Executive Summary

Research on Clinical Practice Guidelines: Theory, Methods, and Empirical Investigation

Vimla L. Patel, Ph.D., D.Sc.
Cognitive Studies in Medicine:
Centre for Medical Education
McGill University
Montreal, QC

A

Research Team

José F. Arocha, PhD
Melissa Diermier, BSc
André Kushniruk, PhD
Laurie Barkun, BA
Tim Branch, BSc
Sarah Kim, BSc
Jo-Ann Wilkins, BSc

Acknowledgments

These studies were supported in part by funds from the National Library of Medicine through the Shared Internet Server for Delivering Guidelines (Columbia University, Edward Shortliffe, PI) subcontract to Vimla L. Patel, McGill University, American College of Physicians-American Society of Internal Medicine, and the Social Science and Humanities Council of Canada to Vimla Patel.

Introduction

The overall purpose of this project is to investigate the process of clinical practice guideline (CPG) development and use, with the aim of improving the design and utilization of information in the guidelines. To this end, we present a scientifically-based framework and methods for the evaluation of guideline-based decision making, including empirical studies in clinical and non-clinical medical settings.

Clinical practice guidelines (CPGs) are aimed at physicians with various degrees of knowledge and experience. The expected result of the use of guidelines is the standardization of clinical practice. However, CPGs can be semantically complex, leading to variation in interpretation In fact, many CPGs are composed of elaborate collections of prescribed procedures, sometimes involving the embedding of procedures within procedures or complicated temporal or causal relations among the various steps in a procedure. Furthermore, CPGs are often supported by evidence that may or may not match the way physicians support their clinical decisions. Understanding the way physicians use evidence will provide important insights that serve to improve the design of CPGs so that they can be more widely accepted and safely used.

This report presents some formal cognitive methods of evaluating clinical practice guidelines and supporting documents. These include, a) methods for analyzing the nature of evidence, its interpretation and utilization and b) formal methods for natural language representation, namely propositional and semantic network analysis (analysis of ideas and concepts and their inter-relationships, respectively), for investigating the conceptual complexity in the understanding of CPGs. Using these methods, it is possible to evaluate various features of the guidelines: (a) understandability, (b) coherence, (c) ambiguity, and (d) complexity. These indicators can be used to assess the quality of the guidelines. In this report, we begin with the outline of the main features of the methods with illustrative examples, followed by results of some empirical studies on guideline use in practice. The document presents six parts dealing with various aspects of guideline evaluation, as follows:

First, we describe our cognitive framework for the study of the nature and utilization of evidence in clinical settings. Specifically, we present (a) a general model of evidence interpretation and utilization: (b) an experimental paradigm to study how individuals evaluate evidence; (c) a coding scheme based on a theoretical model for the classification of the nature of evidence and the strategies people use when evaluating information.

Second, we provide a description of formal methods of analysis of representation of CPGs, based on semantic (meaningful) and pragmatic (practical and contextual) analysis of textual information. These methods are propositional analysis (i.e., analysis of ideas and concepts) and semantic network analysis (i.e., analysis of meaningful relationships between these ideas and concepts).

Third, we present an analysis of a CPG and its representation in a computer language formalism, GLIF (<u>G</u>uide<u>L</u>ine <u>I</u>nterchange <u>F</u>ormat), using methods of propositional and semantic analyses.

The guideline used is "Screening for thyroid disease[1]," developed and published through the ACP-ASIM. In addition, we also illustrate the use of these methods in the analysis of a second CPG, "Pharmacological treatment of acute major depression and dysthymia."

Fourth, we report on two studies of clinical decision making with and without the use of CPGs. In this study, physicians with different levels of expertise (primary care physicians and specialists) solve clinical patient problems, with and without the use of guidelines, which were designed for managing diabetes [2] (study 1) and for screening thyroid disease (study 2).The processing of written and algorithmic versions of the guidelines for making decisions was also investigated.

Fifth, we give a description and preliminary analysis of studies of guideline usability at the "point of care." by electronically tracking when and why physicians use a set of provided guidelines during their clinical practice.

Finally, we give a brief report on our current investigation on reasoning about depression by lay-people (including patients) and psychiatrists, followed by recommendations for guideline evaluation in terms of selection, design and use.

Given the aim of standardizing clinical practice and reducing variability, CPGs should be designed to optimize the understanding of the intended meaning of the guideline and to minimize the possibility of making incorrect inferences. The methods of cognitive evaluation presented in this document provide a formal way of identifying and characterizing the nature of any inferences that are generated (including incorrect ones). Problems can arise because of faulty interpretation of evidence or its utilization to make decisions, because of the semantic and pragmatic complexity of CPGs in either written or algorithmic formats, or because of difficulties in the translation of written guidelines into graphical or to computer-ready formats. A summary of each of the six studies is outlined below. Detailed descriptions supporting these summaries are given as separate documents.

