International Journal of medical Informatics (2001,in press)

Cognitive Psychological Studies of Representation and Use of Clinical Practice Guidelines

Vimla L. Patel, Ph.D., D.Sc.[1,2] (patel@dmi.columbia.edu)
José F. Arocha, Ph.D.[2] (francisc@med.mcgill.ca)
Melissa Diermeier, B.Sc.[2] (mdierm@po-box.mcgill.ca)
Jacques How, M.B., ChB., MRCP[3] (jhow@is.mgh.mcgill.ca)
Christel Mottur-Pilson, Ph.D.[4] (cmotturpilson@mail.acponline.org)

[1]Department of Medical Informatics,
College of Physicians and Surgeons,
Columbia University, New York, NY

[2]Cognitive Studies in Medicine,
Centre for Medical Education,
McGill University, Montreal, Canada

[3]Division of Endocrinology
Montreal General Hospital, Montreal, Canada

[4]American College of Physicians-American Society for Internal Medicine,
Philadelphia, PA

Keywords: Clinical practice guidelines, expertise, decision making, cognitive research

Running Head: Representation and Use of Clinical Guidelines

Corresponding author:
Vimla L. Patel, PhD, DSc
Department of Medical Informatics
Vanderbilt Clinic Blg.,5th Floor
Columbia University
622 West 168th. Street
New York, NY 10032

Abstract

Clinical practice guidelines provide a means to enhance physician performance. This investigation was undertaken in an attempt to understand the nature of impact of guideline use on physician performance. We investigated the impact of (a) algorithmic-based and (b) text-based practice guidelines on clinical decision making by physicians at varying levels of expertise. Data were collected using clinical scenarios and a think-aloud paradigm, both with (primed) and without (spontaneous) the use of the guidelines. The two guidelines used in the study were management of diabetes and screening for thyroid disease.

The results show that guidelines were used as reminders for both experts and non-experts. Guidelines acted as an educational tool for non-experts by assisting in knowledge reorganization, particularly for the non-experts. Text and algorithmic guideline formats were both useful to physician performance depending on the purpose of use—solving clinical problems or learning.

These results provide insights into how guidelines can be fine-tuned for different users and for different purposes. Empirical research, coupled with design principles from the cognitive sciences can form an essential component of guideline design and development.

1 Introduction

Each day physicians must make informed, accurate decisions concerning patient care—a task that requires processing reams of information while simultaneously selecting relevant pieces of evidence to support decisions. This is a formidable challenge considering that the body of medical knowledge is in a state of constant growth and revision [1, 2]. Clinical guidelines are meant to help manage this "information overload" and provide up-to-date, scientifically valid information to aid the practitioner in making more efficient and effective clinical decisions [3]. Furthermore, guidelines can serve to reduce inappropriate and/or ineffective practices while encouraging effective practices, improving quality, and reducing costs of health care [3-6]. However, the outlook for the potential impact of clinical guidelines as positive is threatened by practitioners' unwillingness to make use of them [3-7]. This noncompliance is puzzling because guidelines can enhance physician performance, as some research has indicated [3, 8]. In this paper we report on two laboratory-based studies of interpretation and use of clinical guidelines by physicians. In particular, we investigate clinical decision making as a function of expertise.

2 Expertise and Guideline Utilization

The cognitive study of expertise provides the theoretical framework that sheds light on understanding clinical guideline interpretation and use. Cognitive research on expertise has examined the differences between experts and novices in a variety of domains ranging from physics to social sciences [9, 10]. It has been estimated that it takes approximately ten years of deliberate practice to reach the level of an expert [11, 12]. In the domain of medicine, differences between expert physicians (e.g., domain specialists, such as endocrinologists or cardiologists solving clinical problems that fall within their specialty), subexperts (e.g., general practitioners or specialists working on problems outside their specific domain of expertise), intermediates (e.g., residents in training) and novices (e.g., medical students) have been reported in a number of publications [13, 14]. In this paper, we will be using the term "expert" to refer to any domain specialists working within his or her field of specialization, and the term "non-expert" to refer to any subexpert, intermediate, or novice as defined above.

Expertise research shows that there are fundamental differences in comprehension, problem solving, and decision making as a function of expertise [9, 10, 15]. There are at least five characteristics that differentiate experts from non-experts. First, it has been shown that experts and non-experts use different patterns of reasoning. In routine problems, experts use a data-driven pattern of reasoning where the problem data lead to the hypothesis. In contrast, non-experts use a hypothesis-driven form of reasoning where the hypothesis guides data collection and interpretation [16]. Second, experts and non-expert differ in the organization of their knowledge base. The expert has a more highly organized knowledge base, which allows them to partition the problem into manageable “chunks.” In contrast, although non-experts may possess a large knowledge base, it is not well organized. As a result, they often generate many hypotheses that are unrelated to the problem. Third, this difference in the knowledge base accounts for their patterns of reasoning and the types of errors they generally make. Typically, non-experts make errors of commission because of their inability to discriminate relevant from irrelevant information. Experts, on the other hand, make errors of omission because of overconfidence, or as a result of skipping steps in problem solving [17]. Fourth, they differ in how they approach a clinical problem. During problem solving experts typically generate a small set of relevant hypotheses at a high level of abstraction that they can quickly narrow down to the most accurate one. Furthermore, experts attack a problem using rules of thumb that allow them to find an approximately correct solution in a timely fashion. Non-experts do not have these rules of thumb to guide their problem solving and therefore take longer to reach the solution [13, 18]. Fifth, in practical settings, experts interpret case evidence based on their clinical experience rather than on the available scientific evidence alone. This allows them to be more flexible in responding to the individual patient’s characteristics. Non-experts do not have the same scope of experience and practical knowledge, and therefore they rely more heavily on available scientific evidence which may reduce their flexibility in dealing with unique or novel cases [19].

In the literature on the use of CPGs there is no mention of expertise as a factor in determining the use of guidelines in clinical practice. Ineffective dissemination of guidelines [3], the relevance of guidelines to the care of specific patients [6, 20, 21], and the substantive nature of guideline information [4, 22] are more frequently cited as factors related to the use of guidelines by physicians. Cognitive research on expertise may provide an important clue to explain why some physicians are more responsive to using guidelines than others. Cognitive studies [13, 14] have shown that experts solve problems in diagnosis and management based primarily on their own vast experience, and that unlike non-experts, they rarely rely on scientific studies to support their decisions, especially if these studies are contrary to their personal experience. Expertise may also play a role in the physician's ability to comprehend and use a guideline because of how guidelines are written. Guidelines are generally written by a team of experts in the medical area covered by the guideline. Since experts approach the guideline with a more highly organized knowledge base, as writers, they may inadvertently expect the reader to have the same knowledge base, and be able to make the inferences required to fully comprehend the guideline. A non-expert may not be able to correctly make these inferences, leading to errors or frustration [13, 23]. This evidence leads us to the hypotheses that (1) experts may avoid guidelines because they find that they do not add anything to their practice, and (2) non-experts, may use them inaccurately, if at all.

3 Evidence and Experience in Guideline Use

One of the ways that CPGs are being promoted is through the growing reliance on evidence-based medicine (considered as a fundamental basis for the development of CPGs). The general tenet of this approach is that medical practice can be optimized and standardized by integrating clinical expertise with the best available scientific evidence [2, 24]. The assumption here is that practitioners will incorporate the latest available evidence into their practice, and that this information is sufficient for improving clinical performance.

The availability of the evidence, however, may not be enough. Indeed, another factor likely affecting the use of CPGs, which is seldom mentioned, is the format of guideline presentation [1]. Research in cognition has demonstrated that various forms of presentation of information have different effects on cognition, where they serve the purpose of aiding the organization of problem solving activity [25] and facilitate the making of inferences [26]. In the medical domain, it has been documented that the way different forms of data presentation (e.g., different algorithmic representations) may lead to the making of different decisions [27, 28]. It becomes important then to investigate how the different modalities of information presentation affect the interpretation and use of CPGs. Guidelines are typically presented in text and algorithmic forms. Text guidelines consist of the description of a procedure together with supporting information for the procedure, usually in the form of scientific evidence from clinical studies. The algorithmic guidelines are presented as algorithms depicting the procedure to follow and the conditions of applicability of each procedure.

