Knowledge Representation of Situation-Specific Clinical Practice Guidelines

by

Ronilda Covar Lacson
Diplomate in Internal Medicine, American Board of Internal Medicine, 1996
M.D., University of the Philippines, 1992
B.S. Basic Medical Sciences, University of the Philippines, 1989

Submitted to the Division of Health Sciences and Technology
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Medical Informatics
at the
Massachusetts Institute of Technology

June, 2000

Signature of Author: _____
Harvard-M.I.T. Division of Health Sciences and Technology
May 5, 2000

Certified by: _____
Aziz Boxwala, MBBS, PhD
Instructor, Harvard Medical School
Thesis Supervisor

Certified by: _____

Peter Szolovits, PhD
Professor of Computer Science and Engineering
Thesis Supervisor

Accepted by: _____
Martha L. Gray, PhD
Edward Hood Taplin Professor of Medical and Electrical Engineering
Co-director, Harvard-M.I.T. Division of Health Sciences and Technology

Knowledge Representation of Situation-Specific Clinical Practice Guidelines

by

Ronilda Covar Lacson

Submitted to the Harvard-M.I.T. Division of Health Sciences and Technology
on May 5, 2000 in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Medical Informatics

ABSTRACT

The Institute of Medicine defines clinical practice guidelines (CPGs) as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances." Unfortunately, despite efforts to disseminate CPGs, adherence has been a problem. Failure to adequately address issues related to applicability of evidence for specific recommendations are postulated to account for this dismal acceptance of CPGs. Decision-analytical tools can be used explicitly within CPGs to address this. This experimental study is designed to assess appropriateness of guideline recommendations when decision analytical approaches are incorporated into representations of clinical practice guidelines. This novel model is capable of explicitly representing competing recommendations (i.e. decision nodes in a decision tree) and ranking them based on utilities derived from the decision tree.

The Guideline Interchange Format (GLIF) specification for representing CPG knowledge was extended to support the model. Three different guidelines were encoded to assess adequacy in representing different areas for CPG use (prevention, work-up and treatment). General population-derived utilities were used to represent patient preferences. The recommendations obtained from the guidelines that incorporate decision analysis approaches were compared to those derived from guidelines that do not explicitly include decision analysis when going through the algorithm to produce recommendations for each patient.

Results reveal some differences in recommendations when decision analysis approaches were included in guideline representations. Surprisingly, this variation did not affect physician experts' agreement with these recommendations, regardless of which process was used to represent the guidelines. Similarly, there was no statistically significant difference in physicians' perceptions whether or not they thought patient preferences were taken into consideration. Further study is recommended to determine whether physicians are insensitive to patient preferences or whether the use of utility measures does not appropriately capture the patient's preferences as perceived by physician experts.

Thesis Supervisor: Peter Szolovits
Title: Professor of Computer Science and Engineering, M.I.T.

Thesis Supervisor: Aziz Boxwala
Title: Instructor, Harvard Medical School

Acknowledgements

I would like to thank several people who have been very supportive of my work. I am grateful for the advice and encouragement of my advisors, Dr. Aziz Boxwala and Dr. Peter Szolovits. I also appreciate the help of all my co-workers from the Decision Systems Group at the Brigham and Women's Hospital. I thank Dr. Lucila Ohno-Machado, Dr. Robert Greenes and the InterMed Collaboratory Group for allowing me the honor of participating in the design of GLIF3 and for intellectual support during the model-building process. In addition, I would like to express my gratitude to Dr. Mor Peleg, Mr. Samson Tu, Mr. Colin Gordon and Mr. Jonathan Bury for providing me with samples of guidelines in various representations. I am sincerely grateful to all the expert physicians who participated in this study and gave credence to my work:

1. Dr. Teodoro Santiago, Professor of Medicine, Board-certified Pulmonologist and Sleep Expert, Robert Wood Johnson – UMDNJ, New Brunswick, NJ.
2. Dr. Michael Nolledo, Asst. Professor of Medicine, Board-certified Pulmonologist and Sleep Specialist, Robert Wood Johnson – UMDNJ, New Brunswick, NJ.
3. Dr. Jose Diaz-Buxo, Professor of Medicine and Board-certified Nephrologist, University of North Carolina, Charlotte, NC.
4. Dr. Dinna Cruz, Asst. Professor of Medicine and Board-certified Nephrologist, Yale University, New Haven, CT.
5. Dr. Primo Lara, Asst. Professor of Medicine and Board-certified Oncologist, University of California Davis Cancer Center, Sacramento, CA.
6. Dr. Pauline Lerma, Senior Oncology Fellow and Board-certified Oncologist, Allegheny University Hospital Hahnemann, Philadelphia, PA.

I am indebted to my parents and all my brothers and sisters who have been very supportive of my career. Finally, I express my unbounded appreciation, love and dedication to my husband, Dr. Eduardo Lacson Jr., for his patience and understanding. He provided encouragement, intellectual and moral support during the entire phase of this research study. Thank you all, very much.

Table of Contents

I. Introduction

1.1 Background of the Study

Clinical practice guidelines (CPGs) are defined as “recommendations for patient management that identify one or more strategies for treatment” (1). Adherence to these guidelines is expected to reduce the costs of medical care while improving patients’ outcomes (2,3). This may be accomplished by reducing variability in the clinical practice of medicine and following an algorithm that is based on the best available scientific evidence.

The Evidence-based Medicine Work-Group, responsible for a series of publications in the Journal of the American Medical Association dealing with the use and evaluation of clinical evidence, published an approach to guideline evaluation (4). They suggested that CPGs should address the following questions:

a. What are the risks and benefits (of a recommendation)?
b. How do these compare in different people and with different screening strategies?
c. What is the impact of people’s values and preferences?

Answers to these questions address the applicability of guidelines to specific patients. In particular, few guidelines explicitly address the patient’s values and preferences as noted in the third question above. Therefore, it is important for CPGs to explicitly state whether or not recommendations will be helpful for specific patients or situations.

CPGs often require adaptation to prevalent local practices. Indeed, physician practices and local conditions are important factors to consider when evaluating local CPG

acceptance and adherence. Moreover, some of the variability in clinical practice may in fact, be evidence-based and derived from locally obtained data. Hence, expert opinion and local practices are often determined by heuristics and inference methods based on local or regional clinical experience on top of pathobiologic knowledge of the domain.

The reasoning behind clinical decisions and CPG recommendations may be explicit but are just as often implicit and indirectly stated. One way to formally express the decision-making process is through decision analysis models. Decision analysis has the advantage of previous historical experience and wide utility in many medical applications. Probability theory and Utility theory, both key elements in decision analysis, are valuable tools that can be utilized to represent individualized patient preferences and direct comparisons between competing options for patient care. The risks, benefits and cost-effectiveness of several options for particular situations can be explicitly modeled and specified. These may be perceived as individual steps in a long and complex guideline. Alternatively, representing these competing choices and incorporating a mechanism to rank them based on utilities or beliefs, may be sufficient to model the variations in CPG recommendations for specific patients in specific situations.

1.2 Objectives

I propose to include decision analysis in a CPG model as a method of ranking various strategies for treatment. This enhanced “decision model” would explicitly represent variations in recommendations resulting from uncertainty and variations in local

conditions and/or patient preferences. I will look for answers to the following questions:

a. Will this model be sufficient to express variation in guideline recommendations for specific situations?
b. Will this model be adequate to represent various guidelines?
c. Will the recommendations derived from guidelines expressed using this model be acceptable to expert clinicians?

The main goal of this study is to extend current CPG representations with decision analytic tools and develop an object-oriented model to represent guidelines that can be tailored to specific clinical situations and for specific patient preferences. Specifically, this study is designed to:

1. Model decision-making within CPGs explicitly as decision analytic steps;
2. Extend an existing CPG model to represent decision analysis with the capability to adapt to specific patients in specific situations, issuing recommendations derived from applied decision-making methods;
3. Assess the adequacy of the model in representing various types of CPGs and variations in recommendations according to specific patient preferences; and
4. Compare the acceptability of the recommendations derived from the model compared to the recommendations from the original CPG model that does not incorporate decision analysis.

1.3 Significance of the Study

The proposed representation of patient preferences and situation-specific variation may be used in standard representations for computerized guidelines. More importantly, it reflects an innovation and potentially an improvement over current representation of CPGs. The efforts expended in the development of guidelines are substantial and actually include the decision analytic process, albeit implicitly. Being able to explicitly represent this is a major contribution to CPG representation, especially when applied through computer-based dissemination. This can potentially enhance acceptability of guidelines, one of the major stumbling blocks towards full implementation.

In addition, the proposed model enables supporting information about decision options described in CPGs to be stated specifically and available for further review and scrutiny. Thus, users may review this information to have a better understanding of the rationale for judging specific options as better than other options in a particular clinical situation. Furthermore, this understanding gives the users more flexibility in substituting local practices when recommendations support these options for exceptional circumstances.

Another potential advantage of this decision model is the ease in representing changes for updating CPGs. When there are changes in disease prevalence, when patient preferences vary (e.g. between different patient groups) or when new technology or evidence becomes available, it is easier to update this information on the object-oriented model, as well as trace the consequences downstream in CPG recommendations. This is analogous to determining exactly where in the algorithm a new piece of information fits and how it alters subsequent choices or decisions.

This proposed model is a situation-specific and patient-specific model that is necessary in medical decision-making. The Evidence-based Medicine Working Group is quite cognizant of and actively supports this approach in evaluating CPGs (4). This approach that incorporates decision analysis tools, made explicit, is certainly closer to how decision-making is performed in real-life clinical situations. Therefore, even local adaptation of a "centralized guideline" may be more feasible in this context.

Finally, the study design is able to document how inconsistency in practice may fall within the same evidence base, depending on the source of evidence. Specifying these variables, including patient preferences, will serve to illustrate in greater detail areas of potential disagreement between guideline developers within the CPG decision-making process. These areas of disagreement can, in turn, stimulate further study to clarify and further strengthen the evidence base leading to refinement of CPG recommendations.

1.4 Scope of the Study

This study is limited to modeling competing decisions in a CPG such as for selecting among treatment or diagnostic options. This decision is modeled to support a decision-analytic approach to decision-making, often with evidence derived from the available scientific literature. For each decision point, choices and their corresponding associated expected values (representing favorable patients' outcomes) are represented. The decision tree with the corresponding probabilities and utilities that represent uncertainty

and patient preferences within the model are likewise represented explicitly in the CPG representation.

This model allows representation of recommendations that are specific to the practice environment, local physician practices, or actual patient preferences. Whenever available, a decision analysis technique can be included in the guideline representation to allow for a more specific CPG.

