed learning (Gillam & Bagade, 2006; McLean, Van Wyk, Peters-Futre, & Higgins-Opitz, 2006; Van Berkel & Dolmans, 2006; White, 2006). In order to foster greater student independence regarding their training, several clinical reference systems such as Harrison's Online, UpToDate, and InfoPOEMs have been made freely available to students of most medical schools in North America. These clinical reference systems are comprehensive repositories of medical knowledge that students can consult at their convenience. However, these clinical reference systems are almost exclusively organized by disease. Thus, these systems are helpful when their users already have some initial hypotheses about their patients' diseases and want to acquire more detailed information about those diseases. Because these systems cannot be searched by symptoms, they are not helpful when their users do not have at least one working hypothesis. Consequently, medical students do not get proper support when they need it the most, that is, when they are dealing with cases where they cannot formulate a working hypothesis.

The purpose of this study was to investigate whether specific support to hypothesis generation can enhance medical students' diagnostic reasoning performance. To accomplish that goal, I developed *Abduction*, a clinical reference system that can be searched by symptom. To use *Abduction*, students select multiple symptoms from a symptom list and the system displays a ranked list of possible diseases. More specifically, I investigated whether *Abduction* could increase the likelihood that its users generate and confirm the right diagnostic hypothesis in comparison to the clinical reference systems that are currently available to medical students. The focus of this paper is on clinical problem solving. It assumes that the learning of clinical skills requires guided practice and that learning mostly occurs as students solve problems and transfer the knowledge gained from the solution of those problems to new problems. It also assumes that clinical reference systems, when properly designed, can support medical students during problem solving in the absence of one-to-one mentorship.

The theoretical framework of this paper is based on the model of clinical reasoning proposed by Elstein, Shulman, and Sprafka (1978). It was named hypothetico-deductive method because it depicted clinical reasoning as an iterative process of generation and evaluation of hypotheses. Later, the model proposed by Elstein and colleagues was critiqued by other researchers who claimed that the use of hypothetico-deductive reasoning was characteristic of novice physicians. According to these other researchers, experienced physicians employed more sophisticated reasoning strategies such as forward reasoning (Groen & Patel, 1988; Patel & Groen, 1986, 1991) and pattern recognition (Coderre, Mandin, Harasym, & Fick, 2003; Elstein & Schwarz, 2002; Ridderikhoff, 1993; Schmidt & Boshuizen, 1993b; H. G. Schmidt, Norman, &

Boshuizen, 1990). After presenting and comparing the different views of diagnostic reasoning, I explain why I consider the hypothetico-deductive method the most satisfactory theoretical approach.

Chapter II: Literature Review

This review is part of a study where I investigate how the use of clinical reference systems (e.g., Harrison's Online, UpToDate, eMedicine) affects the diagnostic reasoning process of medical students. More specifically, I investigate how specific support to hypothesis generation affects their diagnostic accuracy. This study requires the review of two types of studies that have relatively distinct histories. One focuses on the reasoning process that physicians go through to diagnose their patients and the other focuses on how physicians and students use clinical reference systems in their daily practice. In the first part of this chapter I review studies on diagnostic reasoning. These studies have been influenced by cognitive research on problem solving and expertise. Depending on their theoretical approach, researchers have reached contradictory conclusions about the nature of diagnostic reasoning. After presenting and comparing the different views of diagnostic reasoning, I explain why I consider the hypothetico-deductive method the most satisfactory theoretical approach. In the second part of this chapter, I review studies on the use of clinical reference systems by physicians and medical students. I conclude the chapter by pinpointing what I consider the main gap in the literature on the use of medical information systems and explain how this study may contribute to the narrowing of that gap.

Diagnostic Reasoning

The work of Elstein, Schuman, and Sprafka (1978) introduced modern cognitive psychology to the field of medical education (Groen & Patel, 1988; Patel, Arocha, & Zhang, 2005). Elstein et al. acknowledge that their work was deeply influenced by

research on problem solving. Subsequently, the findings of Elstein et al. were contested by other studies based on research on expertise such as the pioneering work of Chase and Simon (1973) and de Groot (1978). The opposing researchers argued that expert physicians did not engage in an iterative process of hypothesis generation and evaluation as was suggested by the work of Elstein et al. According to the findings of later studies, experts adopted a data-driven approach rather than a hypothesis-driven approach. Furthermore, it was argued that experts were able to diagnose cases in their areas of expertise using pattern recognition or instance scripts. These contrasting views of the diagnostic reasoning process will be presented in the next subsections.

Hypothetico-Deductive Reasoning

> Given that maximum uncertainty characterizes the initial state of a diagnostic encounter, hypotheses form an essential function: they frame, or constrain, a patient's problem and provide a context for further diagnostic reasoning and exploration. Each diagnostic hypothesis evokes a template of possible clinical findings against which a given patient's findings can be compared. (Kassirer & Kopelman, 1991, p.9)

Based on a set of studies conducted between 1969 and 1973, Elstein et al. (1978) concluded that both expert and novice physicians resorted to hypothetico-deductive reasoning, an iterative process of hypothesis generation and evaluation, to diagnose their patients. The researchers found that most participants started generating hypotheses quite early in the process. Drawing from earlier research on problem solving (Newell, Shaw, & Simon, 1958; Newell & Simon, 1972; Simon, 1969), the researchers attributed this pattern of behavior to the necessity of working around the constraints of one's short-term

memory (STM). That is, the early generation of hypotheses would serve to constrain the size of the space that must be searched in order to find a solution to the problem.

Problem space. The analogy of a problem space was proposed by Newell and Simon (1972) to help explain problem solving strategies. The problem solving model consisted of a problem space with an initial state, a goal state, and a set of operators that the problem solver uses to move from one state to the other (Newell & Simon, 1972). Problem solvers do not necessarily have the whole problem space represented in their minds at one time. Furthermore, some problem spaces are so large that the problem solver cannot search through all possible intermediate states. Consequently, strategies to select the most promising paths are necessary. One of the simplest problem solving strategies is known as *hill climbing*. In hill climbing, the problem solver moves to the next intermediate state that is most likely to lead to the goal state. One limitation of the hill climbing strategy is that, in the absence of a panoramic view of the problem space, a move that appears to lead the problem solver closer to the goal state may in fact lead him or her further from it. A more effective problem solving strategy is *means-ends analysis*. Means-ends analysis is a decomposition or subgoaling strategy: the problem solver starts by tracing intermediate states and subgoals between the initial state and the goal state. These subgoals can be solved with relative independence to the rest of the problem. If a subgoal cannot be solved, it can be further subdivided.

Hypothetico-deductive reasoning is a form of means-end analysis. The early generation of hypotheses is a strategy adopted by the physician to map the most promising paths to the solution of the problem. As in most problems in complex domains,

the goal state (the diagnosis) is unknown, the physician creates a set of reachable subgoals: hypotheses that can be tested.

The ability to generate promising hypotheses is conditional to the physician's domain knowledge. Thus, the more experienced the physician, the higher the likelihood that he or she will generate stronger hypotheses. Once an initial set of hypotheses has been generated, all incoming data is interpreted in light of those hypotheses. If necessary, new hypotheses can be formulated and thereby reconfigure the physician's problem space. Elstein et al. (1978) also found that physicians often start by generating a mix of specific and general hypotheses. The general hypotheses (e.g., infection) are progressively refined into more specific classifications (e.g., infectious mononucleosis).

Subsequent studies supported the case for early hypothesis generation. Focusing on medical students, Gruppen et al. (1993) found that subjects who did not include the correct diagnosis in their initial hypothesis list were significantly less likely to produce an accurate diagnosis. Sisson et al. (1991) found that early hypothesis generation was common practice among physicians and medical students. The difference between the two groups related to the quantity and specificity of hypotheses. Students' hypotheses were significantly more numerous and more specific. Johnson et al. (1981) also found evidence that physicians and medical students start generating hypotheses early in the process. However, contrary to the results obtained by Sisson et al., the researchers found that experts and novices alike generated hypotheses of similar types in similar quantities. Early hypothesis generation was also detected in studies with neurologists (Barrows & Bennett, 1972) and in surgical diagnosis (Dudley, 1970, 1971).

Like the problem solving model proposed by Newell and Simon, the hypothetico-deductive reasoning model has a very general and comprehensive nature. Elstein et al. in fact argue that "hypothetico-deductive processes are ubiquitous in solving complex problems" (p. 79). Subsequently, other researchers have opposed those claims and proposed other approaches to the study of diagnostic reasoning which they argue help to explain issues that Elstein and colleagues have not addressed. These approaches are discussed in the next sections of this chapter.

Forward Reasoning

Patel and Groen (1986) were among the researchers that were not convinced that expert and novice physicians alike were using some form of hypothetico-deductive reasoning to diagnose patients. Some studies in other domains suggested that experts often used pure forward reasoning to solve problems (Larkin, McDermott, Simon, & Simon, 1980).

Forward reasoning means that the physician reasons from the symptoms (clues) to the disease (e.g., if fever, then infection). Conversely, backward reasoning means that the physician reasons from the disease (hypothesis) to the symptoms (e.g., if infection, then fever). Hunt (1989) summarizes the advantages and disadvantages of forward reasoning in the following way:

> Forward-driven problem solving is riskier than goal-based problem solving, because operations are executed (i.e. new states of the problem space are visited) without first checking to see if these operations are likely to be an advance toward the goal. On the other hand, forward-driven reasoning is cheaper, because operator selection is made without contrasting the present state of knowledge to

> the goal state. Thus forward-driven problem solving is preferable if the problem solver knows enough about the problem-solving domain to recognize when certain actions should be taken. This implies that a rational problem solver would use forward-driven reasoning in those (limited) domains with which he or she was familiar. (p. 617)

Patel and Groen (1988) surmise that Elstein et al. (1978) did not find any differences between expert and novice physicians regarding their reasoning methods due to methodological inadequacies. The researchers suggest that the belief in the use of hypothetico-deductive reasoning in medicine is a result of "protocols being overinterpreted, so that the data is made to fit the preexisting theory, rather than a theory being created to fit the data" (p. 289).

Patel and Groen conducted a series of studies with physicians of different levels of expertise, employing the techniques of propositional analysis (Frederiksen, 1975; Kintsch, 1974) to detect whether participants were using forward reasoning, backward reasoning, or a mix of the two. In one of their earlier studies, they examined the explanations of seven cardiologists working on an endocarditis case (Patel & Groen, 1986). Their findings were that the physicians that made accurate diagnoses explained the underlying pathophysiology of the case using pure forward reasoning. Conversely, physicians with inaccurate diagnoses used a mix of forward and backward reasoning.

Subsequent studies with subjects of different levels of expertise yielded similar results: their findings continued to support the idea of "a strong relation between diagnostic accuracy and the use of forward reasoning" (Patel & Groen, 1991). Conversely, the use of backward reasoning or a mix of forward and backward reasoning

was most likely to lead to an inaccurate diagnosis (Groen & Patel, 1988; Patel & Groen, 1986, 1991).

Another one of their findings was that while experts demonstrated to be very good at discerning relevant from irrelevant information in a patient case, novices were not nearly as good at the task (Patel & Groen, 1991). This brings up the issue of relevance, understood here as pertinence or relation to the matter at hand. Something is never relevant per se. That is, something is considered relevant in relation to something else. In the case of clinical diagnosis, information can only be considered relevant in relation to the patient's actual disease. In the absence of any diagnostic hypothesis, all symptoms and signs acquire the same level of relevance since there are no parameters to establish hierarchical differences among them. Consequently, it is unlikely that experts are able to excel at distinguishing relevant from irrelevant information without relying on diagnostic hypothesis. A few studies have shown that the consideration of the correct diagnosis lead to an increase in the chances that the relevant clinical signs will be detected (Berbaum et al., 1986; Brooks, LeBlanc, & Norman, 2000; LeBlanc, Brooks, & Norman, 2002; Leblanc, Norman, & Brooks, 2001; Norman, Brooks, Colle, & Hatala, 1999).

The same research team also conducted a specific study on hypothesis generation and its relation to domain knowledge (Joseph & Patel, 1990). In this study, the clinical case was presented to subjects one segment at a time. The researchers found that the experts generated accurate hypotheses early in the process and spent the rest of the time refining it by explaining the patient cues. These results, rather than confirming the use of pure forward reasoning by experts, support the opposite idea that experts and novices alike use hypothetico-deductive reasoning.

Differences in the granularity of the data collection procedures adopted by Patel and her colleagues is one likely explanation for the contradictory results described in the previous paragraph. Previously, the researchers were trying to capture the diagnostic reasoning process through post-hoc written explanations given by the participants. This method of data collection generates very brief protocols that are unlikely to represent the entire reasoning process of the physicians. The latter method in which the researchers use a gradual presentation of the clinical case combined with the collection of think aloud protocols allowed them to obtain finer-grained data about the actual reasoning process followed by their subjects.

There is another factor that casts doubts about how representative of the reasoning process were the protocols collected by the researchers. It refers to the instructions given to the subjects. After the subjects had read the case presentation and made a list of everything they could recall about the case, they were asked to write an explanation of the underlying pathophysiology of the case. This written explanation was then used to map the reasoning process of the subjects in diagrams similar to flow charts. Consequently, one can accept that the researchers have produced concrete evidence that expert physicians often use pure forward reasoning to explain the functional changes associated with a disease they have diagnosed. However, it is questionable whether those protocols also represent the entire reasoning process used to issue the diagnosis (Eva, Brooks, & Norman, 2002). Lemieux and Bordage (1992) argue that "what Patel and Groen described as pure forward reasoning is more a reflection of their method of investigation than the actual reasoning of the clinician." (p. 201).

The method of presenting the clinical case to subjects one segment at a time was repeated in a study with medical students (Arocha, Patel, & Patel, 1993). The researchers found that second year students ignored or reinterpreted incoming data that did not fit their initial hypotheses. Third year students generated other hypotheses to account for data that did not fit their initial hypotheses. Fourth year students started by generating broad hypotheses and gradually narrowed them toward a diagnostic that explained all the cues.

By the end of the nineties the research group had revisited their position regarding the relation between clinical expertise and hypothesis-driven reasoning. They concluded that "the ability to index and use adequate evidence by physicians, residents and students is a function of the early generation of accurate hypotheses" (Allen, Arocha, & Patel, 1998, p. 91). In addition, they argued for "a characterization of the process of expert medical diagnostic reasoning as a succession of limited comparisons involving related diagnostic hypotheses" (Kushniruk, Patel, & Marley, 1998, p. 255).

Pattern Recognition

It has been argued that expert physicians resort to pattern recognition rather than hypothesis testing unless they are dealing with a challenging patient case (Coderre, Mandin, Harasym, & Fick, 2003; Elstein & Schwarz, 2002; Kempainen, Migeon, & Wolf, 2003). Pattern recognition refers to the act of discerning patterns in the current situation that are then used to search the long-term memory (LTM) for phenomena with similar patterns. Research on expertise has looked into how experts use pattern recognition to solve complex problems (Gobet, 1997; Gobet & Simon, 1996). Let's take

the example of multiplication problems: multiplying 123 by 456 can be considered a paradigmatic problem solving situation in the sense that it has an initial state onto which one applies operators to reach the goal state (the product of the multiplication). However, multiplying 2 by 2 does not entail the same conditions since one can retrieve the solution directly from LTM. In this regard, pattern recognition is similar to what is often known as analogical reasoning where the problem solver jumps from one part of the problem space to another by mapping the solution of a known problem onto the new problem (Dunbar, 1998).

Some researches have taken the concept of pattern recognition in clinical diagnosis even further. Schmidt, Boshuizen and colleagues (Schmidt & Boshuizen, 1993b; Schmidt, Norman, & Boshuizen, 1990) argue that the most advanced form of diagnostic reasoning relies on instance scripts, which are memories of previous patients that are stored as individual entities and not merged in a prototypical form. However, the researchers have not presented any convincing evidence for the claim that instance scripts are a central feature of expertise in medicine. Patel and Groen (1991) have argued against the idea that expert diagnosis is a process of pattern recognition based on their findings that recall is nonmonotonically[2] related to expertise and that directionality of reasoning is an "all-or-none" phenomenon that is more likely related to the two extremes expert-novice rather than a developmental pattern.

Other researchers adopt a more moderate position and argue that medical diagnosis is a categorization task composed by two complementary processes: analytic processing and similarity-based processing (pattern recognition) (Ark, Brooks, & Eva,

[2] Monotonic functions either increase or decrease without reversing directions. For example, the sequence 1, 3, 5, 7 increases monotonically but the sequence 1, 5, 3, 7 is nonmonotonic.

2006; Kulatunga-Moruzi, Brooks, & Norman, 2001). Based on a study with 12 pre-clinical medical students, Kulatunga-Moruzi, Brooks, and Norman concluded that "the relative reliance on analytic and similarity-based processes is amenable to instruction and dependent on expertise" (p. 110). Ark, Brooks, and Eva (2006) found that groups of students instructed to use both analytic and similarity-based strategies performed significantly better (regarding diagnostic accuracy) than groups of students instructed to use either strategy alone. Further the researchers found no significant differences between the performances of the groups instructed to use either analytic or similarity-based strategies. The researchers conclude that there are advantages to teaching medical students to use both strategies.

There have been few empirical studies that have directly addressed the issue of the use pattern recognition in clinical diagnosis. Moreover, different studies have employed the term pattern recognition in slightly different ways. When discussing pattern recognition, some researchers refer to the use of similarity-based reasoning while others refer to the activation of episodic memory. Some studies have found that the exclusive use of pattern recognition has the highest correlation with successful diagnostic performance while others have found that best results are achieved through the combined use of analytical and similarity-based strategies. What we currently know for sure is that solving a case through a pattern recognition approach requires that the problem solver has seen one or more similar cases before and is able to detect the relevant similarities between past and present cases. We can also safely assume that more experienced physicians employ pattern recognition more often because they have more patient cases

stored in their LTM. Claims that venture further than that would still be speculative at this point in time.

Cognitive and Instructional Advantages to Adopting a Theoretical Approach Based on the Hypothetico-Deductive Reasoning

There are specific cognitive and instructional advantages for the adoption of hypothetico-deductive reasoning as a theoretical approach. From an instructional perspective it is problematic to accept forward reasoning or pattern recognition as the ideal form of diagnostic reasoning. Both are conditional to the possession of large amounts of highly structured domain knowledge which, in practice, renders these types of reasoning nearly useless to novice physicians. Rather than the cause for successful diagnoses, these types of reasoning are a consequence of sufficient domain knowledge. That is, physicians will resort to forward reasoning or pattern recognition only with cases that they consider easy. On the other hand, all researchers agree that the hypothetico-deductive method can be used by both novice and expert physicians. In fact, even advocates of forward reasoning and pattern recognition admit that experts do resort to hypothetico-deductive reasoning when dealing with patient cases that they consider challenging.

From a cognitive perspective, the hypothetico-deductive method is more comprehensive than the competing approaches. Rather than considering diagnostic reasoning as a monolithic process, the hypothetico-deductive method encompasses all the relevant stages that a physician might go through when working on a patient case. From a practical perspective, it is hard to accept that expert physicians often work exclusively

from the signs and symptoms to the disease. In a field such as clinical diagnosis where the stakes are so high, some form of backtracking is always required, especially nowadays when hospitals are increasingly guarding themselves against malpractice suits. Forward reasoning and pattern recognition are hypothesis generation strategies. Indeed they are robust strategies to generate hypotheses since they rely on expert knowledge. Nonetheless, every time a physician orders a laboratory test or checks how a patient responds to a drug, he or she is testing a hypothesis about the patient's disease and consequently engaging in hypothetico-deductive reasoning. No matter how confident a physician is in a diagnosis, the nature of the profession requires some form of triangulation.

Another advantage of the hypothetico-deductive method over competing approaches is the consideration of information seeking processes. In a clinical environment, consultation of colleagues and clinical reference systems is common practice since no individual physician can possess all the medical knowledge necessary to solve every patient case that is presented. Studies on diagnostic reasoning have largely ignored this fact. Studies on how incoming information from fellow physicians and reference systems are incorporated and affect the problem solving process are nonexistent. A theoretical approach based on the hypothetico-deductive method is more comprehensive in this regard because it incorporates the information-seeking operations (inquiry) that are often part of clinical practice in naturalistic settings (see figure 1).

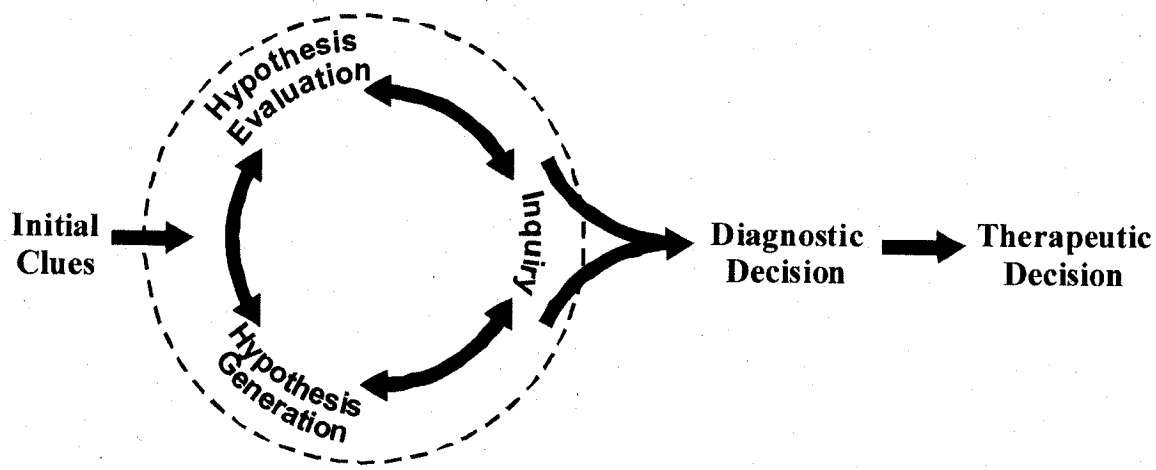

Figure 1. Hypothetico-deductive method of diagnostic reasoning.

The Use of Clinical Reference Systems by Physicians and Medical Students

Information Needs of Physicians and Medical Students

> Much of human experience is characterized by the notion of search; we seek and pursue material objects such as food or shelter, sensual experiences such as adventure or ceremony, and ethereal objects such as knowledge or justice. We are concerned here with the search for information that we will call *information seeking*, a process in which humans purposefully engage in order to change their state of knowledge. [...] The term *information seeking* is preferred to *information retrieval* because it is more human oriented and open ended. Retrieval implies that the object must have been "known" at some point; most often, those people who "knew" it organized it for later "knowing" by themselves or someone else. Seeking connotes the process of acquiring knowledge; it is more problem oriented as the solution may or may not be found. (Marchionini, 1995, p. 1)

There has been a considerable amount of studies conducted on the use of clinical reference systems by physicians and medical students. Most of these studies have concentrated on the information needs and information-seeking behavior of physicians

and medical students. The identification of the information needs of physicians and medical students is not a straightforward task as it might seemingly appear. Sometimes subjects cannot clearly articulate their questions. Other times, the formulated questions prove to be an amalgam of different questions, implicit or not. In these cases, it may be difficult to determine exactly how many questions are being asked and how many answers they require (Gorman, 1995).

A comparison of published studies on the information needs of physicians and medical students is complicated by the fact that these studies have adopted different methodological approaches (ethnographic, interview, mail survey, etc.) and terminology and have produced disparate results (Gorman, 1995). Furthermore, different studies have identified physicians' information needs in different ways. Gorman identifies four approaches: (1) unrecognized needs (Clinician not aware of information need or knowledge deficit); (2) recognized needs (Aware that information is needed: may or may not be pursued); (3) pursued needs (Information seeking occurs; may or may not be successful); and (4) satisfied needs (Information seeking succeeds). Further, not all information needs of physicians and medical students relate to medical knowledge. Osheroff et al. (1991), for example, also counted questions pertaining to patient data (information about a specific person). Moreover, some questions may be wrongly assumed to be related to patient diagnosis or therapy. For example, the question "how do you treat a terminal patient with leukemia?" may be implicitly asking for advice on doctor-patient relationship rather than for clinical information.

Ely et al. (2000) have suggested that a taxonomy of clinical questions can be used to guide the design of medical databases. The taxonomy of generic clinical questions

developed by Ely et al. used 295 questions formulated by 49 primary care doctors to modify a previous taxonomy of 1,101 questions formulated by 103 family doctors (Ely et al., 1999). The purpose of the study was to "determine whether the essence of clinical questions could be captured by a limited number of generic question types" (Ely et al., 2000, p. 429) . The study resulted in a taxonomy of 64 generic types. The second study revisited the results of the first by working with a more heterogeneous group of physicians and coders. Five of the top 10 questions were related to diagnosis (e.g.: what is the cause of symptom X? what test is indicated in situation X?) and five were related to therapy (e.g.: what is the drug of choice for condition X? What is the dose of drug X?).

The potential utility of a taxonomy to guide the redesign of the content of a medical database is partial. Lazoff (2001, May) related her experience using some medical databases to answer the clinical questions of the American Board of Internal Medicine's (ABIM) recertification program. She reported a great variability in the content of medical databases and even the existence of contradictory information. Lazoff's description of her personal experience shows that the information contained in medical databases often provides inadequate support to clinical decision making. It also demonstrates that taxonomic studies alone are not sufficient to provide precise guidelines to the creation of database content. The assessment of how the content of a medical database is actually used by physicians to answer clinical questions is also fundamental. Not only the patterns inherent in data (e.g. location, alphabet, time, category, hierarchy) (Bradford, 1996; Wurman, 1989) but also the patterns in the use we make of the data should be taken into consideration when designing any kind of information system (Nakamura & Lajoie, 2003). In order to determine in which ways information should be

organized within a clinical reference system, it is important to discriminate what kind of information physicians and medical students might need while they try to solve a patient case and how do they search for that information.

Information-Seeking Behavior of Physicians and Medical Students

Questions pursued and questions left unanswered. According to Hersh and Hickam's review of works on the use of medical information retrieval systems published between 1966 and 1998, physicians have an average of two unanswered questions for every three patients and use information retrieval systems an average of 0.3 to 9 times per month (Hersh & Hickam, 1998). Those numbers raise an important question: what parameters do physicians use to decide whether or not to pursue the answer to a clinical question? Gorman and Helfand (1995) have looked into that issue with a study conducted with 49 primary-care physicians with no ties to a medical school. They found that while participants generated many questions regarding optimal patient care, they pursued the answer to about 30% of their questions. Using a multiple regression model they concluded that only two factors were significant predictors of information-seeking: the belief that a definite answer existed, and the urgency of the patient's problem. Covell, Uman, and Manning (1985) found that a physician's self-perceived information needs is another factor that have an impact on information-seeking behavior. In their study, 47 internists answered a questionnaire regarding their information needs. Each participant was then interviewed during half a day after each patient encounter. Although the participants reported to have an average of one clinical question per week, the interviews

showed that for every three patients seen, an average of two questions remained unanswered.

Correlation between information-seeking proficiency and domain knowledge. Some studies on information-seeking behavior (Bates, 1977; Hsieh-Yee, 1993) concluded that there are no statistically significant effects of domain knowledge on information-seeking proficiency. According to these studies, search experience is the factor that has the greatest impact on information-seeking proficiency. However, the results of Hsieh-Yee's study showed that there are interaction effects between domain knowledge and search experience. Other studies (Fenichel, 1979; Wanger, McDonald, & Berger, 1980) produced opposite results, concluding that search experience affects search time but not search outcomes. In other words, inexperienced searchers took longer and made more mistakes than experienced searchers but were still able to complete the tasks in the studies. A direct comparison between the results of the above studies is complicated by the fact that they used different methods to define level of search experience, worked with tasks of different levels of complexity, and assessed the results in different ways. For example, Hsieh-Yee categorized participants as either novices or experienced searchers while Fenichel divided participants into 4 different categories. Different tasks were given to participants according to their level of experience in the former study while the same task was given to all participants in the latter.

The first study to compare the information seeking performance of novice and expert searchers in medicine was conducted by Haynes, McKibbon, Walker, Ryan, Fitzgerald, and Ramsden (1990). One hundred and fifty-eight trainees and attending staff of a university medical center were recruited as novice searchers. One medical librarian

and one clinician experienced with MEDLINE were recruited as experts. Seventy eight searchers performed by the novices were randomly selected to be duplicated by the two experts. The dependent variables of the study were recall[3] and precision[4]. Significant differences in recall and precision were found between the novice and experienced groups. Hersh and Hickam (1998) comment on the problem about using recall and precision as dependent variables: "These measures estimate the quantity of relevant articles retrieved, although it may not always be the most important aspect of a search done for clinical care. Clinicians may instead be interested in how effectively searches answer clinical questions."