1. The nature of evidence in clinical practice guidelines: theory, coding scheme and empirical investigation

This section presents a cognitive science framework and coding scheme for the investigation of the nature of evidence and its utilization during decision making in clinical and non clinical medical settings (Patel & Arocha, 2000). More specifically, the framework and the coding scheme focus on the nature of evidence in the use of clinical practice guidelines (CPGs). The specific coding scheme is illustrated with examples from the history of science and from use of evidence in clinical settings such as primary care, intensive care, and emergency medicine. We

1 http://www.acponline.org/journals/annals/15jul98/ppthyroid1.htm

2 http://www.cma.ca/cmaj/vol-159/issue-8/diabetescpg/index.htm

present (a) a general model of evidence interpretation and utilization; (b) an experimental paradigm to study how individuals evaluate evidence; (c) a coding scheme based on a theoretical model for the classification of the nature of evidence and the strategies people use when evaluating data; and (d) an analysis of the evidence used in two background papers written to support CPGs. This analysis was carried out to identify the types of evidence that are used to support the recommendations made in CPGs. The approaches to evidence classification, interpretation and use described in this document lead to a number of recommendations for following procedures in guideline selection and endorsement, as well as for general principles for the design of clinical practice guidelines. Data from several clinical settings are selected as illustrating the use of coding scheme.

Based on the theory and the empirical investigations we present the following recommendations:

- When interpreting evidence, goals of the guidelines should be an important consideration in their design, as evidence may be interpreted differently depending on the purpose of the guideline (e.g., screening and diagnosis). By clearly establishing the goal of the guideline, evidence can be chosen that best supports such goal.
- Although typically we interpret a procedure in terms of steps to be performed (i.e., what to do), it's important to realize that actions are likely to change from one situation to another. Instead of stressing actions, CPGs should stress *results* of actions. For instance, to specify, not only what procedures or tests to perform, but what *consequences* such tests can have (e.g., what information is to be gained, what undesirable effects they may have).
- As circumstances change and there is no single set of actions that bring about a goal, *flexibility in the ways to achieve the goals* should be designed into CPGs and other evidence-based tools. This circumstantial variability suggests then that there is not one single way to proceed and therefore, evidence-based tools should allow flexibility in the application of recommendations for practice.
- In order to foster CPG utilization, it may be necessary to explore how the structure of an argument constraints the acceptability of guidelines and the quality of decision making when using CPGs. To this end, the argumentative approach can be used as an evaluation tool to improve the acceptance of evidence-based medicine. This information can be used in the design of textual information that optimizes the quality of an argument, by using supporting evidence that is likely to be accepted by guideline users.
- To be effective, information in CPGs should be easily accessible *at the right time*, when it is needed. In this way, *just-in-time* information can make the use of the CPG information more efficient by taking the burden off the user's memory (i.e., it relieves cognitive load).
- At the strategic level, there are two ways in which evidence can be generated and evaluated: by searching and providing (a) supportive or (b) contrary evidence for a hypothesis. Given that supportive evidence is much more likely to be used in natural situations and that contrary evidence is logically stronger in evaluating a hypothesis, evidence-based clinical guidelines

may be improved by providing reminders to information (e.g., patient signs and symptoms) that *disconfirm* the hypothesis.

- There are various ways in which people respond to anomalous evidence. Those responses depend on the quality of the evidence (i.e., how good the evidence is) as well as on the degree of entrenchment of the prior beliefs and knowledge. This suggests that the evidence used in CPGs should be assessed *in the context of many constraints* (e.g., the state of current knowledge, the degree of risk of using the new evidence, patient-specific pattern of signs and symptoms, structure of the argument where the evidence is presented), which influence the evaluation of the evidence, not on the quality of the evidence alone.
- In practice, physicians base their decisions on their prior knowledge, which they bring to bear at the time of making decisions. Prior knowledge in the form of *experiential knowledge* (heuristics, analogies, and reminders) is more pervasive than knowledge obtained indirectly, such as scientific biomedical knowledge. That is, given a discrepancy between experiential knowledge and biomedical knowledge, it is more likely that physicians depend more on the former than on the latter. It is important then to design clinical guidelines that make use of both types of knowledge to be successfully used in decision making (e.g., implemented in decision support systems as reminders).
- Physicians process information at different levels of granularity depending on their expertise in the domain. At higher levels of expertise, physicians can use information that lack in detail but are more general information chunk (e.g., "metabolic state" to account for lack of appetite, gaining weight and feeling cold). Less than experts, in contrast, process information that is more detailed and specific, given their inability to discriminate relevant information from less relevant at higher level of decisions. Information in clinical guidelines should therefore match the level of knowledge granularity with the level of expertise of the user. Evidence at higher levels of knowledge processed by physicians may be inadequate for less experienced physicians.