The Guideline InterChange Format (GLIF), a specification for modeling CPG knowledge, is extended to incorporate the decision model for representation of decisions (5). GLIF uses an object-oriented approach to modeling of CPGs (see Chapter 3). Guidelines are represented as flowcharts consisting of different types of steps. One type of step, the choice step, allows representation of competing alternatives in a guideline.

Three different guidelines, one from each of three medical areas (prevention, work-up and treatment) were encoded using the derived model in the extended GLIF specification. The guidelines were encoded manually in Extensible Markup Language (XML) using a text editor. Likewise, the guidelines were encoded using the Protégé knowledge-modeling tool (6), tailored to represent the extended GLIF specification for modeling guideline knowledge. This created a text file in Resource Description Framework (RDF) format (7). Similarly, the traditional guidelines without the decision-analytic steps involved were also encoded using Protégé.

Various clinical situations based on possible patient presentations and clinical scenarios were developed. These patient presentations were used to check if appropriate recommendations could be obtained from the CPGs. For each of these patient presentations, the actual recommendations that the CPG contained using the traditional approach were recorded. These recommendations are manually copied from the CPGs and are similar for every patient. In comparison, decision analysis models were included in the traditional CPG and by substituting patient utilities into the decision analytic model and ranking the decision choices in the decision tree according to their expected utilities, the recommendations were obtained from the augmented CPGs that incorporated a decision analysis approach. The variations in recommendations between the CPGs represented using the decision model and the traditional model were assessed to assess if substantial differences occur in specific patient situations and specific patient preferences.

The recommendations for the test cases obtained using both models were then sent to a panel of physicians. These were sent together with the patients' preferences and presentations that were used to generate the recommendations. These experts assessed appropriateness of the recommendations. They also recorded their perception of whether or not the patients' preferences were considered in the generation of the guideline recommendations.

1.5 Limitations of the Study

The object-oriented model can explicitly represent variations in disease prevalence, local practices and patient characteristics. However, the evaluation study is limited to varying the patient preferences in an abstract scenario for a group of healthy individuals.

In addition, this study model does not have examples of guidelines that contain cost-effectiveness analysis. This is important when several options are equally preferable, but differ significantly in cost. However, this added dimension is easily incorporated into the current representation model.

The execution engine and user-interface for GLIF are currently under development. In this study design, the recommendations were obtained by manually entering the patient preferences and all the other patient variables in the corresponding decision tree of the guideline representation that contains the decision model. An actual evaluation of the model would necessitate a machine-readable representation, which is executed and presented to a user at the point of patient contact. This would be optimal when used and implemented with real patients in real clinic situations.

The actual decision analysis was hidden from users in this study. As clinicians slowly adapt to evidence-based medicine, there may be an increasing demand for showing the decision tree, especially when faced with an informed patient. It would be much easier to understand the reasoning behind a recommendation or the role of variables such as preferences when the entire path in the decision model is explicitly shown.

One problem in the interpretation of the rationale for some guidelines was having users who are not familiar with probabilistic models and utility theory. Unfortunately, this drawback in adapting CPGs appears to increase in specific sites where disease prevalence is unusual or where skills of practitioners are not average. Worse, these are the specific cases where the use of this model would be most preferred. One potential remedy will be the inclusion of tutorials illustrating specific probability and utility concepts phrased in a practical and concise format. This can be incorporated into the final implementation software when designing the program for clinical test use.

References:

1. Sutton PA. Clinical guidelines evaluation. Final report of Department of Health guideline evaluation project. Department of Public Health Medicine. University of Hull.
2. *Development and implementation of the National Guideline Clearinghouse™*. Fact Sheet, May 19, 1998. Agency for Health Care Policy and Research, Rockville, MD. http://www.ahcpr.gov/clinic/ngcfact.htm.
3. Fang E, BS Mittman, S Weingarten. Use of clinical practice guidelines in managed care physician groups. *Arch Fam Med*, 1996; 5(9): 528-31.
4. Barratt A, L Irwig, P Glasziou, RG Cumming, A Raffle, N Hicks, JAM Gray, GH Guyatt. Users' guide to the medical literature: How to use guidelines and recommendations about screening. *JAMA*, 1999; 281(21): 2029-34.
5. Peleg M, Boxwala A, Ogunyemi O, Zeng Q, Tu S, Lacson R, et al. GLIF3 Technical Specification. In: http://smi-web.stanford.edu/projects/intermed-web/guidelines/GLIF1.htm; 2000.
6. Musen MA, Gennari JH, Eriksson H, Tu SW, Puerta AR. Protégé II: Computer support for development of intelligent systems from libraries of components. MEDINFO 95, The Eighth World Congress on Medical Informatics. Vancouver, B.C., Canada, 1995: 766-70.
7. Lassila O, Swick RR. Resource Description Framework (RDF) Model and Syntax Specification. W3C Proposed Recommendation, 1999 Jan, PR-rdf-syntax-19990105. In: http://www.w3.org/TR/PR-rdf-syntax/; 2000.

II. Clinical Practice Guidelines

2.1 Definition

Clinical Practice Guidelines (CPGs) are defined as “recommendations for patient management that identify one or more strategies for treatment” (1). More specifically, the Institute of Medicine defines CPGs as systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances (2). The main objective of these guidelines is to decrease physician practice variation, slow the rise of healthcare costs, monitor inappropriate care, assist clinicians to stay abreast of new clinical information, set research priorities and eventually, promote better healthcare outcomes (3). Evidence-based methods ensure that these CPGs provide valid and appropriate recommendations. Evidence-based medicine focuses on the use of best available clinical evidence to reach informed decisions in patient care (4).

Guidelines have been developed and used for all aspects of patient care including referral, disease management, risk-assessment, utilization and preventive care (5-9). In this study, guidelines of three different types were encoded to address this variability in usage and rationale of guidelines. The guidelines used in the study included disease management or treatment guidelines, utilization or diagnostic work-up guidelines, and preventive care guidelines. Referral guidelines can already be classified as a type of “diagnostic work-up” guideline. A risk assessment guideline is specifically developed to classify patients into categories of risk and is an inference method in itself. The process of generating

recommendations for these various types of guidelines will be discussed in the succeeding sections.

2.2 Development Process

Five major steps describe the typical current practice of development to actual implementation and integration of CPGs into clinical information systems. These are: (1) developing a guideline in text form, (2) encoding the text guideline into an algorithmic and structured format for computer interpretation, (3) disseminating the rules and protocols in a local environment, (4) integrating these guideline-based rules into the clinical system, and (5) examining the impact of these guidelines on processes and outcomes with subsequent refinement as necessary (10). I will specifically address the first two steps, the main focus of this current study. It is important to note that while the current practice involves creating a guideline in a narrative text format first and later encoding it into a structured format, one can imagine that in the future, guidelines will be created directly in a structured format.

2.2.1 Guideline Content Development

This initial step in guideline development typically involves identifying a problem where there is either widespread disagreement or variability in practice and where something valuable, such as lives saved or cost savings, may be gained from arriving at an ideal recommendation. The core CPG development team then conducts a preliminary analysis of the problem. Subsequently, domain experts are identified and together, an exhaustive search and review for pertinent information in the scientific literature is conducted.

Whenever possible, draft guidelines are created, often requiring multiple revisions thereafter, as other experts and a broader audience gets a chance to provide comments. Often, there is a deficiency in the literature or there is a substantial disagreement on best practice. At this point, a panel offers expert opinions and arrives at a recommendation based on a consensus (11). Several methods have been employed in the development of CPGs over the past decade. The modern foundation for "future" development of CPGs are found in the Institute of Medicine's book (circa 1990) entitled "Guidelines for clinical practice: from development to use" (2). Six priority areas for research in CPG development were identified:

a. Means for setting priorities among topics for guideline development
b. Procedures for securing thoughtful and useful statements of expert judgements
c. Methods for analyzing and rating scientific evidence
d. Techniques for improving knowledge of health outcomes and giving due importance to patient preferences
e. Methods for identifying and projecting the costs of alternative courses of care and comparing their cost-effectiveness
f. Mechanisms for identifying and evaluating inconsistent and conflicting guidelines

Realizing the need for CPGs to evolve as knowledge accumulates, research towards mechanisms for updating CPGs were also recommended. A formal process for revising

the CPGs remains underdeveloped to date. Currently, this process of guideline development and revision has several inconsistencies in the methodologies used.

A British article published in 1993 on "Development of guidelines for general practice care," proposed formal steps in guideline development (11). It distinguishes a centralized from a decentralized approach. In the decentralized or local method, a local group formulates guidelines using the literature, local practices and expertise. In the centralized approach, a group of experts within a broader coverage area (i.e. national or international) develop the CPGs. The major criticism of the former approach is a deficiency in systematic and rigorous literature analysis. Non-adaptability of centrally developed guidelines to specific needs and local situations is a problem associated with the latter approach. In an effort to standardize the guideline development process, several steps were identified and discussed in this article. These steps include:

a. Analysis of literature – A systematic analysis of literature is undertaken, including meta-analyses, local guidelines and relevant studies.
b. Consensus on draft guidelines – A group of experts (5-10 people) is selected based on competence, recognition and "representativeness". The last term referring to the likelihood that the experts' opinions reflect the opinions of the majority of clinicians. These experts review and discuss the scientific literature collected above and draft guideline recommendations with a corresponding scientific justification for each. In the consensus development process, the recommendations have to meet the following criteria: (1) validity,

(2) reliability, (3) clinical relevance and applicability, (4) comprehensiveness and specificity, and (5) flexibility.

c. Consensus in a broader audience – This phase involves the targeted group of clinicians expected to use and adhere to the CPGs. They provide input either through the consensus development process or by sharing their personal opinions. Four different methods for obtaining this information are proposed: (1) consensus conference, (2) survey, (3) group interview or (4) Delphi procedure. It is important that in using any of these methods, the process is structured and well designed.

d. Testing the guidelines – This phase will determine whether the CPGs are applicable in clinical practice, early into the development stage. It can be done through pilot studies in different practices or by developing several prototypical cases and comparing the CPG recommendations to actual opinions by practicing clinicians.

e. Authorization - An independent and well-respected body usually conducts this phase. Here, the development process is certified as accurate and reliable and the recommendations that have been developed are analyzed. The "seal of approval" is expected to improve acceptance of the CPG and withholding certification acts as a safeguard against poorly developed guidelines.