Our approach to research is based on the theoretical and methodological framework provided by cognitive science [29], which looks at the psychological processes underlying performance in a variety of tasks. Much of its emphasis is on providing a detailed characterization of the perceptual and cognitive processes that lead to observable behavior [30]. In particular, our approach has been to investigate how individuals comprehend information, solve problems, and make decisions in the fields of medicine. In general there are two main aims of this type of research: (a) to identify invariances in human cognitive processing [31]; that is, characteristics of cognition that remain stable across all people; and (b) to characterize individual variability in cognitive processing, most notably along the expertise continuum [32]. In any of these cases, the focus is on the detailed characterization of cognition, via the study of individuals, rather than on the study of typical characteristics of groups or populations. This research involves the collection and analysis of a few, but lengthy think-aloud protocols from physicians—sometimes a single subject [33], often 3 or 4 subjects [34]—while solving and explaining clinical problems. From such detailed studies, theories are constructed, often in the form of computer models, that are then put to the test by comparing human performance with that of the model [35, 36]. To the extent to which a match can be found between the model's output and human performance, there is support for the underlying model. The aim of such matching is not to draw generalizations to the population, but rather to characterize in detail the cognitive processes used when completing a task and to provide evidence for or against the cognitive processes hypothesized in the model.

Consistent with the goals of qualitative cognitive research, our aim in the studies reported here is to understand the impact of text and algorithmic guidelines on clinical decision making and problem solving by identifying the cognitive processes used in guideline-based decisions. Like most guidelines, these are directed at general practitioners, who are usually the first and frequently the principal medical contact point for most patients. We will investigate how people at different levels of expertise (general practitioners, internists, and endocrine specialists) solve problems without and with the aid of guidelines.

To pursue this aim, we describe two separate but related studies, both of which attempt to characterize the cognitive processes underlying guideline-supported clinical decision making by endocrine specialists, internists, and general practitioners. The two studies use a similar procedure, whereby the subjects are asked to evaluate a patient scenario both before and after having seen the guideline. The key difference between the studies is that one used a text-based guideline, whereas the other used an algorithmic guideline. We hypothesize that clinical guidelines in text form are difficult to use during practice because of the detailed processing required for written texts [37]. In contrast to written guidelines, algorithmic guidelines help organize the information and allow the problem-solver to proceed with an efficient search of the problem at hand. Theoretically, algorithmic representations of a clinical practice guideline could organize all of the relevant information into a manageable form, and therefore would aid in decision making.

4 Study 1: Investigation of Algorithmic Guidelines

In this study, we examined the use of a clinical guideline in algorithmic form by physicians at different levels of expertise. The study was conducted in two parts. In the first part the subjects were asked to read a problem scenario, solve the problem and provide recommendations for treatment without outside support, assistance, or manipulation from the experimenters (without making use of the guideline). This is called *spontaneous problem solving* (or free reasoning). This first part provides a baseline for comparison. The second part involved directing the subjects towards a specific solution through cued knowledge, using the clinical guideline. This is called *primed problem solving* (or primed reasoning). Generally, algorithms are generated from written guidelines. As a result, algorithms tend to be one step removed from the written text and do not provide evidentiary support for the recommendations. However, they are also easier to use at the time of care [38], given that they can be glanced over rapidly. Despite the fact that these algorithms are designed to be used together with the written guideline, practitioners often look for fast and easy ways of extracting guideline information, which algorithms provide. Specifically, physicians look for tables, charts, and algorithms to quickly assist their decision-making [38]. In this study we looked at how algorithms were represented and used in the management of a patient problem without the support of the text-based guideline.

4.1 Methods

4.1.1 Subjects

A cohort of six subjects, three general practitioners (GPs) and three endocrinologists (ENDOs), participated in this study. All three GPs were working in private practice in the city of Montreal, Canada, at the time of this study. Their clinical experience ranged between 6 and 18 years of practice after medical training. All three ENDOs were affiliated with McGill teaching hospitals, where they had regular part-time practices in the hospital setting and conducted research on various aspects of endocrine disorders.

4.1.2 Materials

An algorithmic representation of a clinical guideline for thyroid screening in women over 50 years old was used. The algorithm is part of the official guideline developed by the American College of Physicians-American Society of Internal Medicine (ACP-ASIM) and is based on the text form of the guideline [39]. The algorithm is available at the ACP-ASIM web site: http://www.acponline.org/ceap/algorithm.htm.

A patient scenario was developed in collaboration with the physician co-author of the paper. The scenario represents a typical screening case for thyroid disorder in women over 50 years old. It describes the case of a 52-year-old woman who had symptoms of excess hair loss, weight gain of 10 pounds, stress, and complained of fatigue. The symptoms in the problem are all representative of the thyroid dysfunction of *hypothyroidism*. The patient scenario is described below:

> A 52-year-old female patient consults with you with complaints of fatigue, weight gain of 10 lbs. and excess hair loss during the past three months. She got divorced recently and has been under stress. The patient also states that her mother is known to have an underactive thyroid gland. Thus, she considers that her complaints must be due to her gland. How would you investigate this case?

4.1.3 Procedure

Subjects were asked to read through the scenario of hypothyroidism and to think aloud while they went through it. They were then requested to make recommendations for patient management (without having seen the algorithm). Next, they were asked to read the algorithm for thyroid screening, to examine the patient scenario for a second time, and to update their decisions regarding the management of the patient, based on their reading of the algorithm. The subjects’ verbalizations were audiotaped and later transcribed for analysis.

4.1.4 Data Analysis

Each subjects' protocol was transcribed and analyzed using formal methods of discourse and protocol analysis [Patel, 1995 #2011; Patel, in press #5379]. The transcription was segmented into clauses where each of the concepts and the relations among the concepts used were identified. Procedural networks were then constructed by associating the concepts together into graphical structures that specify the steps in the procedure. Each graph consists of a non-empty set of nodes and a set of links connecting such nodes [40]. Nodes may represent clinical findings, hypotheses, or steps in a procedure, whereas links represent directed connections between nodes. Such networks convey two types of information: conceptual (i.e., the concepts used to solve a problem) and structural (i.e., how the concepts relate to one another), and thus provides a picture of the whole procedural representation. The reader is referred to other publications [Patel, 1995 #2011; Patel, in press #5379] for detailed explanation of our methods.

In addition, the subjects' responses were coded into the following three categories:

(1) Accuracy of diagnoses. An accurate diagnosis was coded only when the subject gave the correct diagnosis of hypothyroidism, otherwise it was considered inaccurate. We analyzed diagnostic accuracy as a way to provide validity to the scenario and to approach the natural conditions of physicians' practice.

(2) Number of tests requested. These were subdivided into those tests that were directly relevant and those that were irrelevant to thyroid dysfunction. Directly relevant tests were those which, based on the guideline and expert opinion, were deemed necessary for diagnosing hypothyroidism. Tests that were not deemed critical for hypothyroidism were coded as non-relevant tests.

(3) Steps skipped or added to the guideline. These were analyzed only for the primed phase and were coded whenever the physician introduced or failed to mention a step that appeared in the guideline.

4.2 Results

4.2.1 Spontaneous Problem Solving

Diagnostic Accuracy: Subjects were initially asked to evaluate the scenario prior to having seen the algorithm and to provide the diagnosis for the case. All subjects eventually generated the correct diagnosis of hypothyroidism. However, GPs entertained more hypotheses before giving the correct diagnosis. As presented in Table I, the GPs generated differential diagnostic hypotheses that were not directly related to hypothyroidism, namely depression and connective tissue disease. In contrast, the ENDOs generated single hypotheses, with one expert, who provided the specific nature of hypothyroidism, namely Hashimoto's thyroiditis. This demonstrates the intermediate effect, which shows that intermediates (non-specialists) often generate hypotheses that are not directly relevant to the disease.