2. Capturing the meaning in a text: Propositional and semantic analyses

Methods of propositional and semantic network analysis are aimed at the investigation of internal representations in the mind of guideline users by identifying units of thought, such as propositions and their interrelationships (van Dijk TA 1983). This is achieved by characterizing how people understand symbolic information (e.g., written or spoken discourse, graphics, etc.). Understanding involves the generation of meaning from the text itself by transforming the information in the written text (e.g., a CPG) into some semantic form of conceptual message that makes sense in terms of the person's prior knowledge. The person then incorporates this interpreted semantic information into his/her general store of knowledge. Understanding, then, may be regarded as a process whereby a reader attempts to *infer* the knowledge structure of a writer through the text. Thus, identifying the type and quality of inferences is a powerful indicator of understanding the way people comprehend a text, such as a CPG.

Propositional analysis is based on the notion of "proposition," which is an idea underlying a piece of text (a text can be a phrase, sentence, paragraph, etc.). It corresponds to the basic unit of mental representation of symbolic information in human memory. A proposition is composed of two concepts and a relation between the concepts (which is itself a concept). For instance, the sentence " A breast lump was revealed" expresses one proposition that can be analyzed into the concepts of "breast," "lump," and the relation "revealed."

Propositional analysis, however, does not end with the identification of the concepts and relations. It also involves the categorization of the concepts and relations in the text. In the example above, "lump" is the *object* (OBJ) of the action and "breast" is a *location* (LOC). A propositional representation of the sentence may be written as follows:

1.1	Reveal	OBJ: lump; LOC: breast; TNS: past

where the number 1.1 in the first column represents the proposition number (in our example, there is only one proposition); the second column, "Reveal," is the head element or *argument*; and the third column is called the *predicate* (what is said or predicated of the argument). Notice that the propositional representation is always in the present tense (use of "reveal" instead of "revealed") while *tense* information is given in the predicate (TNS).

Although the above example may look simple, far more complex representations are possible. In this context, the method can be used to evaluate the complexity of information (e.g., its level of difficulty, its completeness). One way in which complexity is investigated is given by the identification of three types of propositions that indicate the extent to which propositions relate to one another (i.e., degree of "connectedness" of the ideas in the text). Texts with more connected ideas are generally more complex. Some propositions are simple and express single ideas, whereas other more complex propositions express ideas that include other ideas. Thus, propositions can be categorized into three types, as far as complexity is concerned: *single*, *embedding*, and *linking* propositions.

- *Single* propositions express single ideas. These propositions are self-contained; that is, they do not include or refer to other propositions. For example, the sentence "She lost weight" is represented as a simple proposition, as it comprises a single idea.
- *Embedding* propositions refer to other propositions. An example is given by the sentence "She lost weight relative to her premorbid state." In this case, "premorbid state" is a second proposition to which "She lost weight" relates.
- *Linking* propositions connect two or more propositions and represent likely inferences involving separate parts of a text (e.g., clauses, sentences, or paragraphs). In terms of analysis, the predicate of a linking proposition includes one or more proposition numbers. An example

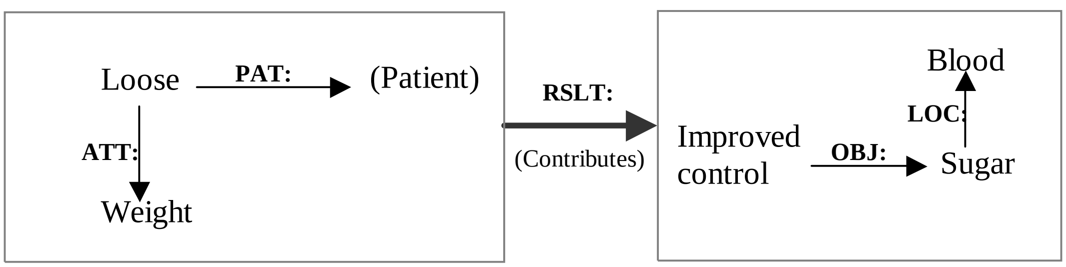
is the sentence “She lost weight and control of her blood sugar improved.” This sentence can be analyzed into three propositions:

(1)“She lost weight,”

(2)“her blood sugar improved”

(3) the relationship between these two propositions, which is signalized by the conjunction “and.” In our case, the conjunction may indicate either of the two following propositions: “Her blood sugar improved *because of* loosing weight” or “Her blood sugar improved *at the same time* that she lost weight.” In the first case, an embedding proposition may indicate a *causal relation*, whereas in the second case, it may indicate a *temporal relation*. Contextual information is used to “disambiguate” these two different meanings.