An indirect method developed for CPG development is the RAND technique (12-13). The RAND Corporation is a research organization in California that proposed a formal consensus approach to the development of "appropriateness criteria." The method

employs a formalized, modified Delphi process that is extensive, evidence-based, and specifically designed to develop these appropriateness criteria. The criteria themselves are not directed towards decision-making, but are utilized for evaluating the appropriateness of certain medical interventions. However, the criteria may eliminate or endorse clinical "best" practices that in turn can be used to develop the actual CPGs (2). Some of the steps in the consensus process are discussed below.

a. Literature search – All pertinent articles are obtained through Medline, including reference lists from authors, papers identified by an expert panel, or papers from relevant scientific meetings. Each is carefully reviewed. Case reports, review articles, letters and editorials are usually excluded *a priori*.
b. Data collection and analysis – Data from the literature are abstracted into evidence tables. All outcome measures are identified and given appropriate weights. In comparing various treatments for a specific indication, comparative data are preferred in constructing the evidence table. In the absence of available comparative studies, qualitative comparisons are allowed by presenting the data from different studies separately. Since these data are not adjusted for confounding variables, statistical comparisons are not feasible.
c. Panel process – A panel of experts is selected, usually prominent leaders of subspecialty organizations and nationally recognized leaders in their fields. Using a modified Delphi approach, these experts are presented with information on the specific topic and then are asked to independently rate the

importance of each of the outcome measures previously identified. The results are subsequently discussed and some new outcomes may be added. The participants again rate the outcomes independently a second time.

A review of the final ratings is conducted to analyze how clinicians may view the results. Further studies such as cost-effectiveness analysis are recommended, as necessary.

Decision analyses contribute towards the development of guidelines. These were initially directed towards utilization of results of decision analysis studies and cost-effectiveness analysis to come up with recommendations. The role of decision analysis was confined to the very early stages. Thus, the decision analytic process itself was not incorporated into the CPG representation. There has been a previous attempt to use decision tables to convert probabilistic data from the decision trees into clinical algorithms (14). Rule sets were identified to create recommendations for CPGs. In the absence of decision tables, text recommendations are available based on results of these studies. This was done in the early 1990's but the potential for its use is yet to be realized.

The latest studies utilize machine learning algorithms for guideline development. Classification algorithms were used to extract rules for classification problems. This is important in risk assessment or in guidelines with at least two competing treatment options. Examples of these algorithms include search space models, Markov models, generalized linear models, artificial neural network and semantic networks. Recently, a two-stage machine learning model was used for the development of a dementia staging

guideline (15). The machine learning component is incorporated in a sequence of steps that lead to the development of useful recommendations. The steps include preprocessing of the data, data mining (using a machine learning algorithm such as identification tree), model selection, post-processing and evaluation. The rule that is extracted from this process may be the basis of a simple classification guideline or may be a part or a step of a more complex guideline. However, the model derived from machine learning is not able to reflect patient and/or physician preferences.

The CPG developed using any of the above methods can then be converted into algorithms or go directly (in text form) into dissemination. Some elements that need to be considered in guideline representation, both from initial development onto implementation and evaluation of performance, are discussed in succeeding paragraphs below.

2.2.2 Development of Algorithm from Text Form

2.2.2.1 Rule-based Approach

Various methods have previously been described for structuring CPG knowledge into computable constructs. Patient-specific reminders are currently in place to alert clinicians about laboratory results or drug interactions or provide advice for daily practice. The Arden Syntax for developing medical logic modules (MLM) is a language for encoding the aforementioned medical knowledge bases. It has been used to generate clinical alerts, diagnostic interpretations, management messages, screening for research

studies and quality assurance. The rules or MLMs are stored as structured text files and can be written using any text editor (16-17). Sets of MLMs have been used to implement CPGs (18-19). This approach, however, is not designed to represent guideline knowledge. CPGs are typically composed of multiple steps that have been arduously designed instead of a string of single-step rules that are artificially combined.

The G-CARE language, developed at the Regenstrief Institute in Indiana utilizes rules as a means of decision support. It allows layers of rules to represent CPGs with tools that support complex interaction with a clinician at the point of patient care (20). The main drawback is in the interaction of various recommendations for each rule. A mechanism for grouping related concepts and placing constraints on the recommendations is still being developed. As with Arden Syntax, G-CARE was not specifically designed to represent CPGs.

At about the same period, in the early 1990's, a frame-based approach to representation of CPG was developed in Germany (21). Facts about patients are organized into predefined structures. CPGs, represented as "medical guideline modules (MGM)," are then applied to a patient. All frames with conditions satisfied by a patient's facts are evaluated and anything new is added to the patient object. Every time patient facts change, the MGM recomputes all derived facts. Finding all frames with conditions satisfied by a patient is then viewed as a classification problem. In short, a patient is classified according to his facts. This is an example of a classification-based, forward-chaining, rule-based system.

2.2.2.2 Tables

Decision tables have also been used to represent CPGs. Rule sets are perceived as recommendations that should be made given every possible combination of situations at any given time. This method assures completeness of information and concurrently allows for a reduction of redundant information. Furthermore, it also allows for deletion of impossible situations and/or unnecessary combinations (22-23). Rules requiring probabilistic data have also been represented as decision tables. Results of decision analysis are transformed into sets of conditions and actions within a decision table. The corresponding probabilities and utilities are stored as additional information. Flowcharts or decision trees are then used to represent information extracted from the tables (24). Decision tables, however, do not capture a large portion of current guidelines.

2.2.2.3 Other representations

One-to-one mapping between concepts in decision analysis and clinical practice guidelines have also been performed (25). This mapping allowed automatic generation of guideline recommendations based on results of decision analysis techniques. The approach has been successful in creating recommendations that are tailored to specific patient and clinical situations. However, it only allows CPG models that are based on decision analytic techniques. Its recommendations are limited to clinical problems that decision models can represent. All the simple forms of representation discussed in this section cannot fully represent the knowledge in CPGs. In the next chapter (Chapter 3),

several guideline models that are specifically developed to represent knowledge in CPGs are discussed.

References:

1. Sutton PA. Clinical guidelines evaluation. Final report of Department of Health guideline evaluation project. Department of Public Health Medicine. University of Hull.
2. Field M, K Lohr (eds.). Clinical practice guidelines: directions for a new agency. Washington, DC. National Academy Press, 1990.
3. Heffner JE. Does evidence-based medicine help the development of clinical practice guidelines? Chest, 1998; 113(3 Suppl): 172S-8S.
4. Lohr KN, Eleazer K, Mauskpf J. Health policy issues and applications for evidence-based medicine and clinical practice guidelines. Health Policy, 1998; 46(1): 1-19.
5. Langley C, A Faulkner, C Watkins, S Gray, I Harvey. Use of guidelines in primary care – practitioners' perspectives. *Family Practice*, 1998; 15:105-11.
6. Schriger DL, LG Baraff, WH Rogers, S Cretin. Implementation of clinical guidelines using a computer charting system: Effect on the initial care of health care workers exposed to body fluids. *JAMA*, 1997; 278(19): 1585-90.
7. Rossi RA, NR Every. A computerized intervention to decrease the use of calcium-channel blockers in hypertension. *J Gen Intern Med*, 1997; 12: 672-8.
8. Knirsch CA, NL Jain, A Pablos-Mendez, C Friedman, G Hripcsak. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system. *Infect Control Hosp Epidemiol*, 1998; 19(2): 94-100.
9. Overhage JM, WM Tierney, CJ McDonald. Computer reminders to implement preventive care guidelines for hospitalized patients. *Arch Intern Med*, 1996; 156: 1551-6.
10. Zielstorff RD. Online practice guidelines. *JAMIA*, 1998; 5(3): 227-36.
11. Grol R. Development of guidelines for general practice care. *Br J of Gen Practice*, 1993; 43: 146-51.
12. Jackson CA, K Pitkin, R Kington. Evidence-based decision making for community health programs. MR-933-MLHS, 1998 (http://www.rand.org/organization/health/healthpubnav.html).
13. SJ Bernstein. Hysterectomy: ratings of appropriateness. MR-592/3, 1997 (http://www.rand.org/organization/health/healthpubnav.html).
14. Shiffman RN, RA Greenes. Use of augmented decision tables to convert probabilistic data into clinical algorithms for the diagnosis of appendicitis. *AMIA Proceedings*, 1992; 686-90.
15. Mani S, WR Shankle, MB Dick, MJ Pazzani. Two-stage machine learning model for guideline development. *Artificial Intelligence in Medicine*, 1999; 16: 51-71.
16. Hripcsak G. The arden syntax for medical logic modules: introduction. *Comput Biol Med*, 1994; 24(5):329-30.

17. Hripcsak G. Writing arden syntax medical logic modules. *Comput Biol Med*, 1994; 24(5): 331-63.
18. Sherman EH, G Hripcsak, J Starren, RA Jenders, P Clayton. Using intermediate states to improve the ability of the arden syntax to implement care plans and reuse knowledge. *AMIA Proceedings*, 1995: 238-42.
19. Cimino, JC. Socratous SA, Clayton PD. Automated guidelines implemented via the world wide web. *AMIA Proceedings*. 1995: 941.
20. Overhage JM, B Mamlin, J Warvel, J Warvel, W Tierney, C McDonald. A tool for provider interaction during patient care: G-CARE. *AMIA Proceedings*, 1995: 178-82.
21. Heinlein C. K Kuhn, P Dadam. Representation of medical guidelines using a classification-based system. *ACM*, 1994: 415-22.
22. Shiffman RN, RA Greenes. Improving clinical guidelines with logic and decision-table techniques: Application to hepatitis immunization recommendations. *Medical Decision Making*, 1994; 14(3): 247-343.
23. Shiffman RN, RA Greenes. Rule set reduction using augmented decision table and semantic subsumption techniques: Application to cholesterol guidelines. *AMIA Proceedings*, 1993: 339-43.
24. Shiffman RN, RA Greenes. Use of augmented decision tables to convert probabilistic data into clinical algorithms for the diagnosis of appendicitis. *AMIA Proceedings*, 1992: 686-90.
25. Sanders GD. Automated creation of clinical-practice guidelines from decision models. Ph.D. Dissertation, Stanford Medical Informatics, SMI Report No. SMI-98-0712, Department of Computer Science Report No. STAN-CS-TR-98-1609. Stanford University, Stanford, CA. July 1998.

III. Guideline Models

Several models have been developed to specifically represent CPGs. The guideline representations I will discuss in this chapter are grouped according to the orientation of the model. The orientation of a model determines the coherence of the steps that make up the guideline representation. For example, a criteria-based model such as Prestige has state-transition criteria that define the appropriate object in the CPG that applies to a specific patient at a given time. In an algorithm-based model such as GLIF, the appropriate object is determined primarily by the previous object, which has a pointer to the next one. I have classified the models into three types: criteria-based, intention-based and algorithm-based. The classification is arbitrary because all these models are able to represent and actually make use of criteria, intentions and algorithms in varying degrees.