===========

Insert Table I about here

===========

Information Requested: Without the guideline, GPs requested more information about the patient (they asked 4-11 questions) than did the experts (0-7 questions). Many of the GPs questions were not directly relevant to thyroid dysfunction (e.g., questions about depression or divorce). Furthermore, as seen in figure 1, GPs requested a greater number of tests (from 4-10) than the ENDOs did. Of these tests, only 45% were specific to thyroid dysfunction. Experts, in turn, asked for fewer tests (2-5), ninety percent of which were specific to thyroid dysfunction.

4.2.2 Primed Problem Solving

When the physicians were asked to evaluate the problem using the guideline, their performance differed markedly from using the problem scenario alone. GPs reduced the number of tests ordered, as compared with spontaneous problem solving (from an average of 6 to 2). Of the tests requested during primed problem solving, 95% were relevant to the correct diagnosis of thyroid disease, as can be observed in Figure 1. In contrast, there was no change in the number of tests requested by the ENDOs when compared with the spontaneous problem solving. In both cases, they requested an average of 4 tests.

With the help of the algorithm, the GPs followed the procedure outlined in the algorithm more closely than the ENDOs, adding on average of only one extra step to the guideline and skipping one step. In contrast, ENDOs added an average of seven steps and skipped an average of three steps. The modifications made by the ENDOs dealt with treatment, drug administration, dosage of medication, and patient follow-up.

==============

Insert Figure 1 about here

============

4.2.3 Analysis of two protocols

As an illustration of the problem solving by the GPs and the ENDOs, we present the semantic networks generated from two physicians' response protocols. Figure 2 presents the network generated from GP#1.

============

Insert Figure 2 about here

============

The figure shows that this physician did not focus on the underlying problem, he ordered tests not directly related to the disease, such as gynecological and breast exams. The information he requested which is related to the disease does not seem to be in any specific order or pattern. This suggests that without the algorithm, this physician had difficulty in organizing the information. This physician first reviewed general information about the patient, including a great deal of medical history. As much of this information was not directly related to thyroid disease (although nevertheless important for the physician to cover), this resulted in the GP ordering many more tests than necessary. In total the physician ordered seven tests. This number is much higher than the two tests recommended in the algorithm. This number is consistent with that of the other GPs, who on average ordered six tests.

When the algorithm was used, GP#1's reasoning became more focused and structured, and the tests followed the precise order outlined in the algorithm. Overall, GPs reduced the number of tests, to an average of 2.5 tests. Figure 3 presents the network derived from GP#1's primed protocol, which followed almost identically the procedure in the algorithm (two out of the three GPs were found to follow the algorithm almost exactly). This suggests that the algorithm influenced the knowledge used and deployed in the task. Also, with the use of the algorithm the GPs were able to recognize immediately the symptoms of a possible thyroid problem, whereas without the algorithm they considered many alternative hypotheses before reaching that conclusion.

===============

Insert figure 3 about here

==============

The semantic representation of one of the endocrine specialist shows a different pattern of results. When no algorithm was provided, the ENDOs generated parsimonious explanations and requested, on average, fewer tests (4.3), and followed the order prescribed in the algorithm (e.g., THS Elevated→FT4 Decreased→Treat), as illustrated in the semantic network representation generated from the protocol of ENDO#2, given in Figure 4.

==============

Insert Figure 4 about here

=============

In this case, the ENDO produced a very focused explanation of thyroid screening, ordering only relevant tests. Overall, the ENDOs' reasoning did not change very much with the addition of the algorithm, the core ideas and organization remained the same. However, with the use of the algorithm, the ENDOs were able to express more ideas as illustrated in Figure 5 (ENDO#2).

===============

Insert Figure 5 about here

===============

The algorithm can be seen as a reminder tool for the expert, whereas it is more useful for the GPs as a means to fill in gaps in knowledge and that as an aid in organizing their knowledge.

Compare for a moment, Figure 3 (GP primed protocol) to Figure 4 (ENDO spontaneous protocol). Notice that with the support of the algorithm the GP's protocol looks a lot more like an expert's, as compared to the GP's spontaneous problem solving, figure 2. However, there are still a few fundamental differences that the algorithm did not reduce. The GP still considered hyperthyroidism as a possibility, whereas the ENDO did not. Although this may suggest a more thorough review of the problem on the GP's part, in the rest of the protocol the GP seemed content to take things at face value, whereas the ENDO shows a lot more concern in checking before diagnosing. In both of these figures, it is shown that the subjects immediately requested a TSH test. However when the test was assumed to come out normal, the ENDO still wanted to look at the pituitary, whereas the GP limited himself to say that there was no thyroid problem. When the results of the test were assumed to be high, the GP recommended either treating the patient immediately or doing a T4 test. The ENDO, on the other hand, expressed the need to do a T3 test and a T4 test, and then test for antibodies, before diagnosing hypothyroidism and providing treatment. The GP would diagnose hypothyroidism when the T4 test is low, without any further testing.

These findings can be explained by taking a closer look at the algorithm. The GPs followed the algorithm almost exactly, including the consideration of hyperthyroidism. The extra tests the ENDO requested are not in the algorithm. Perhaps this shows the non-expert's inability to adapt the rigid guidelines to a specific patient. In this study we have a case where the experts do not even consider hyperthyroidism, and the non-expert's consider it simply because it is in the guideline.

4.3 Discussion

In summary, this study explored expertise-related differences in diagnostic reasoning between general practitioners and endocrine specialists when explaining a clinical case without and with the help of an algorithmic CPG. Examination of the findings show clear qualitative and quantitative differences between the GPs and ENDOs. Consistent with previous research [13, 14], the number of initial diagnostic hypotheses generated during spontaneous reasoning was higher for the GPs than for the ENDOs. Research has consistently demonstrated that whereas experts generate a few selected and related hypotheses to account for a clinical case, non-experts typically generate a larger number of either related, or unrelated hypotheses [13]. The GPs accounted for the case findings by linking them to specific hypotheses. They seemed to have followed a strategy of " when in doubt gather as much information as possible," and this resulted in the generation of relevant as well as irrelevant information. In contrast, the use of the algorithmic guideline appears to have had a large impact on the GPs' reasoning with the clinical scenario. The CPG helped them focus on the relevant information by reducing the number of tests requested, the number of questions asked, and the number of elements in the physical examination. This reduction decreases cognitive load on memory.

Based on our results and the preceding discussion, we can list some recommendations to improve acceptance of guidelines among physicians: Guidelines should be adapted to the user's depth of knowledge. This study showed that the ENDOs used the algorithm as a reminder tool, therefore they may only require a guideline that outlines the main steps in a diagnostic or treatment procedure. On the other hand, the GPs used the algorithm to fill in gaps in their knowledge and to help organize the information so they may require a guideline which provides a greater level of detail. Standardization could be fostered by matching the user's level of expertise and range of experience to the granularity of the guideline. To test this approach, one could explore which form of representation lends itself to quick use. A guideline may be represented differently to various users. However, more straightforward guidelines with less detail up front could be designed for all physicians. Details should be provided at various levels of granularity underlying the overt features such that guidelines are equally usable for generalists as well as for specialists. The best way to present a guideline with different levels of granularity may be to have an electronic guideline which is integrated into the EMR (electronic medical record), this way the physician can choose to look at as little or as much detail as needed. Also, as "reminders" may be necessary during problem solving, these additional details can be accessed when needed. Although in this study all subjects were able to accurately assess the clinical scenario with the use of the algorithm, another way to ensure that details are provided when needed would be to ensure that a text is always provided with an algorithm. In the next study we present findings which indicate that it is beneficial to provide an algorithm with a text-based guideline because of the time constraints that physicians face when dealing with a patient.

5 Study 2: Investigation of Text-based Guidelines

This study examines the effect of a text-based guideline on the clinical reasoning of practitioners at two levels of expertise. As with the previous study, we focus on the understanding of the guideline, as well as on the reasoning strategies employed in making decisions about a patient both with and without the algorithm. The goal of priming in this study was to investigate if and how the guideline affects the subject's reasoning and problem solving. Since the amount and quality of a physician's experience seem to play an important role in clinical performance, we also investigate the effect of the guideline on reasoning processes as a function of expertise. Considering the reported differences in the reasoning patterns of experts and non-experts, it cannot be expected that guidelines will have the same effect on physicians at various kinds and levels of expertise. Therefore the consideration of expertise is a very important facet of the study.