Wildemuth (2004) conducted a study with medical students to evaluate their search tactics and verify whether their tactics were correlated to their domain knowledge. The researcher found some common patterns in the students' search tactics. The most common approach was to specify a concept and then to add one or more concepts in order to narrow down the retrieved set. She also found that students' search tactics became more efficient over time, attributing the changes to the changes in students' domain knowledge rather than changes in their search experience. Wildemuth argues in her conclusions that future research on information-seeking behavior should be especially concerned about the separate and combined effects of domain knowledge and search

---

[3] For a query, *recall* is the proportion of relevant documents retrieved from the database calculated as the number of relevant documents retrieved in the search divided by the total number of relevant documents in the entire database.
One problem with the measure of recall is that the denominator implies that the total number of relevant documents for a query is known, which is impossible for large databases. In this situation, a measure that approximates recall, called *relative recall*, is used. This measure uses in the denominator the total number of unique relevant documents retrieved in 3 or more different searches on the same topic. (Hersh & Hickam, 1998)

[4] *Precision* is the proportion of all retrieved documents that are relevant calculated by the number of relevant documents retrieved in the search divided by the number of documents retrieved. (Hersh & Hickam, 1998)

experience on search behaviors. Interestingly enough, the same researcher arrived at diametrically opposite conclusions in a study conducted nine years before. In her previous study, Wildemuth concluded that there was little evidence of any relationship between personal domain knowledge and searching proficiency (Wildemuth, de Bliek, Friedman, & File, 1995).

More often than not, research on the use of clinical information systems has failed to provide concrete evidence for the claim that these systems can actually improve problem-solving performance. There has been a considerable amount of research on the information needs and the information-seeking behavior of the users of such systems (Cogdill & Moore, 1997; Gorman, 1995; Hersh & Hickam, 1998), but still not much is known about how the retrieved information is interpreted or applied (Hersh & Hickam, 1998). There is a need to move the research on this area beyond measures of relevance of the retrieved information to the assessment of how the use of information systems affects problem-solving performance (Gorman, 1995).

Database-Assisted Diagnostic Reasoning

To the best of my knowledge, there have been only four studies (Berner et al., 2002; de Bliek et al., 1994; Wildemuth, de Bliek, Friedman, & File, 1995; Wildemuth, Friedman, Keyes, & Downs, 2000) that have assessed how the use of medical databases affects clinical reasoning, and two of them do so only indirectly (Berner et al., 2002; Wildemuth, de Bliek, Friedman, & File, 1995).

In 1986, de Bliek, Friedman, Wildemuth, Martz, Twarog, and File, a research group at the University of North Carolina at Chapel Hill (UNC-CH) developed

INQUIRER, a computer database in bacteriology. Eight years later, they published the results of a number of studies on the effects of the use of INQUIRER on the performance of medical students at UNC-CH (de Bliek et al., 1994).

Data for the study was collected on three different occasions during a period of nine months. The first assessment occurred during the term students were taking the bacteriology class. The second assessment occurred three months later, and the third assessment, five months later. On each occasion students were assessed twice, with and without the assistance of INQUIRER. The sample consisted of 36 first year medical students that were assessed on all three occasions. In the first pass of each assessment occasion, participants were presented with six clinical case problems and were asked to answer three to six questions pertinent to each case. In the second pass, participants were asked to answer six questions selected from the pool of questions left unanswered in the first pass with the assistance of INQUIRER. Questions for the cases were related to both diagnosis and treatment. The independent variable in this study was the test occasion. The dependent variables were *personal knowledge score* (proportion of questions answered correctly) and *database-assisted score*. Data were analyzed with MANOVA, associated univariate analyses, and trend analysis. The analyses show significant differences in personal knowledge scores and database-assisted scores across the 3 assessment occasions. Mean personal knowledge scores were low at the first assessment (X=13.1). In the second assessment they increased to 50.2 but decreased again in the third assessment to 24.2. Database-assisted scores increased linearly, from a mean score of 36.9 in the first assessment occasion, to 51.7 in the second, and again to 74.1 in the third. One of the main contributions of this study was to show that the use of an

information system can convert the parabolic trend in medical students' clinical reasoning skills to a linear increasing trend. This seems to indicate that, in the long run, there is a synergistic relation between clinical reasoning and the use of medical databases.

Wildemuth et al. (1995) conducted a study on the relationship between domain knowledge, information searching proficiency, and database assisted problem-solving performance. Sixty-four first-year medical students participated in the study. Participants were assessed in four different occasions (between Fall 1990 and Spring 1992) in three different domains (bacteriology, pharmacology, and toxicology). The methodology for this study was analogous to the one followed by de Bliek et al. (1994). An expanded version of INQUIRER was used in this study. The study's primary findings were that there is little correlation between domain knowledge and information searching proficiency. The secondary findings show a correlation between information searching proficiency and successful use of information in problem solving.

Wildemuth once again replicated the study by de Bliek et al. (1994) as part of a larger study (B.M. Wildemuth, Friedman, Keyes, & Downs, 2000) that included the use of two different database interfaces. Similar results were found: "Personal Knowledge Scores varied by occasion, being highest just after and lowest just before the course. The Database-Assisted Scores were similar just before and after the course, but were higher six months after the course." The three studies conducted at UNC-CH show that, in the long run, there are advantages to encouraging students to use medical databases as there seems to exist a synergistic relationship between clinical reasoning and the use of these databases. It is important to stress, however, that the measured effects were a result of the

use of a custom-made database. The study presented next examined the effects of more general information resources on medical students' performance.

Berner et al. (2002) conducted a study on the effects of information retrieval instruction on medical students' information retrieval skills and their ability to apply the retrieved information to solve patient cases. In this study students were presented with patient cases in pediatrics and asked to consult MEDLINE and other World Wide Web resources of their choice, evaluate the accuracy of the information, and prescribe a course of action to the patient case. The treatment group (instructed group) performed significantly better in four of the seven tasks.

Despite the different results of the studies presented in this chapter, researchers seem to agree that information-seeking operations demand more time than most physicians would like to spend consulting clinical reference systems. As medical knowledge grows, physicians and medical students will have to increasingly rely on clinical reference systems, whether they like it or not. However, these systems are currently clumsy external memory devices. Cohorts function far better as external memory. Cohorts are far better at understanding our questions and information needs. And if they do not know the answer to our question, at least they do not take several minutes of our time to say so. But cohorts are not always available when they are needed. The additional cognitive demands that these clinical reference systems impose on their users create sharp usability constraints. If we want databases to really function as cognitive tools then they should be adapted to the users' needs and reasoning processes and not the other way around.

There is a need to push the development of consultation systems to a more user-centered approach where the systems complement instead of duplicate the users' skills. To complement rather than duplicate, these systems must trust the user's competence whereas traditional consultation systems are often based on some mistrust (Buscher, Baumeister, Puppe, & Seipel, 2005). Maybe then these systems will live up to their potential as cognitive artifacts (Norman, 1991) or cognitive tools (Lajoie, 2000; Lajoie & Derry, 1993) that allow students to engage in activities that would be out of their reach otherwise, including the generation and evaluation of hypotheses in the context of problem solving.

Purpose of the Study and Research Questions

Purpose of the Study

The purpose of this study was to assess the effects of the use of clinical reference systems on the diagnostic reasoning process of medical students. Previous studies have shown that the use of these systems have positive effects on medical students' diagnostic reasoning performance (de Bliek et al., 1984, Friedman et al., 1999, Wildemuth et. al., 1995, Wildemuth et al., 2000). However, these studies do not show exactly in which ways the use of these systems affect the students' reasoning process. In this study I specifically assessed the effects of different clinical reference systems on hypothesis generation and hypothesis evaluation. This choice derives from the assumption that hypothesis generation and hypothesis evaluation, although complementary, are different types of cognitive processes and, therefore, require different kinds of support.

The second assumption made in this study is that, during hypothesis generation, the physician reasons from the symptoms to the disease. That is, based on an initial set of symptoms, the physician generates hypotheses about which diseases may be causing the patient's symptoms. Conversely, it is assumed that, during hypothesis evaluation, the physician reasons from the disease to the symptoms. That is, the physician tries to determine which diseases best account for the patient's symptoms. The physician seeks to confirm or disconfirm the generated diagnostic hypotheses by seeking additional information (e.g., questioning and examining the patient, ordering laboratory tests, and consulting other physicians, books, or medical information systems) (see Figure 2).

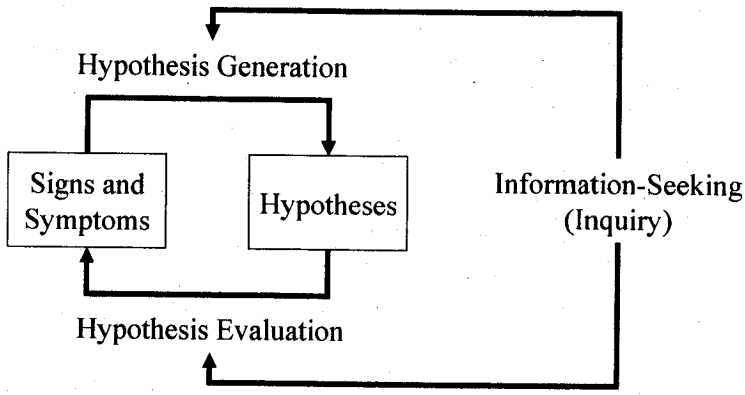

Figure 2. Diagram showing how information-seeking operations can support either hypothesis generation or hypothesis evaluation.

The clinical reference systems that are mostly used by medical students such as Harrison's Online, UpToDate, and eMedicine are exclusively indexed by disease. That is, they can be searched by disease but not by symptom. Due to this choice of indexation, it is hypothesized that clinical reference systems may provide support to hypothesis evaluation but not to hypothesis generation.

In order to test whether differentiated support to hypothesis generation can improve diagnostic accuracy of medical students, *Abduction* was developed. *Abduction* is a clinical reference system that provides diagnostic suggestions based on the selection of multiple symptoms by the user. More detail on *Abduction* will be provided in the next chapter.

Medical students who participated in this study were asked to try to solve two medical cases in two different conditions. In both conditions participants were allowed to consult the clinical reference systems of their choice (the ones that they most regularly use). In one of the conditions participants were asked to use *Abduction* before they began using the clinical reference systems of their choice.

Research Questions

RQ #1: Will medical students' use of *Abduction* increase the likelihood that the correct diagnosis is included among their working hypotheses?

RQ #2: Will medical students' use of the clinical reference systems that they use on a regular basis increase the likelihood that the correct diagnosis is included among their working hypotheses?

RQ#3: Does *Abduction* provide better support to hypothesis generation than the clinical reference systems that are most used by medical students?

RQ #4: Can the clinical reference systems that medical students use on a regular basis be used to confirm the correct diagnosis (in case it was generated)?

RQ #5: In cases where the correct hypothesis is generated but not confirmed, is it due to the student's inability to interpret and apply the retrieved information or to inadequacies in the clinical reference systems?

Chapter III: Methods

Participants

The sample consisted of 16 medical students, four from one American college and 12 from one Canadian university. Eight participants were second-year and eight were fourth-year medical students (both schools have four-year programs). Four of the eight fourth-year students were from an American college (see Table 1). This was a convenience sample: it was opted to work with second- and fourth-year medical students due to difficulties in recruiting 16 students from the same cohort. Further, participants were recruited in two different institutions due the difficulty of recruiting 16 students in the same institution. Differences in clinical skills between the fourth-year students of the two institutions were not anticipated. In order to avoid effects from students' level of experience, the two cohorts were counterbalanced. More details on the counterbalancing procedures will be provided in the Design section.

Table 1

Participants

Participants	University X	College Y	Total
Med-2	8	0	8
Med-4	4	4	8
Total	12	4	16

Design

The main purpose of this study was to investigate whether specific support to hypothesis generation (in this case, through the use of *Abduction*) can increase diagnostic

accuracy. Accordingly, the main factor being tested in this study was condition (A: with *Abduction*; and B: without *Abduction*). To maximize the use of the sample of 16 participants, each participant was asked to solve two cases (Friedrich's ataxia and Kennedy's disease). To minimize the effects caused by possible differences in case complexity, cases and conditions were counterbalanced: half of the participants solved the Friedrich's ataxia case in condition A and the Kennedy's disease case in condition B and the other half solved the Friedrich's ataxia case in condition B and the Kennedy's disease case in condition A. To avoid carry over effects, the order of presentation of cases and conditions was also counterbalanced (see Figure 3). Because the sample included second- and fourth-year medical students, the student cohorts were also counterbalanced.

PARTICIPANT		FIRST CASE	SECOND CASE
MED-2	MED-4		
1 & 5	9 & 13	Kennedy's Disease	Friedrich's Ataxia
2 & 6	10 & 14	Kennedy's Disease	Friedrich's Ataxia
3 & 7	11 & 15	Friedrich's Ataxia	Kennedy's Disease
4 & 8	12 & 16	Friedrich's Ataxia	Kennedy's Disease

Condition A: Consultation of *Abduction* and chosen Clinical Reference Systems

Condition B: Consultation of chosen Clinical Reference Systems

Figure 3. Research design.

In summary, 16 participants solved two cases under two different conditions, totaling 32 occasions and eight possible combinations. Thus, the resulting design model would be:

$$Subj_4 (CaseOrder_2 \times CondOrder_2) \times Cond_2$$

The order in which the cases were presented (CaseOrder) and the order in which the conditions were applied (CondOrder) are the between groups factors. Condition was the repeated factor.

Procedure

At the beginning of each session, participants were asked to read and sign the informed consent form (see Appendix B), retaining a copy for themselves. Subsequently, participants received an oral explanation of the all activities to be performed by them during the session, including how to use *Abduction*. More details on *Abduction*'s operation is provided in the Materials section. No warm up problem was given. Participants were informed that they could ask questions at any time during the session. No time limit was imposed although participants were informed that the session would take an average of two hours (based on data from the pilot).

In condition A the participants went through the following steps:

1. Read case presentation;
2. Create initial hypothesis list. Each hypothesis consists of three compulsory elements:
   A. Name of the hypothesis (e.g., Multiple Sclerosis);
   B. Confidence level (ranging from 0 to 100%);

C. Supporting evidence (e.g., muscle weakness, hyperthermia)

3. Use *Abduction* to revise hypothesis list;
4. Evaluate hypotheses by ordering laboratory tests and consulting the clinical reference systems of their choice;
5. Write case summary.

In condition B participants followed the same procedures, skipping step 3 (consultation of *Abduction*).

Figure 4 illustrates the steps described above for the two conditions.

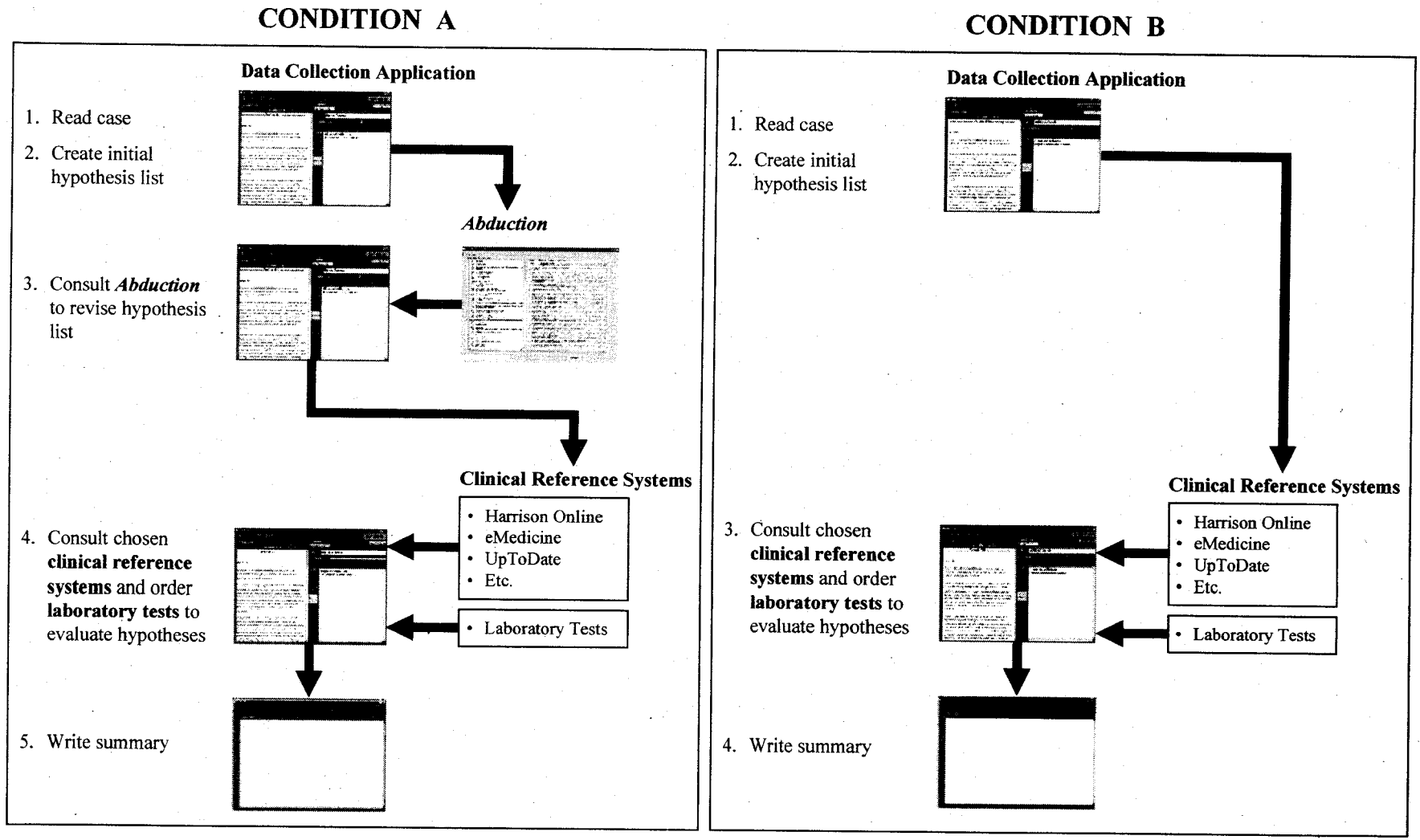

Figure 4. Procedures.

Participants were free to create as many or as few hypotheses as they wished (the application allows a maximum of 13 hypotheses). Participants were forced to enter all three components for each hypothesis they created (name, confidence level, and supporting evidence). Participants were allowed to add new hypotheses and update the three components of existing hypotheses until the end of the case. Participants could not delete existing hypotheses, but they could take the confidence level of a hypothesis to zero to show that, from that moment on, that hypothesis was abandoned.

A list of supporting evidence for each hypothesis was demanded only to make sure that participants had concrete evidence to back up their hypotheses. The ability to distinguish between relevant and irrelevant information to the solution of a medical case is indeed an important clinical skill which has been studied by other researchers (Patel & Groen, 1991). However, the quality of the lists of supporting evidence was not analyzed in this study since such analysis would not directly contribute to answering any of the research questions.

Participants were asked to write down on the provided electronic notepad their questions and retrieved answers every time they consulted a clinical reference tool.

When participants decided they had gone as far as they could with the patient case (whether or not they thought they had reached a diagnosis), they proceeded to the last stage where they wrote a summary of their thinking process for the solution of the case.

At the end of each session, participants were asked to fill out a post-test questionnaire. More information on the questionnaire is provided in the Materials section of this chapter.

Participants worked on a laptop computer with internet access so that they could access the clinical reference systems of their choice. The only task that they did not perform on the computer was the ordering of laboratory tests, which was done by filling out the *Laboratory test order form* (see Materials section). Both patient cases and their respective laboratory test results were based on real cases. Some laboratory test results were not available because they were not performed on the patient. Upon submitting a written laboratory test order, participants received a printout with the results of the ordered tests, if the test was performed. Participants did not know which tests were actually performed. They were informed that a given laboratory test was not performed only if they ordered the test. All available test results were printed in advance.

All sessions were recorded using Camtasia, a commercial application that records screen activity and converts it into a video file. The video files were used mainly to time the duration of each problem-solving stage (e.g., generation of hypotheses, revision of initial hypothesis list, evaluation of hypotheses) and to track information-seeking activities.

Materials

Cases

Two patient cases were used for this study. The two cases were solved by all participants. Both patient cases were extracted from the website of the Department of Neurology at Baylor College of Medicine (http://www.bcm.edu/neurol/index.html), an open source library of neurological cases. The diseases were Kennedy's disease (X-

linked spinobulbar muscular atrophy) and Friedrich's ataxia. The presentation and solution of the two cases are reproduced in Appendix C.

Indeed clinical diagnosis is not an exact science. That is, sometimes it is not possible to identify with undeniable precision the causal relationships associated with certain health conditions. Both cases used in this study were chosen mainly for methodological reasons. Kennedy's disease and Friedrich's ataxia are both neurological diseases that can be detected by specific laboratory tests. The use of cases that can yield a precise diagnosis is a methodological necessity in a study involving diagnostic accuracy.

Laboratory Test Results and Laboratory Test Request Forms

Results from laboratory tests (available at the same website from where the cases were taken) were printed out and made available to participants upon their request through the *Laboratory test request form*. Participants were required to justify in the *Laboratory test request form* (see Appendix D) every test or study they ordered to avoid guessing. Justifications were not analyzed.

Data Collection Application

Participants worked on the cases using an application especially developed for this study (see Figure 5). I designed the application with the assistance of a colleague who also programmed it using *Revolution* (http://www.runrev.com/). The case is presented in the left half of the window. Hypotheses are managed on the right half of the window. The case presentation box has a highlighting tool that allows participants to mark parts of the text that they judge important. The application allows the creation of a maximum of 13 working hypotheses. Each hypothesis is composed of three elements: name (e.g., multiple sclerosis); confidence level (ranging from 0 to 100%); and

supporting evidence (e.g., progressive muscle weakness, fasciculation, hyperthermia). Participants use the application in a linear fashion, going through four stages:

1. Generation of an initial hypothesis list;
2. Revision of hypothesis list (using *Abduction*);
3. Evaluation of hypothesis list[5] (using clinical reference systems of participant's choice); and
4. Case summary.

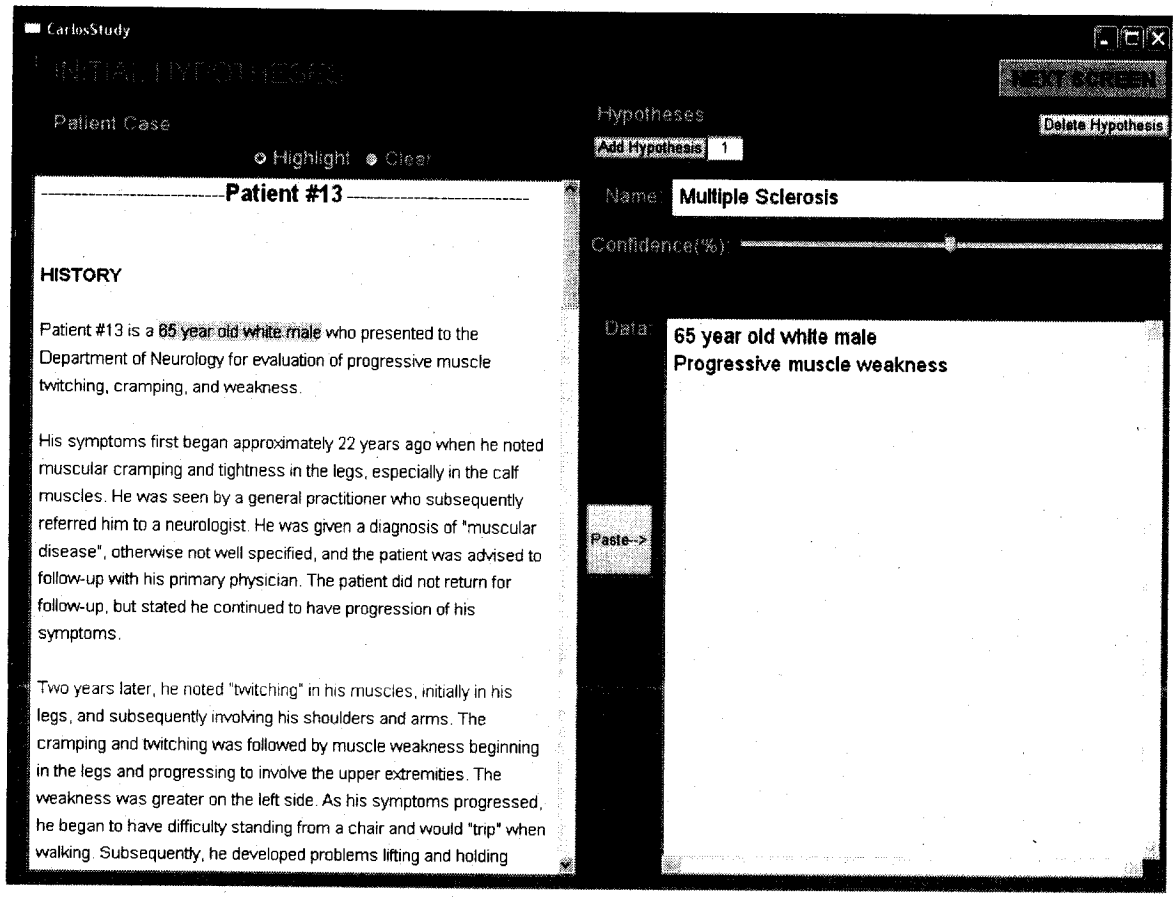

Figure 5. Data collection application

[5] At this stage additional hypotheses still may be generated as participants consult the clinical reference systems of their choice

The screens for hypothesis generation, revision, and evaluation are identical. The distinction between these three stages was made for analytical purposes: to allow the researcher to analyze how the hypotheses change after consultation of *Abduction* and the chosen clinical reference systems.

At the end of a session, the application produces a log-file containing all information entered by the participant. Appendix E presents a detailed description of how the data collection application works.

Abduction

Abduction, the core material in this study, is a computer application intended to support medical students to generate diagnostic hypotheses. *Abduction* is based on a database of diseases and associated symptoms that can be searched by symptom (see Figure 6). Currently, the database is populated only with neurological cases. The list of symptoms is organized by body systems (e.g. neurological, cardiovascular, endocrine). When users select multiple symptoms, the system returns all diseases associated with the selected symptoms. The diseases that best match the selected symptoms are displayed at the top of the list. In front of the name of each disease, the system displays (in percentage) the level of the match. For example, if five symptoms were selected and a disease matches four of those symptoms, that disease would be displayed as an 80% match. The matching and non-matching symptoms are displayed to the right of the name of the disease, the former in green and the latter, in red. *Abduction*'s database currently has 71 diseases, 158 symptoms, and 750 connections between diseases and symptoms. A detailed description of *Abduction* operation is provided in Appendix F.

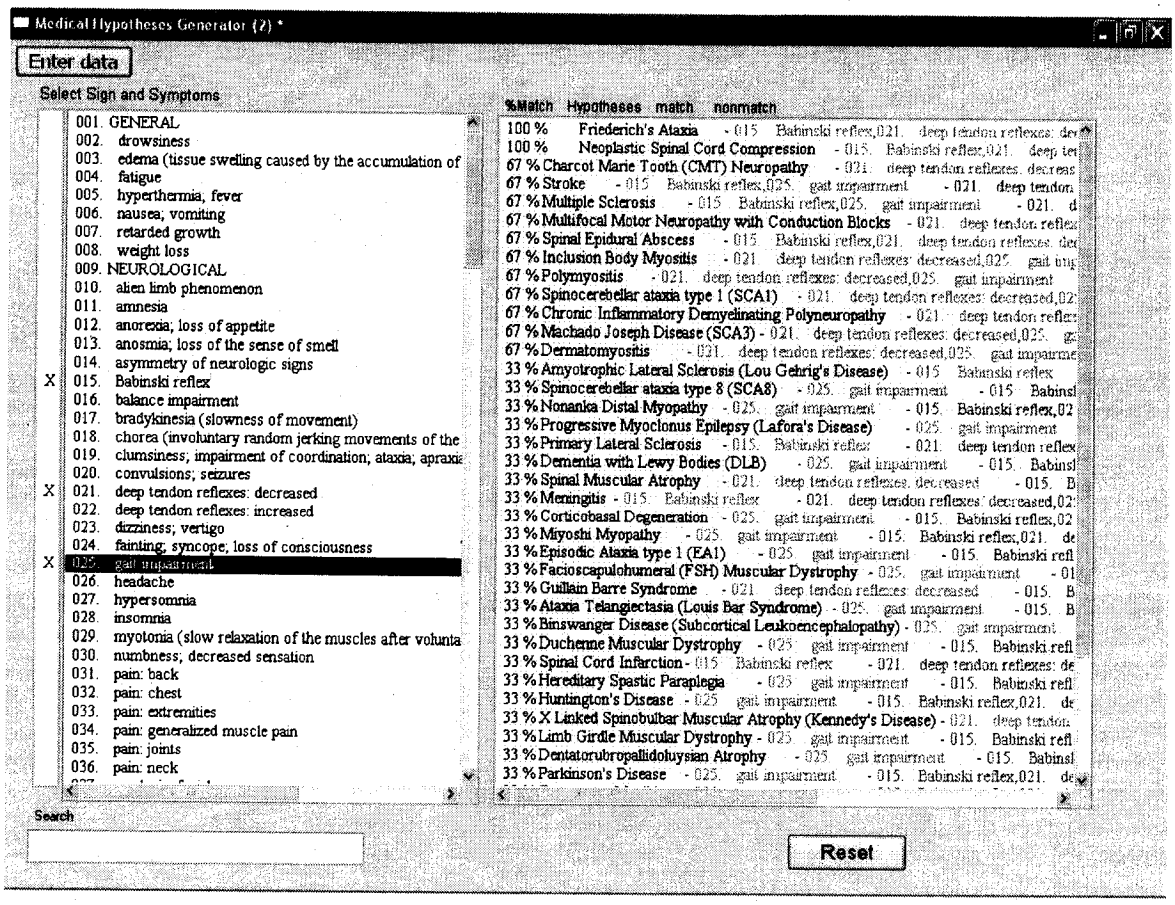

Figure 6. Abduction

Abduction, as the name itself implies, was conceived to provide support to abductive reasoning and it does that in two complementary ways. First, by functioning as an extension of one's declarative knowledge. Second, by allowing users to visualize the existing patterns in a patient case. In this regard, *Abduction* may be considered a class of cognitive tool (Lajoie, 2000; Lajoie & Derry, 1993) that goes beyond simple mnemonic extension, allowing its users to perform some pattern recognition operations that have been the prerogative of medical experts.