3.1 Criteria-Based Models

3.1.1 Prodigy

Prodigy stands for "**P**rescribing **R**ati**O**nally with **D**ecision-support **I**n **G**eneral-practice stud**Y**." The project is funded by the National Health Service of the United Kingdom and was specifically designed to develop a guideline-based decision support system that can assist general practitioners in choosing appropriate therapies for their patients (1). In the Prodigy model, a guideline recommendation is believed to be appropriate for various combinations of patient states when a patient has a particular diagnosis. These states are referred to as scenarios (see Figure 1). In turn, a scenario may have different strategies for treatment. Each of these strategies can then be further broken down into more detailed instructions.

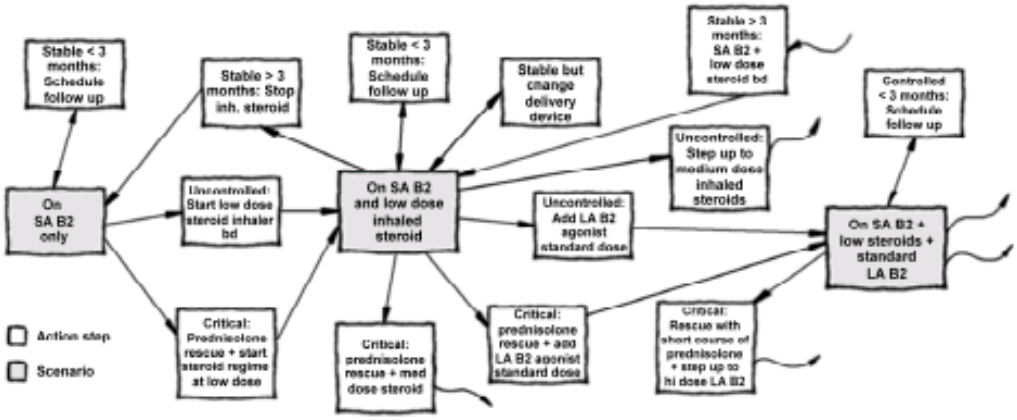


Figure 1. A section of the asthma guideline showing scenarios and action steps

Figure1: PRODIGY representation of an asthma guideline. This guideline is based predominantly on the recent, as yet unpublished, 1999 update of the 1996 North of England Evidence Based Guideline for the primary care management of asthma in adults. (Reprinted with permission from Ms. Mor Peleg, Stanford Medical Informatics, SMI-1999-0790. In: http://smi-web.stanford.edu/pubs/SMI_Abstracts/SMI-1999-0790.html.)

The model has been developing since 1995 and is presently commissioned to address chronic disease management. It has two distinct sections, the consultation template and the management guideline (2). The consultation template is associated with a scenario and is used to gather more data relevant to the therapeutic decision. It is an optional feature for improving decision support. The management guideline is represented as a network of three basic objects: scenarios, action steps and subguidelines. A scenario is usually a combination of diagnosis and specific therapy that corresponds to a specifically recognisable patient state. An example would be “an asthmatic on a short acting beta2 agonist and low-dose inhaled steroid.” (2) A scenario may have several outcome assessments, which would sum up how the patient turned out prior to a visit (e.g. controlled or uncontrolled). This would in turn determine the choice of potential management strategy (e.g. continue low-dose steroids or use high dose steroids) available for that scenario and outcome assessment.

Boolean expressions that are implemented to select both the appropriate scenarios and actions at every given consultation are referred to as “criteria”. Criteria are used to eliminate options that are not viable (e.g. presence of penicillin sensitivity for the action step that would prescribe penicillin). When the system is unable to rule out particular action steps or scenarios, the user is presented with the complete list and allowed to choose manually. However, the system may still depend on criteria to rank the remaining options and express preference for a particular scenario or action, when information is available. The goal is to assist the user in classifying the patients into an appropriate scenario. Consequently, after choosing from several outcome assessments, the user is again guided to select preferred actions from the appropriate set of potential actions. This set of actions implemented in a particular visit then determines the most likely scenario for a subsequent visit.

Lastly, it is noteworthy to mention that Prodigy distinguishes actions that are instantaneous from activities that may persist over time and may be modified at subsequent consultation(s).

3.1.2 Prestige

Prestige is a structured set of subprojects. It evolved from various sites in Europe aimed at guideline and protocol-based care, as well as resource management in the daily practice of healthcare (3). The conceptual model and architecture grew from and extends DILEMMA (protocol knowledge representation), NUCLEUS (clinical act management)

and GALEN (concept representation and term manipulation). These are all projects supported by the European Commission for healthcare telematics (4,5). DILEMMA will be discussed in more detail as it forms the basis for guideline modeling in Prestige.

DILEMMA is a project within the 1991-94 AIM program of the European Commission to develop computerized decision support, particularly for drug prescribing and the use of clinical guidelines and protocols (4). It is an object model and contains an activity hierarchy, with the specified state transition entailed by these activities. Some basic concepts include:

1. protocols (means of achieving a goal),
2. protocol implementation (decision to apply a protocol for a suitable focus),
3. subjects (objects that a protocol is designed to address, such as patients),
4. subject characteristic type (e.g. something which may be observed of a subject, such as blood pressure),
5. agents (objects with the power to act, such as physicians),
6. actions (tasks performed in the real world), and
7. procedures (type of action resulting from implementation of the protocol).

A protocol typically consists of “would be” actions (see Figure 2). A protocol may also be a component of another protocol. When a protocol is implemented, a procedure,

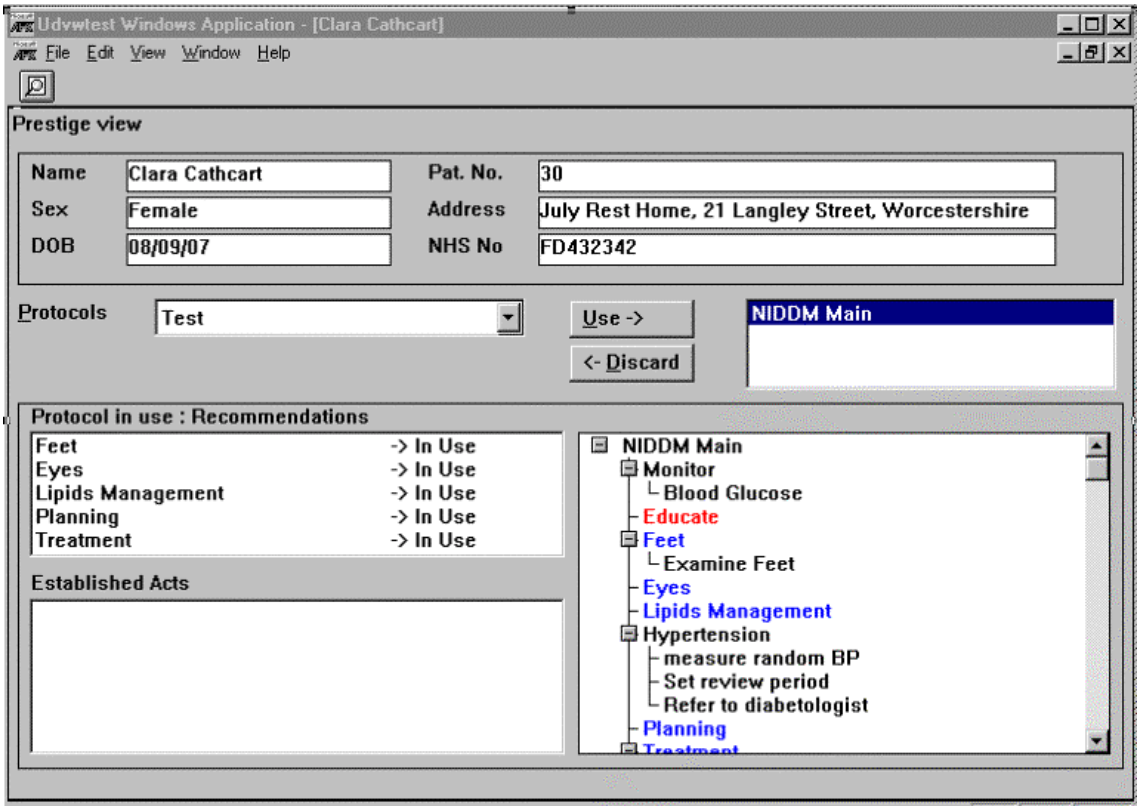


Figure 2: Screenshot of a protocol for diabetes management using Prestige. (Reprinted with permission from Mr. Colin Gordon.)

defined as a type of action, may be generated. This procedure assumes one of many action states (e.g. relevant, established, requested, accepted, cancelled). Subsets of transitions between states are pre-determined (e.g. from requested to completed). This is further controlled by the state transition criterion, which may contain one or more criteria before the state transition is allowed. The state transition criterion is responsible for controlling the sequence in which protocols are implemented. Using logic similar to Prodigy, criteria are likewise evaluated to select the protocols that are relevant at any given time.

When choosing the best relevant protocol, pros and cons are evaluated to mandate or preclude each one. Ranking parameters are available for protocols with a covered attribute that can be compared (e.g. average cost of implementation) between them.

3.1.3 PROforma

PROforma is another language that has been developed over the last few years for specifying clinical guidelines and protocols (6,7). It provides a formal mechanism for specifying the patient data, medical knowledge and various tasks that represent CPGs. The basic tasks or objects supported by PROforma include a plan, a decision, an action and an enquiry step (see Figure 3). Templates are available in building the knowledge required for a task to be defined in a CPG. A graphical knowledge editor for the creation of CPGs complements the knowledge representation model. The editor performs consistency and completeness checks prior to creating embeddable objects or steps. An enactment engine for testing and execution is also available as a stand-alone application or for use with a hospital information system.

The template in PROforma distinguishes it from the typical algorithmic approach to CPG modeling. The model supports sequenced tasks (e.g. protocols) by the use of pre- and post-conditions, both of which are Boolean conditions or criteria. Decisions are defined in terms of a set of options with an argumentation mechanism to choose between these alternatives. This is implemented by ranking competing options according to the number of criteria that are triggered to support an option as long as a contraindication is not triggered.