5.1 Methods

5.1.1 Subjects

The sample consisted of six subjects. Subjects selected were personal acquaintances of the senior author, where the recruitment paid particular attention to a balance of domain specialists, internists and general practitioners. No particular incentives were provided for participation, except for the challenge of the task. The subjects were two general practitioners (GPs), two internists, and two endocrinologists (ENDOs). The endocrinologists were experts (specialists)in the area of diabetes. The GPs worked in a university clinic and they were 35-40 years of age. The internists and the endocrinologists worked in a university hospital. The internists were 45-50 years of age, and the endocrinologists were 50-55 years of age. The GPs had the least number of years of experience, while the endocrinologists had the most.

5.1.2 Materials

The materials for the study involved the clinical guideline entitled the "1998 Clinical practice guidelines for the management of diabetes in Canada" and a patient scenario, which presented a short clinical case problem. The CPG was developed by the Canadian Medical Association, and it is available at http://www.cma.ca/cmaj/vol-159/issue-8/diabetescpg/index.htm. We used a guideline for the management of *Diabetes Mellitus* for several reasons. First, diabetes mellitus is a condition that virtually all practitioners have seen, yet a much higher level of knowledge is required of specialists treating this condition. Second, the guideline was of recent publication and therefore offered new and updated management recommendations. Third, this guideline is widely available to physicians in Canada. Finally, its text-based form is relatively short and therefore suitable for research purposes (i.e., it takes relatively little time to go through).

The short patient scenario, involving the description of a patient with diabetes mellitus and risk of cardiovascular disease, was developed by the physician co-author, and was verified by two other specialists in the area of diabetes. This scenario was based on actual clinical cases. The scenario is as follows:

> A 46-year-old male executive requests a complete physical examination to determine his health status and, in particular, his risk of cardiovascular disease. He is overweight, has a family history of type 2 diabetes, smokes a pack of cigarettes a day and complains of fatigue.

5.1.3 Procedure

The materials were presented in three parts. The first part involved spontaneous, *free problem solving*. The subject was given only the patient case scenario and was asked to explain the case, assess the risk for diabetes and cardiovascular disease, and provide a detailed treatment and management plan. The second part involved *primed problem solving.* The subjects were asked to revise their evaluation of the patient, including the management plan, based on the information presented in the guideline. The third part involved presenting the subjects with a set of specific recommendations, which are part of the original CPG. The subjects were asked to integrate these recommendations into their evaluation and management plan. Finally, they were given a 5-point Likert scale (1 - not useful to 5 - very useful), which asked them to rate the extent to which they found the guideline useful.

All subjects were asked to think-aloud while explaining their evaluation of the problem. Subjects were permitted to speak for as long as they felt necessary and when they stopped verbalizing the researcher prompted them to continue [41]. Their explanations were audiotaped. All subjects' protocols were later transcribed verbatim for analysis.

5.1.4 Data Analysis

As in Study 1, the transcripts of the generated verbal responses were analyzed using the techniques of discourse and protocol analysis [Patel, in press #5379]. The data represented in the networks were analyzed for (1) the nature and types of concepts and relationships generated by the subjects while explaining the clinical problem (spontaneous problem solving); (2) the nature of their reasoning including the hypotheses generated for making decisions, including procedures in the guideline; and (3) the updating of their explanations of the patient problem after the guideline was introduced (primed problem solving). This allowed for the identification and comparison of the procedures specified in the guideline with those generated by the subjects.

5.2 Results

5.2.1 Spontaneous Problem Solving

During spontaneous problem solving, we found that there was an increase in the number of concepts generated (requests and procedural steps) as a function of expertise. The GPs generated an average of 19.5 concepts, while the Internists generated a mean of 33 concepts and the ENDOs, a mean of 50 concepts. A qualitative evaluation of these concepts indicates that GPs and Internists generated generic information about the patient condition. This is in contrast to the experts who generated far more specific information. For example, whereas the GPs recommended "weight loss," the ENDOs recommend checking for "weight distribution" and "waist circumference," before any dietary changes are implemented. In the guideline, data about the age of the patient (45 years of age)" is identified as the age at which all people should begin periodic screening for diabetes. Although all the subjects suggested screening for diabetes, only the ENDOs considered the age of the patient as a clinically significant factor.

In terms of diagnostic hypotheses, we observed that the internists generated more diverse hypotheses than either the ENDOs or the GPs. Internists generated an average of five hypotheses, whereas the GPs generated two and the both ENDOs generated one hypothesis. If we consider the internists as intermediates between the GPs and the ENDOs, this finding can be interpreted as an instance of the "intermediate effect." This refers to an apparent decrement in performance by intermediates when compared to novices and experts [42]. Such an effect has been observed in many studies of expertise [19, 43, 44]. This decline in performance occurs because intermediates (in this case, the Internists) typically possess a great deal of information, but this information is not yet organized in an efficient way (as compared to the domain expert). The consequence of this is that when solving a given problem, many relevant and irrelevant hypotheses are activated in long-term memory, leading to difficulty in separating the two types of information.

5.2.2 Primed Problem Solving

The effect of the guideline on reasoning patterns varied across subjects. First, one internist used the guideline to a greater extent than the other subjects. For this subject, the guideline was effective in focusing his attention on the recommended procedures. Second, for both the internists and the GPs, the guideline served mainly as a reminder for procedures that were not readily accessible during spontaneous problem solving. The guideline allowed them to state the specific tests that they would request, and in particular, it appears that they used the guideline as a reference for determining diagnostic criteria and blood sugar levels. Figure 6 shows the network of a GP, which is characteristic of the rest of the GPs and internists.

=============

Insert figure 6 about here

============

Third, neither the GPs nor the internists recommended doing anything differently than what was advised in the guideline, although both internists said that they would not necessarily follow the guideline word for word. Fourth, only the endocrinologists gave reasons for not following some of the guideline recommendations, although for the most part they were in agreement with the guideline advice. For example, one ENDO disregarded a guideline procedure that recommended "foot examinations annually" on the basis that, in his experience, they are not typically done because of the time they take to perform the procedure. This subject also commented that he was typically more aggressive in screening and prescribing oral anti-hyperglycemic medications, and that he did it earlier than recommended by the guideline, because in his experience, this was more effective in the management of the patient. Overall, the guideline provided little benefit to this expert. However, for another ENDO, it was used as a reminding device and as an organization tool. This is illustrated on Figure 7.

==============

Insert Figure 7 about here

=============

5.2.3 Physicians' Evaluation of the Guideline

Despite the potential usefulness of the guideline, if only as a reminder tool, most subjects found that it failed to meet their need of timely delivery. Both GPs commented that they would read the entire guideline, although this would be dependent on time constraints. One GP declared that he would make decisions about patient management for this patient based on his experience and common sense. He further stated that he would go home to read the guideline when he had the time. Both internists remarked that they would not read the guideline entirely, but instead would look for a chart, or an algorithm that could give them quick and easy access to the information. Both the GPs and the internists perceived guidelines to be useful as memory aids, but said that typically guidelines reflect ideal patients and are difficult to use effectively with specific patients. On their part, the ENDOs said that they found the guideline to be very useful although neither one felt that it adds much to their clinical practice, one ENDO said that he would skip or change a recommendation if his clinical experience indicated that this would be a reasonable step.

When the subjects were asked to rate the usefulness of the guidelines, one GP and both endocrinologists found them to be very useful (gave a rating of 5); one internist consider it useful (gave a rating of 4) and the other was unsure about its usefulness (i.e., gave a rating of 3). One GP did not rate the guideline usefulness, but he comments that unless he had plenty of time to read over the guideline it could not be used with a given patient. However, he did find the tables and recommendations useful. It is interesting to note, however, that the usefulness of the guideline that we have observed did not necessarily reflect the ratings given by the subjects. For example, an internist— who made great use of the guideline and whose reasoning became more focused during primed problem solving—gave the guideline a rating of 3, the lowest rating given. In contrast, an endocrinologist who did not appear to make any use out of the guideline gave it the highest possible rating.