Beta-testing of Abduction. To test whether *Abduction* can actually give good advice to its users, a beta-test of the system was conducted. In it, the subject (a person non-related to the health sciences) was presented with seven patient cases (including the

two cases used in the study) and asked to select the relevant symptoms from each case and input those symptoms in *Abduction*. For six of the seven cases, the correct diagnosis came at the top of the list of diseases. For one of the cases, the correct diagnosis was the second in the list. The only case which was not displayed at the top of the list happened to be one of the cases that was actually used in the study (Kennedy's disease).

Abduction's purpose is not to prescribe the solution to the case, but to increase the likelihood that the user considers the correct diagnosis when formulating a hypothesis list. *Abduction* was designed based on user-centered principles. That is, it was designed to give the user total control over which symptoms to input in the system and which suggestions offered by the systems to accept. Consequently, the exact position of the correct diagnosis in the ranked list of diseases has a relative importance. It should be only high enough in the rank to persuade the user to take it into consideration. The beta test of the system showed that *Abduction* ranks the correct diagnosis considerably high. The main study should show whether the correct diagnosis is displayed high enough in the list to be considered by the participants.. The beta-test was conducted independently of the pilot study.

Electronic Notepad

In order to evaluate participants' information-seeking behavior, participants were asked to write down their questions and retrieved answers every time they consulted a clinical reference tool. Annotations were made on Memento, a freeware electronic notepad (http://www.guyswithtowels.com/downloads.html).

Post-Test Questionnaire

Participants were asked to fill a post-test questionnaire (see Appendix G) with questions about their information-seeking habits and their opinion about the level of difficulty of the cases and about the usefulness of *Abduction.*

Pilot Study

A pilot study with three second-year medical students was conducted to check for problems in the data collection procedures and materials, evaluate the level of difficulty of the patient cases, and obtain an estimate of problem-solving times. Participants in the pilot spent an average of two hours to work on the two cases. The same cases were used for the pilot and the study.

In order to see the effects of the use of different information systems, it was necessary to work with cases that participants would have difficulty solving on their own. Consequently, the design of this study required using patient cases with a high level of diagnostic complexity. None of the participants of the pilot study were able to generate the correct diagnosis without the support of an information system. The participants' choice for systems to consult were: Harrison's, eMedicine, UpToDate, and Google.

The three participants thought *Abduction* was helpful and easy to use. When asked if they would use *Abduction* if it were available online, the answer was positive.

Analysis

Main Outcome Measures

Two binary measurements for assessing participants' hypothesis lists were used: presence of correct diagnosis in hypothesis list (present or not present) and confirmation of correct hypothesis in case it was generated (confirmed or not confirmed).

The main outcome measures were used to answer the four first research questions. The 32 occasions (16 participants vs. two cases) were used as the unit of analysis to provide a directly interpretable representation of the effects of the consulted systems on participants' diagnostic reasoning.

For the first research question, the number of times the correct diagnosis appeared in the participants' hypothesis list before and after the use of *Abduction* were compared.

For the second research question, the number of times the correct diagnosis appeared in the participants' hypothesis list before and after the use of the clinical reference systems of their choice were compared.

For the third research question, the McNemar's test for correlated proportions was conducted to detect differences between the effects of *Abduction* and the consulted clinical reference systems. The rationale for using the McNemar's test instead of a chi-square test is better explained by Levin and Serlin (2000):

> In a nonrepeated-measures research context, one may wish to compare the performance of specially instructed participants with their matched-pair control counterparts on a dichotomously scored item or on a pass-fail mastery test. In that regard, it should be noted that in situations where matching has been employed, comparing the proportions of "successful" instructed and uninstructed participants via a two-sample chi-square test of homogeneity is not statistically appropriate -- just as an independent samples *t* test would not be appropriate for assessing a difference in means between the two matched samples. (¶ 3)

For the fourth research question, the number of times the participants managed to confirm the correct diagnosis was tallied. The correct diagnosis was considered confirmed by the participant in either of the following situations:

1. If the correct diagnosis was associated with a confidence level of 90% or more in the confidence meter (and there were no competing hypotheses associated with higher confidence levels);
2. If the participant explicitly named the correct diagnosis as his/her final diagnosis in the case summary.

For the fifth research question, the information-seeking operations conducted by successful and unsuccessful participants were compared in order to investigate whether failure in confirming the correct diagnosis in situations where it was generated was due to inadequacies in the consulted clinical reference systems or to participants' inability to interpret and apply the retrieved information (see Figure 7).

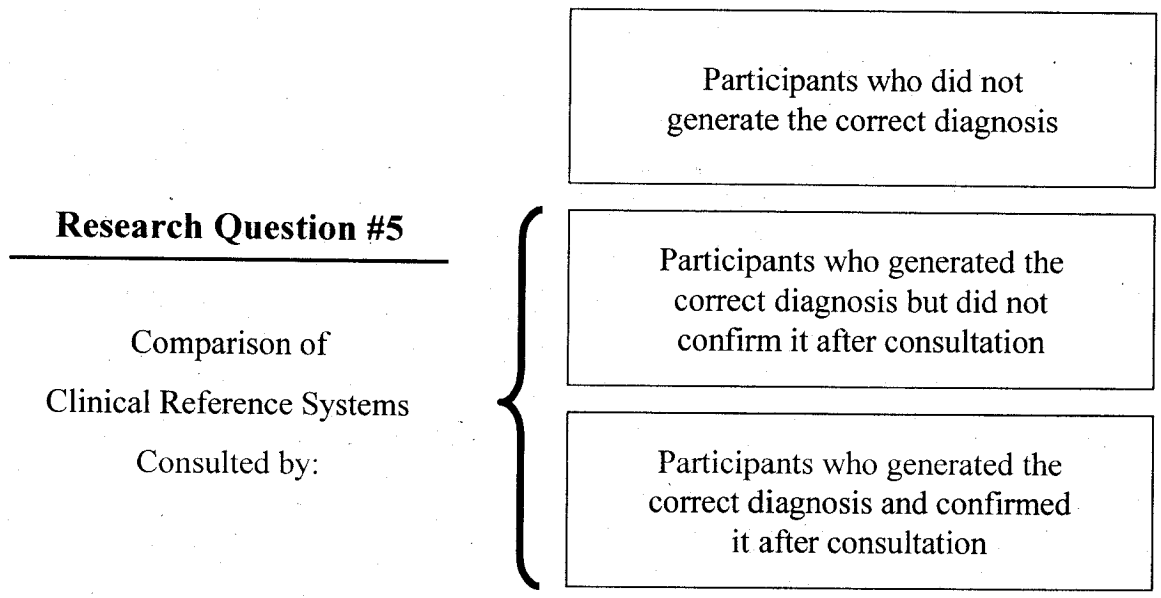

Figure 7. Analysis of research question #5

Table 2 shows the purpose of each type of data that was collected in this study.

Table 2

Summary of Data Collected and Analyzed

Data Source		**Used to:**
Hypothesis lists	1. Initial	Verify if participant can generate correct diagnosis without assistance
	2. After consultation of *Abduction*	Verify if participants can generate the correct diagnosis with the support of *Abduction* (Condition A)
	3. After consultation of chosen clinical reference systems	Verify if participant can generate the correct diagnosis with the support of the chosen clinical reference systems (in Condition B) Verify if participant can confirm the correct diagnosis, in case it was generated, with the support of the chosen clinical reference systems (in Conditions A & B)
Elements of each hypothesis	Name	Verify if participant can generate the correct diagnosis
	Confidence level	Rank hypotheses Identify final diagnosis in case it was not explicitly stated in summary
	Supporting evidence	Discourage guesswork (not analyzed)
Summary		Identify final diagnosis (not analyzed as verbal protocols)
Laboratory Tests Ordered		Not Analyzed
List of questions to be searched		Identify what questions students generate when trying to solve a patient case and whether or not they are able to find the answers to those questions
Post-test questionnaire		Identify which systems participant consult regularly Obtain participant's opinion about the usefulness of *Abduction*
Video file		Calculate problem solving time Track information seeking operations Verify if participant has any difficulties using *Abduction*

Chapter IV: Results

Results by Research Question

RQ #1: Can medical students use Abduction to increase the likelihood that the correct diagnosis is included among their working hypotheses?

In Condition A (consultation of *Abduction* and chosen clinical reference systems), none of the participants were able to generate the correct diagnosis before consultation. After consulting *Abduction*, all 16 participants (100%) were able to include the correct diagnosis in their hypothesis list.

Table 3 displays a summary of the main outcome measures in this study. A table containing all 32 measurements is provided in Appendix H. A table containing the subtotals by condition, case, and cohort is also included in Appendix H.

Table 3

Main Results: Number of Times Correct Diagnosis Was Generated and Confirmed in Each Condition

	Hypothesis Generation ↓		**Hypothesis Evaluation** ↓
	Correct Diagnosis Generated?		
	Initial Diagnosis List (Unassisted)	Revised Diagnosis List (Assisted)	Correct Diagnosis Confirmed?
Condition A (N = 16)	0	16	13
Condition B (N = 16)	0	3	1
Total (N = 32)	0	19	14

Before consultation, participants generated a mean of 3.38 hypotheses per patient case. After consulting *Abduction*, participants added a mean of 2.69 hypotheses to their hypothesis list (total: 6.19 hypotheses). That means that, with the support of *Abduction*, medical students can insert the correct diagnosis in their hypothesis list by adding less than three hypotheses on average. Figure 8 shows the number of hypotheses generated by each participant before and after consultation. The hypotheses are organized in a way to give a better sense of the proportion between hypotheses created with and without assistance at the level of the individual participants as well as at the level of the sample of participants as a whole.

When consulting *Abduction*, participants often initiated an iterative process of selecting symptoms, analyzing the suggestions given by the system, and adding or changing hypotheses in their hypothesis list. Consequently, due to the dynamic nature of the process, it was not possible to produce an average of the correlation between the ranking of a disease in *Abduction* and the likelihood of it being selected by the participant.

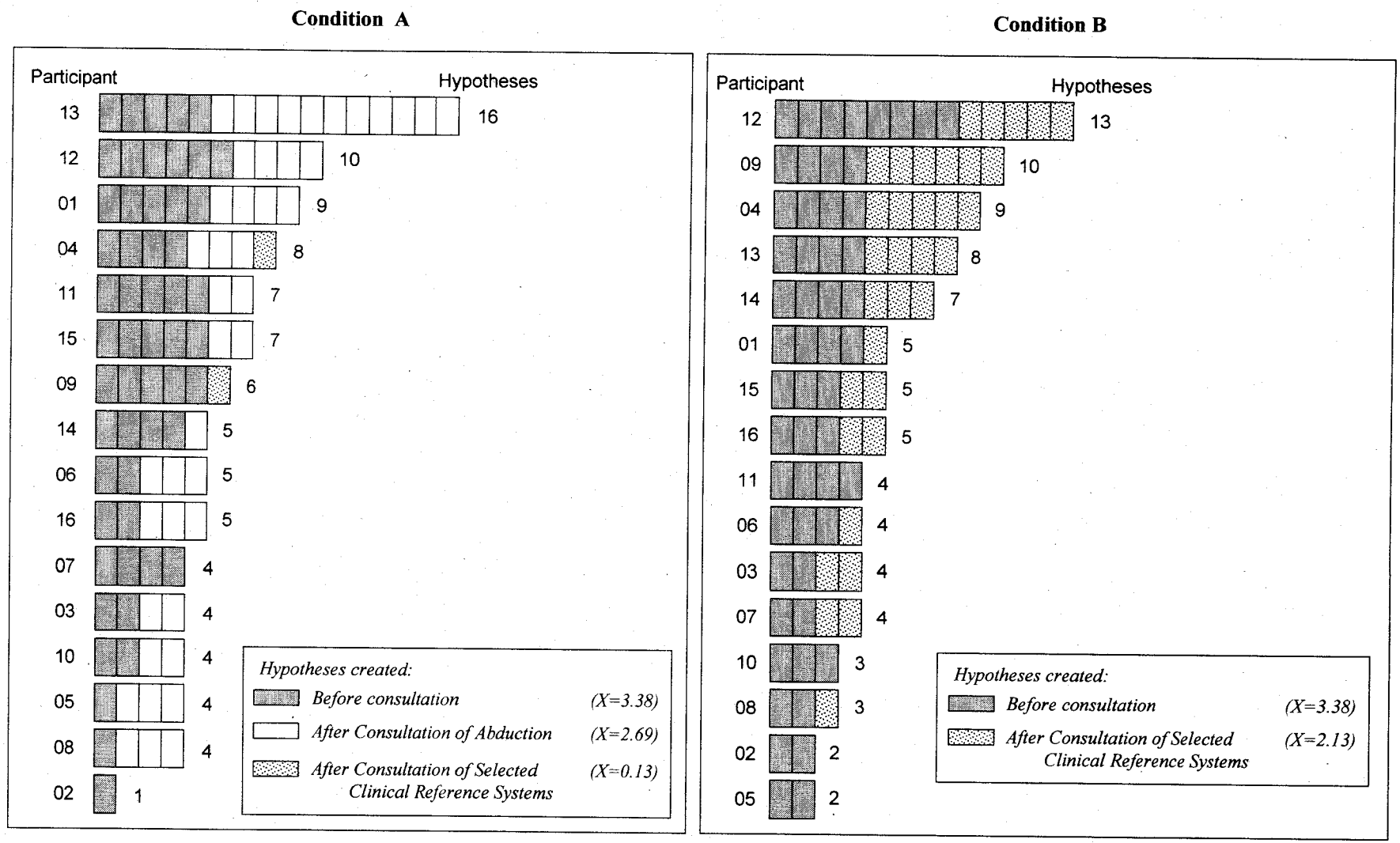

Figure 8. Number of Hypothesis Generated for Conditions A & B

RQ #2: Can medical students use the clinical reference systems that they use on a regular basis to increase the likelihood that the correct diagnosis is included among their working hypotheses?

In Condition B (consultation of chosen clinical reference systems), none of the participants were able to generate the correct hypothesis before consultation. After consulting the clinical reference systems that they regularly use, three participants (18.75%) were able to include the correct diagnosis in their hypothesis list.

Before consultation, participants generated a mean of 3.38 hypotheses per patient case. After consulting the clinical reference systems of their choice, participants added a mean of 2.13 new hypotheses to their initial hypothesis list (total: 5.50 hypotheses) (see Figure 8). After consulting the chosen clinical reference systems, participants added on average 0.56 less hypotheses than when consulting *Abduction*. However, they were, on average, 5.33 times less successful in inserting the correct diagnosis in their hypothesis list.

For the Kennedy's disease case, six new diseases were added to the hypothesis list after participants consulted the clinical reference systems that they regularly use. Kennedy's disease was not among the added hypotheses. For the Friedrich's ataxia case, 19 new diseases were added to the hypothesis list. Friedrich's ataxia was included three times.

RQ#3: Can Abduction provide better support to hypothesis generation than the clinical reference systems that are most used by medical students?

A McNemar's test for correlated proportions (see Table 4) was used to test for differences between Conditions A and B (with and without the use of *Abduction*) regarding their effects on hypothesis generation. In this case, the null hypothesis was that there were no significant differences between the two conditions (H_0: $\pi_1 = \pi_2$).

The critical value for the contingency table in testing H_0 at the .01 level of significance is ${}_{.99}\chi_1^2 = 6.64$. Since the computed chi-square of 13 exceeds the critical value of 6.64, H_0 is rejected at the .01 level.

More sensitive tests, such as a paired t-test could also be performed in this situation. However, since significant differences were already found through a nonparametric test, I opted for not performing further tests.

Table 4

McNemar's Test for Correlated Proportions

		Condition B (without *Abduction*)		
		Correct diagnosis included	Correct diagnosis not included	
Condition A (with *Abduction*)	Correct diagnosis not included	0	0	0
	Correct diagnosis included	3	13	16
		3	13	16

$$\chi^2 = \frac{(n_{22} - n_{11})^2}{n_{22} + n_{11}} = \frac{(13-0)^2}{13+0} = \frac{169}{13} = 13$$

RQ #4: Can medical students use the clinical reference systems that they use on a regular basis to confirm the correct diagnosis (in case it was generated)?

Sixteen participants attempted to solve two patient cases, totaling 32 occasions[6]. The correct diagnosis was generated on 19 occasions (59.4%), 16 occasions in Condition A (with *Abduction*) (100%), and three occasions in Condition B (without *Abduction*) (18.7%). Participants were able to confirm the right diagnosis on 14 out of the 19 occasions in which the correct hypotheses was generated (73.7%). It can be concluded that the effectiveness of clinical reference systems in supporting hypothesis testing is in the order of 70%, provided that the correct diagnosis was already generated (see Figure 9).

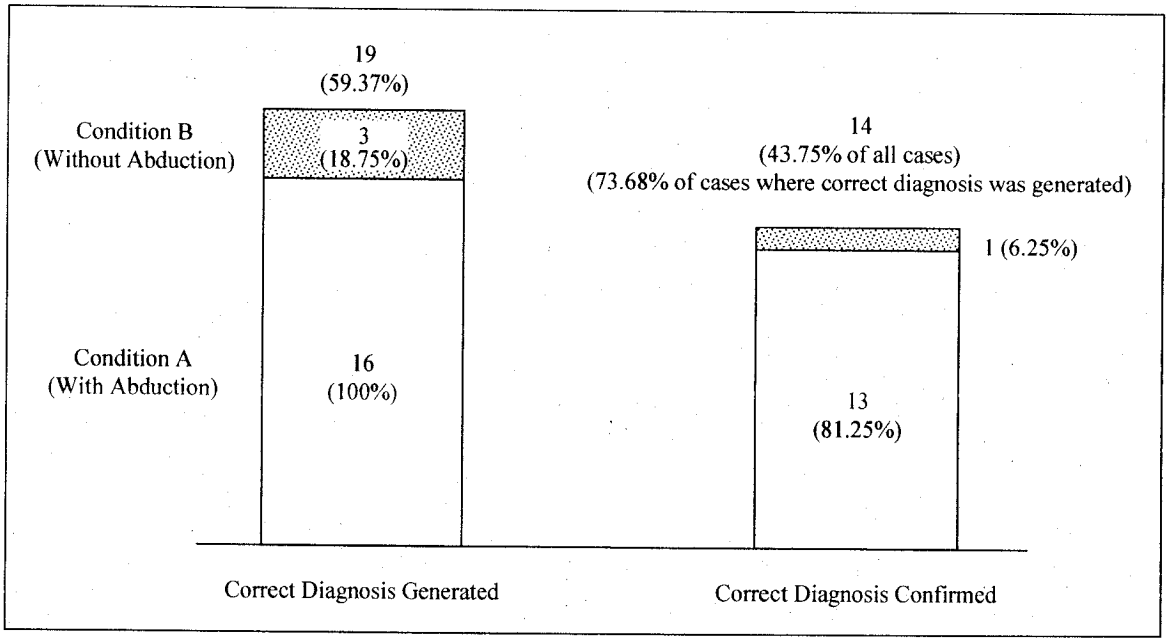

Figure 9. Number of occasions where the correct diagnosis was generated and confirmed.

[6] An occasion refers to each event where a participant solves a case.

RQ #5: In cases where the correct hypothesis was generated but not confirmed, is it due to the student's inability to interpret and apply the retrieved information or to inadequacies in the clinical reference systems?

The correct diagnosis was generated on 19 occasions. In five of those occasions the correct diagnosis was not confirmed. In all of the five occasions the case was Friedrich's Ataxia. Three of those occasions were in Condition A and two of them were in Condition B (see Figure 10). One way to investigate whether the failure in confirming the right diagnosis is due to the student's inability to interpret and apply the retrieved information or to inadequacies in the clinical reference systems is by comparing the systems consulted by the participants who succeeded at solving the case with the ones that did not. Table 5 shows which systems were consulted by the successful and unsuccessful participants on the topic of Friedrich's Ataxia.

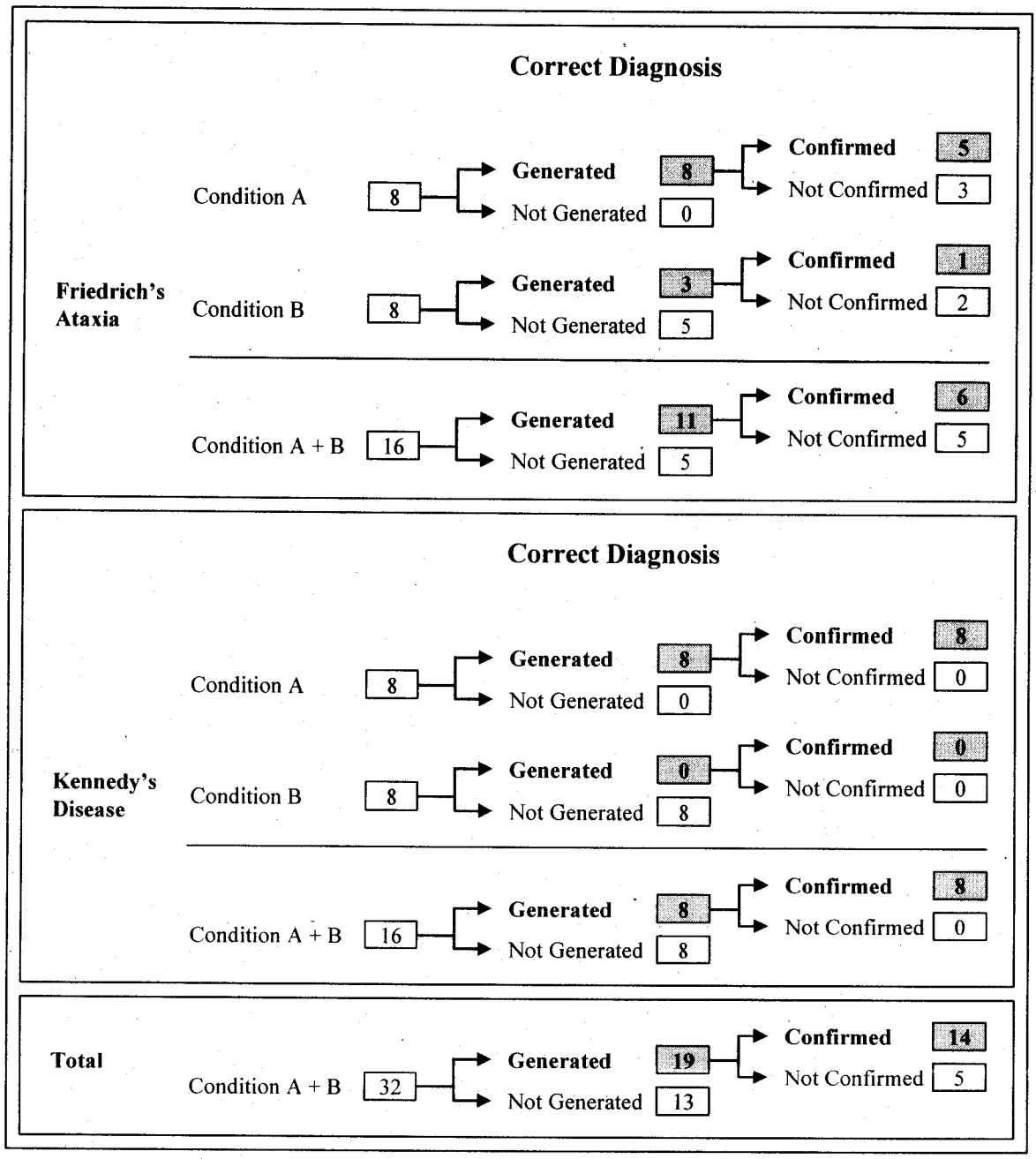

Figure 10. Number of times the correct diagnosis was generated and confirmed by patient case.

Table 5

Clinical Reference Systems Used by Successful and Unsuccessful Participants When Working on Friedrich's Ataxia Case

Participant	Information System			
	Harrison's	eMedicine	UpToDate	Google
Successful				
2*				
7	x	x	x	
10			x	
14		x	x	
15			x	
13	x			x**
Total	2	2	4	1
Unsuccessful				
3	x	x		
6		x	x	
11	x			
9			x	
16			x	
Total	2	2	3	

* Participant 2 chose to issue a final diagnosis immediately after consulting *Abduction*.

** The search performed on Google took the participant to the T. J. Samsom website (http://tjsamson.client.web-health.com/).

Table 5 shows that the clinical reference systems used by the successful and unsuccessful participants were basically the same. Harrisons's Online was used twice by both successful and unsuccessful participants. eMedicine was also used twice by both successful and unsuccessful participants. UpToDate was used four times by successful participants and three times by unsuccessful participants. The search engine Google was used by one successful participant but was not used by any unsuccessful participant. This suggests that failure in solving the case is most probably due to the student's inability to interpret and apply the retrieved information rather than to inadequacies in the consulted information systems. In other words, enough additional information to allow for proper hypothesis evaluation was provided by the consulted information systems but not properly used by the unsuccessful participants.

A more detailed analysis of the written protocols related to the information-seeking operations during hypothesis evaluation was not possible due to great variations in the protocols across participants. A few participants produced very detailed protocols but most of them produced incomplete protocols (sometimes forgetting to write down a question or an answer) rendering the data unreliable. Although I monitored each participant closely and reminded them several times during the course of the sessions to write down every question and answer, I was not able to prevent these gaps in the protocols. Consequently, data were not sufficiently detailed to allow me to perform the kind of analysis that might have identified without doubt which factor or factors prevented participants from confirming the correct diagnosis.

Additional Results

Problem-solving times. When participants used *Abduction* in addition to the clinical reference systems they regularly use, they experienced a considerable boost in their problem solving performance without any increase in problem solving time. The mean time for case resolution was 57:12 minutes in Condition A and 58:74 minutes in Condition B (not including time spent writing case summary). This is an important factor to consider since the amount of time spent with medical information systems and decision support systems has been frequently cited one of the major barriers to their adoption.

No comparisons of unassisted versus assisted problem solving times will be offered because participants in this study knew that they would be allowed to consult several clinical reference systems after the initial stage where they worked on the case without any support. Therefore, the actual time they spent working on the cases on their own was probably shorter than the time they would dedicate to the problem if they knew no consultation would be allowed afterwards.

Participants' perception of Abduction. The relative effectiveness of a system and its perceived usefulness are not necessarily the same. Medical students' predilections for certain information systems may not be proportional to those systems' actual impact on their problem solving performance. In the post-test questionnaire, participants were asked to rank the usefulness of *Abduction* in a four-point Likert scale (very useful, useful, slightly useful, not useful). Fourteen participants (88%) ranked *Abduction* as *very useful*, one participant (6%) ranked it as *useful*, and one participant (6%) ranked it as *slightly useful*. None of the participants ranked *Abduction* as *not useful*. These results suggest a

certain correspondence between the actual effectiveness of *Abduction* and its perceived usefulness.

Chapter V: Discussion

Hypothesis Generation

This study was born from the argument that hypothesis generation and hypothesis evaluation are two complementary but quite different cognitive processes in clinical reasoning. In hypothesis generation, one reasons from the symptoms to the disease. In hypothesis evaluation, one reasons from the disease to the symptoms. Being organized by disease, existing information systems do not provide adequate support to hypothesis generation. To test this assumption, *Abduction*, a hypothesis generation tool, was developed to test the effects of specific support to hypothesis generation in comparison to popular clinical reference systems.

Sixteen medical students were asked to solve two diagnostically challenging patient cases. None of the participants (0%) were able to generate the correct hypothesis on their own. When allowed to use *Abduction*, all participants (100%) were able to generate the correct diagnosis. When using the clinical reference systems of their choice, only three participants (18.75%) were able to generate the correct diagnosis.

The above results show that there are significant advantages in providing specific support to hypothesis generation. *Abduction* proved to be five times more effective at supporting hypothesis generation than the clinical reference systems that are regularly used by medical students.