The output of propositional analysis consists of list of numbered propositions. However, it is useful to represent the propositional structure of a text as a **semantic network**. A semantic network is a graphic representation of the propositional structure. Similarly to the propositional representation, a semantic network is based on the idea of the triplet, e.g., a concept, a relation, and another concept; this time represented in graphical form. In the example below, the arrow represents the relationship between the nodes. Arrows can be unidirectional (e.g., weight loss contributes to improved control of blood sugar), as in the illustration, or bi-directional (e.g., when there is mutual influence).



In addition to information about directionality, propositional and semantic network analyses convey structural aspects, which include **directionality of reasoning (e.g., hypothesis-driven or data –driven) coherence, causality or conditionality of relationships, and temporal sequene** of events. A semantic network captures the varying complexity and overall coherence of the text. It illustrates the simple, embedded, and linked aspects of the sentence, and clearly demonstrates the semantic and logical structures. Temporal relations are captured with the help of propositional types. For instance, such relations can be expressed directly in the text (e.g., the patient felt dizzy *before* collapsing) or they can be inferred from the temporality present in the text (e.g., from the sequence of steps in a procedure), in which case propositional analysis links different parts of the textual information. This analysis can be used to identify semantic information generated by physicians when solving problems they may encounter when using CPGs.Some issues about how these methods of cognitive analysis aid in the development and utilization of guidelines:

- CPGs are developed by a number of highly qualified domain experts. Physicians and trainees require different levels of detail of information for guideline use. An expert physician requires less information than a novice does. It is also true that the guidelines developed by experts are not so well understood by someone who is less than an expert. CPGs may require a large number, and different types of inferences to be generated by various users for their proper interpretation. These generated inferences provide coherence to the representation and may be crucial in the adequate understanding of the guideline. However, it is essential that the user make the correct inferences. In addition, the "embedding" and "linking" propositions provide coherence and aid in developing a macrostructure for the guideline text.
- Propositional analysis can also be used to assess various types of semantic and pragmatic information in texts, such as guidelines. Some of this information is implicit and has to be inferred by relating propositions that are explicitly stated in the text. One such type is the *temporal* sequence of events, as illustrated in the example presented above. Other types are *causal* and *conditional* relations, which are related to the directionality of reasoning. Further relations that are captured by propositional analysis are algebraic relationships in comparing features of different dimensions (e.g., greater than, less than) and that are often presupposed by the explicit information.
- When trying to represent clinical guidelines in automated systems (e.g., Web-based, DSS), the resulting user representation may fail to include information that is contained in the linking or embedding propositions. The complexity of CPGs can however be lessened by converting complex propositions to sets of single propositions easing the need for generating inferences that make guideline interpretation more variable. To tune the information to different levels of user expertise, it may be possible to make this information optional. By browsing through the CPG on the computer, the expert physician may skip through this information (and therefore make the necessary inferences himself or herself) whereas the inexperienced physician may inspect the information at will.

These methods of propositional and semantic analysis provide a general methodological tool that allows us to compare how the meaning intended in the guideline by the designer is interpreted by the users. Propositional analysis of guidelines could allow us to identify points of misunderstanding and errors, and either improve the clarity and accuracy of the information to be delivered or to tailor it to the characteristics of specific users (e.g., according to their levels of expertise).

2.1. Analysis of a clinical practice guideline and its representation in a computer language, GLIF: Screening for thyroid dysfunction

We applied our methods to the analyses to the thyroid screening guideline and compared it to its GLIF representation, generated electronically. In this comparison, four aspects of the analysis of the thyroid screening guideline were used: the guideline algorithm (ACP-ASIM), the guideline's semantic network (developed by researchers), the GLIF representation of the guideline (developed by a physician-computer scientist), and the feedback given by the physician while using the guideline to create the GLIF representation. The focus of this comparison was to study the nature of guideline representation, comprehension, and use, with the specific example of the thyroid screening guideline. This provides a baseline data for evaluating the points where potential problems are likely to arise at various stages of guideline interpretation and use.