5.3 Discussion

In summary, the study presented an investigation of the problem solving process of a clinical case by general practitioners, internists, and endocrinologists during both spontaneous and primed problem solving. The results during spontaneous problem solving show that all subjects seemed to rely on their experience. However, GPs and internists appeared to be less precise in their explanation of the problem. They tended to give use generic terms (such as "blood-work") when no guideline was given and more specific terms (such as "fasting glucose") when the guideline was provided. This may reflect a lack of knowledge on the part of the physician or it may emphasize the role the guideline plays as a reminding tool. Did the physician not know which test to order or could he not remember the name of the test? The ENDOs, in contrast, were more precise in their explanation during spontaneous problem solving. They gave specific names to the procedures or management steps needed for problem solution.

During primed problem solving, the subjects differed more markedly. Whereas the GPs attempted to follow the CPG very closely, the internists were willing to depart from it. One

ENDO did depart from the guideline whenever he felt necessary and added his own reasons to follow a different path. Reasons for both groups to depart from the recommendations were all based on practical clinical experience. However, only in the ENDO's protocol did experiential knowledge take precedence over what should be done based on the recommendations, the internists were still willing to put trust in the guideline and change their views accordingly. In this study the guideline did not appear to change how experts approached a clinical case. However, it did appear to be useful as a reminder tool. The GPs and internists tended to use and see the guideline as a knowledge source and a memory tool.

The results can be explained in light of prior research. In actual practice, research suggests that physicians do not use CPGs at point of care because these are very time consuming to go through and unsuitable for just-in-time use [Shahar, 1998 #4529]. When they are used, it is probably when the physician has time to review the case, before or after patient interviews. When they are used they can have a beneficial effect on decision making by acting as reminders of procedures that may have been forgotten or ignored, despite the physicians' evaluation on the lack of usefulness of the CPG. However, the fact that typically guidelines are time-consuming to read, make the development of new forms of guidelines that are readily accessible a necessity. Current CPGs impose limits to when and how guidelines are read, but the inclusion of CPGs as part of decision-support systems or instructional systems may be a way to overcome such limitations [45, 46].

The subjects of this study expressed a desire to see guidelines that provide faster access to pertinent information. Many guidelines are accompanied by algorithms, which represent practice recommendations in a diagrammatic form. This makes them easier to read in a short period of time and extract prescribed recommendations at once [38]. A potential drawback is that because of the condensed nature of the algorithms, they are often rigid [47] and cannot provide all the information present in text-based guidelines, which may be needed by non-expert physicians. Also, they may be seen as ambiguous because there is no room for explanation of counter-intuitive advice. There seem to be promising attempts to remedy this situation [48].

6 General Discussion and Implications

Our studies suggest that practice guidelines, either in text or diagrammatic form present some problems for both general practitioners and experts alike. Guidelines seem to serve different functions for the two groups of people, depending on the prior knowledge. Reliance on experience makes expert physicians more confident about their decisions and not easily affected by the use of CPGs. Expert decision-making and behavior is difficult to change because experts have established ways of approaching and solving problems which, in their experience, have been successful in the past. The finding that experts performance did not change after reviewing the case with the guideline may be t he result of their reliance on what they know. They have developed explanations of how things should be done, which have been validated over a long period of practice. In fact, expert practitioners tend to rely on a heuristic of *explanatory sufficiency*. That is, they work through the solution of a problem until a good-enough solution is found, rather than attempting to reach a "perfect" solution. Experts may generate explanations for a patient problem only to the extent to which they are satisfactory ("good enough" explanations). In other words, experts are "satisficers" not "maximizers." However, this requires possessing knowledge of when a satisfactory solution has been achieved, something that probably requires a great deal of experience. This characteristic of experts was first identified by Herbert Simon in his study of decision making in organizations [49]. In general, decision-makers do not try to provide a perfect solution to a problem or maximize personal gains in making decisions, but rather they typically settle for "good-enough" solutions only. In medicine, the same phenomenon occurs, when physicians generate explanations that satisfactory and not necessarily optimal. They will generate an explanation that matches what they habitually know from practice. However, it is this reliance on experience that contributes to the addition and skipping of steps. This may sometimes lead to errors of omission [16].

One goal of clinical guidelines is to standardize clinical practice. This would result in the non-expert performing in a similar way to the expert. However, to reach the same level, non-expert would need a different type of support from guidelines than the type of support that an expert would find useful. This requires adaptation of guidelines to the specific level of expertise of the user. Guideline delivery forms that are dynamically generated and adaptable to the specific expertise level of the user may be more effective. Neither text-based nor diagrammatic forms alone could serve this adaptive function. For clinical practice guidelines to be successfully used, either as decision making or as educational tools, they need to be included in decision support systems (e.g., EMR) or instructional systems that help focus the non-expert physician on relevant information or remind the expert physician of important steps. In this way, patients could receive the best care in a timely manner. Newly acquired knowledge must be fully integrated into existing knowledge structures before changes in practice habits can be seen. The problem is that whenever new information is introduced into a stable system, it often creates confusion, because integration involves more than concatenating additions to previous knowledge. Integration of knowledge requires a change in the conceptual structure of the (knowledge) system. This re-structuring requires time and cognitive effort, as well as an environment in which to learn—all of which are currently lacking in the lifestyle of the medical practitioner. It would be an interesting question to ask under what conditions physicians learn from using guidelines. Since GPs are required to take continuing medical education courses, guidelines can be turned into useful continuing education tools that remain useful during their clinical practice. Similarly, with residents and medical students, guidelines can be incorporated in their training program as a part of the electronic medical record or decision support systems.

Acknowledgement

The studies reported in this paper were supported in part by funds from the American College of Physicians-American Society of Internal Medicine (Education and Scientific Policy), and from the National Library of Medicine through “A Shared Internet Server for Delivering Guidelines” (PI: Edward. Shortliffe, Columbia University), subcontract to Vimla Patel. (McGill University). We greatly thank Jo-Ann Wilkins and Sarah Kim for their efforts in collection and analysis of the data related to studies one and two, respectively. Suggestions and critical comments from Herbert Waxman (ACP-ASIM) and Edward Shortliffe (Columbia) on the earlier versions of the paper are graciously acknowledged

7 References

[1] Wyatt J. Use and sources of medical knowledge. Lancet. 1991;338.

[2] Sackett DL, Rosenburg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. British Medical Journal. 1996;312:71-72.

[3] Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluations. Lancet. 1993;342:1317-1322.

[4] McAlister FA, Teo KK, Lewanczuk RZ, Wells G, Montague TJ. Contemporary practice patterns on the management of newly diagnosed hypertension. Canadian Medical Association Journal. 1997;157:23-30.

[5] Lomas J, Anderson GM, Dominik-Pierre K, Vayda E, Enkin MW, Hannah W. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. The New England Journal of Medicine. 1989;321:1306-1311.

[6] Gupta L, Ward JE, Hayward RSA. Clinical practice guidelines in general practice: A national survey of recall, attitudes and impact. Medical Journal of Australia. 1997;166:69-72.

[7] Tunis SR, Hayward RSA, Wilson MC, Rubin HR, Bass EB, Johnston M, Steinburg EP. Internists' attitudes about clinical practice guidelines. Annals of Internal Medicine. 1994;120:956-963.

[8] Weingarten S. Translating Practice Guidelines Into Patient Care : Guidelines at the Bedside. Chest. 2000;118:4S-7S.

[9] Ericsson YA, Smith J. Toward a general theory of expertise: Prospects and limits. . New York: Cambridge University Press; 1991.

[10] Feltovich PJ, Ford KM, Hoffman RR. Expertise in context. Menlo Park, CA: AAAI/MIT Press; 1997.

[11] Chase WG, Simon HA. Perception in chess. Cognitive Psychology. 1973;4:55-81.

[12] Ericsson A, Smith J. Prospects and limits of the empirical study of expertise: An introduction. In: Ericsson A, Smith J, eds. Toward a general theory of expertise: Prospects and limits. New York, NY: Cambridge University Press; 1991:1-50.

[13] Patel VL, Arocha JF, Kaufman DR. Diagnostic reasoning and medical expertise. The Psychology of Learning and Motivation: Advances in Research and Theory. 1994;31:187-252.