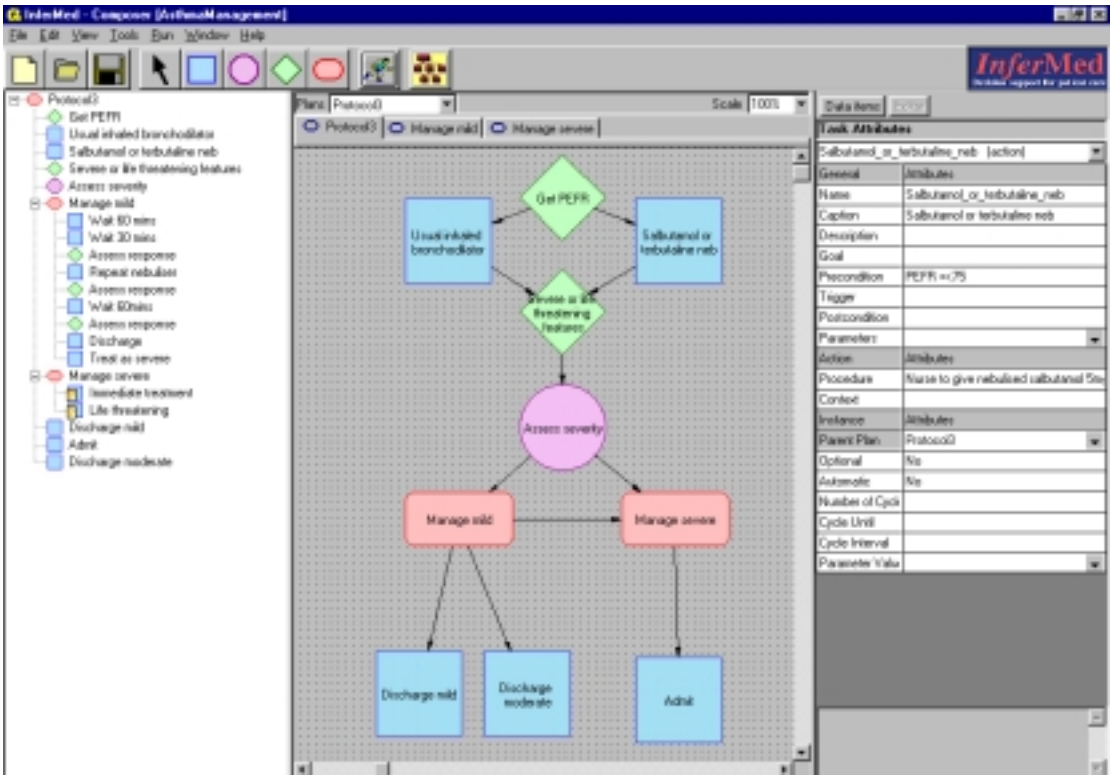


Figure 3: Screen shot of the Task-based representation of the British Thoracic Society's Guidelines for the Management of Acute Asthma in Emergency Departments, using Infermed's Composer software. (Screen shot printed with permission from from Mr. Jonathan Bury.)

In the remainder of this chapter, I describe three models being developed in the United States for representation of CPGs. The last two models are algorithmic in nature, a different approach when compared to the three models previously described above.

3.2 Intention-Based Model

3.2.1 Asbru

Asbru is an intention-based language that was developed at Stanford University for the representation of CPGs (see Figure 4). This permits explicit representation of CPG intentions, patient states and prescribed actions, all of which have temporal patterns (8).

```
(PLAN observing-NID-GDM
(DOMAIN-DEPENDENT TIME-ASSIGNMENT
  (TIME-SHIFTS DELIVERY <- 38 WEEKS)
  (TIME-POINT CONCEPTION
      <- (ask (ARG “conception-date?”)))

(PREFERENCES
    (SELECT-METHOD EXACT-FIT))

(INTENTION: INTERMEDIATE-STATE
  (MAINTAIN STATE(blood-glucose)
    (NORMAL|SLIGHT-HIGH) GDM-Type-II
    [[24 WEEKS, 24 WEEKS],
    [DELIVERY, DELIVERY], [_,_],
    CONCEPTION]))

(SETUP-PRECONDITIONS
  (PLAN-STATE one-hour-GTT COMPLETED
    [[24 WEEKS, 24 WEEKS],
    [26 WEEKS, 26 WEEKS], [_,_],
    CONCEPTION]))
(FILTER-PRECONDITIONS
  (one-hour-GTT (140, 200) pregnancy
    [[24 WEEKS, 24 WEEKS],
    [26 WEEKS, 26 WEEKS], [_,_],
    CONCEPTION]))
(SUSPEND-CONDITIONS (OR STARTED RESTARTED)
  (STATE (blood-glucose) HIGH GDM-Type-II
    [[24 WEEKS, 24 WEEKS],
    [DELIVERY, DELIVERY], [4 DAYS,_],
    CONCEPTION]
    (SAMPLING-FREQUENCY 30 MINUTES)))
(ABORT-CONDITIONS
  (OR STARTED SUSPENDED RESTARTED)
    (insulin-indicator conditions TRUE **
    (SAMPLING-FREQUENCY 30 MINUTES)))
(COMPLETE-CONDITIONS (OR STARTED RESTARTED)
  (delivery TRUE GDM Type II *
    (SAMPLING-FREQUENCY 30 MINUTES)))
(DO-ALL-TOGETHER
  (glucose-monitoring)
  (nutrition-management)
  (observe-insulin-indicators)))
```

Figure 4: A portion of a guideline for the management of gestational diabetes mellitus represented using Asbru. (Reprinted with permission from AMIA Proceedings, Inc.) (8)

Actions (specific tasks to be performed), plans (specifications of guidelines and their individual components) and intentions (goals at various levels of the guideline to be achieved, maintained or avoided) comprise the basic concepts (see Figure 4). A plan is composed of a set of smaller plans, which is referred to as an action when it is no longer decomposable into further plan steps. Plans have states (started, completed, suspended,

restarted, aborted). State transition criteria specify transition between states, as described for Prestige. The plans may be sequential, parallel, or cyclical.

Conditions or criteria determine transition between neighboring plan states and enable a plan to get started. In addition, preferences constrain the selection of a plan (e.g. resource restriction).

Finally, intentions, states and actions are temporal patterns with a time annotation and support multiple time lines (e.g. different zero points). This enables expression of time interval-based intentions, patient states and actions.

3.3 Algorithm-Based Models

3.3.1 Eon

EON is a component-based representation system being developed at Stanford University (9). It contains knowledge structures to represent parts of a CPG (see Figure 5). These include protocol steps, intervention states, conditions, eligibility criteria and revision rules. Patient data are obtained from a database with a specified temporal database manager. Input can also be obtained from a clinician while the system is used at run-time. Recommendations are then generated based on the specific CPG. Additions to EON include the development of specialized guideline models (e.g. consultation guidelines, management guidelines) to allow for a certain degree of model flexibility (10). This modification permits different sub-models to represent specific requirements of a guideline.

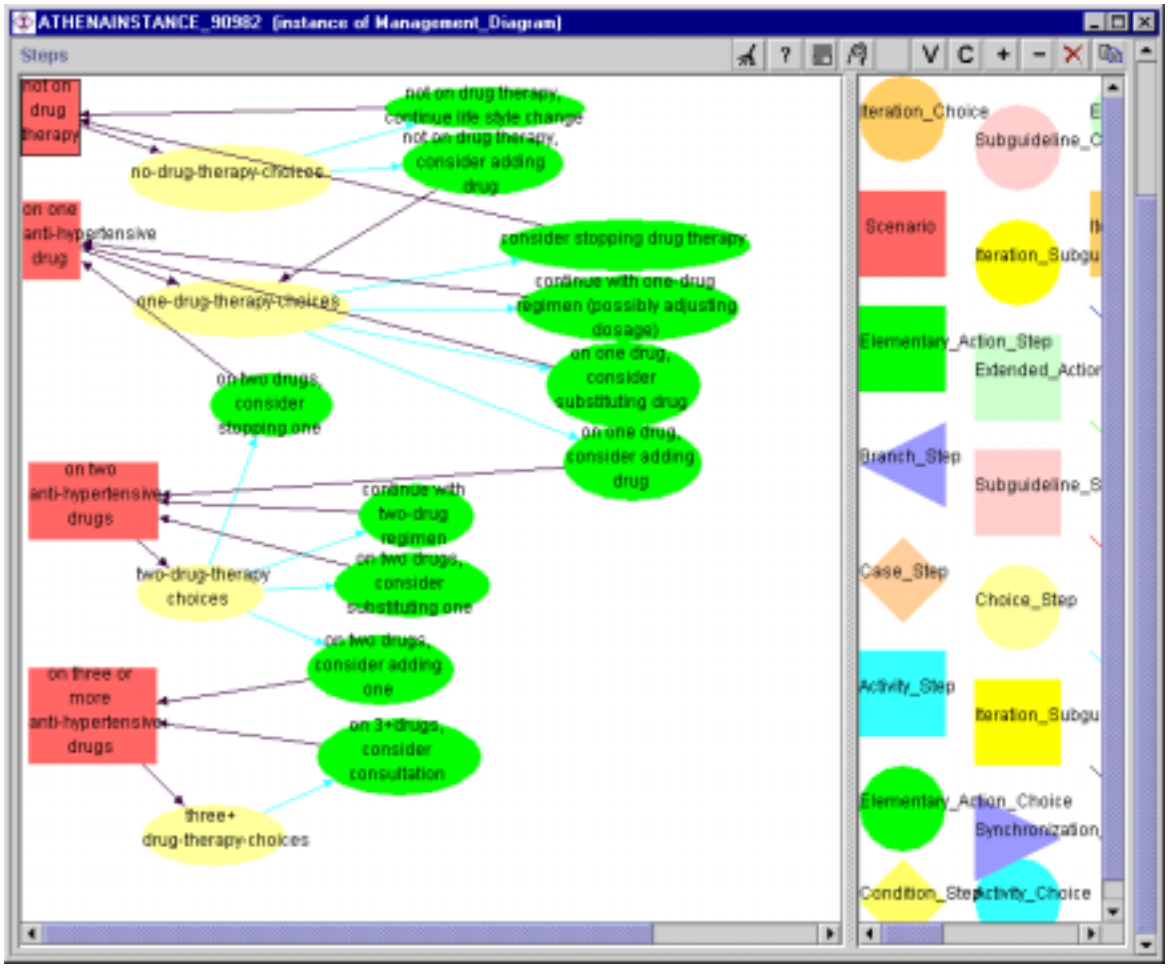


Figure 5: Revised version of the JNC-VI Hypertension Guideline, The Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure, November 1997 represented in EON using Prestige. (Printed with permission from Samson Tu, who is responsible for the content with Brian Hoffman and Mary Goldstein.)