- The algorithm was found to be highly simplified compared to the written guideline, as expected. The written guideline expands on information and generally gives definitions where the algorithm does not. For example, the algorithm asks if the patient with subclinical hyperthyroidism has any signs and symptoms. The written guideline explains what these signs and symptoms are and also informs the clinician that if the answer is ‘yes’, then the patient should be referred to an endocrinologist, whereas the algorithm does not provide this kind of additional information. Having detailed information in the text and surface procedures in the algorithm seems an appropriate way to “hide” the complexity in the layers. These “invisible” layers can be referred to as needed when the guideline and the algorithm are used together as they are meant to be.
- In comparing the text version of the thyroid screening guideline to its algorithmic form, three inconsistencies were identified:
    - The algorithm has a path, which tells the physician that if the sensitive TSH test is normal then no further testing is required. However, the algorithm does not define “normal;” it only says that a level of greater than or equal to 10 mU/L is elevated. In contrast, in the written guideline, markedly elevated TSH levels are defined as greater than or equal to 10mU/L, and mildly elevated TSH levels are defined as 6 to 9 mU/L. Someone using the algorithm alone may assume that a normal level is anything between undetectable and 9 mU/L. On the other hand, someone who refers to the text will assume that normal levels are between undetectable and 5 mU/L. However, there is a glossary available on the ACP-ASIM website that provides details of the normal values, as well as other definitions. A link between the algorithmic chart to the glossary could be inserted to remind users of this resource or at least a reminder to indicate that glossary is available, if needed.
    - The algorithm consistently refers to an FT4 test and FT4 levels, while the written guideline never mentions FT4 and only refers to “free thyroxine tests” and “free thyroxine levels.” While it is certain that a physician will know that FT4 is the same as free thyroxine, it is advisable that terms are used consistently throughout both documents. Often the written guideline and its algorithm are used by computer-scientists to develop computer representations of the guidelines. Inconsistency in the terms used may cause errors to be generated as the guideline is translated into the computer representation.

- The third inconsistency arises in the recommendation of how to deal with a patient who has subclinical hypothyroidism. Here, the algorithm gives concrete instructions about how to deal with these patients (they should be re-tested every 2 to 5 years). But the written guideline does not give these same instructions. Instead, the guideline provides a lot of useful background information without ever explicitly giving instructions on treatment. This problem arose because of insufficient evidence to recommend for or against the treatment. In fact, both the treatment or the observation are reasonable approaches. In this sense, the advice given in the algorithm reflects a compromise. The guideline also tells physicians who are concerned about the possible progression to overt hypothyroidism that they can prescribe l-thyroxine to asymptomatic patients, but this advice is not provided in the algorithm. There is an assumption here that everybody will refer back to the written guideline. However, a reminder is needed in the algorithm to notify the user that more information is provided at that particular point in the written guideline, if needed. On a computer-based guideline this could easily be achieved by making links that could be clicked on to take the user to the relevant place in the text.

These types of inconsistencies are difficult to deal with in guidelines, as the algorithm by nature has to be much simpler than the written text. Ideally, written guidelines are supposed to be used with the algorithm. This is a valuable support when one is in education or learning mode, but this can be difficult during the point of solving patient problems and making decisions, where reminders would be very useful. Efforts are being made to develop guidelines that are computer-based. Currently, these computer-based guidelines are developed from the algorithmic and the written guidelines. There are many intermediary stages in guideline development, and translation into computer format where errors could be made. To reduce the possibility of error or ambiguity, it would be ideal to develop the computer-based guideline from the beginning, rather than translating them from paper-based guidelines. It is hoped that computer-based guidelines would allow the simplicity, and ease of the algorithms to be coupled with the detail of the written text in such a way that the physician could quickly and easily use the guidelines without losing valuable information. Use of authoring tools is recommended.

Conclusions: Given that the algorithm is a simplified and a generic version of the written guideline, it is not only important to provide detailed written guideline support at various levels of granularity (for users at different levels of expertise), but also, to attach reminders or some indicators that alert the user that information is available if needed. Monitoring end-users as they interact with the algorithm and written papers could provide valuable feedback into identifying problems such as inaccurate inferences, nature of misinterpretations, and any navigational problems that may arise. For computer implementation, a more explicit procedure is needed that the computer can "understand." This translation problem could be resolved by generating guidelines directly in computable form, with the help of guideline authoring tools.