[14] Patel VL, Groen GJ, Patel YC. Cognitive aspects of clinical performance during patient workup: The role of medical expertise. Advances in Health Sciences Education. 1997;2:95-114.

[15] Chi MTH, Glaser R, Farr MJ. The nature of expertise. . Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

[16] Patel VL, Ramoni M. Cognitive models of directional inference in expert medical reasoning. In: Feltovich P, Ford K, Hoffman R, eds. Expertise in Context. Menlo Park, CA: AAAI Press/MIT Press; 1997:67-99.

[17] Leprohon J, Patel VL. Decision making strategies for telephone triage in emergency medical services. Medical Decision Making. 1995;15:240-253.

[18] Joseph G-M, Patel VL. Domain knowledge and hypothesis generation in diagnostic reasoning. Medical Decision Making. 1990;10:31-46.

[19] Lesgold A, Rubinson H, Feltovich P, Glaser R, Klopfer D, Wang Y. Expertise in a complex skill: Diagnosing x-ray pictures. In: Chi MTH, Glaser R, Farr MJ, eds. The nature of expertise. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1988:311-342.

[20] Hoey J. The one and only Mrs. Jones. Canadian Medical Association Journal. 1998;159:241-242.

[21] Miettinen OS. Evidence in medicine: invited commentary. Canadian Medical Association Journal. 1998;158:215-221.

[22] Kenny NP. Does good science make good medicine? Incorporating evidence into practice is complicated by the fact that clinical practice is as much an art as science. Canadian Medical Association Journal. 1997;157:33-36.

[23] Arocha JF, Patel VL, Patel YC. Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. Medical Decision Making. 1993;13:198-211.

[24] Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. British Medical Journal. 1995;310:1085-1086.

[25] Koedinger KR, Anderson JR. Abstract planning and perceptual chunks. Cognit Sci. 1992;14:511-550.

[26] Larkin JH, Simon HA. Why a diagram is (sometimes) worth ten thousand words. Cognit Sci. 1987;11:65-99.

[27] Elting L, Martin C, SB. C, EB. R. Influence of data display on physician

investigators' decisions to stop trials: prospective trial with repeated measures. British Medical Journal. 1998;317:1527-1531.

[28] Cole WG, Stewart JG. Human performance evaluation of a metaphor graphic display for respiratory data. Methods of Information in Medicine. 1994;33:390-396.

[29] Hunt E. Cognitive science: Definition, status, and questions. Annual Review of Psychology. 1989;40:603-629.

[30] Kluwe RH. Single case studies and models of complex problem solving. In: Funke J, Frensch P, eds. Complex problem solving. Hillsdale, NJ: Lawrence Erlbaum Associates; 1995:269-294.

[31] Simon HA. Invariants of human behavior. Annual Review of Psychology. 1990;41:1-19.

[32] Glaser R. On the nature of expertise. In: Klix F, Hagendor H, eds. Human memory and cognitive capabilities: Mechanisms and performances. Amsterdam: Elsevier Science; 1986:915-928.

[33] Kuipers BJ, Kassirer JP. Causal reasoning in medicine: Analysis of a protocol. Cognit Sci. 1984;8:363-385.

[34] Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task difficulty. Memory & Cognition. 1990;18:394-406.

[35] Richman HB, Gobet F, Staszewski JJ, Simon HA. Perceptual and memory processes in the acquisition of expert performance: The EPAM model. In: Ericsson A, Smith J, eds. Toward a general theory of expertise: Prospects and limits. New York, NY: Cambridge University Press; 1996:167-187.

[36] Richman HB, Simon HA. Context effects in letter perception: Comparison of two theories. *Psychological Review*. 1989;96:417-432.

[37] Medin DL, Ross BH. The specific character of abstract thought: Categorization, problem solving, and induction. In: Sternberg RJ, ed. Advances in the psychology of human intelligence. Hillsdale, NJ: Lawrence Erlbaum Associates; 1989:189-223.

[38] Hadorn DC. Use of algorithms in clinical guideline development in clinical guideline development: Methodology perspectives. AHCPR Pub. No. 95-0009. Rockville, MD: Agency for Health Care Policy and Research; 1995:93-104.

[39] Hefland M, Redfern C. Screening for thyroid disease. Annals of Internal Medicine. 1998;129:144-158.

[40] Sowa JF. Conceptual structures: Information processes in mind and machine.

Reading, MA: Addison-Wesley; 1983.

[41] Ericsson KA, Simon HA. Protocol Analysis: Verbal Reports as Data. Revised Edition.: Cambridge, MA: MIT press; 1993.

[42] Groen GJ, Patel VL. Medical problem-solving: Some questionable assumptions. Medical Education. 1985;19:95-100.

[43] Arocha JF, Patel VL. Novice diagnostic reasoning in medicine: Accounting for clinical evidence. Journal of the Learning Sciences. 1995;4:355-384.

[44] Schmidt HG, Boshuizen HP. On the origin of intermediate effects in clinical case recall. Memory & Cognition. 1993;21:338-351.

[45] Sanders GD, Nease RF, Jr., Owens DK. Design and pilot evaluation of a system to develop computer-based site-specific practice guidelines from decision models. Medical Decision Making. 2000;20:145-59.

[46] Dayton CS, Ferguson JS, Hornick DB, Peterson MW. Evaluation of an Internet-based decision-support system for applying the ATS/CDC guidelines for tuberculosis preventive therapy. Medical Decision Making. 2000;20:1-6.

[47] Kassirer JP, Kopelman RI. Diagnosis and decisions by algorithms. . Hospital Practice (Office Edition); 1990:23-4, 27, 31.

[48] Owens DK, Shachter RD, Nease RF, Jr. Representation and analysis of medical decision problems with influence diagrams. Medical Decision Making. 1997;17:241-62.

[49] Simon HA. A behavioral model of rational choice. Quarterly Journal of Economics. 1955;69:99-118.

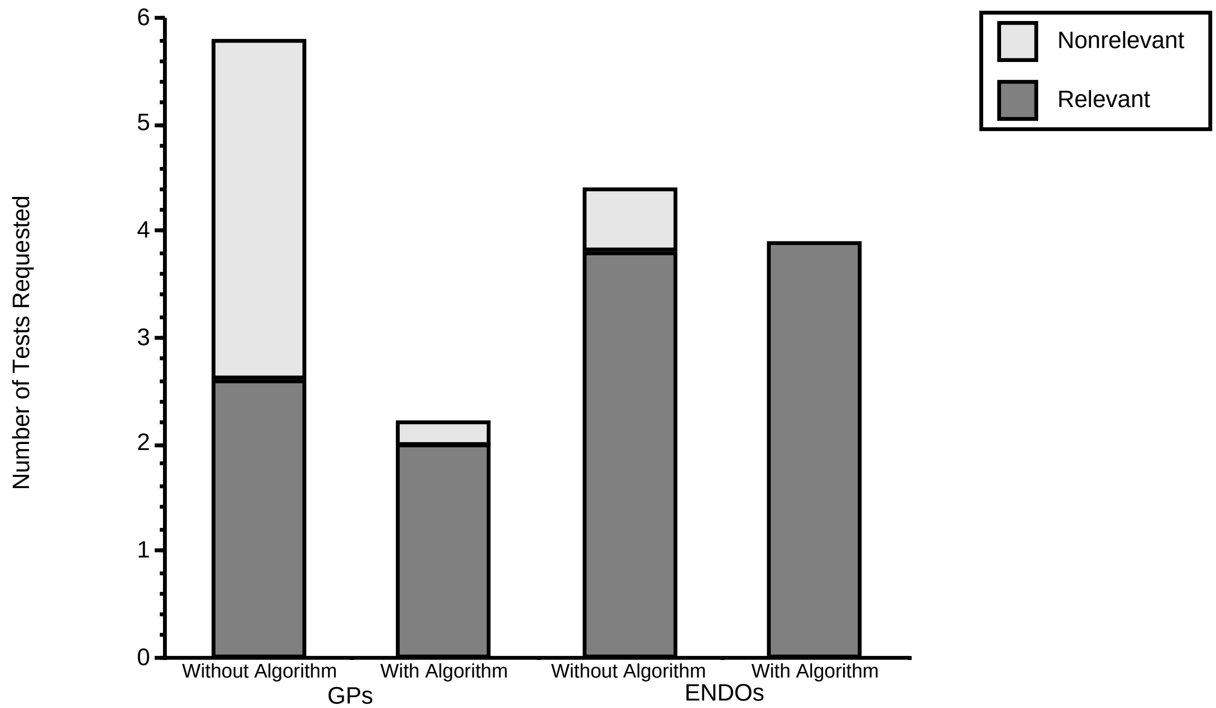

Figure 1. Total number of tests (critical and less relevant) requested by general practitioners (GPs) and endocrinologists (ENDOs) during spontaneous and primed problem solving situations. Critical tests are defined as those, which were deemed necessary for diagnosing hypothyroidism.