When medical students consult information systems to help them solve a patient case, they do not necessarily make a clear distinction between hypothesis generation and hypothesis evaluation. Nonetheless, some of their questions demonstrate that they require support when they are reasoning from the symptoms to the disease, as can be seen in the

following questions extracted from the questions sheet they were asked to fill out when conducting online searches.

- What diseases present with cramps, twitch, weakness?
- Diseases with "Stocking and Glove" loss of sensation?
- What are the most common diagnoses in young patients with gait disturbances?
- What can decreased iliopsoas strength be indicative of?
- What is the differential diagnosis of gait ataxia in a patient of this age?

One noticeable aspect of this study regarding hypothesis generation is that students benefited far more from *Abduction* than from the systems that they use on a regular basis despite the fact that they were using *Abduction* for the first time. This shows that there is almost no learning curve associated with the use of *Abduction*. Further, the use of *Abduction* did not increase the mean time for case resolution in comparison with the exclusive use of other systems. From an implementation perspective, these are important outcomes since two major reasons cited by physicians for not using information systems on a regular basis are that they are too time-consuming and cumbersome (Covell, Uman, & Manning, 1985).

Participants demonstrated one unexpected behavior when using *Abduction*. They used it not only to generate new hypotheses but also to evaluate existing ones. On occasion, they would verify how the hypotheses they had already created were ranked in *Abduction* and, in some cases, even change the confidence level of a hypothesis after consultion. These unexpected maneuvers show that, although *Abduction* was specifically

designed to support hypothesis generation, some users will try (and eventually succeed) to extract more help than the application was intended to provide.

There is an inevitable gap between 'function' and 'use'; between what a tool was designed to do and the ways it is actually used (Lillehaug & Lajoie, 1998). Depending on how it is designed, a tool can discourage people to use it to its full potential or it can incite users to take it beyond its original design. User-centered consultation systems (Buscher et al., 2005) aim for the latter. These systems operate by trusting in the users' competence. They give the user full control over the process. In the case of Abduction, users decide which set of symptoms they will work with and which suggestions from the system they will accept. The process is transparent to the user: the list of diseases is instantly updated as symptoms are selected and deselected. These are the main advantages of *Abduction* over clinical reference systems and over decision support systems (DSS), which mostly work by forcing the user through a decision tree based on a probabilistic algorithm and, at the end, produces a ranked list of diseases. Further, there are no DSS that are freely available to medical students, which partially explains their low popularity. None of the participants in this study use a DSS on a regular basis.

Hypothesis Evaluation

For hypothesis evaluation (confirmation of the right diagnosis), the clinical reference systems that are regularly used by medical students were effective in 73.68% of the cases in which participants had generated the correct diagnosis. However, if a distinction is made between cases solved with and without specific support to hypothesis generation, the numbers change to 81.25% and 33.33%, respectively. These differences indicate that the effects of specific support to hypothesis generation carry over to the

hypothesis evaluation stage. Consequently, it is fair to conclude that differentiated support to hypothesis generation and hypothesis evaluation have positive effects on the diagnostic accuracy of medical students for diagnostically challenging cases. Furthermore, when analyzing the effects of computer-based support to diagnostic reasoning, the distinction between hypothesis generation and hypothesis evaluation does seem to provide us with more precise methods to evaluate the effects of those support systems.

For diagnostically challenging cases, the combined use of *Abduction* and other clinical reference systems has a greater effect on diagnostic accuracy than decision support systems. The review of studies on decision support systems by Hunt, Haynes, Hanna, and Smith (1998) have reported that there is still a lack of evidence regarding the effects of decision support systems on patient outcomes. A later study by Friedman et al. (1999) on two decision support systems (Iliad and QMR) showed net gains in diagnostic accuracy of 6%.

Information-Seeking

Unlike decision support systems, clinical reference systems do not provide suggestions or answers. Rather, users must find the answers to their questions through keyword searching and/or browsing. This search results in longer consultations depending on the user's information-seeking proficiency and does not always produce clear-cut answers. In the results section of this study it was concluded that failure in confirming the right diagnosis was mostly due to inability to interpret or apply the information contained in the consulted clinical reference systems since successful and unsuccessful participants consulted the same systems. However, since there was an equal

proportion of successful and unsuccessful participants, it can also be argued that the consulted systems contained the necessary information but that the information was not necessarily organized in the most optimal way to support hypothesis evaluation. A more detailed analysis of the written protocols related to the information-seeking operations during hypothesis evaluation was not possible due to great variations in the protocols across participants.

Other researchers have argued for the need to move the research on medical information systems beyond measures of relevance of the retrieved information to the assessment of how the use of information systems affects problem-solving performance (Gorman, 1995). This study has addressed that issue by evaluating how the use of two different types of systems, namely, *Abduction* and clinical reference systems, affect the diagnostic accuracy of medical students. An absolute answer could not be reached. This study has shown that clinical reference systems offer better support to hypothesis evaluation than to hypothesis generation. However, these systems were not designed to provide optimal support to hypothesis evaluation either. It is possible to argue that the use of a consultation system specifically designed to support hypothesis evaluation could have produced a higher percentage of diagnostic accuracy. Such a system would be an abridged version of a clinical reference tool with a different organization of its content. Florance (1996) has proposed such a system based on content extracted from Medline, a bibliographic database. However, her system was intended to offer support to experienced physicians. For medical students, a system based on the content of full-text databases such as the ones used by the participants in this study would probably be more effective.

Chapter VI: Conclusion

Contributions

This study was intended to contribute to scholarship at both the research and development levels. At the research level, it has argued against the study of diagnostic reasoning as a monolithic process. Instead, it was based on the often criticized hypothetico-deductive method, which makes a clear distinction between the complementary but directionally opposite processes of hypothesis generation and hypothesis evaluation. This choice of model enables the researcher to regard the changes in the directionality of reasoning (from clues to hypotheses or from hypotheses to clues) as an essential process in diagnostic reasoning. This understanding cannot be reached if the researchers adopt a conceptual approach based on dichotomies such as forward versus backward reasoning or weak versus strong methods.

The second advantage in the adoption of the hypothetico-deductive model is that it incorporates the auxiliary processes of information-seeking or inquiry. Studies on diagnostic reasoning have systematically ignored information-seeking as a component of diagnostic reasoning although it is widely recognized that it is common practice among physicians and medical students. Conversely, studies on the information needs and information-seeking behavior of physicians never went as far as analyzing how the retrieved information affects the original problem solving situation. This study fills this gap by integrating both lines of research.

At the development level, *Abduction* has proved to be highly effective in the support of hypothesis generation and, at the same time, easy and quick to use. It combines the advantages of decision support systems (e.g., PKC, Isabel, Iliad, QMR) and

differential diagnosis PDA tools (e.g., Diagnosaurus, First Consult). Decision support systems are considered cumbersome and time-consuming because they require the input of extensive patient information. DDx PDA tools, on the other hand, are quick and easy to use but they do not work with multiple symptoms simultaneously. *Abduction* has the same friendly user-interface of DDX PDA tools plus the added benefit of working with multiple symptoms.

Abduction could also have a positive impact on medical education. The problem based learning (PBL) approach widely adopted to teach clinical diagnosis can benefit from the use of a consultation system that is easy and quick to use. A PBL class usually relies on intense scaffolding by the instructor. That is, the instructor guides the students and also fills-in the gaps in their medical knowledge so that they can complete the patient case they are trying to solve. However, the benefits of PBL are conditional to intense practice. Students must practice solving cases far beyond the limited hours that they spend under the instructor's direct supervision. The problem is that, without the supervision, students loose not only the guidance but also the influx of relevant medical information that is provided by the instructor. The use of *Abduction* could address the latter issue by offering students some support to allow them to practice solving patient cases on their own.

Limitations of the Study

This study has two classes of limitations worth discussing. One of them refers to the limitations of *Abduction.* In its current version, the prototype of *Abduction* is populated exclusively with neurological diseases. Consequently, the results obtained in

this study cannot be indisputably generalized to other medical domains. One could also make the argument that, because *Abduction* is not based on any sophisticated probabilistic model such as Bayesian networks, its effectiveness would significantly decrease when the database is expanded to include diseases from other medical domains. The algorithm I used to populate the database is supposed to be robust against changes in magnitude. However, the only sure way to verify that is to further populate the database and conduct new studies.

Regarding the study itself, there are a few points that need to be considered. Given the sample size of eight second-year and eight fourth-year medical students, a statistical comparison between the effects of *Abduction* and other clinical reference systems on different student cohorts could not be performed. Because only two cases were used, it was not possible to detect variations of the effects of Abduction and other clinical reference systems depending on the level of difficulty of the patient case being solved. Finally, a more detailed analysis of the written protocols related to the information-seeking operations during hypothesis evaluation was not possible due to great variations in the protocols across participants.

Future Research

The results obtained in this study suggest some possibilities for further research. Future studies could benefit from larger sample sizes where differences between student cohorts can be analyzed. Future studies should also consider working with a broader spectrum of patient cases so that the benefits of the consultation systems can be analyzed in relation the level of difficulty of the patient cases. For more in-depth qualitative analyzes of information-seeking operations, audio recording of verbal protocols (rather

than written ones) based on prompts by the researcher may produce data that is more consistent and also more homogeneous across participants. A qualitative analysis could reveal why the unsuccessful participants failed in applying the retrieved information to reach the correct diagnosis. That could be achieved by prompting the participants to state:

1. The goal behind each information seeking operation;
2. Their interpretation of the retrieved results;
3. Their assessment of whether or not the retrieved information answers their question;
4. Their opinion about what would be the ideal answer in terms of organization and presentation of the information;
5. In case they cannot find the information, which combination of factors led them to stop searching.

A post-task interview where participants are asked to provide a retrospective reflection about their searches could be used as triangulation data. This method is indeed quite intrusive and would therefore demand strong interviewing skills from the researcher to make sure that participants provide all the necessary information and, at the same time, do not become frustrated or defensive which could have an impact on their problem solving performance. The results of such a study could produce great insights on how to modify consultation systems to provide optimal support to hypothesis evaluation.

Regarding *Abduction*, before more studies are conducted, it will be necessary to further populate its database including other categories of diseases. The process of scaling up *Abduction*'s database would imply some modifications in the user interface. One possible way to maintain the current level of user-control would be to offer the user

the option of choosing which categories of diseases to display in the disease list. That could be accomplished by adding checkboxes that can be selected in any combination (see Figure 11).

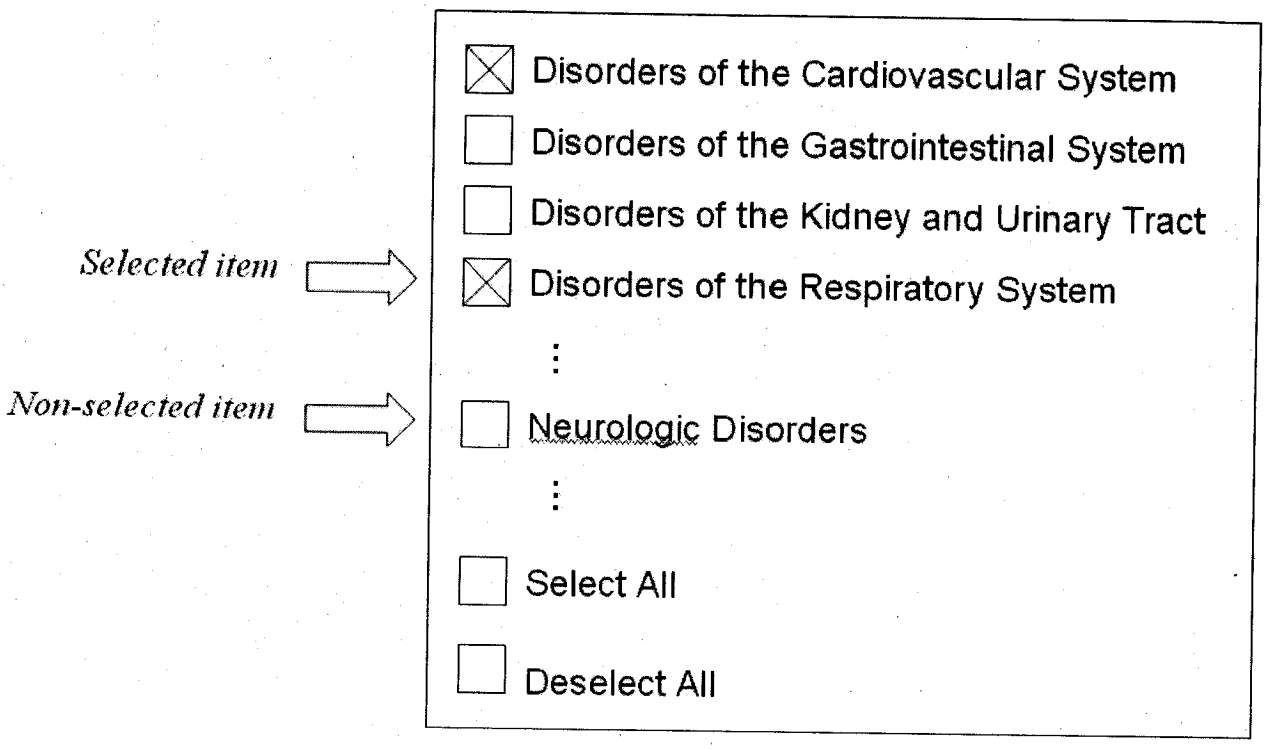

Figure 11. Example of the use of checkboxes to give the user control over which categories of disease the system will display.

Summary

The hypothetico-deductive method, which involves an iterative process of hypothesis generation and evaluation, has been used for decades by physicians to diagnose patients. This study focuses on the levels of support that medical information systems can provide during these stages of the diagnostic reasoning process. The physician initially generates a list of possible diagnoses (hypotheses) based on the patients' symptoms. Later, those hypotheses are examined to determine which ones best account for the signs, symptoms, physical examination findings, and laboratory test

results. Hypothesis generation is especially challenging for medical students because the organization of knowledge in medical school curricula is disease-centered. Furthermore, the clinical reference systems that are regularly used by medical students (such as Harrison's Online, UpToDate, and eMedicine) are mostly organized by disease. To address this issue, *Abduction*, a hypothesis generation tool, was developed for this study. Sixteen medical students were asked to solve two patient cases in two different conditions: A (support of clinical reference systems chosen by the participant and *Abduction*) and B (support of clinical reference systems chosen by the participant). In condition B, participants were able to generate the correct diagnosis in three out of 16 occasions and were able to confirm it once. In condition A, participants were able to generate the correct diagnosis in all 16 occasions and were able to confirm it in 13 occasions. These results show that there are significant advantages in providing specific support to hypothesis generation. *Abduction* proved to be five times more effective at supporting hypothesis generation than the clinical reference systems that are regularly used by medical students. These results also indicate that the effects of specific support to hypothesis generation carry over to the hypothesis evaluation stage. Consequently, when analyzing the effects of computer-based support to diagnostic reasoning, the distinction between hypothesis generation and hypothesis evaluation does seem to provide us with more precise methods to evaluate the effects of those support systems.

References

Allen, V. G., Arocha, J. F., & Patel, V. L. (1998). Evaluating evidence against diagnostic hypotheses in clinical decision making by students, residents and physicians. *International Journal of Medical Informatics, 51*(2-3), 91-105.

Ark, T. K., Brooks, L. R., & Eva, K. W. (2006). Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Academic Medicine, 81*(4), 405-409.

Arocha, J. F., Patel, V. L., & Patel, Y. C. (1993). Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. *Medical Decision Making, 13*(3), 198-211.

Barrows, H. S., & Bennett, K. (1972). The diagnostic (problem solving) skill of the neurologist. Experimental studies and their implications for neurological training. *Archives of Neurology, 26*(3), 273-277.

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning. An approach to medical education.* New York: Springer.

Bates, M. J. (1977). Factors affecting subject catalog search success. *Journal of the American Society for Information Science, 28*, 161-169.

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Barloon, T., Ell, S. R., Lu, C. H., et al. (1986). Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. *Investigative Radiology, 21*(7), 532-539.

Berner, E. S., McGowan, J. J., Hardin, J. M., Spooner, S. A., Raszka, W. V., Jr., & Berkow, R. L. (2002). A model for assessing information retrieval and application skills of medical students. *Academic Medicine, 77*(6), 547-551.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16*, 153-184.

Boshuizen, H. P. A., Schmidt, H. G., Custers, J. F. M., & Van de Wiel, M. W. (1995). Knowledge development and reestructuring in the domain of medicine: The role of theory and practice. *Learning and instruction, 5*, 269-289.

Bradford, P. (Ed.). (1996). *Information architects*. Zurich, Switzerland: Graphis Press.

Brooks, L. R., LeBlanc, V. R., & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychological Science, 11*(2), 112-117.

Buscher, G., Baumeister, J., Puppe, F., & Seipel, D. (2005, October). *User-centered consultation by a society of agents.* Paper presented at the 3rd international conference on Knowledge capture, Banff, Alberta, Canada.

Charlin, B., Tardif, J., & Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning. *Academic Medicine, 75*(2), 182-190.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37*(8), 695-703.

Cogdill, K. W., & Moore, M. E. (1997). First-year medical students' information needs and resource selection: Responses to a clinical scenario. *Bulleting of the Medical Library Association, 85*(1), 51-54.

Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: Are they being met? *Annals of Internal Medicine, 103*(4), 596-599.

de Bliek, R., Friedman, C. P., Wildemuth, B. M., Martz, J. M., Twarog, R. G., & File, D. (1994). Information retrieved from a database and the augmentation of personal knowledge. *Journal of the American Medical Association, 1*(4), 328-338.

de Groot, A. D. (1978). *Thought and choice in chess*. The Hague: Mouton.

Dudley, H. A. (1970). The clinical task. *Lancet, 2*(7687), 1352-1354.

Dudley, H. A. (1971). Clinical method. *Lancet, 1*(7688), 35-37.

Dunbar, K. (1998). Problem solving. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 289-298). Malden, MA: Blackwell.

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *British Medical Journal, 324*(7339), 729-732.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Ely, J. W., Osheroff, J. A., Ebell, M. H., Bergus, G. R., Levy, B. T., Chambliss, M. L., et al. (1999). Analysis of questions asked by family doctors regarding patient care. *British Medical Journal, 319*(7206), 358-361.

Ely, J. W., Osheroff, J. A., Gorman, P. N., Ebell, M. H., Chambliss, M. L., Pifer, E. A., et al. (2000). A taxonomy of generic clinical questions: Classification study. *British Medical Journal, 321*(7258), 429-432.

Eva, K. W., Brooks, L. R., & Norman, G. R. (2002). Forward reasoning as a hallmark of expertise in medicine: Logical, psychological, phenomenological inconsistencies. In S. P. Shohov (Ed.), *Advances in psychology research* (Vol. 8, pp. 41-69). New York: Nova Science Publishers.

Fenichel, C. H. (1979). *Online information retrieval: Identification of measures that discriminate among users with different levels and type of experience*. Doctoral dissertation, Drexel University, Philadelphia, PA. Retrieved on June 3, 2005, from http://dspace.library.drexel.edu/handle/1860/40.

Florance, V. (1996). Clinical extracts of biomedical literature for patient-centered problem solving. *Bulletin of the Medical Library Association, 84*(3), 375-385.

Frederiksen, C. H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7*, 371-458.

Friedman, C. P., Elstein, A. S., Wolf, F. M., Murphy, G. C., Franz, T. M., et al. (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation: A multisite study of 2 systems. *Journal of the American Medical Association, 282*(19), 1851-1856.

Gillam, S., & Bagade, A. (2006). Undergraduate public health education in UK medical schools – struggling to deliver. *Medical Education, 40*(5), 430-436.

Gobet, F. (1997). A pattern-recognition theory of search in expert problem solving. *Thinking and Reasoning, 3*(4), 291-313.

Gobet, F., & Simon, H.A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31*, 1–40.

Gorman, P. N. (1995). Information needs of physicians. *Journal of the American Society for Information Science, 46*(10), 729-736.

Gorman, P. N., & Helfand, M. (1995). Information seeking in primary care: How physicians choose which clinical questions to pursue and which to leave unanswered. *Medical Decision Making, 15*(2), 188-189.

Groen, G. J., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 287-310). Hillsdale, NJ: Erlbaum.

Gruppen, L. D., Palchik, N. S., Wolf, F. M., Laing, T. J., Oh, M. S., & Davis, W. K. (1993). Medical student use of history and physical information in diagnostic reasoning. *Arthritis Care and Research, 6*(2), 64-70.

Haynes, R. B., McKibbon, K. A., Walker, C. J., Ryan, N., Fitzgerald, D., & Ramsden, M. F. (1990). Online access to MEDLINE in clinical settings. A study of use and usefulness. *Annals of Internal Medicine, 112*(1), 78-84.

Hersh, W. R., & Hickam, D. H. (1998). How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *Journal of the American Medical Association, 280*(15), 1347-1352.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science, 44*(3), 161-174.

Hunt, E. (1989). Cognitive science: Definition, status, and questions. *Annual Review of Psychology, 40*, 603-629.

Hunt, D. L., Haynes, R. B., Hanna, S. E., & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *Journal of the American Medical Association, 290*(15), 1339-1346.

Johnson, P., Duran, A., Hassebrock, F., Moller, J., Prietulla, M., Feltovich, P., et al. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science, 5*, 235-283.

Joseph, G. M., & Patel, V. L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making, 10*(1), 31-46.

Kassirer, J. P., & Kopelman, R. I. (1991). *Learning clinical reasoning*. Baltimore, MA: Williams & Wilkins.

Kempainen, R. R., Migeon, M. B., & Wolf, F. M. (2003). Understanding our mistakes: A primer on errors in clinical reasoning. *Medical Teacher, 25*(2), 177-181.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine, 13*(2), 110-116.

Kushniruk, A. W., Patel, V. L., & Marley, A. A. (1998). Small worlds and medical expertise: Implications for medical cognition and knowledge engineering. *International Journal of Medical Informatics, 49*(3), 255-271.

Lajoie, S. P. (Ed.). (2000). *Computers as cognitive tools* (Vol. 2). Hillsdale, NJ: Erlbaum.

Lajoie, S. P., & Derry, S. J. (Eds.). (1993). *Computers as cognitive tools*. Hillsdale, NJ: Erlbaum.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*(4450), 1335-1342.

Lazoff, M. (2001, May). Online clinical References. *Medical Computing Today*. Retrieved May 22, 2005, from http://www.medicalcomputingtoday.com/0nvclinref.html

LeBlanc, V. R., Brooks, L. R., & Norman, G. R. (2002). Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Academic Medicine, 77*(10), S67-S69.

Leblanc, V. R., Norman, G. R., & Brooks, L. R. (2001). Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Academic Medicine, 76*(10), S18-20.

Lemieux, M., & Bordage, G. (1992). Propositional versus strucural semantic analyses of medical diagnostic thinking. *Cognitive Science, 16*, 185-204.

Levin, J. R., & Serlin, R. C. (2000). Changing students' perspectives of McNemar's test of change. *Journal of Statistics Education, 8*(2). Retrieved on July 20, 2006 from http://www.amstat.org/publications/jse/secure/v8n2/Levin.cfm

Lillehaug, S. I., & Lajoie, S. P. (1998). AI in medical education: Another grand challenge for medical informatics. *Artificial Intelligence in Medicine, 12*(3), 197-225.

Mandin, H., & DesCoteaux, J. G. (1998, October). Knowledge structure and effective clinical problem-solving. *Thought Exchange*. Retrieved on December 2, 2004, from http://www.med.ucalgary.ca/education/ume/thought-exchange/articles/article12.htm

Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge: Cambridge University Press.

Maudsley, G. (1999). Do we all mean the same thing by "problem-based learning"? A review of the concepts and a formulation of the ground rules. *Academic Medicine, 74*(2), 178–185.

McLean, M., Van Wyk, J. M., Peters-Futre, E. M., & Higgins-Opitz, S. B. (2006). The small group in problem-based learning: More than a cognitive 'learning' experience for first-year medical students in a diverse population. *Medical Education, 28*(4), e94-103.

Nakamura, C., & Lajoie, S. P. (2003). Using information architecture to support problem-based learning environments. In D. Lassner & C. McNaught (Eds.), *Proceedings of ED-MEDIA 2003: World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 3001-3004). Honolulu, HI.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*, 151-166.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1991). Cognitive artifacts. In J. M. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface* (pp. 17-38). New York: Cambridge University Press.

Norman, G. R., Brooks, L. R., Colle, C. L., & Hatala, R. M. (1999). The benefit of diagnostic hypotheses in clinical reasoning: Experimental study of an instructional intervention for forward and backward reasoning. *Cognition and Instruction, 17*(4), 433-448.

Norman, G. R., & Schmidt, H. G. (1992). The psychological basis of problem-based learning: A review of the evidence. *Academic Medicine, 67*(9), 557-565.

Norman, G. R., & Schmidt, H. G. (2000). Effectiveness of problem-based learning curricula: Theory, practice and paper darts. *Medical Education, 34*(9), 721-728.

Osheroff, J. A., Forsythe, D. E., Buchanan, B. G., Bankowitz, R. A., Blumenfeld, B. H., & Miller, R. A. (1991). Physicians' information needs: Analysis of questions posed during clinical teaching. *Annals of Internal Medicine, 114*(7), 576-581.

Patel, V. L., Arocha, J. A., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. Holyoak (Ed.), *Cambridge handbook of thinking and reasoning* (pp. 2298-2538). Cambridge: Cambridge University Press.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*, 91-116.

Patel, V. L., & Groen, G. J. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 287-310). Hillsdale, NJ: Erlbaum.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93-125). Cambridge: Cambridge University Press.

Ridderikhoff, J. (1993). Problem-solving in general practice. *Theoretical Medicine, 14*(4), 343-363.

Rikers, R. M., Schmidt, H. G., Boshuizen, H. P., Linssen, G. C., Wesseling, G., & Paas, F. G. (2002). The robustness of medical expertise: Clinical case processing by medical experts and subexperts. *American Journal of Psychology, 115*(4), 609-629.

Schmidt, H. G. (1993). Of problem-based learning: Some explanatory notes. *Medical Education, 27*, 422–432.

Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On the origin of intermediate effects in clinical case recall. *Memory and Cognition, 21*(3), 338-351.

Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On acquiring expertise in medicine. *Educational Psychology Review, 5*(3), 205-221.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine, 65*(10), 611-621.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Sisson, J. C., Donnelly, M. B., Hess, G. E., & Woolliscroft, J. O. (1991). The characteristics of early diagnostic hypotheses generated by physicians (experts) and students (novices) at one medical school. *Academic Medicine, 66*(10), 607-612.

Stillman, P. L., & Hanshaw, J. B. (1989). Education of medical students: Present innovations, future issues. *Mayo Clinic Proceedings, 64*(9), 1175-1179.

Van Berkel, H. J., & Dolmans, D. H. (2006). The influence of tutoring competencies on problems, group functioning and student achievement in problem-based learning. *Medical Education, 40*(8), 730-736.

Vernon, D. T., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine, 68*(7), 550-563.

Wanger, J., McDonald, D., & Berger, M. C. (1980). *Evaluation of the online search process: A final report*. Santa Monica, CA: Cuadra Associates (PB81-132565).

White, C. B. (2006, June 10). Smoothing out transitions: How pedagogy influences medical students' achievement of self-regulated learning goals. *Advances in Health Science Education.* Retrieved August 10, 2006 from http://www.springerlink.com/content/e5r1208220142312/

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactics formulation. *Journal of the American Society for Information Science and Technology, 55*(3), 246-258.

Wildemuth, B. M., de Bliek, R., Friedman, C. P., & File, D. D. (1995). Medical students' personal knowledge, searching proficiency, and database use in problem solving. *Journal of the American Society for Information Science, 46*(8), 590-607.

Wildemuth, B. M., Friedman, C. P., Keyes, J., & Downs, S. M. (2000). A longitudinal study of database-assisted problem solving. *Information Processing and Management, 36*, 445-459.

Wurman, R. S. (1989). *Information anxiety*. New York: Doubleday.

Appendix A: Certificate of Ethical Acceptability of Research Involving Humans

McGill

Faculty of Education – Ethics Review Board
McGill University
Faculty of Education
3700 McTavish; Room 230
Montreal H3A 1Y2

Tel: (514) 398-7039
Fax: (514) 398-1527
Ethics website: www.mcgill.ca/rgo/ethics/human

Faculty of Education – Review Ethics Board
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 603-1005

Project Title : ***Database-Assisted Diagnostic Reasoning***

Applicant's Name: Carlos Nakamura **Department**: ECP

Status: Ph.D student **Supervisor's Name: Susanne Laoie**

Granting Agency and Title (if applicable):

Type of Review: Expedited ✓___ Full _____

This project was reviewed by: Stapley/McAlpine

Approved by

R. Bracewell Nov. 8, 2005

Signature/Date
Robert Bracewell, Ph.D.
Chair, Education Ethics Review Board

Approval Period: Nov 8/05 **to** Nov. 8/06

All research involving human subjects requires review on an annual basis. An Annual Report/Request for Renewal form should be submitted at least one month before the above expiry date. If a project has been completed or terminated for any reason before the expiry date, a Final Report form must be submitted. Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received. This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement on the Ethical Conduct for Research Involving Human Subjects.