3. Representation and use of clinical guidelines by physicians with different expertise: diabetes management and screening for thyroid disease

The goal of these studies is to investigate clinical decision making as a function of expertise both without and with the use of CPGs, presented in both diagrammatic and textual formats. These studies illustrate how the cognitive framework presented above can be used for investigating and analyzing the nature of evidence, its representation and utilization during clinical decision making in medical practice. Our specific aims are the following: 1) To understand the impact of (a) text-based CPG, and (b) guideline algorithm by physicians of different levels of expertise. 2) To examine the nature of evidence and reasoning strategies used with and without the use of guidelines in making decisions.

This investigation is divided into two studies. Study A uses the 1998 Canadian Guideline for the Management of Diabetes Mellitus, which is in a textual format. This study looks at how physicians at two levels of expertise interpret and make decisions about a clinical problem both without and with the written guideline. Study B uses the ACP-ASIM Screening for Thyroid Disease in its diagrammatic format. This second study examines how physicians, with two different levels of expertise, understand an algorithm, and the reasoning strategies they deploy in making decisions about patients.

- **Methods**

A multifaceted -approach was used to capture the data, these include semi-structured interviews, and simulated conditions using developed clinical scenarios. The technique of "think aloud " was used, all the data were tape recorded and transcribed for analysis of the nature of evidence, its interpretation, and decisions. Both studies asked the physicians to engage in spontaneous problem solving (without guideline), followed by primed problem solving (with the guideline). The generated responses, without the use of the guideline and with the guideline added were analyzed for concepts and were semantically represented in network form to investigate the nature of knowledge organization.

- **Summary of Findings - Study A (written guideline)**

During the free reasoning the non-experts generated generic information compared to the experts who generated specific information. The non-experts missed crucial information in the clinical case which was picked up by the expert physicians. During the primed reasoning, using the guideline as support, the non-expert physician was reminded of specific procedures that were not accessible during free-reasoning. In this part the experts did not follow the guideline but gave many reasons based on experience why deviations from the recommendations are advisable. It was also found that the non-experts had difficulty selecting the relevant sections of the guideline and need some support in filtering the relevant information from the irrelevant.

- **Summary of Findings - Study B (algorithmic guideline)**

In the free-reasoning part of this study all the subjects made the correct diagnosis. The non-experts took longer than the experts to generate the correct diagnosis and they generated differential diagnostic hypotheses which were not directly related to the correct diagnosis. The experts generated single hypotheses. During the reasoning the non-experts requested more information about the patient than the experts did, much of this requested information was not directly relevant to the correct diagnosis. The non-experts also requested more tests, and a higher percentage of irrelevant tests as compared to the experts. Networks of the non-expert's protocol showed that they had difficulty organizing information. The algorithm helped to focus and structure the reasoning of the non-expert leading to improved organization of information in the primed-reasoning part of the study. In this part the non-experts followed the procedures outlined in the algorithm more closely than the experts did. The algorithm also reduced the number of tests the non-experts requested and the percentage of irrelevant tests requested. The algorithm did not effect the number of tests requested by experts, nor did it influence their reasoning structure.

- **Discussion and Implications**

These studies show how CPGs will be used differently by people of different levels of expertise. For this reason guidelines should be adapted to the user's depth of knowledge. Also, because the algorithm serves to structure and focus the non-experts reasoning and diagrammatic guideline should be included with every written guideline. Guidelines should be used as educational tools in continuing medical education courses and in resident training. For CPGs to be successfully used, whether as decision making or as educational tools, they need to be included in decision support or instructional systems that will help focus the non-expert physician on relevant information, or remind the expert physician of important steps.

It seems obvious that guidelines for patient problem solving need to be presented in an easily accessible and usable manner, if they are to be used at the point of care. Graphical representations of guidelines have the advantages over text-based representations in that they are easily scanned for important information, although they require advanced knowledge in their interpretation (e.g., knowing when to depart from the guideline). Text-based guidelines are extremely useful for learning purposes, provided that they are implemented in instructional or teaching systems and when there are no time constraints. Current models of learning and instruction support a differentiation of two stages in learning: problem solving and reflection on the problem solved, as crucial for effective mastery of domain tasks. Using both algorithmic and text-based guidelines tools can be designed to support instructional systems that are based on sound cognitive research.

4. Studies of guideline usability at the "point of care"

These studies describe an analysis of the process of guideline translation from its text and algorithm form to a GLIF representation. The approach taken in this research involved video recording of a physician's interaction with an advanced computer system, as he encoded the guideline into the GLIF representation language. The purpose was to characterize the cognitive processes involved in the encoding of guidelines in the GLIF language and assess our methodology for analysis of the process of guideline encoding. Our analyses include identification of the following: (1) potential problems with the GLIF language and its representations, (2) difficulties originating from the underlying guidelines and (3) issues related to the interaction between the encoder and the computer tools designed to support the encoding process. The data obtained from these sessions were initially analyzed to determine the main tasks involved in encoding the guidelines and was later analyzed in further detail to identify problems ranging from difficulty in representing guideline steps to problems arising from the underlying guideline.