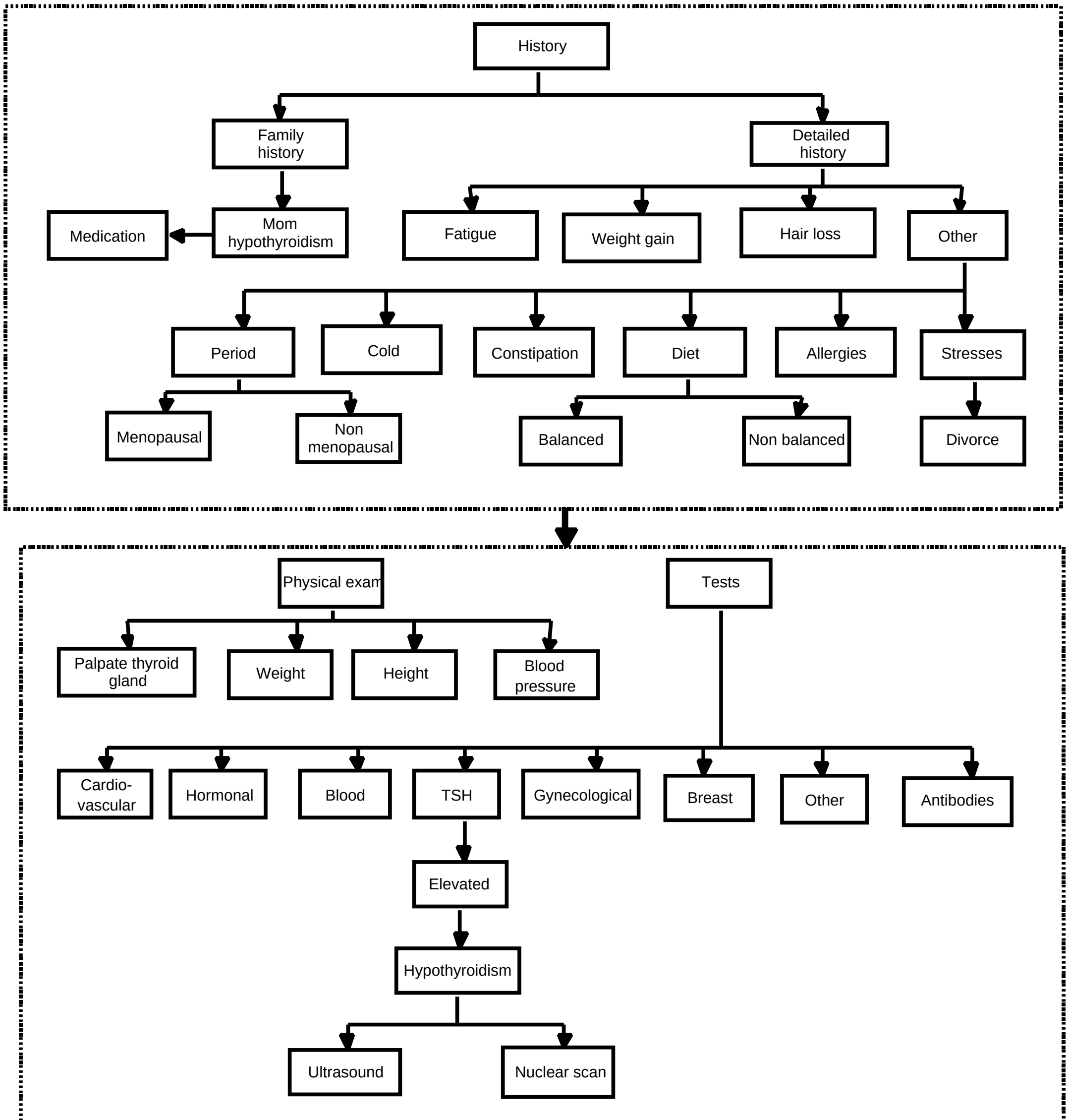

Figure 2. Schematic representation of a representative protocol of a general practitioner's (GP#1) reasoning during spontaneous problem solving. The top portion presents the history information and the bottom portion gives the physical examination and laboratory tests requested. Nodes represent concepts and arrows represent the relationships among concepts.

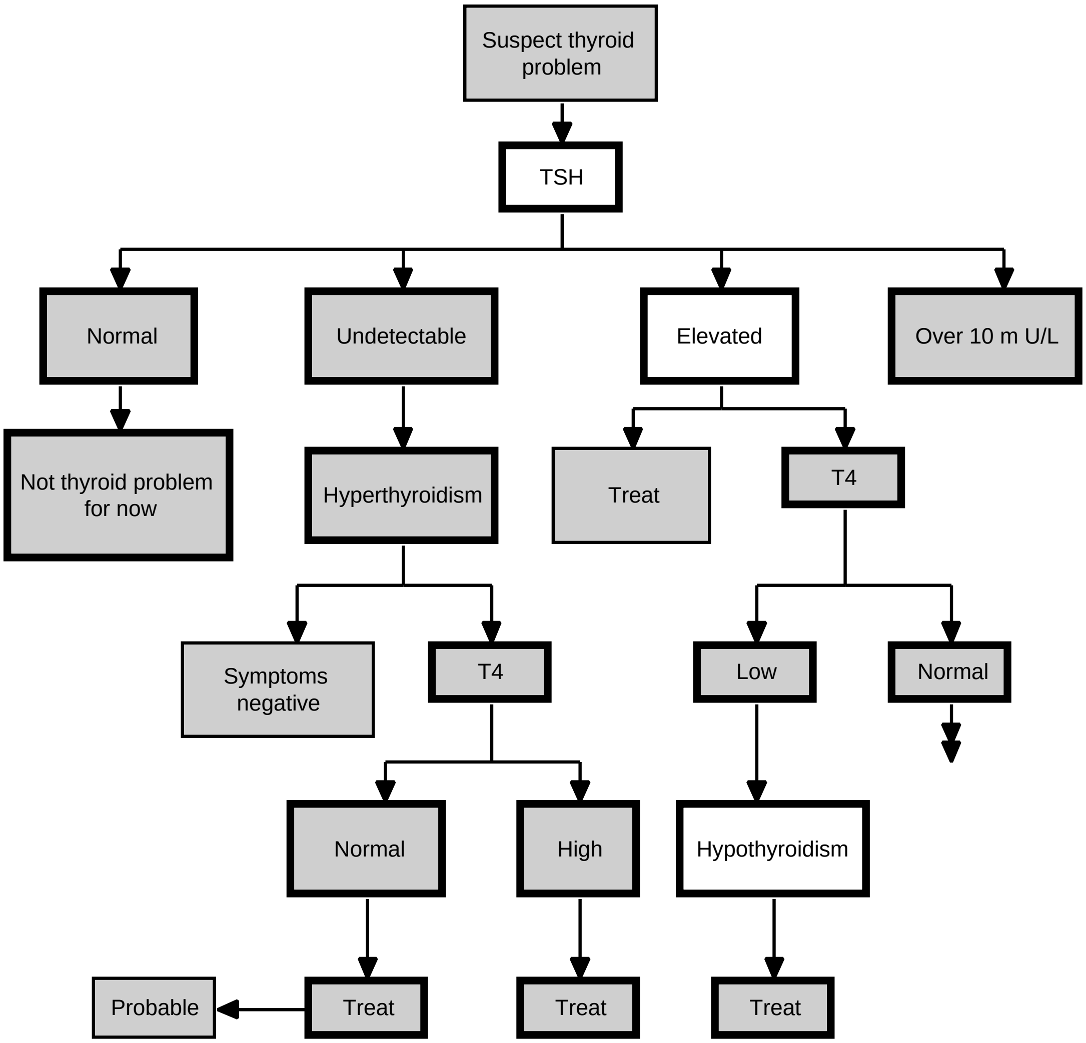

Figure 3. Schematic representation of a protocol of a general practitioner's (GP#1) reasoning when revising the clinical scenario of a patient with the help of the algorithm during primed problem solving. Boxes with heavy border indicate information that appears in the algorithm. Grey boxes indicate new information generated in the primed protocol that was not in the spontaneous protocol. Double arrows represent further reasoning by the subject not included in the figure.