Appendix B: Informed Consent to Participate in Research

INFORMED CONSENT FORM TO PARTICIPATE IN RESEARCH

Title: **Database-Assisted Diagnostic Reasoning**

Investigator: **Carlos Nakamura** Supervisor: **Susanne P. Lajoie**

Dear Participant,

This consent form is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more details about something mentioned here, or information not included here, please ask or contact me by e-mail (cnakam@po-box.mcgill.ca). Take the time to read this carefully and to understand any accompanying information. You will receive a copy of this form.

The primary purpose of this research is to study the cognitive processes used by medical students as they attempt to diagnose medical cases. I would like to invite you to participate in my study. If you agree, your participation will contribute to our understanding of the cognitive processes involved in medical problem solving as applied to diagnostic reasoning. The results of the study will be used for my doctoral dissertation.

You will be asked to solve two patient cases presented in writing. You will be requested to follow a pre-established annotation system as you work on each case. The annotation system consists of tables where you state your hypotheses, confidence levels, supporting evidence, rationale, and questions regarding the patient case. I might also ask you a few questions as you work on the cases. The two cases will be solved in the same session. The duration of the session will mostly depend on how long it will take you to produce a diagnosis for the two cases. All data from this study will be kept confidential and will be used strictly for research purposes. Participants' performance will in no way affect their academic standing or performance rankings. Each participant will be assigned a code number and referred to by that number in all discussion of the results. Participation entails no costs or foreseeable dangers to the participants. You are free to refuse to participate in the study without fear of any negative consequences and are free to withdraw your consent and discontinue your involvement in the study at any time, without any penalty.

I have read this consent form and understand what my participation involves. My questions have been answered to my satisfaction. I consent to participate in the study.

Participant's name (Please print)	Signature and date
Carlos Nakamura, PhD Candidate Dept of Educational and Counselling Psychology McGill University 3700 McTavish Street, B199 Montreal, QC, H3A 1Y2, CANADA	Signature and date

The Faculty of Education at McGill University has a Research Ethics Board which regulates all research projects involving human subjects conducted by staff and students of the Faculty of Education. Complaints or problems concerning research projects may and should be reported if they arise. Telephone: (514) 398-7039.

Appendix C: Presentation and Solution of Patient Cases Used in Study

Case Presentation: Kennedy's disease (X-Linked Spinobulbar Muscular Atrophy)

HISTORY

Patient #13 is a 65 year old white male who presented to the Department of Neurology for evaluation of progressive muscle twitching, cramping, and weakness.

His symptoms first began approximately 22 years ago when he noted muscular cramping and tightness in the legs, especially in the calf muscles. He was seen by a general practitioner who subsequently referred him to a neurologist. He was given a diagnosis of "muscular disease", otherwise not well specified, and the patient was advised to follow-up with his primary physician. The patient did not return for follow-up, but stated he continued to have progression of his symptoms.

Two years later, he noted "twitching" in his muscles, initially in his legs, and subsequently involving his shoulders and arms. The cramping and twitching was followed by muscle weakness beginning in the legs and progressing to involve the upper extremities. The weakness was greater on the left side. As his symptoms progressed, he began to have difficulty standing from a chair and would "trip" when walking. Subsequently, he developed problems lifting and holding objects, and stated that presently he was unable to lift more than 20 pounds. The degree of weakness varied during the day, did not seem to follow a specific temporal pattern, and appeared to worsen with cold weather.

Over the course of the years, he continued to have muscle cramping, but reported the cramping had improved recently. The twitching and cramping improved when taking potassium tablets, which he had been treating himself with for the past 5 to 6 years.

He stated he had occasional pain from a fall several years ago that "shattered" both calcanei bones in his feet. Surgical history included a left leg fracture with pinning approximately 23 years prior. He had no history of hypertension, diabetes, pulmonary, renal, coronary artery, or cardiovascular disease. The patient denied constipation, bowel or bladder incontinence, and numbness or tingling. He also denied swallowing or speech difficulties, blurred vision, double vision, or emotional incontinence. However, he stated that recently he had noted problems in remembering people's names.

He reported no known drug allergies. Family history was positive for hypertension and coronary artery disease in his father, negative for neurological disorders. He was married with four children, and gave a negative history of tobacco, alcohol, or drug abuse. His medications included ativan, doxepin, zantac, and potassium p.r.n.

PHYSICAL EXAMINATION

Vital signs: Temp-97.9, BP-134/90, P-72, RR-18
General: Well-nourished, well-developed, white male in no acute distress.
HEENT: Normocephalic, atraumatic. No oropharyngeal lesions noted. Neck supple, no JVD, bruits, or lymphadenopathy. Thyroid gland was not appreciable to palpation.
Chest: Clear to auscultation and percussion bilaterally. Mild gynecomastia, right more than left.
CV: Regular rate and rhythm, normal S1 and S2 without murmurs, rubs, or gallops.
Abdomen: Soft, non-tender, and non-distended with positive bowel sounds.
Extremities: No cyanosis, clubbing or edema. Mild atrophy in the scapular and hand muscles. Fasciculations were noted during the examination in the upper extremities.

NEUROLOGICAL EXAMINATION

Mental status: The patient was awake, alert, and oriented to person, place, and time. Mini-mental exam was 29/30. Speech was fluent with normal repetition and comprehension. Normal labial with slight impairment of lingual and guttural components. A mild nasal component in his speech was noted.

Cranial nerve function:

II: Pupils were 3 mm and bilaterally reactive to light and accommodation. Visual acuity within normal limits with no evidence of visual field deficits. Optic discs clear with sharp margins.
III,IV,VI: Extraocular movements intact. No nystagmus.
V: Positive corneal reflex bilaterally. Normal facial sensation to light touch, pinprick, and temperature in the V1-3 distribution. Temporalis and masseter 5/5 bilaterally.
VII: Diminished labial strength, otherwise no focal deficits. Fasciculations were noted around the mouth.
VIII: Intact to finger rubs bilaterally.
IX,X: Symmetrically elevating palate with positive gag reflex.
XI: Sternocleidomastoid and trapezius muscles 5/5 bilaterally.
XII: Tongue moderately atrophic with no apparent fibrillations.

Motor examination:

Mild atrophy in the proximal and distal upper and lower extremities. Normal tone.

Neck flexion/extension	5-/5
Deltoids	4-/5
Biceps	4-/5
Triceps	5-/5
Wrist extensors	4-/5
Wrist flexors	5-/5
Finger extensors	3+/5
Finger flexors	4+/5
Interossei	4-/5
Iliopsoas	4-/5
Quadriceps	5-/5
Hamstrings	5-/5
Ankle dorsiflexion	4-/5
Plantar flexion	5-/5

Reflexes:

- Absent in biceps, triceps, brachioradialis, patellar, and ankle.
- Negative Hoffman's sign, negative jaw jerk.

Sensory Examination:

- Slightly diminished pinprick along the sole of the right foot.
- Diminished light touch bilaterally at the sole of both feet, which the patient felt occurred secondary to his fall and his calcaneal fracture.
- Proprioception was intact.
- Vibration was slightly diminished distally in the right lower extremity greater than the left lower extremity.
- Romberg was negative.

Cerebellar testing: No dysmetria on finger-to-nose and heel-to-shin bilaterally. Normal rapid alternating movements.

Gait: Intact with normal arm swing with normal toe and tandem walking. Mild difficulty in heel-walking.

Case Presentation: Friedreich's Ataxia

HISTORY

Patient # 25 is a 32 year old right handed male who presented with a progressive gait disorder.

Three years ago, the patient first noticed difficulty walking in a straight line but did not think much of the problem. Over the next two years, however, this difficulty progressed to the point where he could no longer deny it. By this time, his gait was unsteady and "drunken", causing him to trip frequently especially when turning. He noted marked difficulty negotiating stairs, especially when walking down. At night, in the dark, his unsteadiness worsened. He was diagnosed with a peripheral neuropathy by another neurologist and sent to our center for further investigation and treatment.

The patient denied any muscle wasting, weakness, fasciculations, muscle stiffness, tingling, numbness, visual disturbance, dysarthria, dysphagia, diplopia, incontinence, or memory disturbance. He is able to walk up to three miles a day, but his legs fatigue easily.

Past Medical History:	Unremarkable
Past Surgical History:	None
Allergies:	NKDA
Medications:	Tylenol, as needed
Social History:	Unmarried tire salesman with no history of alcohol, tobacco, or drug abuse. He denied any HIV risk.
Family History:	No neurological disorders; specifically, no gait abnormalities.

PHYSICAL EXAMINATION

B.P.:	132/60; pulse 70; temperature 97.6F; respiration 18.
General:	Well developed, well nourished male in no acute distress
HEENT:	Normocephalic, atraumatic; sclerae anicteric; conjunctivae pink; oropharynx clear, moist without lesions; neck supple without lymphadenopathy, thyromegaly, bruits
Cardiovascular:	Regular rate and rhythym without rubs, gallops, or murmurs; PMI not displaced
Chest:	Clear to auscultation and percussion bilaterally
Abdomen:	Soft, nontender, without visceromegaly.
Skin:	No significant hyper- or hypo- pigmented lesions.
Extremities:	No cyanosis, clubbing, or edema.

NEUROLOGICAL EXAMINATION

Mental status: The patient was alert and fully oriented. MMSE was 30/30. Attention was intact, and speech was fluent without paraphasic errors. Comprehension, naming, repetition, reading, and writing were all intact. Short-term memory was intact, as well as constructional ability.

Cranial nerve function:

II:	visual acuity 20/20 OU; visual fields full to confrontation; pupils 3mm and reactive to light and accomodation;
III, IV, VI:	fundoscopic exam WNL extraocular movements full without nystagmus or ptosis; no ocular dysmetria
V:	intact sensation in all three divisions bilaterally; intact masseter and temporalis strength
VII:	smile symmetrical
VIII:	hearing intact to finger rub bilaterally; Weber non-lateralizing; air>bone conduction
IX, X:	palate elevates in midline; gag intact bilaterally
XI:	SCM and trapezius strength intact bilaterally
XII:	Tongue midline without atrophy or fibrillations

Motor examination: Tone was normal. Muscle bulk was normal. There was no cogwheel rigidity or tremor. Strength in the neck flexors and extensors was 5/5.

Strength in the upper extremities was:

	Right	**Left**
Deltoids	5/5	5/5
Biceps	5/5	5/5
Triceps	5/5	5/5
Wrist flexors	5/5	5/5
Wrist extensors	5/5	5/5
Finger flexors	4+/5	4+/5
Finger extensors	4+/5	4+/5
Hand intrinsics	4+/5	4+/5

Strength in gluteus maximus was 4-/5 bilaterally and strength in the hip abductors and adductors was 4+/5 bilaterally.

Strength testing of the lower extremities was:

	Right	**Left**
Iliopsoas	4+/5	4+/5
Knee flexors	5-/5	5-/5
Knee extensors	5/5	5/5
Ankle extensors	5-/5	5-/5
Ankle flexors	5/5	5/5

Sensory Examination: Decreased pinprick, temperature, and light touch in a symmetrical stocking distribution up to the mid thigh on the leg and a glove distribution up to the mid arm; absent vibration and proprioception in the toes bilaterally; decreased vibration and proprioception in all four limbs; Rhomberg was positive.

Reflexes:

	Right	**Left**
Biceps	1	1
Triceps	1	1
Brachioradialis	1	1
Patellar	0	0
Ankle	0	0

Babinski's were present bilaterally. There was no Hoffman's sign. Palmomental and snout signs were present.

Cerebeullum: Mild dysmetria on finger-nose-finger; moderate dysmetria on heel-shin test; slight truncal titubation; rebound test positive; mild dysdiadokokinesia in the fingers, more pronounced on toe tapping.

Gait: Unsteady, slow, wide-based, with irregular stride length; arm swing normal; the patient attempted to turn on a pivot but was very unsteady; able to walk on toes and heels in an unsteady manner; unable to tandem

Case Solution: Kennedy's disease (X-Linked Spinobulbar Muscular Atrophy)

Italo Linfante, M.D.
Resident, Department of Neurology

The patient's symptomatology started 22 years ago with tightness and cramps in his lower extremities. Subsequently, he developed fasciculations and weakness in proximal as well as distal muscles in both upper and lower extremities. Although he was initially thought to have a primary disease of muscle, this diagnosis was unlikely because of the history of fasciculations and the pattern of muscle weakness; namely, the involvement of both proximal (arising from the chair, lifting weights) and distal muscle groups (tripping on his feet, holding objects tightly).

Most adult onset generalized myopathies result in proximal muscle weakness (i.e. polymyositis, dermatomyositis, limb girdle, late onset Becker's). A minority present with a more prevalent distal involvement (i.e. myotonic dystrophy, inclusion body myositis). Other myopathies affect selected muscle groups (scapuloperoneal, fascioscapulohumeral, or oculopharyngeal musculature).

In our patient, the involvement of proximal and distal muscle groups in both upper and lower extremities does not follow the pattern of any of the above mentioned myopathies. Furthermore, fasciculations are not a significant feature of primary muscle disease.

The presence of weakness in proximal and distal muscle groups as well as fasciculations, suggested the diagnosis of motor neuropathy or lower motor neuron disease. The physical examination confirmed the pattern of muscular weakness suggested by the history and indicated involvement of the bulbar musculature with tongue atrophy and perioral weakness. In addition, reflexes were absent. EMG confirmed the involvement of the lower motor neuron with widespread chronic denervation in the extremities, thoracic paraspinal muscles, and tongue. The normal motor conduction velocities in both arms and legs made motor neuropathy less likely. The muscle biopsy confirmed a long standing denervation-reinnervation process as evidenced by type I and type II fiber grouping, and ruled out the presence of an inflammatory process.

The clinical and EMG findings of lower motor neuron involvement suggested a diagnosis of spinal muscular atrophy. The spinal muscular atrophies of adult onset are a clinically well-defined group due to diverse genetic defects. They are characterized by late onset (3rd-5th decade), slowly progressive bulbo-spinal muscle weakness with atrophy, fasciculations, and areflexia. Proximal muscles of the shoulders and pelvic girdles are affected earlier and to a greater extent then distal muscles. Intrinsic hand muscles are affected later in the course of the disease. Bulbar involvement is manifested initially as weakness and fasciculations of the oro-mandibular musculature, and subsequently by atrophy and fibrillations of the tongue. Reflexes are usually absent. The clinical syndrome is transmitted by three distinct modes of inheritance: autosomal-

dominant, autosomal-recessive, and X-linked recessive. Patients with these three distinct modes of inheritance present with a similar phenotype.

Recent studies of the autosomal recessive forms of SMA have defined a single locus on chromosome 5q11.2-13.3 in acute infantile, late infantile, juvenile, and adult onset types.[1] The genes for two proteins in this region, Survival Motor Neuron (SMN) and Neuronal Apoptosis Inhibitory Protein (NAIP), have been identified as possibly involved in SMA. Deletions in SMN have been described in greater than 98% of SMA cases.[2] However it is still unclear how alterations in SMN give rise to the clinical syndrome of SMA, and what the role of NAIP is in this disease process.

Patients with the X-linked recessive form (Kennedy's disease) present with bulbar and spinal muscular involvement with absent reflexes, as well as gynecomastia, testicular atrophy and non-insulin dependent diabetes mellitus.

Our patient presented with evidence of both bulbar and spinal muscular atrophy as well as a mild degree of gynecomastia, but no testicular atrophy. The combination of symptoms and signs, including the prominent facial fasciculations, suggested the possible diagnosis of Kennedy's disease. Genetic testing was performed, and PCR analysis revealed 43 CAG repeats which is consistent with the diagnosis of Kennedy's disease.

A Review of Kennedy's disease
(X-linked recessive bulbo-spinal muscular atrophy)

In 1968, William R. Kennedy and co-workers described a distinctive "slowly progressive spinal and bulbar muscular atrophy of late onset" in 11 affected males from 2 families.[3] According to his original work, the disease is a separate entity from other lower motor neuron disorders because of "the late age of onset; consistent involvement of bulbar, proximal, and distal muscle groups; sex-linked recessive inheritance; and normal life expectancy". Since the original description, the disease has been reported in several kindred, in particular in the U.K., Harding et al.[4]

The disease is thought to be rare, but the true incidence is not known, and probably underestimated, since many patients are undiagnosed or misdiagnosed. The widespread availability of genetic testing will provide a more accurate estimate in the next several years. Onset of the disease is variable from 15 to 59 years of age, although most commonly these patients seek medical attention between the fourth and fifth decade of life.

Most authors agree that the clinical features include muscle weakness, atrophy, and fasciculations. The proximal muscle groups are affected at an earlier stage, subsequently followed by atrophy and weakness of the intrinsic muscles of the hands and peroneal muscles. Weakness and fasciculations of the oro-mandibular musculature, as well as perioral fasciculations, are present early and are subsequently followed by facial weakness, atrophy, and fibrillations of the tongue. Some authors also describe a *nasal* component to the speech, most likely due to facial weakness. Deep tendon reflexes are

usually absent. Sensation may be abnormal. Gynecomastia is present in many of the cases, but can be quite variable, as are testicular atrophy and reduced fertility. EMG and muscle biopsy typically show chronic neurogenic atrophy with reinnervation. Post mortem examination of the spinal cord reveals marked loss and/or atrophy of anterior horn cells.

Genetics

In 1991, La Spada et al[5] mapped the genetic mutation that causes Kennedy's disease to the first exon of the androgen receptor (AR) gene on the proximal long arm of the X chromosome. PCR analysis revealed amplification of CAG triplet nucleotide repeats. DNA sequence analysis showed that the average CAG repeats was 21 ± 2 in 75 normal controls versus a range of 40 to 52 in the 24 patients with Kennedy's disease. There was no overlap between the 2 groups. The enlarged band segregated with disease in 15 Kennedy's disease families, with no recombination in 61 meioses. The maximum odds ratio (lod score) of this mutation being the cause of the disease was determined to be 13.2 at 0 centimorgans, which is highly significant. Since the original report, the mutation has been confirmed by many groups.[6-7] The CAG repeats encode for glutamine residues in the amino-terminal domain of the AR receptor. It has been reported that the size of the amplified CAG repeats correlates with age of onset.[8-9] The larger the number of CAG repeats, the earlier the age of onset. This phenomenon is present in other neurological disorders with trinucleotide repeats. Expansion of trinucleotide repeats has been found to be present in several inherited neurological disorders. The trinucleotide repeats that have been discovered are:

1. CAG: Expansion of CAG repeats has been found in spinocerebellar atrophy type 1 (SCA1) on chromosome 6p22-23; Machado-Joseph/ spinocerebellar atrophy type 3 (SCA3) on chromosome 14q24-32; Dentato Rubro Pallido Luysian Atrophy (DERPLA) on chromosome 12p; Huntington's disease (HD) on chromosome 4p16.3; and Kennedy's Disease (KD) on chromosome Xq21.3. Recent studies also suggest that spinocerebellar atrophy type 2 (SCA2) on chromosome 12q is possibly associated with an expansion of trinucleotide repeats.
2. CTG: Expansion of CTG repeats has been found in myotonic dystrophy on chromosome 19q16.3
3. CGG: Expansion of CGG repeats has been found in fragile X syndrome, on chromosome Xq27.3
4. GAA: Expansion of GAA repeats has been found in Friedreich ataxia in intron 1 of the gene X25 on chromosome 9 encoding for frataxin, a protein of 210 amino acids.

While GAA repeats are not transcribed into mRNA, CAG repeats in the coding region are transcribed and translated into a peptide of polyglutamine. The size of the polyGlu tract is determined by the number of repeats. The polyGlu can bind to DNA, mRNA, and cellular proteins. Therefore, it has been hypothesized that the binding of

polyGlu tract with nucleic acids or proteins could derange cellular functions ultimately inducing cell death. For example, in HD the polyGlu tract binds to a protein called Huntingtin Associated Protein (HAP1), which is highly expressed in brain. HAP1 may be associated with microtubule-mediated transport.[10] The authors speculate that the binding of polyGlu with HAP1 could result in a toxic gain of function leading to apoptosis.[10] More recently, Burke et al. reported that the polyGlu tract could also bind to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in brains of HD and DERPLA patients.[11] GAPDH, besides being a fundamental step in the glycolytic pathway, can bind to uracyl DNA glycosylase and tubulin.

The mutation present in Kennedy's disease involves the AR gene. The AR is a member of a superfamily of DNA binding proteins that includes steroid, Vitamin D, and retinoic acid receptors. The AR is a nuclear transcription factor that mediates the steroid dependent activation of several genes necessary for the biologic action of steroids. It has been shown that CAG encodes for the glutamine residues of the NH_2 terminal domain of the AR protein.[6] The NH_2 terminal is the same site where another protein, the receptor accessory factor (RAF), binds to enhance the binding of the AR to the DNA.[12] Therefore, it has been hypothesized that the polyGlu tract interferes with the complex AR, RAF, and DNA impairing the efficiency of the hormonal mediated AR action.

The possibility that an alteration of the AR gene could be associated with motor neuron disease has led many investigators to study the potential relationship of androgens with motor neuron injury. However, the exact mechanism of disease is unknown. Many reports have described mutations of the AR gene.[14] Such mutations cause a variety of defects of virilization, but none result in motor neuron disease. Moreover, the deletion of the entire AR gene (present in some of the patients affected by androgen insensitivity syndrome), results in severe virilization abnormalities, but does not give rise to motor neuron disease.[15] In defects of virilization, the decreased binding properties of AR do not always correlate with the severity of the disease.[14] However, in Kennedy's disease, the decreased binding affinity of the AR not only correlates with the mild degree of gynecomastia and testicular atrophy, but also with the size of the CAG repeats.[13]

The low binding affinity of the AR in Kennedy's disease does not correlate with the degree of weakness.[13] Thus, it seems that there are two independent components in Kennedy's disease: gynecomastia and testicular atrophy, which are mild and androgen dependent; and lower motor neuron disease, which is the main feature of the disease and is androgen independent. The major question still exists as to how the CAG repeats in the AR gene lead to motor neuron injury.

References

1. Brzustovicz, L.M., Lehner, T., Castilla, L.H. et al. Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3 Nature, 344:540-541, 1990.

2. Lefebvre, S., Burglen L., Rebuller, S. et al. Identification and charecterization of spinal muscular atrophy-determining gene. Cell, 80:155-165, 1995.

3. Kennedy, V.T.R., Alter, M., and Sung, J.H. Progressive proximal spinal and bulbar muscular atrophy of late onset: a sex-linked recessive trait. Neurology, 18:671-680, 1968.

4. Harding, A.E., Thomas, P.K., Baraitser, M. et al. X-linked recessive bulbospinal neuronopathy: a report of ten cases. J. Neurol. Neurosurg. Psychiatry, 45:1012-1019, 1982.

5. La Spada, A.R., Wilson, E.M., and Lubahn, D.B. Androgen receptor gene mutations In X-linked spinal and bulbar muscular atrophy. Nature, 352:77-79, 1991.

6. Amato, A.A., Prior, T.W., Barohn, R.J. et al. Kennedy's disease: a clinicopathologic correlation with mutations in the androgen receptor gene. Neurology, 43:791-794, 1993.

7. Biancalana, V., Serville, F., Pommier, J. et al. Moderate instability of the trinucleotide repeat in spinobulbar muscular atrophy. Hum. Mol. Genet., 1:255-258, 1992.

8. La Spada, A.R., Roling, D.B., Harding, A.E. et al. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. Nature Genet, 2:301-304, 1992.

9. Igarashi, S., Tanno, Y., Onodera, 0. et al. Strong correlation between the number of CAG repeats in androgen receptor genes and the clinical onset of features of spinal and bulbar muscular atrophy. Neurology, 42:2300-2302, 1992.

10. Xiao-Jlang, L., Shi-Hua, L., Sharp, A. et al. A huntingtin-associated protein enriched in brain with implication for pathology. Nature, 378:398-402, 1995.

11. Burke, J.R., Enghild, J.J., Martin, M.E. et al. Huntingtin and DERPLA proteins selectively interacts with the enzime GAPDH. Nat. Med. 2:347-350, 1996.

12. Kupfer, S.R., Marschke, K.B., Wilson, E.M., and French, F.S. Receptor accesory factor enhances Specific DNA binding of androgen and glucocorticoid receptors J. Biol. Chem., 268:17519-1752, 1993.

13. McClean, H., Choi, W.T., Rekaris, G., Warne, G.L., and Zajac, J.D. Abnormal androgen receptor binding affinity in subjects with Kennedy's disease (spinal and bulbar muscular atrophy). J. Clin. Endocrinol. Metab., 80:508-516, 1995.

14. Griffin, J.E. Androgen resistance-the clinical and molecular spectrum. N. Engl. J. Med., 326:611-618, 1992.

15. Quigley, C.A., Friedman, K.J., Johnson, A. et al Complete deletion of the androgen receptor gene: definition of the null phenotype of the androgen insensitivity syndrome and determination of carrier status. J. Clin. Endocrinol., 74:927-933, 1992.

Case Solution: Friedreich's Ataxia

Patient #25 presented with progressive gait and limb ataxia, mild distal symmetrical sensory loss, dimished deep tendon reflexes, weakness of the gluteal muscles, and bilateral extensor plantar responses. These findings indicate cerebellar, peripheral nerve (or dorsal root ganglion), and corticospinal involvement and imply a multisystem degenerative disease. There was no evidence of autonomic dysfunction, and eye movements were left unaffected. Extrapyramidal involvement was not present arguing against one of the Multi-System Atrophies (MSAs), such as olivopontocerebellar atrophy (OPCA).

The most salient feature in this case is the patient's marked progressive ataxia. Ataxia may be due either to cerebellar or proprioceptive dysfunction, though it is rarely difficult to distinguish the two. However, when both are present, diagnostic difficulties arise. The findings in this case - gait and limb ataxia, titubation, loss of check response, and dysdiadochokinesia - point to an abnormality in the cerebellar system. This patient also had evidence of a peripheral neuropathy with involvement of the posterior columns evidenced by decreased position and vibratory sense. The degree of proprioceptive abnormality was not sufficient to explain the marked gait disturbance, however. While this patient exhibited dysfunction in both the cerebellar and proprioceptive systems, the cerebellar involvement was most impressive. The primary defect, therefore lies somewhere in the connections to, from, or within the cerebellum.

We typically separate causes of cerebellar dysfunction by the age of onset and acuity of presentation. Typical causes of acute cerebellar dysfunction in childhood include drug ingestion, infection (cerebellitis), several genetic disorders, brain tumors (cerebellar astrocytomas,etc.), postinfectious immune syndromes, migraine, and cerebellar hemorrhage/stroke. In adults, acute causes of cerebellar dysfunction are largely restricted cerebellar stroke or hemorrhage. One must also consider demyelinating diseases, such as multiple sclerosis, and posterior fossa tumors. Postinfectious cerebellitis is an uncommon, but well-recognized cause in adults as well as children. Chronic or progressive ataxias in children are typically caused by posterior fossa tumors, structural abnormalities such as basilar impression, Chiari malformations, the Dandy-Walker malformation and other cerebellar aplasias, or a hereditary form of ataxia. Hereditary ataxias in childhood include the typical cerebellar degenerative diseases, such as Machado-Joseph disease, Olivopontocerebellar atrophy, Ramsay-Hunt syndrome, Friedreich's ataxia, and ataxia-telangiectasia among a host of metabolic disorders, such as abetalipoproteinemia, Hartnup disease, juvenile GM2 gangliosidosis, juvenile sulfatide lipidosis, maple syrup urine disease, Marinesco-Sjogren syndrome, Refsum disease, pyruvate dysmetabolism, and sea-blue histiocytosis. Adreoleukodystrophy and Leber optic atrophy may also cause cerebellar dysfunction. In adults, chronic ataxias are due usually to one of the spinocerebellar atrophies or toxin exposure (e.g. alcohol).

Given this bewildering array of possible diagnoses, one might consider it impossible to reach a final diagnosis. Ataxia is a common finding, so it, in itself, cannot be used to define a specific disease entity. One must consider other clinical information

in arriving at a diagnosis. For instance, a history of intermittent ataxia and metabolic acidosis should raise the possibility of amino acid or organic acid dysmetabolism. Retinits pigmentosa, sensorineural deafness, neuropathy, and ataxia are very suggestive of Refsum disease. Cranial nerve dysfunction associated with signs of increased intracranial pressure and ataxia should prompt a search for a posterior fossa tumor. In this case, the constellation of posterior column signs, ataxia, diminished deep tendon reflexes, lower extremity weakness, and extensor plantar responses raise the possibility of Friedreich's ataxia, though the age of onset and absence of other supporting findings (kyphoscoliosis, pes cavus, deafness, etc.) makes this diagnosis less likely. Unfortunately, separating the causes of hereditary ataxia in adults is extremely difficult on purely clinical grounds. Most of the spinocerebellar degenerations have relatively late onsets (beyond the age of puberty), and most demonstrate some degree of cerebellar, pyramidal, and peripheral nerve involvement. However, most of these entities are also autosomal dominant, so a family history is often very instructive. In this case, there was no family history to suggest an autosomal dominant mode of inheritance. This may be due to a new mutation arising in this individual, or to an autosomal recessive inheritance pattern (as in Friedreich's ataxia), wherein neither parent would be affected.