The algorithm is represented as a set of scenarios, action steps, decisions, branching, synchronization, and repetition control nodes that are connected by a “followed_by” relation. The decision and action steps may have goals (represented as Boolean criteria) associated with them. A scenario, similar to that described in Prodigy, is a partial state of a patient. It allows the clinician to properly classify a patient into an appropriate state in

the CPG. A scenario is followed either by a decision or by an action step. Action and activity are implemented in a manner similar to Prodigy.

Decision-making is represented as a set of choices following the decision step. There are two ways by which this is implemented in Eon. Initially, an "if-then-else" condition may determine the next choice. Alternatively, a second way to make decisions is through a "rule-in" and/or a "rule-out" condition. A rule-in condition that evaluates to true would mark the option as being the preferred choice while a rule-out condition rejects an alternative option. If there are no rule-in conditions that evaluate to true and the rule-out condition does not eliminate all but one option, a default preference determines the optimal choice between the alternatives.

3.3.2 GLIF

Guideline Interchange Format (GLIF) is a model for representing guidelines developed by the InterMed Collaboratory Group (11). The main goal is to develop a standard for representing guidelines and facilitate guideline sharing. The GLIF specification consists of an object-oriented model and corresponding text syntax. The GLIF model contains a set of classes that represent CPG contents, with their corresponding data types and attributes. Collections of steps are linked together in a directed graph to form the logic of a CPG. There are four types of guideline steps in the GLIF 2.0 specification: action steps, conditional steps, branch steps and synchronization steps (see Figure 6).

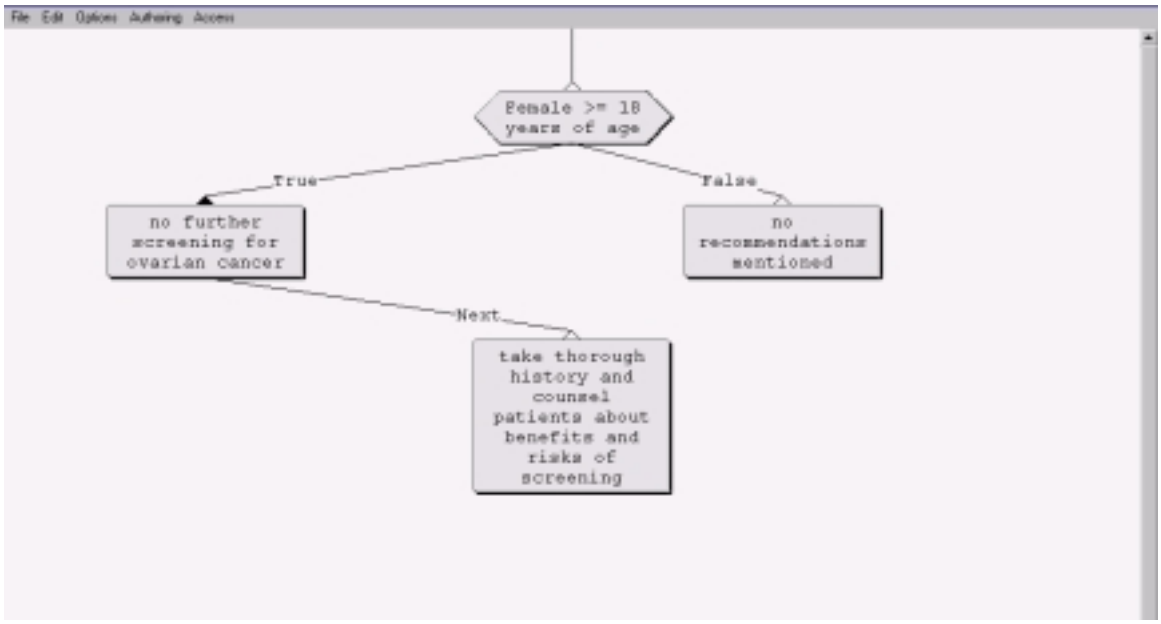


Figure 6: Ovarian cancer screening guideline from the American College of Preventive Medicine (12) in GLIF 2.0 using a GLIF authoring tool from the Decision Systems Group, Brigham and Women's Hospital (Boston, MA).

The InterMed Collaboratory is developing specifications for GLIF 3.0 as described at the GLIF Workshop in Boston, MA (March, 2000) (13). New features of GLIF 3.0 include subguidelines, which are nested guidelines within decision and action steps. Another effort to control complexity of guidelines is with the use of a macro step, which abstracts and encapsulates a set of underlying GLIF steps that follow a structural pattern (e.g. risk assessment). Other new features in GLIF 3.0 include a structured grammar for specifying expressions and criteria, domain ontology support, and representations for iterations, events, exceptions and a patient-state step.

The decision model in GLIF 3.0 has been extended to support a hierarchy of decision step classes. This includes case steps (that can be automatically executed) and choice steps (where multiple decisions for consecutive options have to be made by the user and cannot

be automated). The choice step, in particular, is flexible in allowing for ways to decide between competing alternatives in a guideline. One subclass contains rule-in and rule-out conditions to decide between competing options, similar to Eon. Another subclass uses k-of-n criteria or ranks options according to the total number of criteria that evaluates to true for that specific option, this time similar to PROforma.

The comparisons between the six guideline models previously discussed are shown in Table 1. Each one allows extension of the decision process in the model to allow ranking if not choosing the best among competing options, which is the essence of decision-making. I have chosen to extend the GLIF3 model for this study to support ranking of competing options using decision analysis because I am involved in the continuing evolution of this model.

Table 1: Comparison between six guideline models

Models	GLIF	Eon	Asbru	PROforma	Prodigy	Prestige
Algorithmic	Yes	Yes	No	Not primarily	Not primarily	Not primarily
Criteria-Based	Not primarily	Not primarily	Yes (temporal)	Yes	Yes	Yes
Intentions and Goals Supported	No	Yes	Yes	No	No	No
Ranking of	Yes	No	No	Yes	No	Yes

Options Supported						
Time Interval Support	No	Yes	Yes	No	No	No
Explicitly Models Patient Preferences	No	No	No	No	No	No

A key element for any CPG model is verifying the validity of the recommendations contained therein. For CPG recommendations based on results of clinical studies, users are oftentimes pressed to read the actual reference(s) in the medical literature from which a recommendation was derived to know the specific application of a CPG for a particular situation. Implicitly, it is assumed that the source is documented and immediately accessible through the guideline, the user knows how to interpret clinical trial results, and that time is available to critically evaluate the literature. Unfortunately, these assumptions are often unsubstantiated. The lack of source information in a usable and understandable format has been identified by physicians as a barrier to understanding the validity of CPGs. (14) Therefore, a guideline representation format needs to ensure that a simplified interpretation of the clinical trial data will be available at the point of care to address this issue. This is partially addressed by rule-in criteria for decision options that state facts to support a recommendation against another competing option. However,

there is no heuristic to weigh these criteria. One way to formalize the decision-making process is through the principles of decision analysis. In addition, decision analysis with the use of utility theory addresses the importance of explicitly including patient preferences when decisions are made. This topic is addressed in greater detail in Chapter 4.

References:

1. Purves IN, Sugden RC, Booth NS, Sowerby M. The Prodigy project- the iterative development of the release one model. Proceedings of the 1999 AMIA Fall Symposium, Washington, DC, 1999.
2. Johnson PD, Tu SW, Booth N, Sugden B, Purves IN. A Guideline Model for Chronic Disease Management in Primary Care, SMI-1999-0790, 1999. In: http://smi-web.stanford.edu/pubs/abstracts_by_author/Purves,I.papers.html.
3. Gordon C, Veloso M. The Prestige project: Implementing guidelines in healthcare. Proceedings of MIE96, Copenhagen, 1996.
4. Gordon C, Veloso M. Knowledge representation and clinical practice guidelines: the DILEMMA and PRESTIGE projects. Proceedings of MIE96, Copenhagen, 1996.
5. Herbert SI. Informatics for care protocols and guidelines: Towards a European knowledge model. In: Health Telematics for Clinical Guidelines and Protocols. C Gordon and JP Christiansen (eds), IOS Press, 1994.
6. Fox J, R Thomson. Decision support and disease management: a logic engineering approach. *IEEE transactions on Information Tech in Biomed*, 1998; 2(4): 1-12.
7. Fox J, N Johns, A Rahmanzadeh. Disseminating medical knowledge: the PROforma approach. *AI in Medicine*, 1998; 157-81.
8. Shahar Y, S Miksch, P Johnson. An intention-based language for representing clinical guidelines. *AMIA Proceedings*, 1996: 592-6.
9. Musen MA, SW Tu, AK Das, Y Shahar. EON: a component-based approach to automation of protocol-directed therapy. *JAMIA*, 1996; 2: 367-88.
10. Tu SW, MA Musen. A flexible approach to guideline modeling. Proceedings of the 1999 AMIA Fall Symposium, Washington, DC, 1999.
11. Ohno-Machado L, JH Gennari, SN Murphy, NL Jain, SW Tu, DE Oliver, E Pattison-Gordon, RA Greenes, EH Shortliffe, GO Barnett. The guideline interchange format: a model for representing guidelines. *JAMIA*, 1998; 5: 357-72.
12. Ferrini R. for the American College of Preventive Medicine. Am J Prev Med, 1997; 13(6): 444-6.
13. Peleg M, Boxwala A, Ogunyemi O, Zeng Q, Tu S, Lacson R, et al. GLIF3 Technical Specification. In: http://smi-web.stanford.edu/projects/intermed-web/guidelines/GLIF1.htm; 2000.
14. Langley C, A Faulkner, C Watkins, S Gray and I Harvey. Use of guidelines in primary care – practitioners' perspective. *Family Practice*, 1998; 15(2): 105-11.

IV. Decision Analysis and Utility Theory

4.1 Role in Clinical Practice Guidelines

Decision Analysis is a systematic approach to decision making under conditions of uncertainty (1). It was not originally designed to formulate solutions for problems in patient care. However, it has been widely applied and used in the medical domain where the physician is constantly dealing with choices under uncertainty (2-4). Uncertainty develops from one or more of several sources: (a) errors in clinical data, (b) ambiguity in clinical data and variations in interpretation, (c) uncertainty in medical knowledge about relationships between clinical information and presence of disease and (d) uncertainty about the effects of treatment (5). Another source of uncertainty include missing clinical data.

The decision tree is the fundamental analytic tool used in decision analysis. Other representations include influence diagrams, spreadsheet models, and state-transition models. These represent variables in a decision-making process and their probabilistic relationships. An example of a decision tree is shown in the next page.