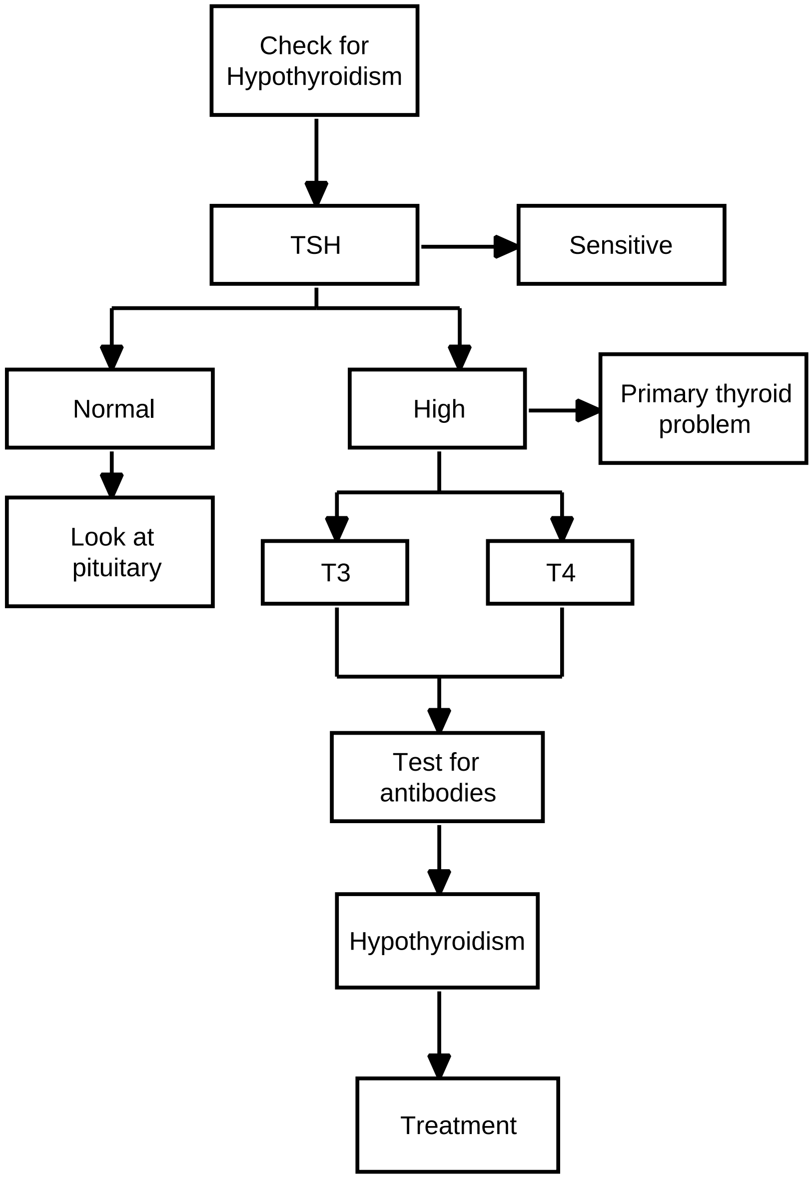

Figure 4. Schematic representation of a protocol by an endocrinologist (ENDO#2) during spontaneous problem solving.

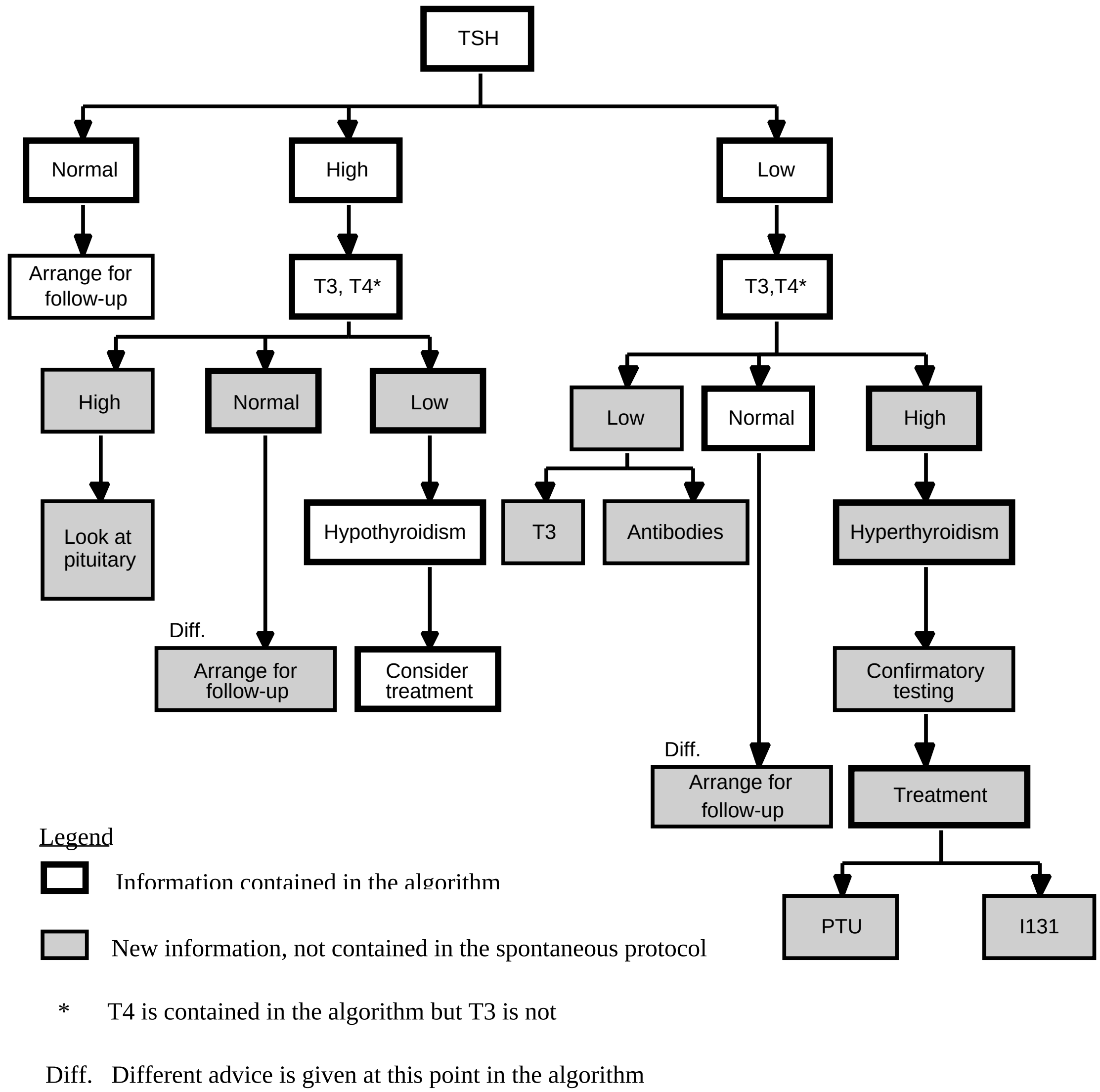

Figure 5. Schematic representation of a protocol of an endocrinologist (ENDO#2) reasoning when revising the clinical scenario of a patient with the help of the algorithm during primed problem solving.