The patient was found to carry the genetic defect responsible for Friedreich's ataxia, an unstable trinucleotide repeat (GAA) expansion in the first intron of the gene coding for frataxin. The triplet repeat expansion was only moderate, most likely accounting for the relatively late onset of the disease and mild symptomatology in this case. He was treated with physical therapy to assist in his gait. No other treatment options are available for this progressive disease. The patient was also told of the genetic nature of his disease.

A review of Friedreich's Ataxia
Gholam K. Motamedi, M.D.
Resident, Department of Neurology

Introduction

Nicholaus Friedreich described the "degenerative atrophy of the posterior columns of the spinal cord" that now bears his name in 1863. The initial reports were met with skepticism, but the disease has since been well accepted. Until recently, controversy continued to surround this entity, largely because of the sizeable array of degenerative ataxias and difficulty in categorizing them. The Quebec Collaborative Group provided diagnostic criteria in 1976, and Anita Harding updated these criteria in the 1980's. The original description included an age of onset before 20 years, while Harding's categorization included an age of onset before 25 years. These differences reflect the well-recognized variability in disease severity and age of onset that is more common with Friedreich's ataxia than other recessive neurological diseases. Recent findings regarding the genetic underpinnings of this disease provide a more secure diagnostic test.

Friedreich's ataxia (FRDA) is an autosomal recessive neurodegenerative disorder that represents the most common hereditary ataxia (accounting for at least 50% of cases). The estimated prevalence of FRDA is 1-2/50,000 in North American and European populations. All races are affected and males and females are equally affected. Because it is autosomal recessive, parents are usually asymptomatic (in contradistinction to other adult onset hereditary ataxias which are dominantly inherited), though the consanguity rate is high. Transmission risk from affected parents to offspring is 1 in 220.

Genetics

FRDA results from an unstable expansion of a polymorphic GAA repeat in the first intron of the X25 gene located on chromosome 9. The gene encodes a mitochondrial protein, frataxin, with unclear functions, but thought to regulate iron homeostasis. Yeast strains lacking a homologous protein develop mitochondrial dysfunction, resulting in excessive mitochondrial iron accumulation and defective oxidative metabolism. Patients with FRDA demonstrate similar abnormal mitochondrial iron deposition and pathological involvement of postmitotic tissues (though frataxin is not clearly recognized as a mitochondrial protein in humans).

All known cases of FRDA are caused by abnormalities resulting in decreased or absent transcription of the frataxin gene. Point mutations are a rare cause of FRDA, accounting for only 2% of recognized cases. Four different point mutations have been described, with all patients being heterozygous for the mutation. All result in a truncated form of frataxin. It is unknown whether homozygous point mutations have not been described because of their relative rarity, or if a homozygous mutation is lethal. The remainder of FRDA cases are due to the GAA expansion with 94% of cases homozygous for the expansion. In the vast majority of cases (98%), the GAA repeat expansion occurs in the first intron of the frataxin gene. FRDA, therefore, represents a novel genetic entity - it is the first disease recognized to result from a genetic abnormality within an uncoded region of DNA. In all other triplet repeat expansion diseases causing neurological dysfunction (myotonic dystrophy, Huntington's disease, Fragile X syndrome, Kennedy's syndrome, etc.), the defect has occurred within a coded region of DNA (an exon), generally resulting in a polyglutamine tail or other intervening stretch of abnormal amino acid sequences. This usually results in a gain of function (the altered protein may take on a new function) that results in the disease process. In FRDA, however, the abnormality results in decreased transcription, and hence translation, of frataxin. It is the decreased protein production, or loss of function, that is responsible for the disease phenotype. The exact mechanism by which this triplet repeat expansion within an intron results in decreased transcription of the gene is unknown. The sequence codes for a DNA segment with all purines on one strand and all pyrimidines on the other. This structure is thought to wind back down the major groove of the DNA helix and interfere with the transcriptional process, either by blocking the promoter region or by blocking transcriptional elongation. The expansion size is inversely correlated with transcriptional output. In patients with larger expansions, less frataxin mRNA is produced. This likely accounts for the observation that patients with larger expansion sizes show an earlier age of onset and more profound disabilities than those with smaller expansion sizes.

Clinically, this poses a slight problem since patients with relatively small expansion sizes tend to present later in life, have a slower progression, are less likely to demonstrate evidence of a cardiomyopathy, and may retain deep tendon reflexes. These patients may be considered to have another of the late onset spinocerebellar degeneartions unless FRDA is considered in the differential diagnosis.

Clinical Presentation

Friedreich originally described this syndrome in 9 members of 3 sibships with an age of onset near puberty. Ataxia and dysarthria were prominent; sensory loss and weakness were late findings in these cases. He also described nystgmus, scoliosis, foot deformity, and cardiac abnormalities in these patients. Erb later described loss of deep tendon reflexes in 1875.

Harding's diagnostic criteria include:

- Autosomal recessive inheritance
- Age of onset before 25 years

Within 5 years from onset

- Limb and trunk ataxia
- Absent tendon reflexes in the legs
- Extensor plantar responses
- Motor NCV > 40 m/s in upper limbs with small or absent SNAPs

After 5 years from onset

- Above plus dysarthria

Additional criteria, not essential for daignosis (present in 2/3)

- Scoliosis
- Pyramidal weakness of the legs
- Absent reflexes in the upper limbs
- Distal loss of joint and position sense in lower limbs
- Abnormal EKG

Other features, present in <50%

- Nystagmus
- Optic atrophy
- Deafness
- Distal weakness and wasting
- Pes cavus
- Diabetes mellitus

Symptoms usually begin between the ages of 8 and 15 years, but the range extends between 18 months and 25 years. Several reports describe families with later onset (between 20 and 30 years of age), but fulfilling all other diagnostic criteria. These families generally have smaller GAA repeat sizes and less severe disease courses. Gait ataxia is the most common first presenting symptom, although some patients fist evidence scoliosis or cardiac symptoms. Dysarthria, areflexia, pyramidal weakness of the legs, and distal loss of joint position sense are inconstant findings at presentation, but generally are present at some point during the course of the disease. Extensor plantar responses are present in 90% of patients.

Early childhood presentation differs slightly from the typical disease course. Children may be slow in learning how to walk, and, when they can ambulate, they do so in a clumsy and awkward manner. Early in the disease course children may demonstrate motor restlessness similar to chorea. Pseudoathetosis is also sometimes evident.

Flexor spasms are common, but muscle tone is usually normal (though some patients develop hypotonia later in the disease course). Muscle wasting, particularly in the upper limbs occurs in approximately 50% of patients. Symmetrical, slowly progressive weakness affects the lower extremities, particularly pelvic girdle muscles. In the majority of patients, the first significant weakness appears in the hip extensors, followed in a variable manner by weakness in other lower limb muscles. Upper extremity and trunk strength remain nearly normal until late stages of the disease.

The majority of cases show loss of vibration and position sense. Rhomberg's sign is usually present at the time of diagnosis. Decreased pain and touch perception may be present. Scoliosis is common and may be severe, especially in early onset cases. Nearly half of the patients have pes cavus and/or equinovarus deformity of the feet. Peripheral cyanosis of the lower limbs is common and most patients complain of cold feet. Optic atrophy is present in 25% of cases, but visual acuity is rarely severely reduced. Only 20% show evidence of nystamus. Extraocular movements are usually abnormal, with impaired saccadic and smooth pursuit movements (but frank ophtalmoplegia does not typically occur). Ten percent develop sensorineural deafness and one-third develop diabetes mellitus or a mild carbohydrate intolerance. At least 2/3 of patients with FRDA show evidence of cardiomyopathy. Symptoms are rare with the exception of exertional dyspnea. Angina and palpitations may occur but are rare. There may be clinical evidence of ventricular hypertrophy, systolic ejection murmurs, and third or fourth heart sounds in asymptomatic patients. Cardiac failure and arrhythmias occur as a preterminal event. Nearly 65% of patients with FRDA have an abnormal ECG, most commonly showing evidence of ventricular hypertrophy or widespread T-wave inversion. Symmetric concentric hypertrophy is the most common finding by echocardiography.

Pathology

Histologically, there is extensive degeneration, especially within the cervical spinal cord, of the posterior columns and the cell bodies (large neurons of the dorsal root

ganglion and Clarke's column) supplying this region. There is also extensive sclerosis of the lateral columns (corticospinal tracts) and spinocerebellar tracts (especially in the lumbar spine). The brain, cerebellum, and brainstem are left relatively unaffected by this disease, with the exception of occasional patchy loss of Purkinje cells and mild degenerative changes in the brain stem nuclei and optic tract.

Neurophysiology

Electrophysiologically, the above pathology is reflected in the delayed, dispersed somatosensory evoked potentials recorded in the sensory cortex, and abnormal central motor conduction. Nerve conduction studies are helpful in differentiating FRDA from CMT. FRDA shows normal or minimally slow conduction velocities with absent or severely reduced sensory nerve action potential consistent with axonal degeneration in contradistinction to CMT which shows a typical demyelinating pattern. Visual evoked responses are usually reduced in amplitude and show delayed latencies.

Disease Course

Patients progress at variable rates. The mean age at which time a wheelchair becomes necessary is 18.2 years. Most patients are unable to walk by 20 years of age. Weakness is not the primary cause for lack of ambulation, but rather cerebellar dysfunction. Reported mean ages of death are variable and depend on age of presentation and rapidity of disease progression. Patients may survive into their seventh decade, though the mean age of death is typically around the mid thirties. Death generally occurs early in patients with significant cardiac disease and/or diabetes mellitus. There is no effective treatment for the disease.

Appendix D: Laboratory Test Request Forms

Patient 13

Lab Tests

			JUSTIFICATION
01	AChR AB (Antibody to human acetylcholine receptor)	☐	
02	ANA (Antinuclear Antibodies)	☐	
03	Anti-Convulsant Levels	☐	
04	Anti-dsDNA	☐	
05	Anti-GM_1 AB	☐	
06	Antiphospholipid AB	☐	
07	Apo E (Apolipoprotein E)	☐	
08	B12	☐	
09	CBC with Diff. Platelets (Complete Blood Count)	☐	
10	CD4 (T4 count)	☐	
11	Chemistries (Blood chemistries)	☐	
12	Clotting Times	☐	
13	Collagen Vasc. Labs	☐	
14	CPK (Creatine Phosphokinase)	☐	
15	Folate	☐	
16	Glycohemoglobin	☐	
17	GTT (Glucose Tolerance Testing)	☐	
18	HIV screen	☐	
19	Homocysteine	☐	
20	HTLV-1 (Human T-Cell Leukemia Virus Type 1)	☐	
21	Infec. Dis.	☐	
22	Iron Studies	☐	
23	L.E. Prep (Lupus-Erythematosus-Cell Preparation)	☐	
24	Lipid Profile	☐	
25	Lupus Anticoag	☐	
26	Lyme Profile	☐	
27	Mono Spot	☐	
28	P-ANCA (Perinuclear neutrophil cytoplasmic antibody)	☐	
29	Parathyroid Hormone	☐	
30	Paraneoplastic AB	☐	
31	Pituitary Studies	☐	
32	Porphyrins	☐	
33	PPD (Purified Protein Derivative: TB)	☐	
34	Rheumatoid Factor	☐	
35	Sedimentation Rate	☐	
36	SOD1 Activity (Superoxide Dismutase)	☐	
37	SPE with HRE (Serum Protein Electrophoresis)	☐	
38	Syphilis Serologies	☐	
39	Thyroid Function	☐	
40	U/A (Urinalysis)	☐	

Studies/Tests

			PLEASE DISCRIMINATE	JUSTIFICATION
01	Angiogram	☐		
02	Audiometry	☐		
03	Biopsy	☐		
04	Bone Scan	☐		
05	Cardiac Evaluation	☐		
06	Cisternogram	☐		
07	CFS Studies (Chronic Fatigue Syndrome)	☐		
08	CT	☐		
09	EEG (Electroencephalogram)	☐		
10	EMG/NCS (Electromyography/Nerve Conduction Studies)	☐		
11	ENG (Electroneurogram)	☐		
12	Evoked Potentials	☐		
13	Genetic Tests	☐		
14	Ischemic Exercise Test	☐		
15	MRA	☐		
16	MRI	☐		
17	Muscle Biochem	☐		
18	Myelogram	☐		
19	Neuropsych Testing	☐		
20	Pulmonary Function	☐		
21	Schilling Test	☐		
22	SPECT Scan	☐		
23	Tensilon Test	☐		
24	Visual Field	☐		
25	X-Rays	☐		

Patient 25

Lab Tests

			JUSTIFICATION
01	AChR AB (Antibody to human acetylcholine receptor)	☐	
02	ANA (Antinuclear Antibodies)	☐	
03	Anti-dsDNA		
04	Anti-GM_1 AB	☐	
05	Antiphospholipid AB	☐	
06	Apo E (Apolipoprotein E)	☐	
07	B12	☐	
08	CBC with Diff. Platelets (Complete Blood Count)	☐	
09	CD4 (T4 count)	☐	
10	Chemistries (Blood chemistries)	☐	
11	Clotting Times	☐	
12	CPK (Creatine Phosphokinase)	☐	
13	Collagen Vasc. Labs	☐	
14	Folate	☐	
15	Glycohemoglobin	☐	
16	GTT (Glucose Tolerance Testing)	☐	
17	HIV screen	☐	
18	Homocysteine	☐	
19	HTLV-1 (Human T-Cell Leukemia Virus Type 1)	☐	
20	Infec. Dis.	☐	
21	Iron Studies	☐	
22	Lipid Profile	☐	
23	Lupus Anticoag	☐	
24	Lyme Profile	☐	
25	Mono Spot	☐	
26	P-ANCA (Perinuclear neutrophil cytoplasmic antibody)	☐	
27	Parathyroid Hormone	☐	
28	Paraneoplastic AB	☐	
29	Pituitary Studies	☐	
30	Porphyrins	☐	
31	PPD (Purified Protein Derivative: TB)	☐	
32	Rheumatoid Factor	☐	
33	Sedimentation Rate	☐	
34	Serum Amino Acids	☐	
35	SOD1 Activity (Superoxide Dismutase)	☐	
36	SPE with HRE (Serum Protein Electrophoresis)	☐	
37	Syphilis Serologies	☐	
38	Thyroid Function	☐	
39	U/A (Urinalysis)	☐	

Studies/Tests

			PLEASE DISCRIMINATE	JUSTIFICATION
01	Angiogram	☐		
02	Audiometry	☐		
03	Biopsy	☐		
04	Bone Scan	☐		
05	Cardiac Evaluation	☐		
06	Cisternogram	☐		
07	CFS Studies (Chronic Fatigue Syndrome)	☐		
08	CT	☐		
09	EEG (Electroencephalogram)	☐		
10	EMG/NCS (Electromyography/Nerve Conduction Studies)	☐		
11	ENG (Electroneurogram)	☐		
12	Evoked Potentials	☐		
13	Genetic Tests	☐		
14	Heavy Metal Screen	☐		
15	Ischemic Exercise Test	☐		
16	MRA	☐		
17	MRI	☐		
18	Muscle Biochem	☐		
19	Myelogram	☐		
20	Neuropsych Testing	☐		
21	Ophthalmology Consult	☐		
22	Pulmonary Function	☐		
23	Schilling Test	☐		
24	Serum/CFS Serologies	☐		
25	SPECT Scan	☐		
26	Tensilon Test	☐		
27	Visual Field	☐		
28	X-Rays	☐		

Appendix E: Operation of Data Collection Application

First stage: Creation of initial hypothesis list

Participants start by reading the case presentation from beginning to end. The case is presented in a text box occupying the left half of the screen. Participants can highlight parts of the text that they judge important using the highlight tool (see Figure E1). The highlight tool only exists to facilitate the participant's task. Highlighted text was not analyzed.

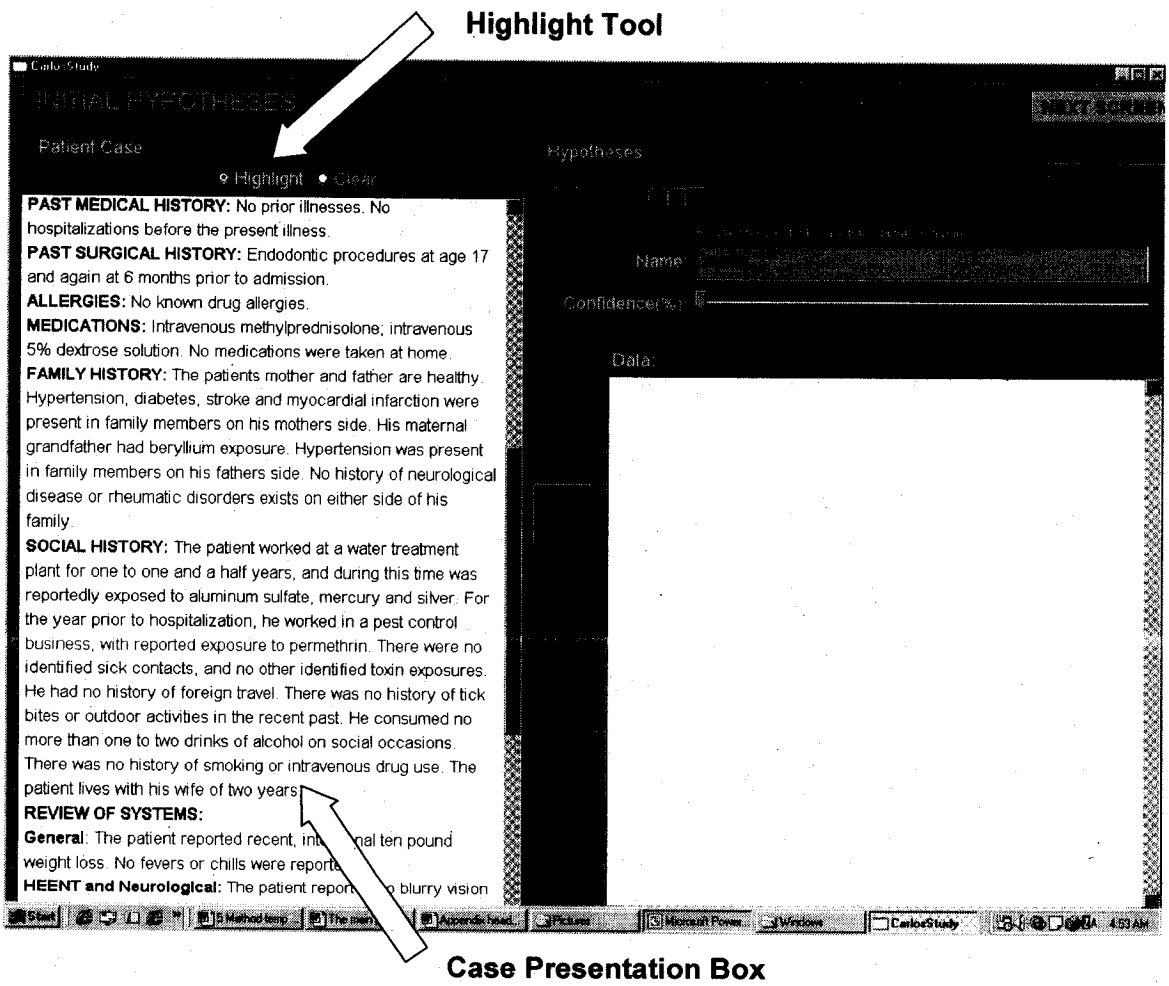

Figure E1. Case presentation box and highlight tool

Next, participants start generating hypotheses. Each hypothesis is composed of: (1) name; (2) confidence level; and (3) supporting evidence. Participants create a new hypothesis by clicking on the *Add Hypothesis* button (see Figure E2). A new tab appears for each hypothesis created. Participants toggle between hypotheses by clicking on the tabs.

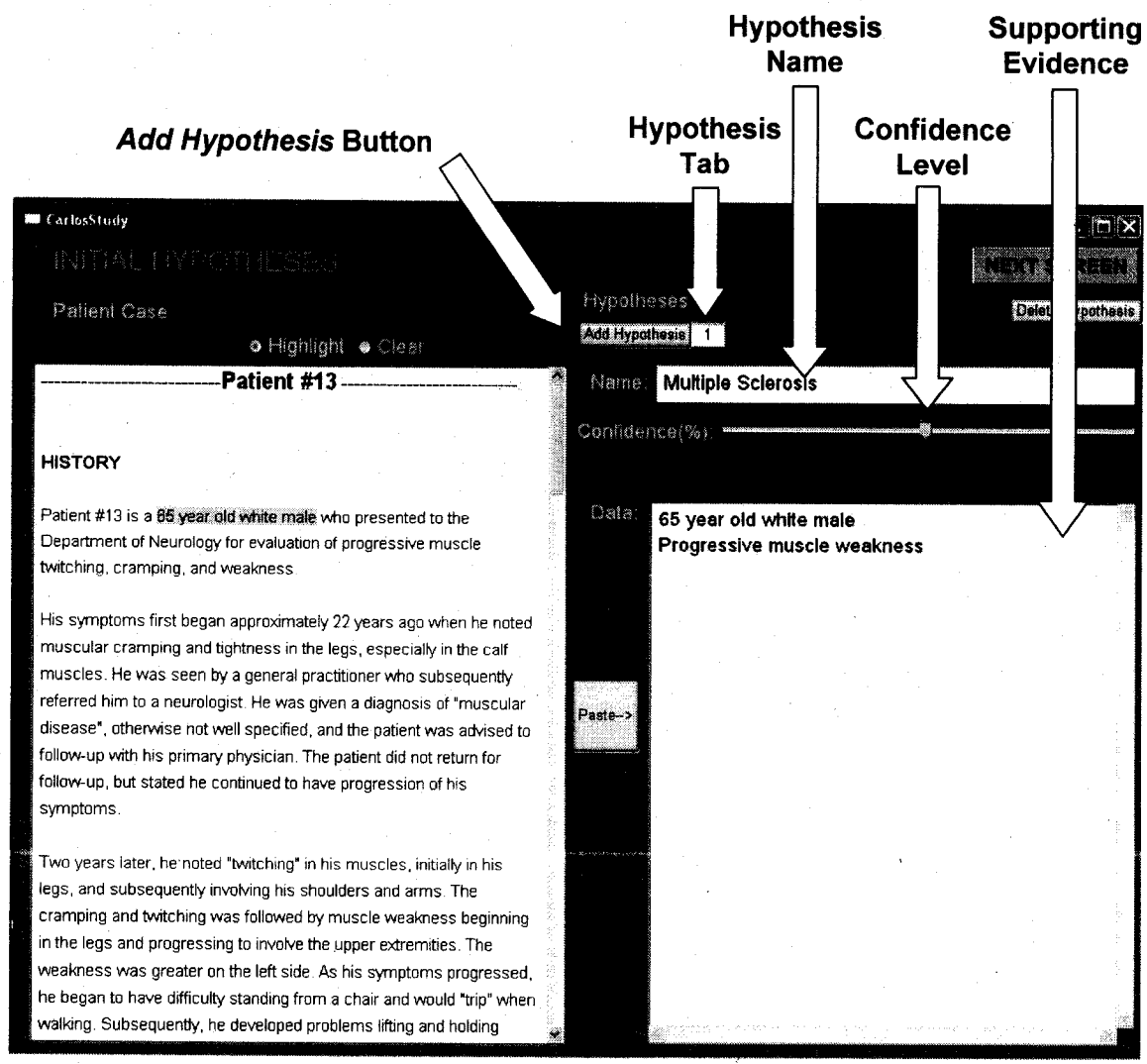

Figure E2. Hypothesis creation and management.

Once participants have finished their initial hypothesis list, they move to the next stage by clicking on the *Next Screen* button, at the top left of the screen (see Figure E3).

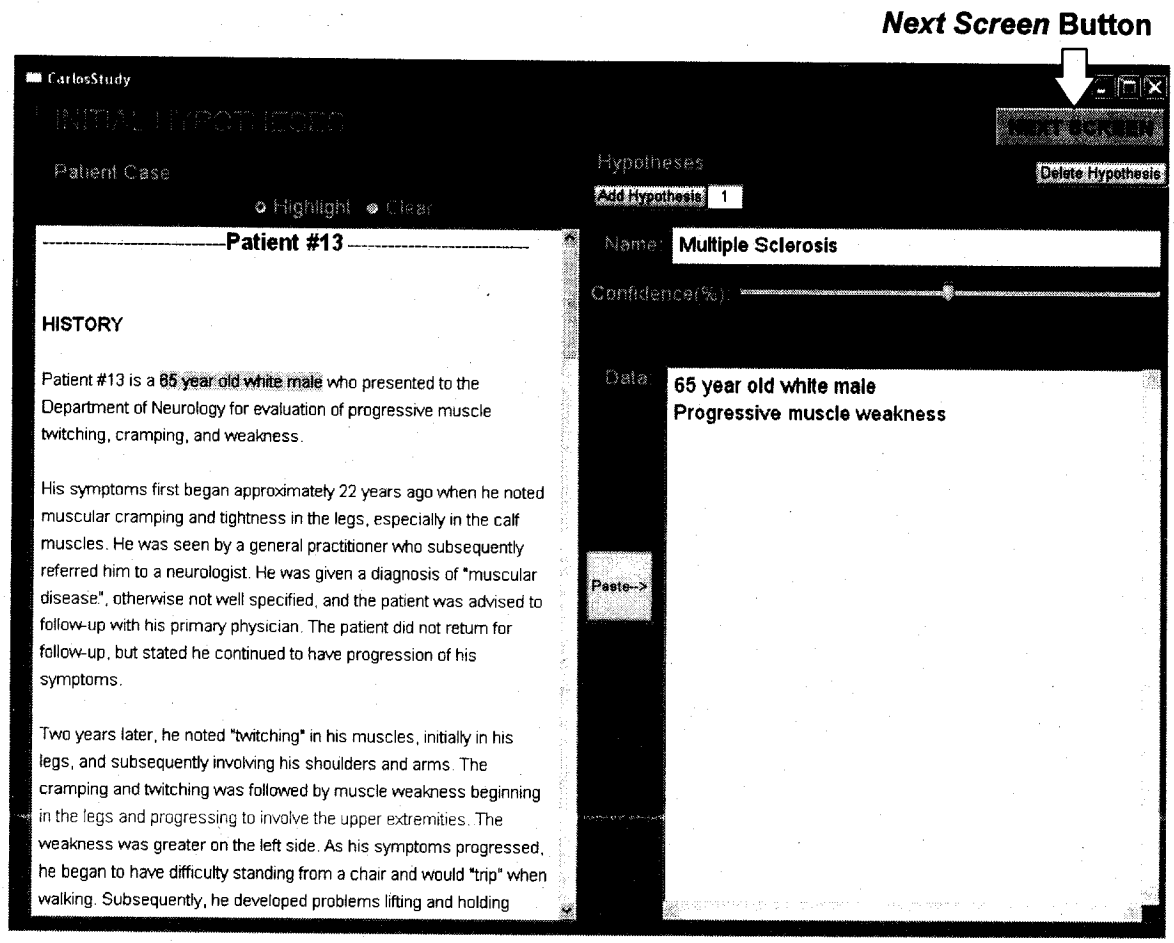

Figure E3. Moving to the next stage

Second stage: Revision of hypothesis list with the use of Abduction.

Participants working on a case in condition B do not complete this stage, proceeding directly to the next stage. The screen elements in this stage are identical to the first stage. The only difference is the title at the top left of the screen that now reads "REVISING HYPOTHESES: Post Hypothesis Generator" (see Figure E4). At this stage, participants use *Abduction* to revise their initial hypothesis list. Participants can add new hypotheses as well as change existing hypotheses.