Legend:

□ - choice node

O - chance node

∇ - terminal node

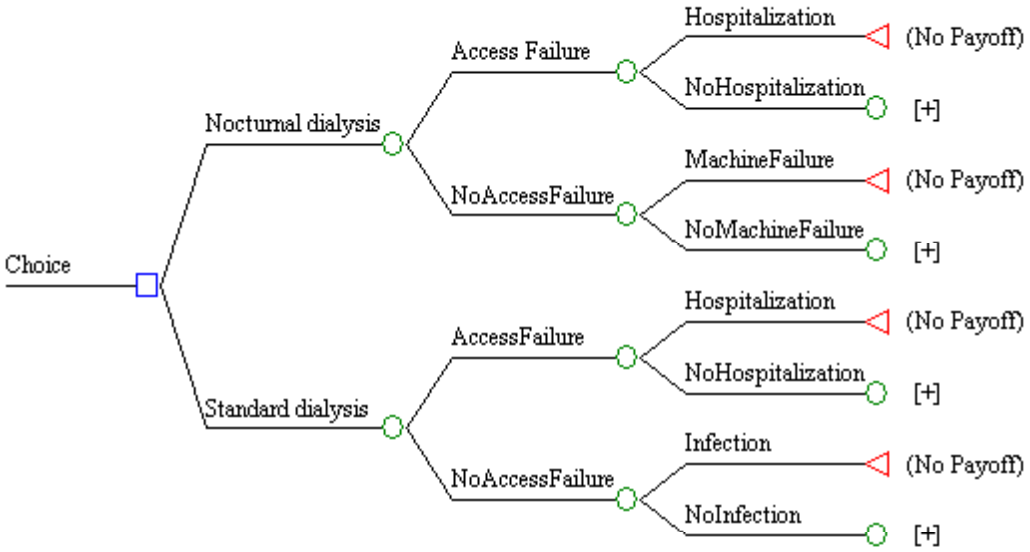


Figure 1: Schematic representation of a decision tree for choosing a dialysis therapy. The tree is explained in detail in the text.

The decision tree above is a schematic display of the temporal and logical structure of the clinical situation where a decision has to be made between alternative options, which are nocturnal dialysis and standard dialysis. In this illustration, the physician or patient must choose which dialysis approach he should take for renal replacement therapy.

A subset of decision analysis includes cost-effectiveness analysis and cost-benefit analysis. Cost-benefit analysis requires that health outcomes be valued in monetary units in order to calculate the net economic benefit of a program. In contrast, cost-effectiveness analysis requires only that a quantitative measure of health effectiveness, commonly referred as utility, be defined. Typically, effectiveness is measured by life years or quality-adjusted life years (QALYs) (6).

Several guidelines have been developed on the basis of decision analysis (7-9). As previously mentioned, work has also been done to automate the conversion of decision analytic trees to CPGs using direct mapping (10) or decision tables (11). This creates a small subset of guidelines that are based solely on decision analytic approaches. Other CPGs may have little uncertainty and therefore have no practical use for decision analytic technique. Some other guidelines contain other inference models to make decisions, or use decision analysis in combination with other inference methods. Furthermore, there are highly complicated guidelines that contain decision analysis as a portion of a well thought out comprehensive management plan (e.g. chronic disease guideline). These last two guidelines described would need representation for a decision model as part of the guideline knowledge representation and will be described in greater detail in the succeeding chapter.

4.2 Components of a Decision Tree

Decision analysis is explicit, quantitative and prescriptive. It forces the decision-maker to break a problem into smaller parts, each of which needs to be analyzed individually and separately (5). Four basic steps are enumerated for the development of a decision tree: (a) Identify and bound the decision problem by creating appropriate choices; (b) Structure the decision problem over time and identifying appropriate variables ; (c) Characterize the information needed to fill in the decision structure; and (d) Choose a preferred course of action based on the result of the decision structure. It is in the first

step where alternative courses of actions are identified when faced with the decision problem.

There are three building blocks of a decision tree (5). 1. A choice node (or decision nodes) denotes a point in the temporal sequence where the decision-maker can select one of several alternative courses of action. This corresponds to the square (□) node in Figure 1 where a choice can be made between standard and nocturnal dialysis. 2. Chance nodes (or probabilistic nodes) denote several points in the temporal sequence where one of several possible events, all beyond the control of the decision-maker, may take place. Each chance node suggests uncertainty, including risks of mortality or therapeutic complication. An example would be the node denoted by a circle (O) in the decision tree for nocturnal dialysis with two branches ("access failure" and "no access failure") coming out from the node. These branches are mutually exclusive and occur probabilistically. 3. Outcomes (terminal nodes) describe what happens at the end of each path of the tree in terms of attributes held to be of value (e.g. health status). It represents what ultimately happens to the patient. An example of a terminal node in Figure 1 is the triangular node (∇) representing hospitalization. Other key concepts include paths (also called scenarios) and strategies. A path in a decision tree is defined as a sequence of events and actions beginning with a particular choice node, and following a particular event or choice at each subsequent node, from left to right. A strategy is a sequence of decisions made over time, based on information made available at the time a decision is to be made.

I have incorporated the decision analysis concepts described above with the exception of strategies into the GLIF model. A decision step in GLIF represents the choice node and only one choice can be made at that step. However, subsequent choices in the decision tree may be included in a CPG as another decision step elsewhere in the guideline.

The outcome of a decision analysis path allows a method to represent QALYs and utilities for treatment selection. As discussed in the succeeding section, this enables modeling of patient-specific variables in a decision-making process.

4.3 Use of Utilities to Represent Patient Preferences

Utility theory has been in use since the 1940's. Utilities represent the strength of an individual's preference for particular outcomes (12). In our case, this is a single numerical measure that reflects the total improvement in health that can be expected from each of several alternative courses of action. An important thing to remember about utilities is the fact that this is not how an individual actually makes decisions in the face of uncertainty. Rather, this is a prescriptive model that has been developed on how the user should make decisions if they wish to act rationally and consistently. It is specifically designed for individual decision-making. However, it may be aggregated to provide a group utility function as well.

QALYs were developed in the 1970's. It is a method that could represent total health improvement for a group of individuals, capturing improvements in both quantity and quality of life. Utilities may be used as the quality-adjustment weights for QALYs. The

use of utilities as weights leads to a variation of cost-effectiveness analysis, known as cost-utility analysis.

4.4 Methods to Elicit Utilities

In order to elicit utilities for clinical decision-making, all possible outcomes of various treatment alternatives are enumerated. These include various health states, both temporary and chronic, for which utilities are required. Patients rank health outcomes by order of preference. Subsequently, health utility values are assigned to each of the ranked health outcomes using one of several methods described below.

The visual analog scale uses a straight line with markers ranging from 0 to 1. A patient is asked to mark where in this line they would value their current state to be. The utility is measured as the number corresponding to that point in the scale between 0 and 1. This is usually used to perform the preference ranking as described in the previous paragraph. However, since the values obtained are developed under certainty instead of under risk, they are usually not considered true utilities. Instead, they are referred to as values and are not thought to be good approximations of utilities. I used the visual analog scale to present the utilities elicited using standard gamble to the panel of expert physicians who determine the appropriateness of recommendations based on these patient preferences. The use of the visual analog scale to represent utilities has the advantage of being easier to explain to physicians.

Another method used to elicit utilities is the time-tradeoff technique. The goal is to find the fraction of a year of perfect health that is perceived equivalent to one whole year in the current health-state. This has been shown in some studies to have results comparable to standard gamble (see below) (13).

The benchmark method used to elicit utilities is by standard gamble. There are four steps to elicit a patient's utility. First, a ranking of health states is obtained. Second, the worst possible health state is assigned a utility of 0 and the best state a utility of 1. Third, standard reference gambles are performed on the intermediate outcomes. A diagram that illustrates this gamble is shown in Figure 2 below. Fourth, the break-even probability is obtained where a patient is indifferent between being in the intermediate health-state and accepting the gamble for potentially being in the best vs. the worst health-state. This was the method used in eliciting the patient preferences that were used in generating recommendations for this study.

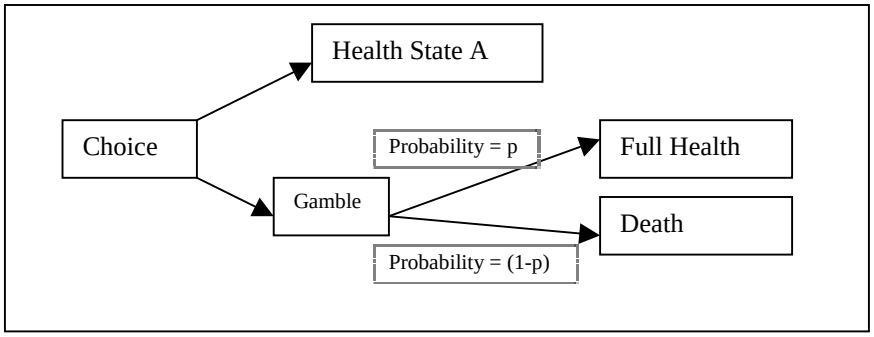


Figure 2: Diagram illustrating the standard gamble technique for determining the value or utility for one health state, A, by eliciting the break-even probability, p, where a patient is indifferent between being in health-state A and accepting the gamble for potentially being in the best (Full health) vs. the worst (Death) health-state.

Whatever methods are used to measure utility, it is important that the health state descriptions be precise and unambiguously written with the interviewer trained to be

consistent. Lastly, it is also important to identify the subjects whose utilities will be measured. This may include patients, health care professionals, health care planners/administrators, and/or members of the general public. There are advantages and disadvantages associated with each choice that have been discussed elsewhere (12).