The subject for the work described in this report was a clinician computer scientist researcher at Harvard University. The recording was conducted by remotely logging on to the user's application (i.e. the encoding software) via NetMeeting, at our evaluation site in Montreal. We recorded computer screens of the subject's interaction with computer system during the task of representing guidelines into GLIF by inputting the resulting screens (accessible using NetMeeting in Montreal) into a PC-Video recorder which then outputs to a VCR. In addition, the subject encoding the guideline at the Harvard site was instructed to "think aloud" as he entered the guideline into GLIF. The subject's verbalizations were audio recorded at the Montreal site and merged with the corresponding video of the computer screens. The instructions to the subject was to use the computer system to encode two ACP-ASIM guidelines into GLIF, working from a graphic representation of the guideline and having access to the on-line text of the guideline.

- The approach taken led to the identification of both generic aspects of the process of guideline encoding as well as specific problems in doing this task. For example, difficulties were encountered by the subject studied in deciding on the type of computer-based representation to apply in:
  - Choosing the appropriate type of computer step for modeling a statement in the guideline.
  - Representing informational statements not associated with a specific step or condition.
- Regarding the guideline material used by the subject in carrying out this task, key background information (e.g. regarding normal ranges) was found to be needed in order to encode the guideline.
- Given the amount and richness of information obtained from this study of a single subject encoding two guidelines, the resulting information from such study can be directly fed back into: (a) the design of information contained in the guidelines, and (b) the fine-tuning of both the underlying representational scheme used in GLIF and the user interface to the encoding

software. However, it would be useful to investigate these issues in more detail and how systematic these findings are. This would allow us to make modifications to our conceptual framework that guides our research.

5. Current research: Reasoning about depression by lay-people (patients and non-patients) and psychiatrists

Since the March 31, 2000, we have concentrated on evaluating existing clinical-practice guidelines and their modes of use. In particular, working from a guideline for depression diagnosis and management from the ACP-ASIM, we have carried out formal studies to investigate (a) physicians' and lay people's beliefs about depression and its diagnosis and treatment, and (b) the interpretation of depression guidelines by physicians and patients. Focal areas of examination are the following: (1) what beliefs about depression people hold and how their beliefs affect the decisions they make about treatment, and (2) how the prior beliefs held by both physicians and patients affect the way they interpret the guidelines intended for their use and, in turn, how that affects their subsequent interaction. Results from these studies continue to inform the guideline development process and the InterMed team as we develop authoring environments and make knowledge-representation decisions for the underlying GLIF model.

6. Recommendations for guideline selection, design, use, and evaluation

These recommendations are targeted at people such as (a) members of CEAS (Clinical Efficacy Assessment Subcommittee) at ACP-ASIM who have responsibility for selecting guideline developers and for endorsing guidelines generated by ACP-ASIM and (b) others such as PIER, who are responsible for developing technological support for the development and use of guidelines both within and outside the organizations. Furthermore, these recommendations can be useful in monitoring and evaluating guidelines for the nature of evidence and their utility at the point of care.

- Evidence-based clinical guidelines may be improved by providing reminders to information (e.g., patient signs and symptoms) that disconfirm a hypothesis. This is based on the fact at the strategic level there are two ways in which evidence can be generated and evaluated: by searching and providing (a) supportive evidence or (b) contrary evidence for a hypothesis. However, disconfirmatory evidence is logically stronger than confirmatory evidence, and less likely to be spontaneously used. Guidelines that foster the use of disconfirmatory evidence can be beneficial, especially in cases where the consequences of clinical actions can have severe repercussions (e.g., for patient safety).