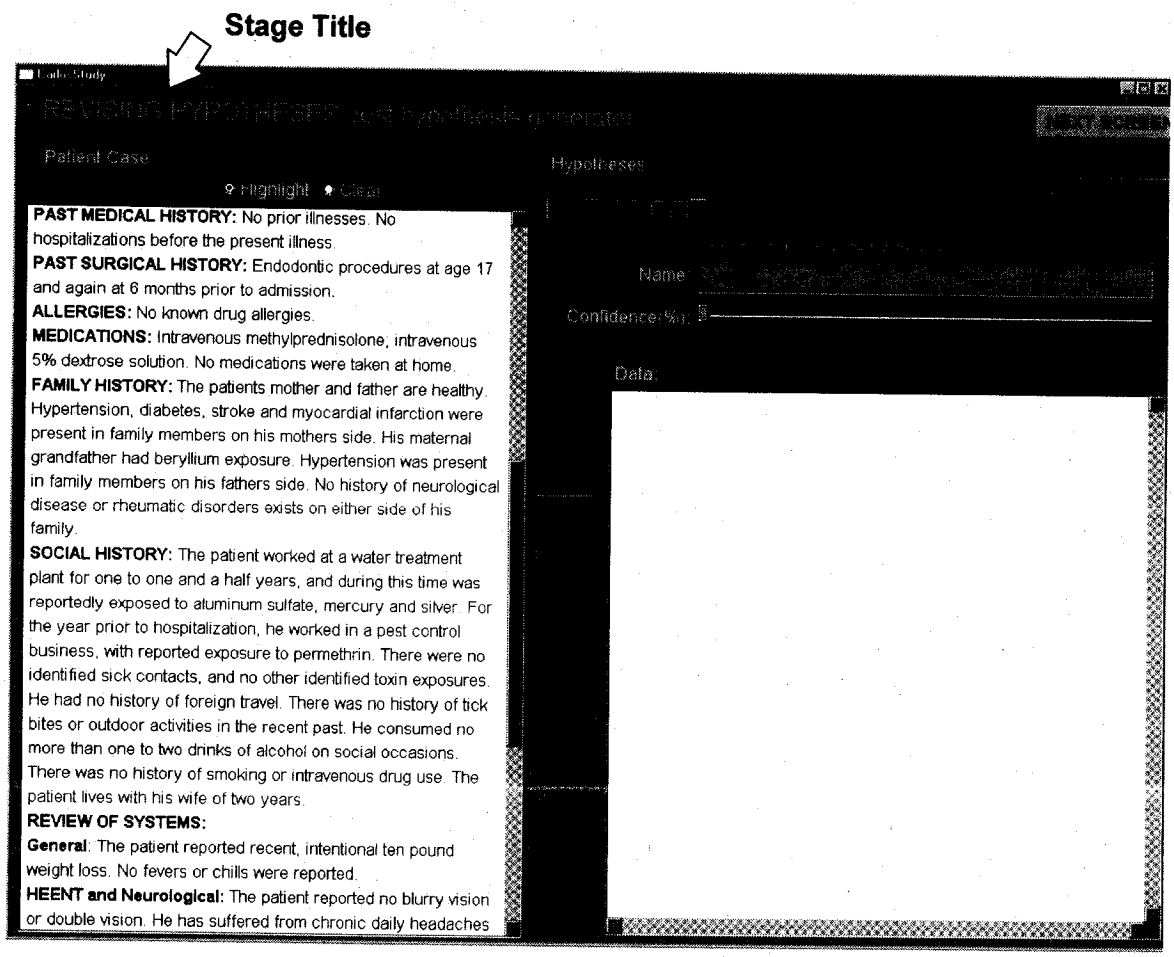

Figure E4. Second stage: hypothesis revision

Once participants have finished revising their hypothesis list, they move to the next stage by clicking on the *Next Screen* button, at the top left of the screen.

Third stage: Evaluation of hypothesis list

In this stage participants can order laboratory tests to evaluate their hypotheses. Participants are also allowed to consult the clinical reference tools of their choice. The screen elements in this stage are identical to the two previous stages. The only difference is the title that now reads "REVISING HYPOTHESES: Post search" (see Figure E5). Participants can add new hypotheses as well as change existing hypotheses.

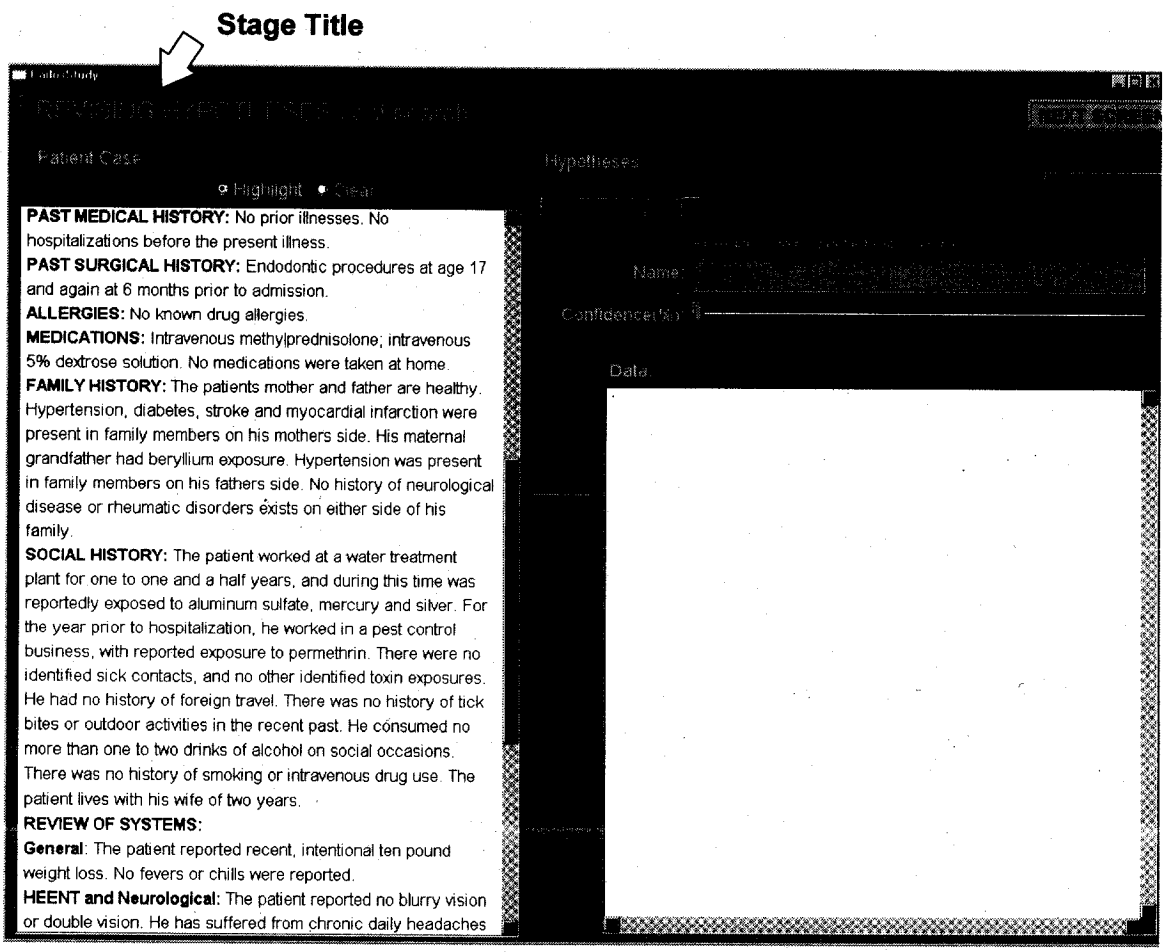

Figure E5. Third stage: hypothesis evaluation

When participants decide they have gone as far as they could with the patient case (whether or not they think they have reached a diagnosis), they proceed to the next and final stage.

Forth stage: Case summary

In this stage, participants state the diagnosis (if they have reached one) and write a summary of their thinking process (see Figure E6).

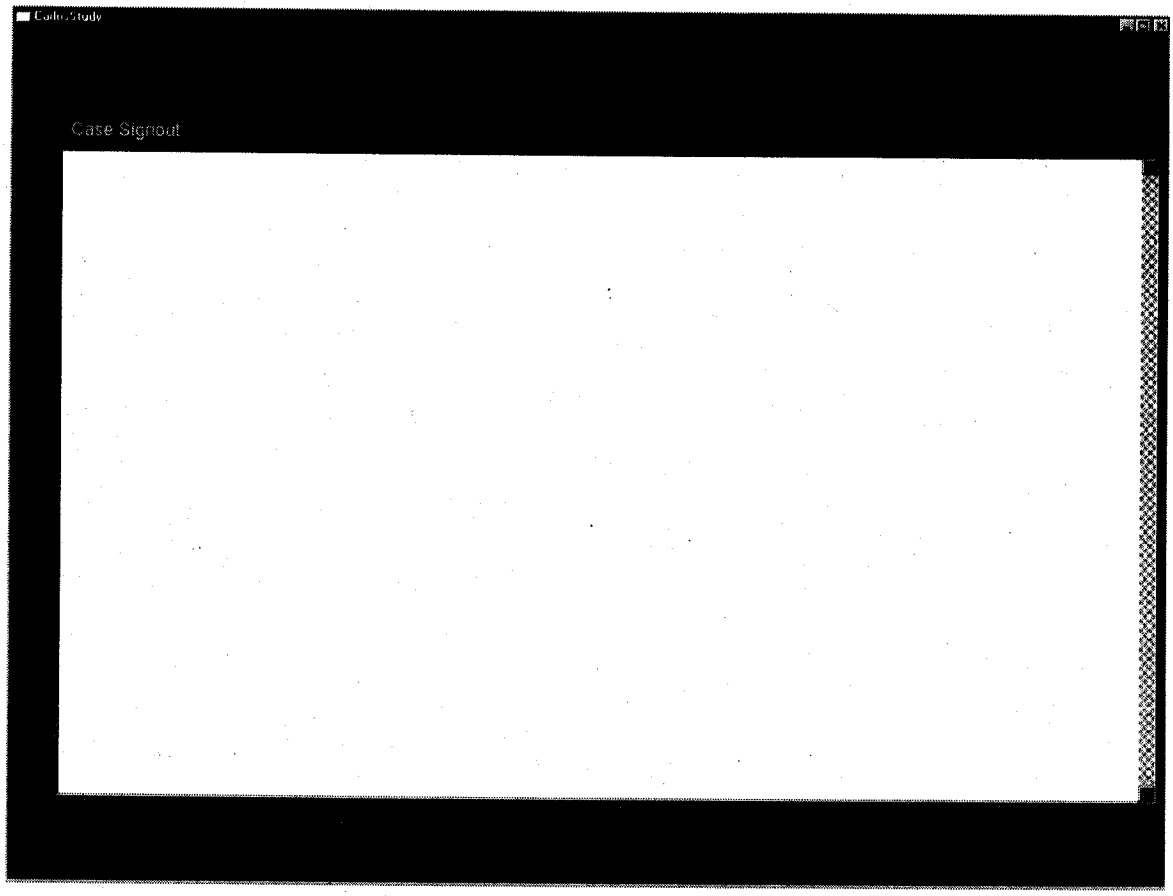

Figure E6. Forth stage: case summary

Appendix F: Operation of *Abduction*

Medical students use Abduction by simply selecting symptoms from the symptom list (see Figure F1). The symptom list is organized by body systems (e.g. neurological, cardiovascular, dermatological, etc.). Users can also locate a symptom by typing it in the Search box at the bottom left of the screen.

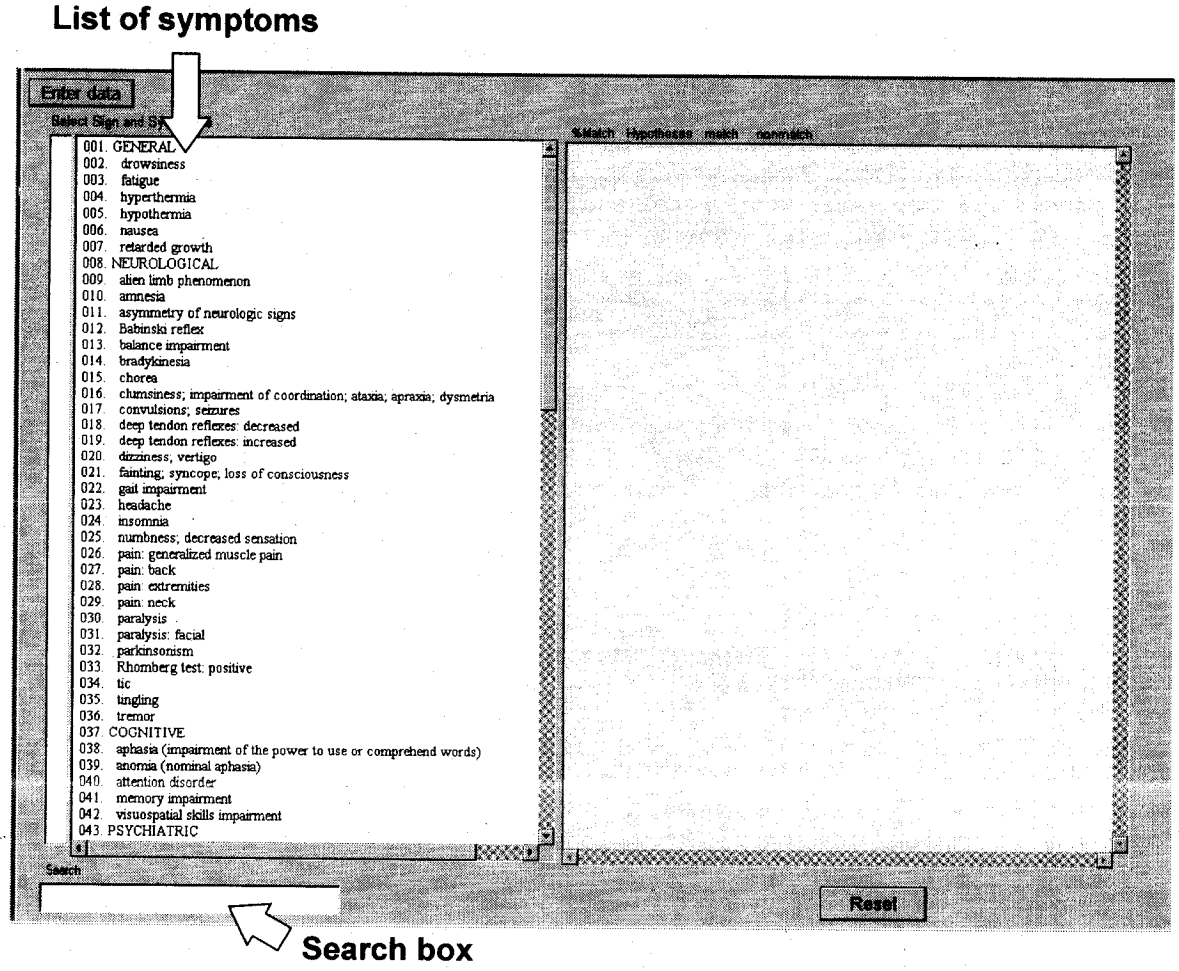

Figure F1. Abduction: symptom list.

When one symptom is selected, a list of matching diseases appears at the right half of the screen (see Figure F2). The percentage in front of the name of each disease indicates the match level between the disease and the selected symptoms. The list of diseases is organized by match level in descending order.

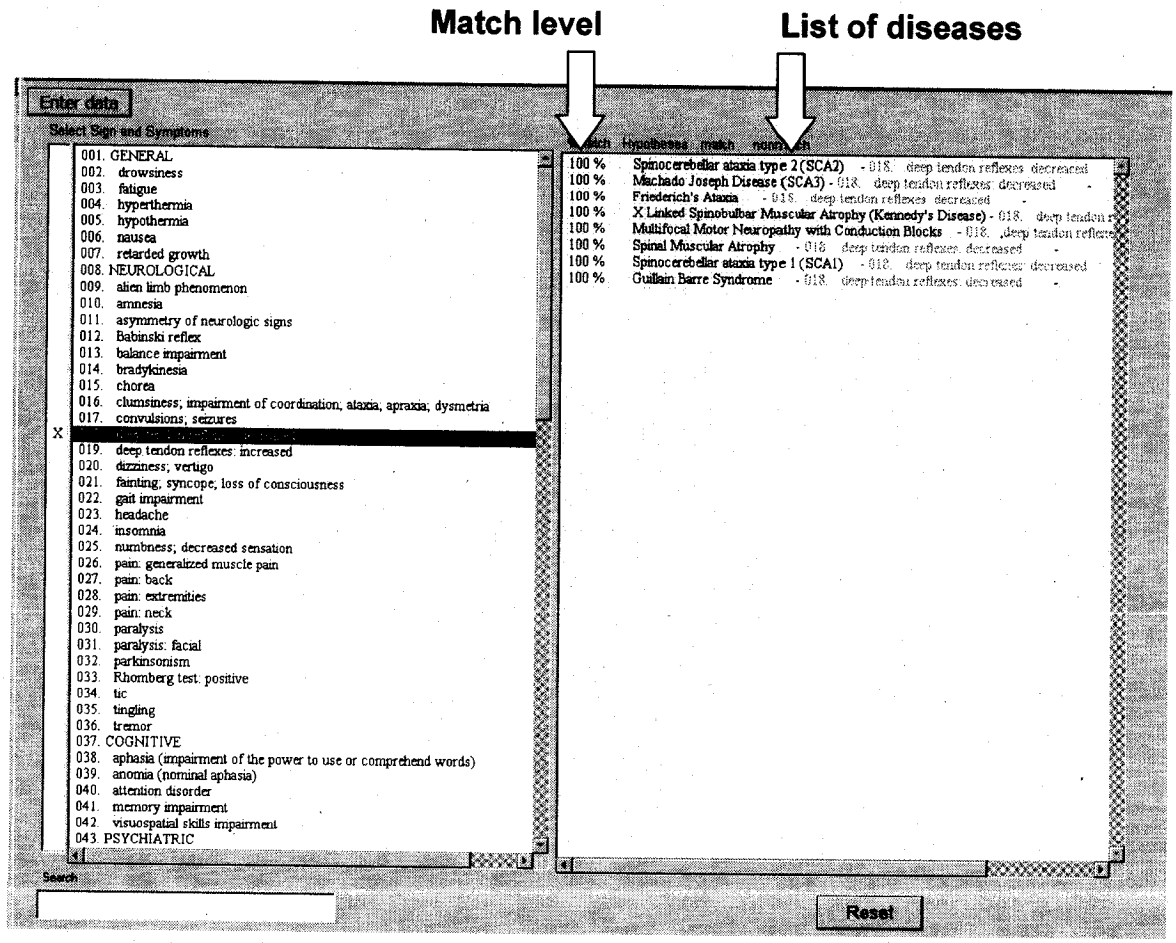

Figure F2. Abduction: list of diseases.

Each disease is followed by a list of matching and non-matching symptoms. Matching symptoms are displayed in green. Non-matching symptoms are displayed in red (see Figure F3).

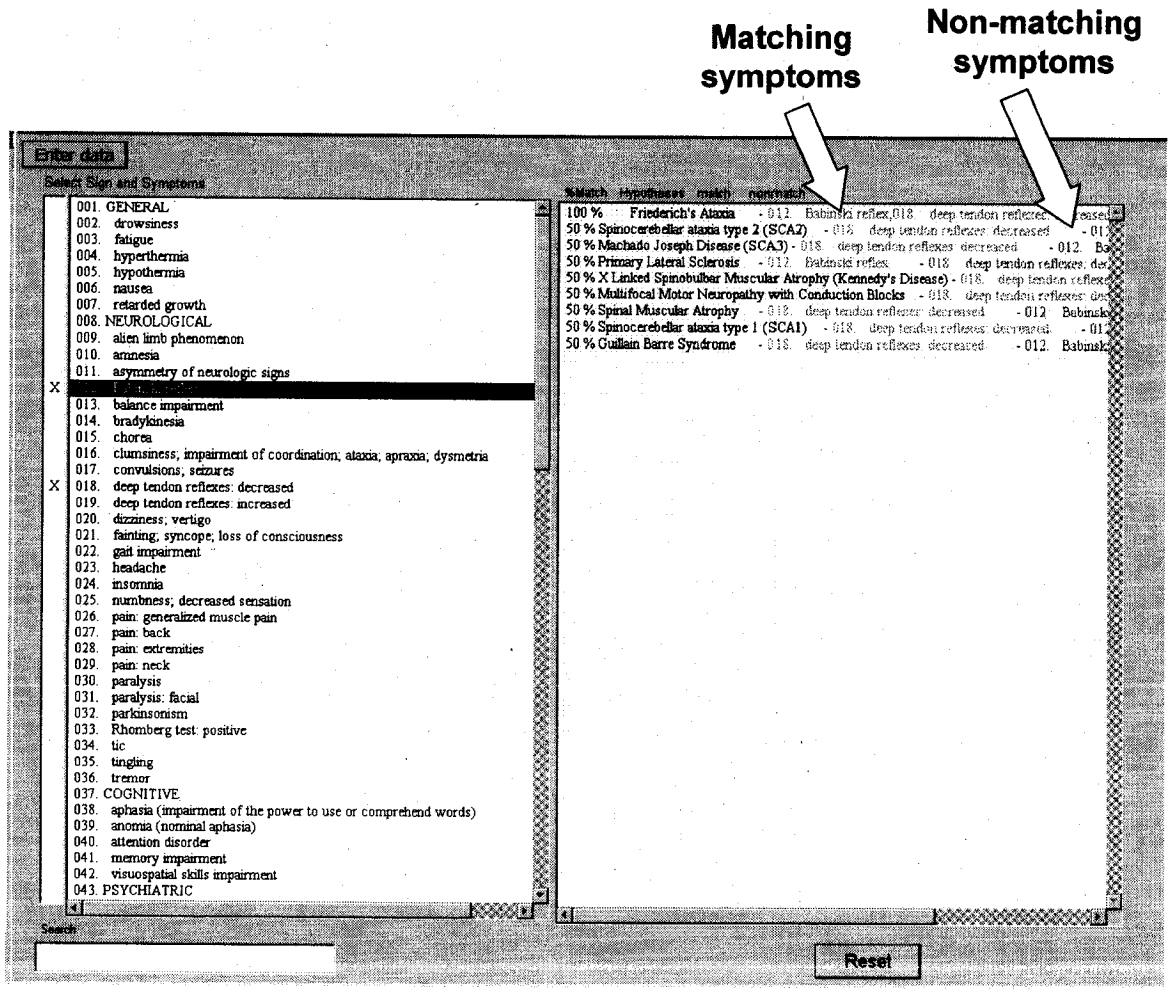

Figure F3. Abduction: matching and non-matching symptoms.

To deselect a symptom, the user must click on the X symbol to the left of the symptom. To deselect all symptoms at the same time, the user must click on the *Reset* button at the bottom of the screen (see Figure F4).

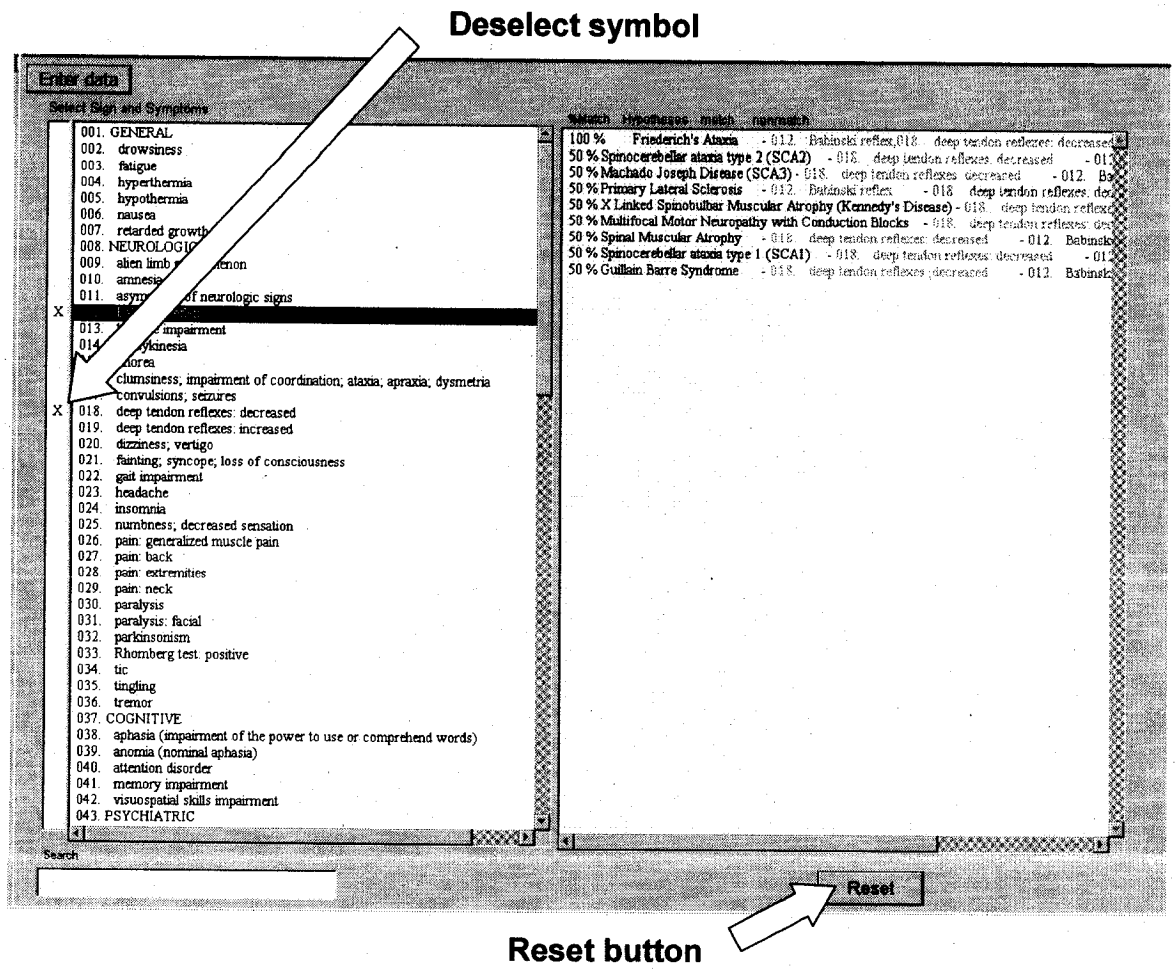

Figure F4. Abduction: deselecting symptoms.

Appendix G: Post-test Questionnaire & Descriptive Summary of Answers

CONFIDENTIAL INFORMATION SHEET

Date: _____

1. Participant's information

Name: _____
Contact: _____

2. Proficiency using information systems

Which information systems (e.g. Medline, Harrison's Online, etc.) do you use on a regular basis?

_____
_____

How often do you use these systems?
☐ Every day ☐ 2-4 times/week ☐ Once a week ☐ Every other week ☐ Once a month

How often do you find the information you are looking for?
☐ 0%-20% ☐ 21%-40% ☐ 41%-60% ☐ 61-80% ☐ 81%-100%

How quickly do you find the information you are looking for?
☐ It usually takes me a lot of time to find the information I need
☐ It depends on the problem I have but usually it is time-consuming
☐ It depends on the problem I have but usually it does not take me much time
☐ I usually find the information I need quickly

3. Comments on your participation

How challenging were the cases to you?

First Case:	☐ very easy	☐ easy	☐ average	☐ difficult	☐ very difficult
Second Case:	☐ very easy	☐ easy	☐ average	☐ difficult	☐ very difficult

How do you feel about the task and the steps you were supposed to follow?
☐ It was clear and easy to follow
☐ I was able to follow the procedures but not without some effort
☐ It was confusing

What did you think of the prompts to justify the lab tests you ordered and your use of the information systems?
☐ They interfered with my reasoning
☐ They did not interfere with my reasoning
☐ They facilitated my reasoning

How useful was the Hypothesis Generator to you?
☐ not useful at all ☐ slightly useful ☐ useful ☐ very useful

We would appreciate any suggestions/criticisms you may have regarding the cases, the applications, and the data collection procedure.

_____
_____
_____
_____

Descriptive Summary of Answers

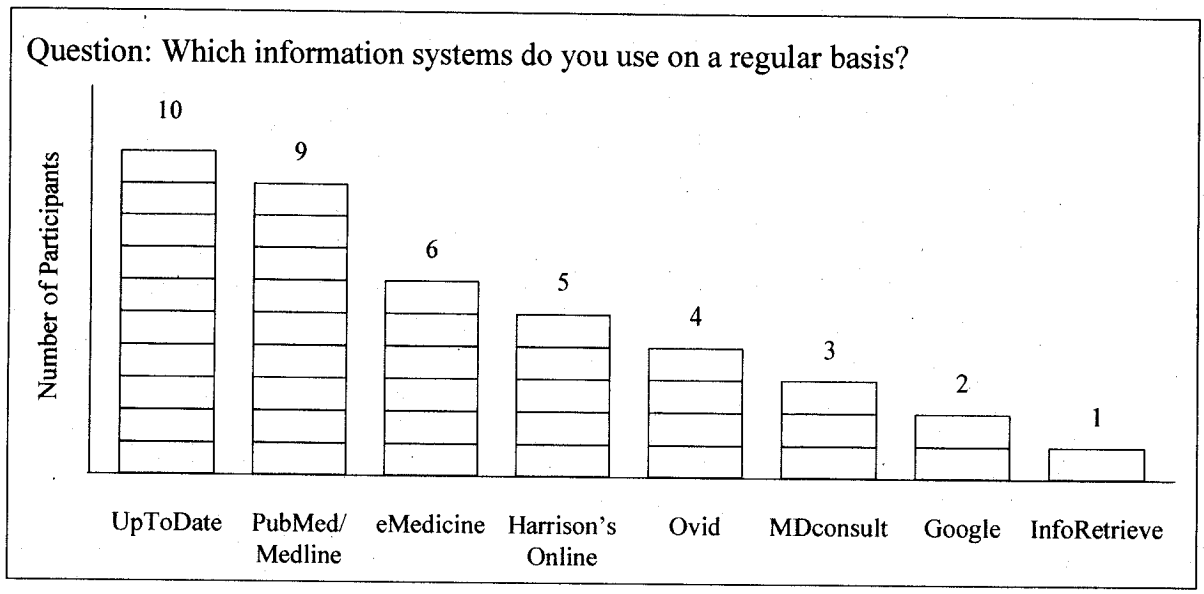

Figure G1. Clinical reference systems consulted on a regular basis by the participants.