- An argumentative approach to evidence in clinical guidelines could be used to improve the acceptance of evidence by designing argument structures that people find more convincing. This way, evidence-based evidence (i.e., use of scientific evidence as a basis for the practice of medicine) can be incorporated in the design of textual information that optimizes the quality and "convincingness" of the information. However, the argumentative approach has not been subjected to rigorous testing with end-users of medical information. Such tests will be required.
- Clinical guidelines should be designed to make use of both experiential and biomedical evidence to be successfully used in decision making (e.g., implemented in decision support systems as reminders). In practice, physicians base their decisions on their prior knowledge, which they bring to bear in decision making. Prior knowledge in the form of experiential knowledge (e.g., heuristics, analogies, and reminders) is more pervasive than knowledge obtained indirectly, such as biomedical knowledge. That is, given a discrepancy between experiential knowledge and biomedical knowledge, it is more likely that physicians will rely more on the former than on the latter.
- Clinical guidelines should identify supporting expert heuristics (e.g., reasoning and evidence evaluation strategies) that have been shown to be effective in patient diagnosis and management. Emphasis should be given to negotiating these heuristics with patients' values and the impact on their life style evaluated. This way, guidelines may be flexible enough to allow for idiosyncrasies of individual patients. This will require us to develop a model of the patient as a part of decision making that could be translated into a set of procedures in developing a guideline. This topic is currently under investigation.
- Investigation of how people (physicians and patients alike) understand and make decisions using guidelines may be necessary for a wider acceptance of CPGs by end users. Methods that uncover the knowledge structures and the cognitive strategies used during guideline interpretation, such as propositional and semantic network analysis, should be made part of the standard methodology used in guideline design and evaluation. In the design process, the application of such analysis allows us to identify the level of complexity of guidelines, as well as their coherence. In addition, during the use at the point of care, problems in CPG interpretation (e.g., misunderstanding) as well as in their utilization (e.g., ineffective decision making) can be identified. This should benefit both guideline comprehension and acceptability.
- Guidelines in written form are not suitable for use in problem solving at the "point of care," given their information density, which requires extensive search for relevant information. They are useful for learning purposes, as when reflecting on a case already seen. This is in agreement with current models of learning and instruction, which identify two stages in learning: problem solving and reflection on the problem solved as crucial for effective mastery of domain tasks.

- Guidelines in diagrammatic form, such as flowcharts, are useful for problem solving provided that the user possesses adequate knowledge of the content domain, as they aid easy access to the relevant information. Flowcharts convey only the major steps in a procedure while leaving room for physicians to adapt the guideline to particular patients.
- Information in clinical guidelines should match their level of knowledge granularity (e.g., their level of detail) with the level of expertise of the user. At higher levels of expertise, physicians can use information that lacks in detail but that is more general. Generalists, in contrast, process information that is more detailed and specific. Evidence at higher levels of knowledge processed by physicians may be inadequate for less specialized physicians.
- Experts are more likely to make errors of omission whereas generalists more errors of commission. In this respect, guidelines can help experts in the form of reminders (e.g., of a step skipped in a procedure). In turn, they can help the generalists by providing guides that constrain extra irrelevant processing (e.g., suggesting more a focused approach to treatment).
- Matching the level of granularity in keeping with the level of the user is of utmost importance when there is anomalous data, as people respond to anomalous evidence in different ways. Physicians when encountering anomalous data, should be able to back-track to another, more detailed, level of information that provides some explanation for the anomaly. For the generalist, presence of anomalous data may also be opportunity for seeking consultation (e.g., through critiquing, reminders or calling a specialist).
- Guideline presentation format should match the purpose of use. A guideline can have a heuristic role when it is used in decision making, or a supporting role, when it is used to justify one's decisions or in teaching and learning. When used to make decisions, guidelines may contain only the information needed to reach the decision. However, information at a more detailed level may be needed when guidelines are used as learning or teaching tools.
- Guidelines for patient problem solving need to be presented in an easily accessible, and ideally usable for "just-in-time" format, that can be integrated into prior knowledge and the general heuristics of the user. There should be very little room for errors. Guidelines for learning should encourage reflection and understanding deeper meaning of clinical situations and reasons for procedures by allowing ways to double-check on decisions taken and thus permitting making errors (and learning from them).
- Clinical guidelines could be designed such that they specify not only what procedures or tests to perform but what consequences such tests can have (e.g., what information is to be gained, what undesirable effects they may have). Although typically we interpret a procedure in terms of steps to be performed (i.e., what to do), it is important to realize that actions are likely to change from one situation to another. For instance, two patients with the same diagnosis may respond differently to the same treatment. What worked with one patient may not work with the other or what worked with a patient once may not work a second time. As circumstances change, actions have to change to adapt to those circumstances and bring about the desired

goal. This circumstantial variability suggests then that there is not "a right way" to proceed and therefore, evidence-based tools should allow flexibility in the application of recommendations for practice. Flexibility is also inherent in the timeliness of guideline information.

- Automation, e.g., in guidelines embedded in electronic record systems, should be promoted that allows a better implementation of both learning and problem solving/decision making strategies and procedures. Although guideline automation will likely play an increasing role in the future, the challenge will be to develop an integrated system that will support these two types cognitive activities.