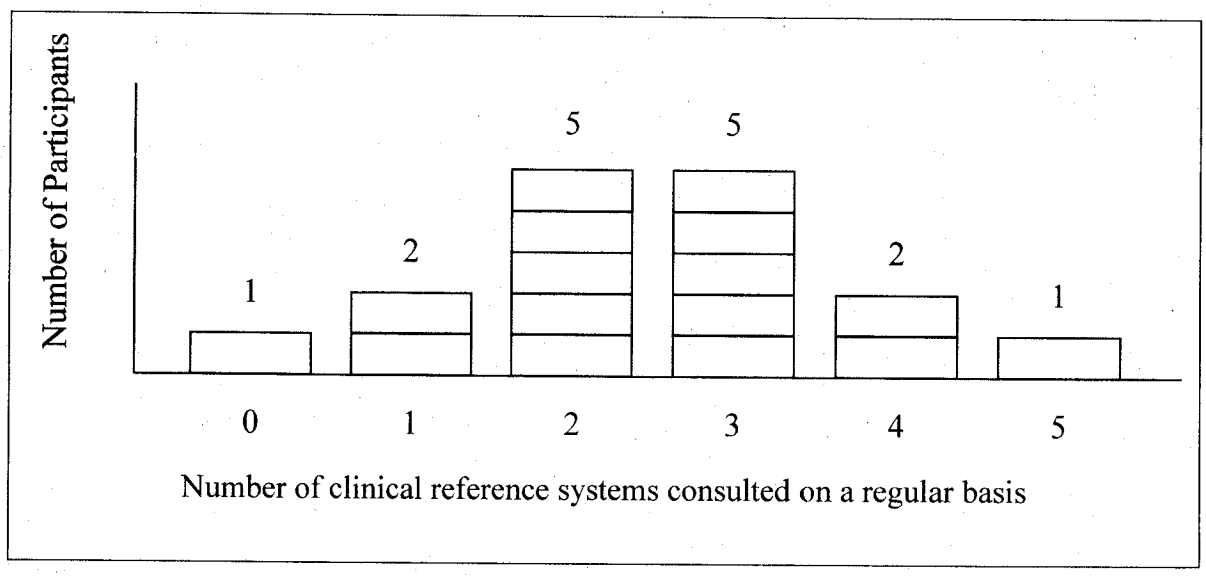

Figure G2. Number of clinical reference systems consulted on a regular basis by the participants.

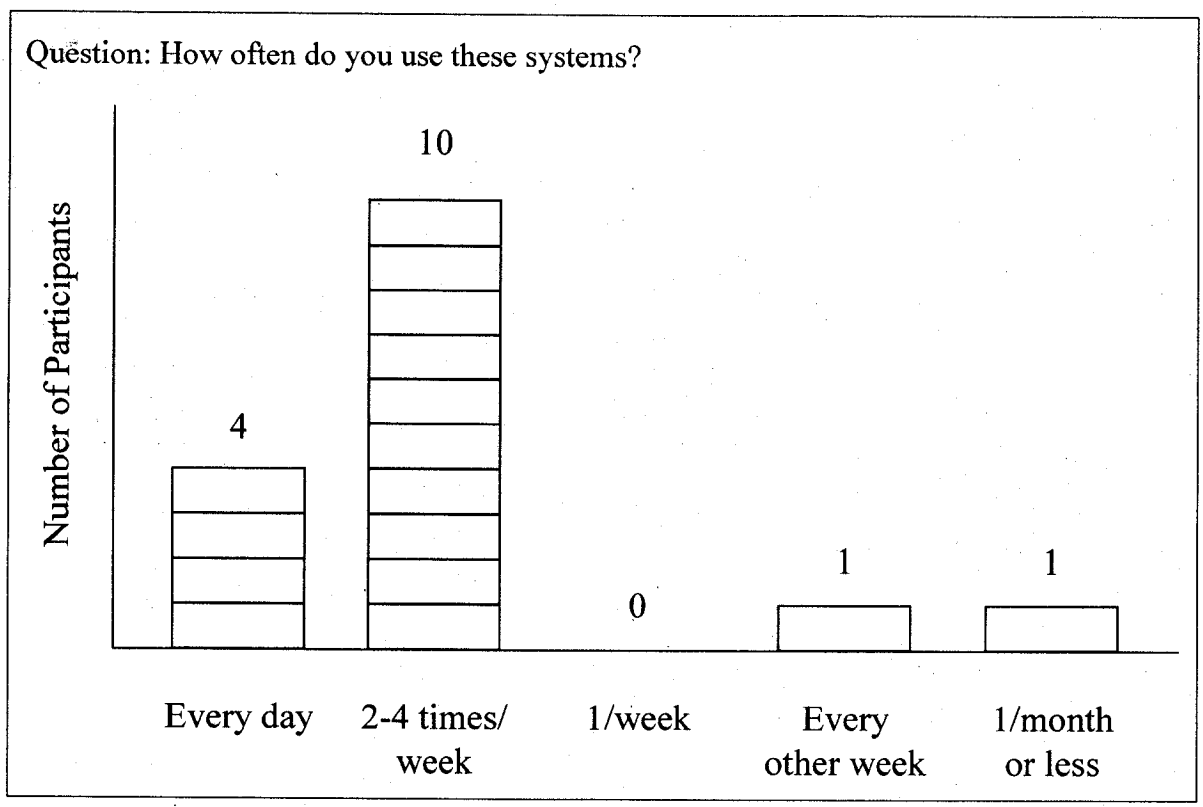

Figure G3. Frequency of use of clinical reference systems in total.

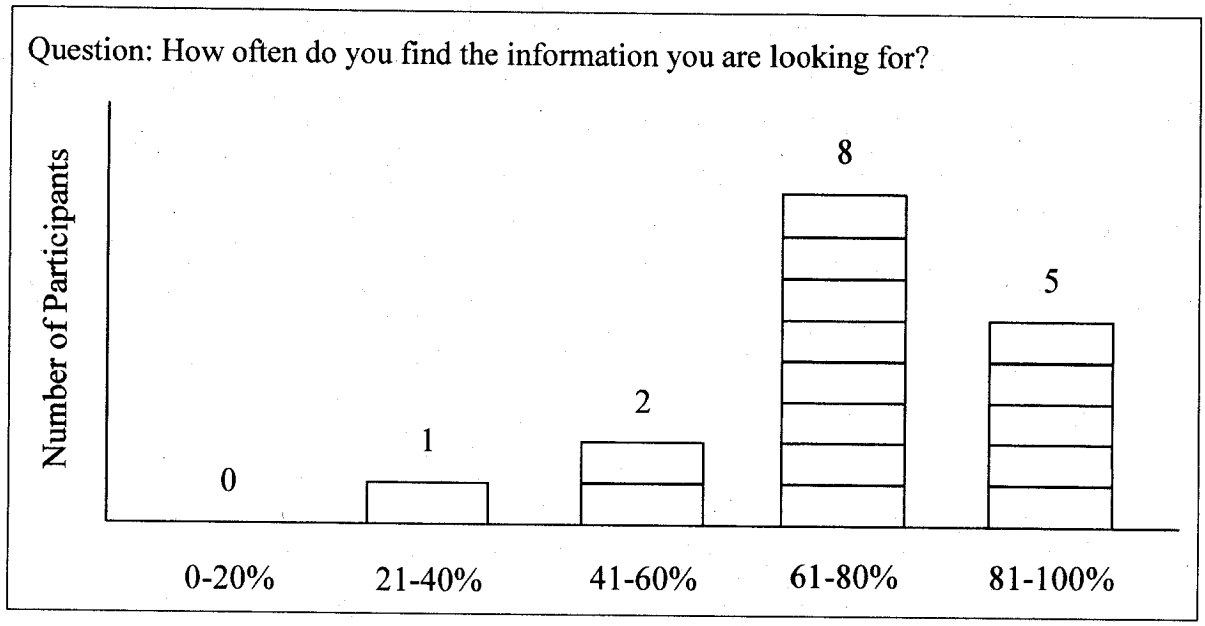

Figure G4. Percentage of successful information-seeking operations.

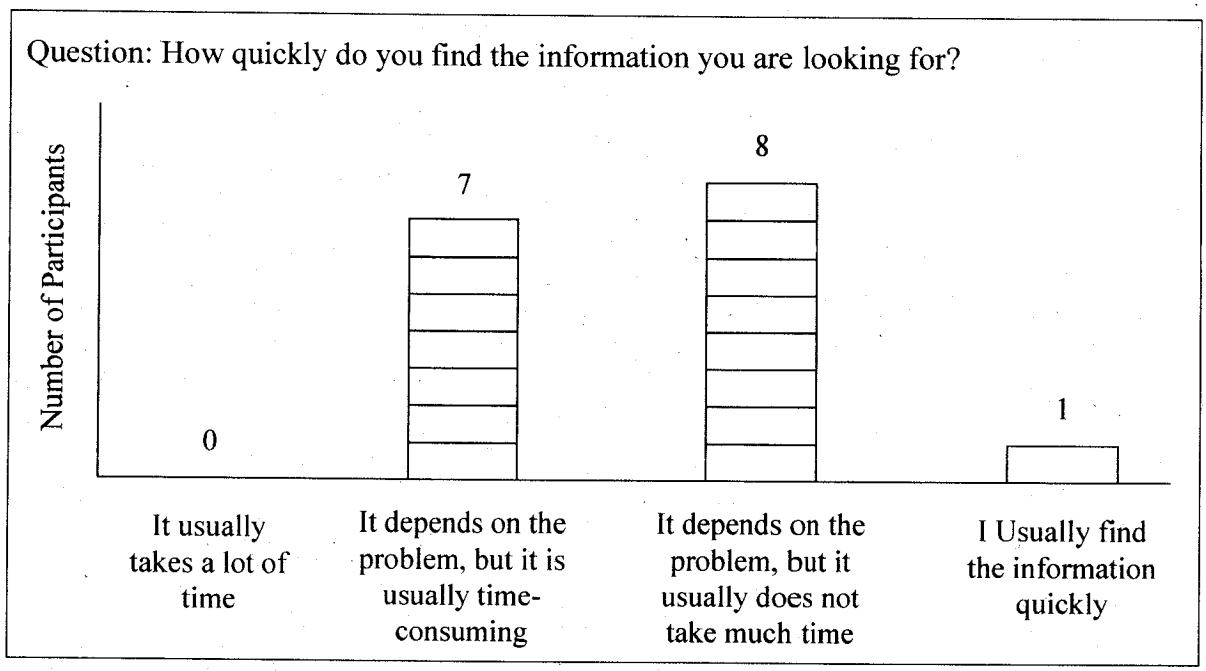

Figure G5. Amount of time spent with information-seeking operations.

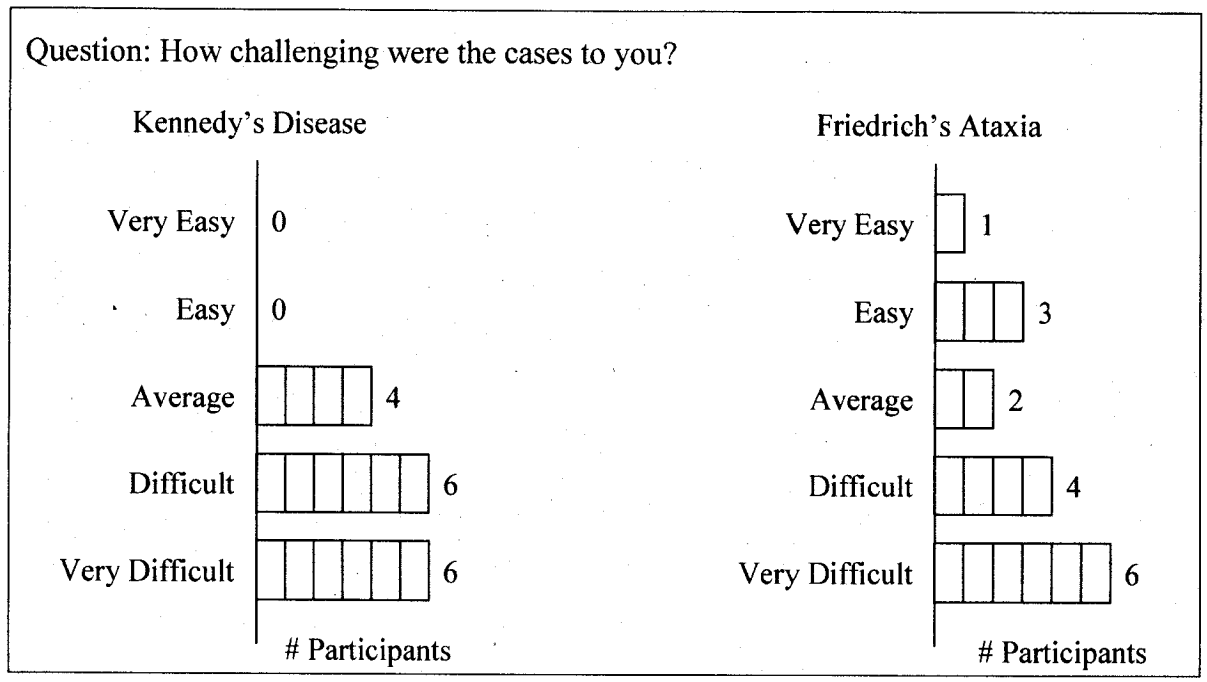

Figure G6. Participants' perception of difficulty level of the cases.

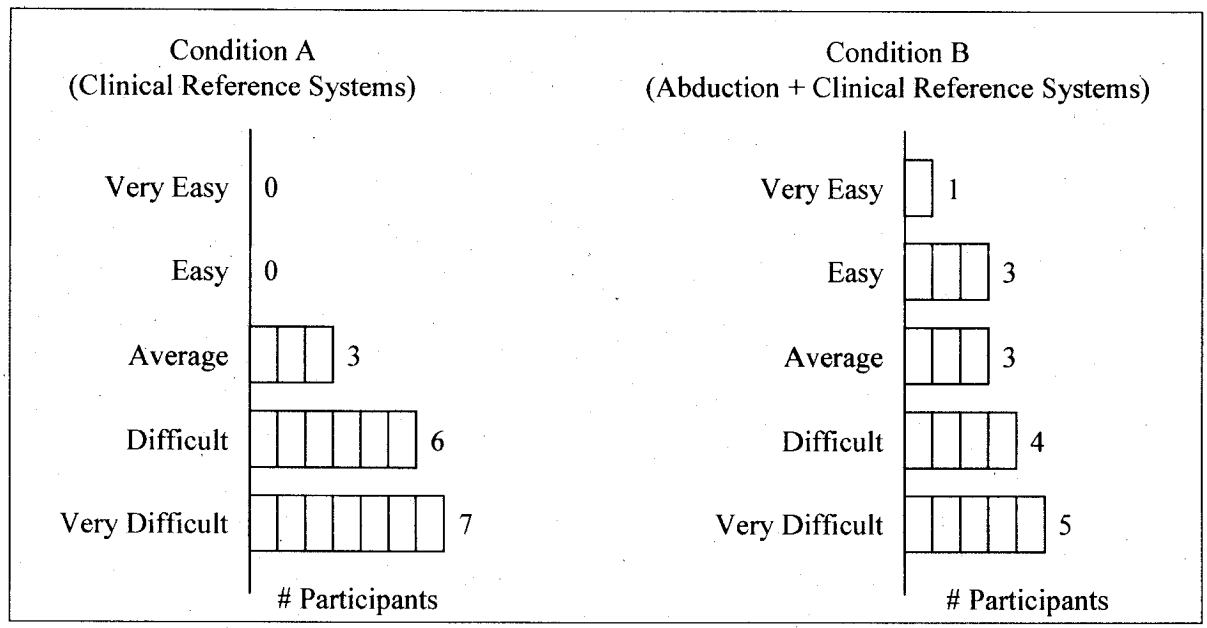

Figure G7. Participants' perception of difficulty level of the cases by condition.

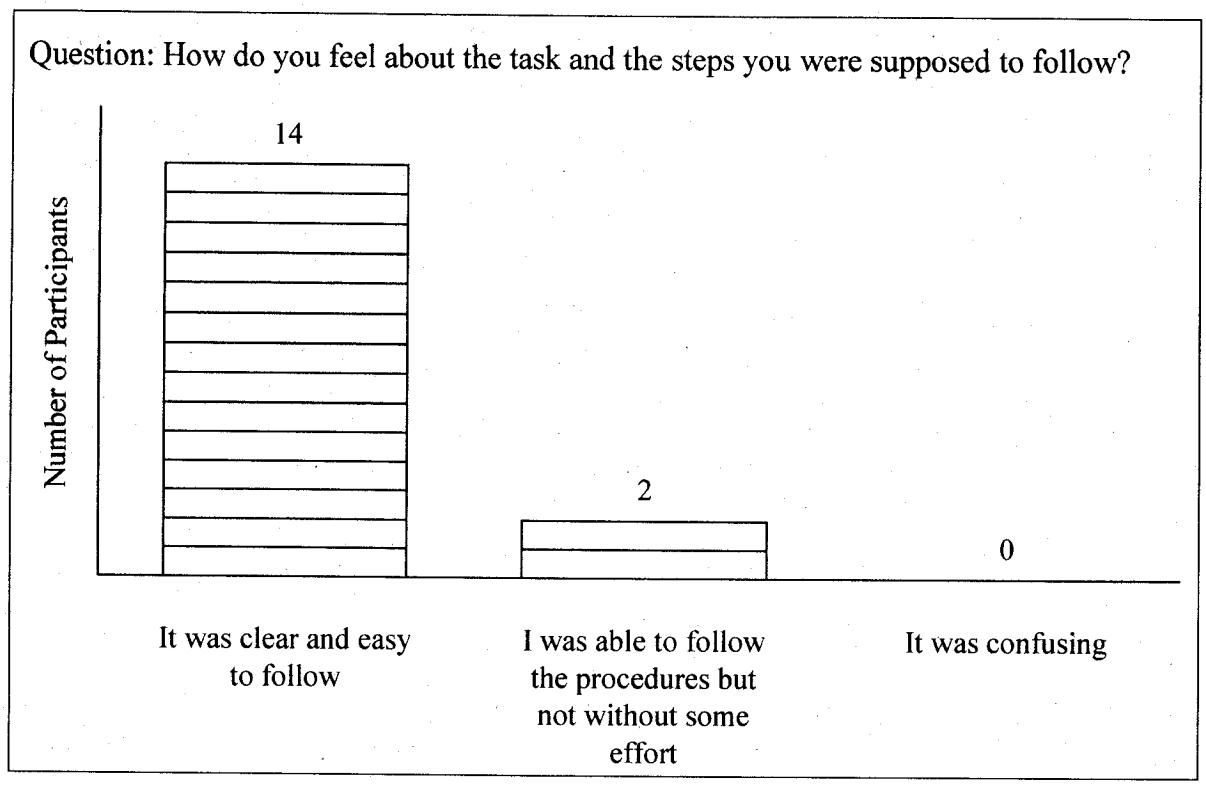

Figure G8. Participants' perception of the task. and procedures.

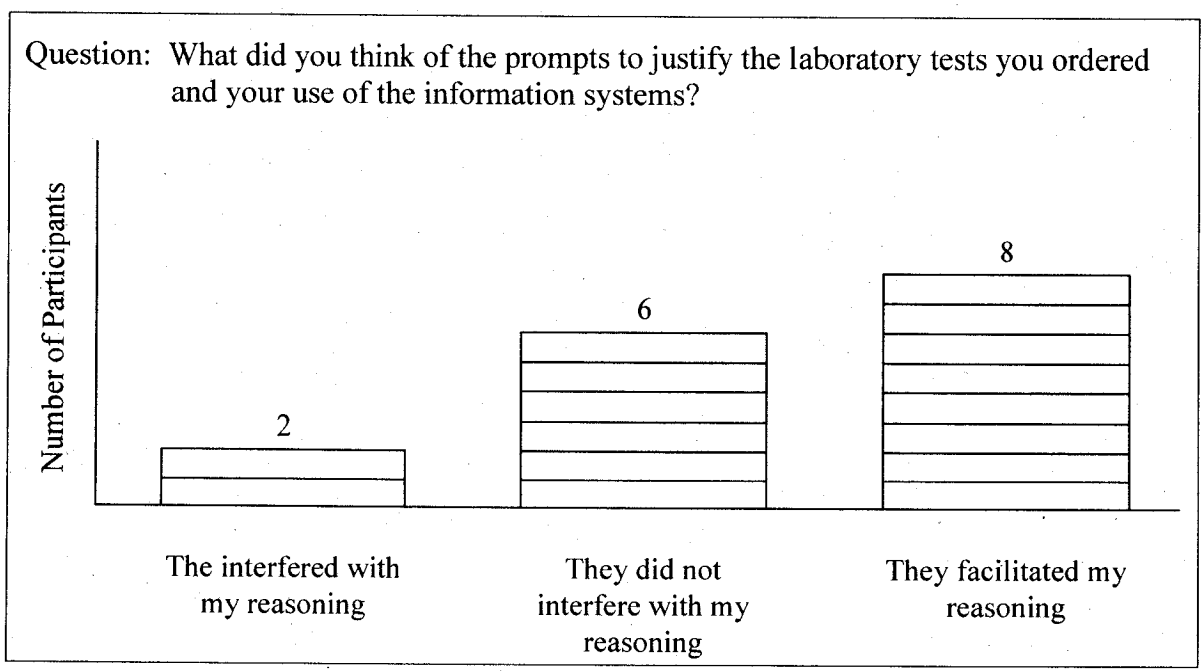

Figure G9. Participants' perception of the researcher's prompts.

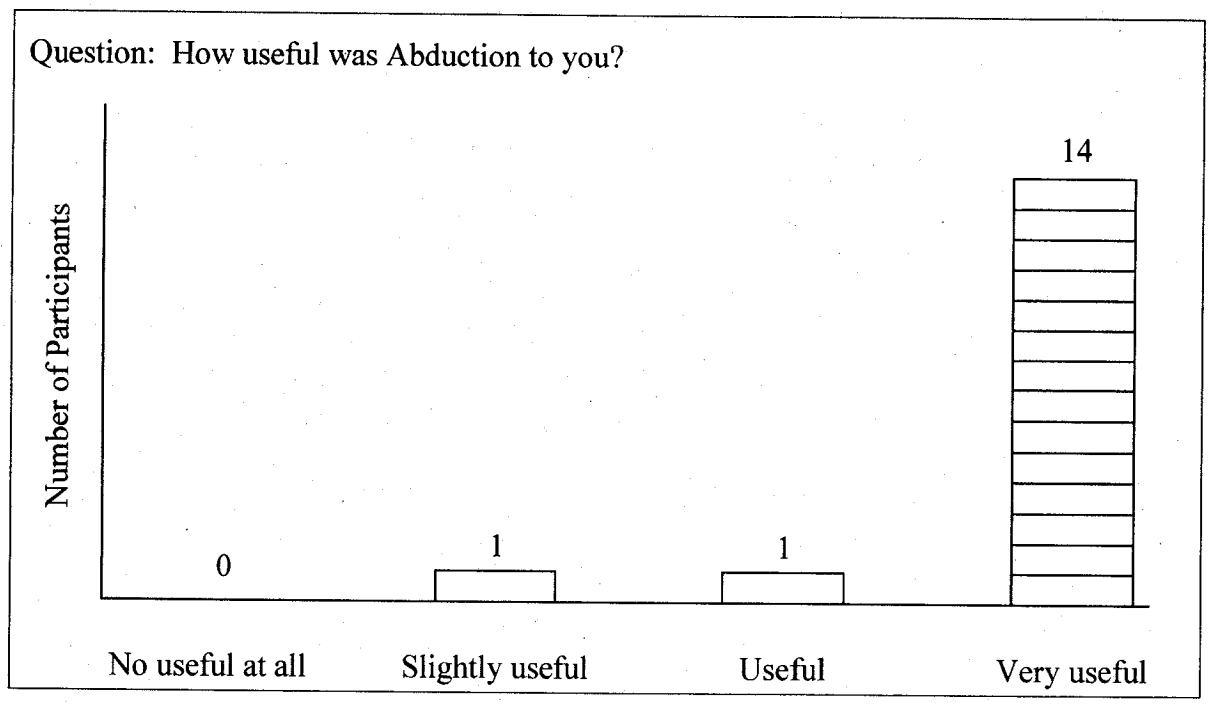

Figure G10. Participants' perception of *Abduction*'s usefulness.

Appendix H: Outcome Measurements and Subtotal by Condition, Case, and Cohort

Table H1

Main Outcome Measurements

Occasion	Participant	Cohort	Order	Case	Condition	Correct Hypothesis Generated Before Consultation	Correct Hypothesis Generated After Consultation	Correct Hypothesis Confirmed
01	01	Med-2	First	Kennedy's disease	A (with Abduction)	No	Yes	Yes
02	01	Med-2	Second	Friedrich's ataxia	B (without Abduction)	No	No	--
03	02	Med-2	First	Kennedy's disease	B (without Abduction)	No	No	--
04	02	Med-2	Second	Friedrich's ataxia	A (with Abduction)	No	Yes	Yes
05	03	Med-2	First	Friedrich's ataxia	A (with Abduction)	No	Yes	No
06	03	Med-2	Second	Kennedy's disease	B (without Abduction)	No	No	--
07	04	Med-2	First	Friedrich's ataxia	B (without Abduction)	No	No	--
08	04	Med-2	Second	Kennedy's disease	A (with Abduction)	No	Yes	Yes

Table H1 (Continued)

Main Outcome Measurements

Occasion	Participant	Cohort	Order	Case	Condition	Correct Hypothesis Generated Before Consultation	Correct Hypothesis Generated After Consultation	Correct Hypothesis Confirmed
09	05	Med-2	First	Kennedy's disease	A (with Abduction)	No	Yes	Yes
10	05	Med-2	Second	Friedrich's ataxia	B (without Abduction)	No	No	--
11	06	Med-2	First	Kennedy's disease	B (without Abduction)	No	No	--
12	06	Med-2	Second	Friedrich's ataxia	A (with Abduction)	No	Yes	No
13	07	Med-2	First	Friedrich's ataxia	A (with Abduction)	No	Yes	Yes
14	07	Med-2	Second	Kennedy's disease	B (without Abduction)	No	No	--
15	08	Med-2	First	Friedrich's ataxia	B (without Abduction)	No	No	--
16	08	Med-2	Second	Kennedy's disease	A (with Abduction)	No	Yes	Yes

Table H1 (Continued)

Main Outcome Measurements

Occasion	Participant	Cohort	Order	Case	Condition	Correct Hypothesis Generated Before Consultation	Correct Hypothesis Generated After Consultation	Correct Hypothesis Confirmed
17	09	Med-4	First	Kennedy's disease	A (with Abduction)	No	Yes	Yes
18	09	Med-4	Second	Friedrich's ataxia	B (without Abduction)	No	Yes	No
19	10	Med-4	First	Kennedy's disease	B (without Abduction)	No	No	--
20	10	Med-4	Second	Friedrich's ataxia	A (with Abduction)	No	Yes	Yes
21	11	Med-4	First	Friedrich's ataxia	A (with Abduction)	No	Yes	No
22	11	Med-4	Second	Kennedy's disease	B (without Abduction)	No	No	--
23	12	Med-4	First	Friedrich's ataxia	B (without Abduction)	No	No	--
24	12	Med-4	Second	Kennedy's disease	A (with Abduction)	No	Yes	Yes

Table H1 (Continued)

Main Outcome Measurements

Occasion	Participant	Cohort	Order	Case	Condition	Correct Hypothesis Generated Before Consultation	Correct Hypothesis Generated After Consultation	Correct Hypothesis Confirmed
25	13	Med-4	First	Kennedy's disease	A (with Abduction)	No	Yes	Yes
26	13	Med-4	Second	Friedrich's ataxia	B (without Abduction)	No	Yes	Yes
27	14	Med-4	First	Kennedy's disease	B (without Abduction)	No	No	--
28	14	Med-4	Second	Friedrich's ataxia	A (with Abduction)	No	Yes	Yes
29	15	Med-4	First	Friedrich's ataxia	A (with Abduction)	No	Yes	Yes
30	15	Med-4	Second	Kennedy's disease	B (without Abduction)	No	No	--
31	16	Med-4	First	Friedrich's ataxia	B (without Abduction)	No	Yes	No
32	16	Med-4	Second	Kennedy's disease	A (with Abduction)	No	Yes	Yes

Table H2

Subtotals by Condition, Case, and Cohort

Subtotals	Correct Hypothesis Generated Before Consultation	Correct Hypothesis Generated After Consultation	Correct Hypothesis Confirmed
By Condition			
A (with Abduction)	0	16	13
B (without Abduction)	0	3	1
By Case			
Friedrich's ataxia	0	11	6
Kennedy's disease	0	8	8
By Cohort			
Med-2	0	8	6
Med-4	0	11	8
Total	**0**	**19**	**14**