References:

1. Raiffa H. Decision Analysis: Introductory lectures on choices under uncertainty. Reading, MA: Addison-Wesley Publishing Co, Inc, 1968.
2. Col NF, MH Eckman, RH Karas, SG Pauker, RJ Goldberg, EM Ross, RK Orr and JB Wong. Patient-specific decisions about hormone replacement therapy in postmenopausal women. *JAMA*, 1997; 277(14): 1140-7.
3. Koff RS. Cost-effectiveness of treatment for chronic hepatitis C. *J Hepatol*, 1999; 31 (Suppl 1): 255-8.
4. Kerlikowske K, P Salzmann, KA Phillips, JA Cauley, SR Cummings. Continuing screening mammography in women aged 70 to 79 years: impact on life expectancy and cost-effectiveness. *JAMA*, 1999; 282(22): 2156-63.
5. Weinstein MC, Fineberg HC, et. al. Clinical Decision Analysis. Philadelphia, PA: W.B. Saunders, 1980.
6. Eisenberg JM. A guide to the economic analysis of clinical practices. JAMA, 1989; 262(20): 2879-86.
7. Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening: a decision analysis model. Annals of Internal Medicine, 1993; 118(11): 838-43.
8. American College of Physicians. Screening for cervical cancer. In: Eddy DM (ed). Common Screening Tests. Philadelphia, PA: American College of Physicians, 1991.
9. American College of Physicians. Screening for osteoporosis in perimenopausal women. In: Eddy DM (ed). Common Screening Tests. Philadelphia, PA: American College of Physicians, 1991.
10. Sanders GD. Automated creation of clinical-practice guidelines from decision models. Ph.D. Dissertation, Stanford Medical Informatics, SMI Report No. SMI-98-0712, Department of Computer Science Report No. STAN-CS-TR-98-1609. Stanford University, Stanford, CA. July 1998.
11. Shiffman RN, RA Greenes. Use of augmented decision tables to convert probabilistic data into clinical algorithms for the diagnosis of appendicitis. *AMIA Proceedings*, 1992: 686-90.
12. Torrance GW, D Feeny. Utilities and quality-adjusted life years. Intl J of Technology Assessment in Health Care, 1989; 5: 559-75.
13. Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. Socio-Economic Planning Sciences, 1976; 10: 129-36.

V. Model that Incorporates Decision Analysis Concepts in Guidelines

5.1 GLIF3 and Decision Steps

Guideline Interchange Format, Version 3.0 (GLIF3) contains decision steps that direct flow conditionally from one guideline step to another (1). A decision step contains several "decision option" objects to represent each of the options of a decision. Each decision option contains a "decision condition" object to determine the flow of control to the next step. This decision step is actually a decision hierarchy. It has two further subclasses: a case step and a choice step, explained below.

A case step contains decision options that are mutually exclusive. When the condition in the case step matches a condition in one of its options, the flow moves deterministically to the next step specified by that option. A default next step may be specified in case there is no condition among the available options that matches the condition of the case

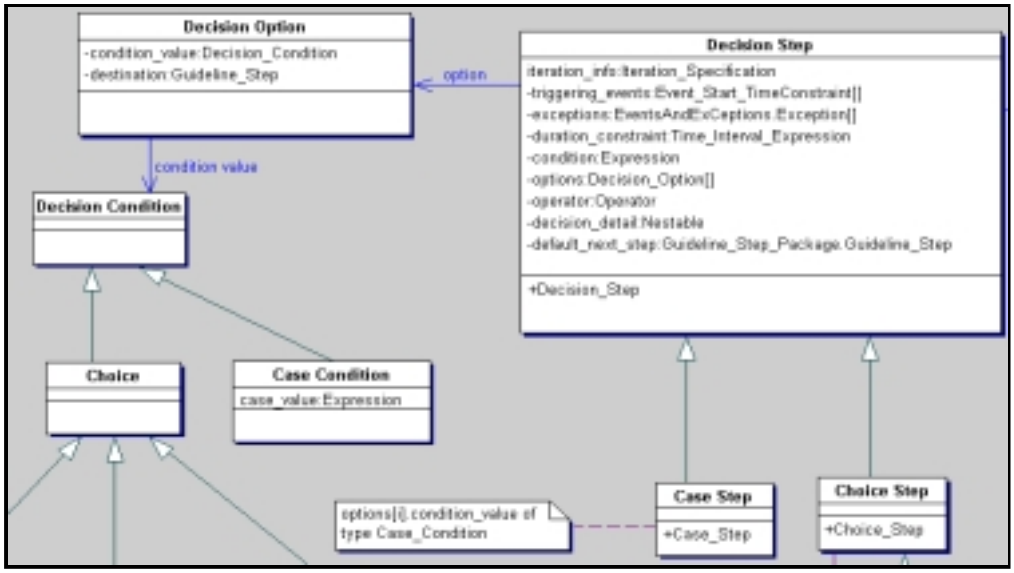


Figure 1. GLIF3 Decision Hierarchy Without the Utility Model. The case step and choice step are subclasses of a decision step. A decision step contains decision options. Each decision option contains a decision condition. In a case step, this decision condition has to be of class case condition.

step. The decision condition of each one of the decision options of a case step has to be of subclass “Case Condition” (see Figure 1).

A choice step may contain options that are not mutually exclusive. Each one of the options specified in this step is evaluated. These options may be ranked following evaluation based on the class of decision condition contained in the decision option. The decision condition of each one of the decision options of a choice step has to be of the “Choice” class or any of its subclasses (e.g. RuleIn Choice, kOfN Choice). Furthermore, the decision conditions of all decision options of a choice step (or its subclasses) should belong to the same subclass. For example, a RuleIn Choice Step should have all decision options with a decision condition of subclass “RuleIn Choice” (a subclass of “Choice”). This is used in the evaluation and presentation of options to the guideline user at execution time. To assist the user in choosing among the competing options, they may be ranked according to a degree of preference for each option. The user may select only one of the provided options.

The model I created extends the choice step by creating a subclass called a utility choice step. This extended model has been incorporated into the draft GLIF3 specification. As seen in Figure 2, the model I designed contains a decision step that has a choice step that is further sub-classed into a Utility Choice Step. The model was created using Together/J (TogetherSoft LLC, Raleigh, NC), an object-oriented modeling tool employing the Java language.

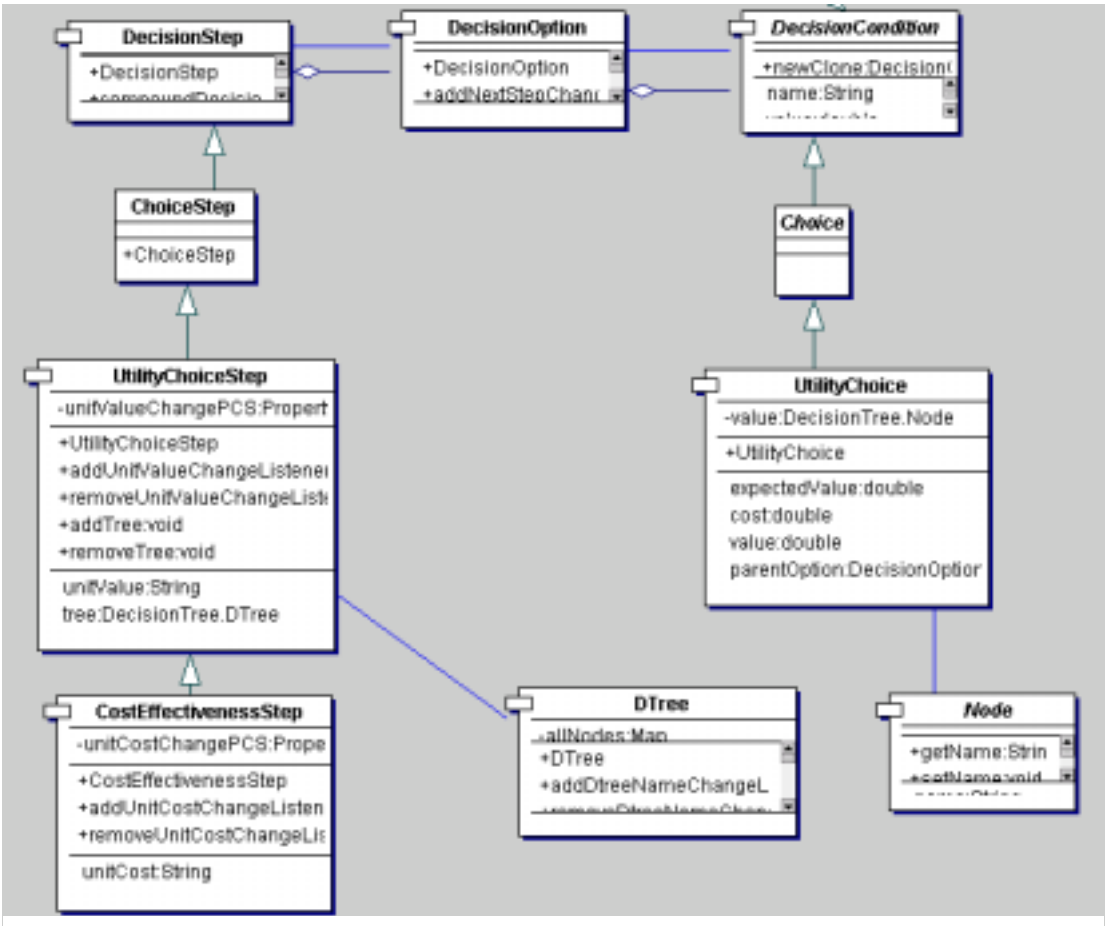


Figure 2. Extended Decision Model. A utility choice step inherits from a choice step. It contains an attribute of class Dtree which represents a Decision Tree. The decision condition of each of the decision options of a utility choice step is of class utility choice, which inherits from the choice class. A utility choice contains an attribute of class node, which represents a branch of a choice node in the decision tree contained in the utility choice step.

A utility choice step contains decision options that should have decision conditions of class "utility choice." Furthermore, a utility choice contains the branches of choice nodes in the decision model (e.g. usually the initial nodes of a decision tree model). However, it is not limited to the initial nodes of a tree; subsequent decision steps in a guideline may include branches of other choice nodes that are contained further down a path of a specified decision tree.

As previously mentioned, the utility choice class contains a node from the decision tree. This node has an expected utility computed from the path probabilities and the utility of the outcome. The utility values are comparable to other corresponding nodes at the same level of a decision tree. More importantly, the utility values allow ranking of these choices to be performed. The ranked list may then be presented to a user during execution time.

5.2 Decision Nodes as Options in a Choice Step

For this dissertation, I elected to use decision trees to model decision analysis because it is widely available and frequently used in the medical domain. As previously stated, it contains an expected utility value obtained from evaluating the entire decision tree. In this model, ranking may be specified in increasing or decreasing order, depending on the type of utility specified by the user at execution time. To illustrate, if the utility measures QALYs, an option that contains a node with the highest utility value often indicates the most preferred option. Conversely, if the utility measure is mortality, then the lowest utility value is most preferred.

To put this in an object-oriented perspective, the decision tree is a class that contains three types of nodes, a choice node, a chance node, and a terminal node. Each of these nodes is a distinct object and inherits from the class "Node" (See Figure 3). During the development process of a guideline, the author identifies which node of a decision tree matches a specific utility choice. It is noteworthy to mention that this assignment process is limited to nodes that are branches of a choice node.

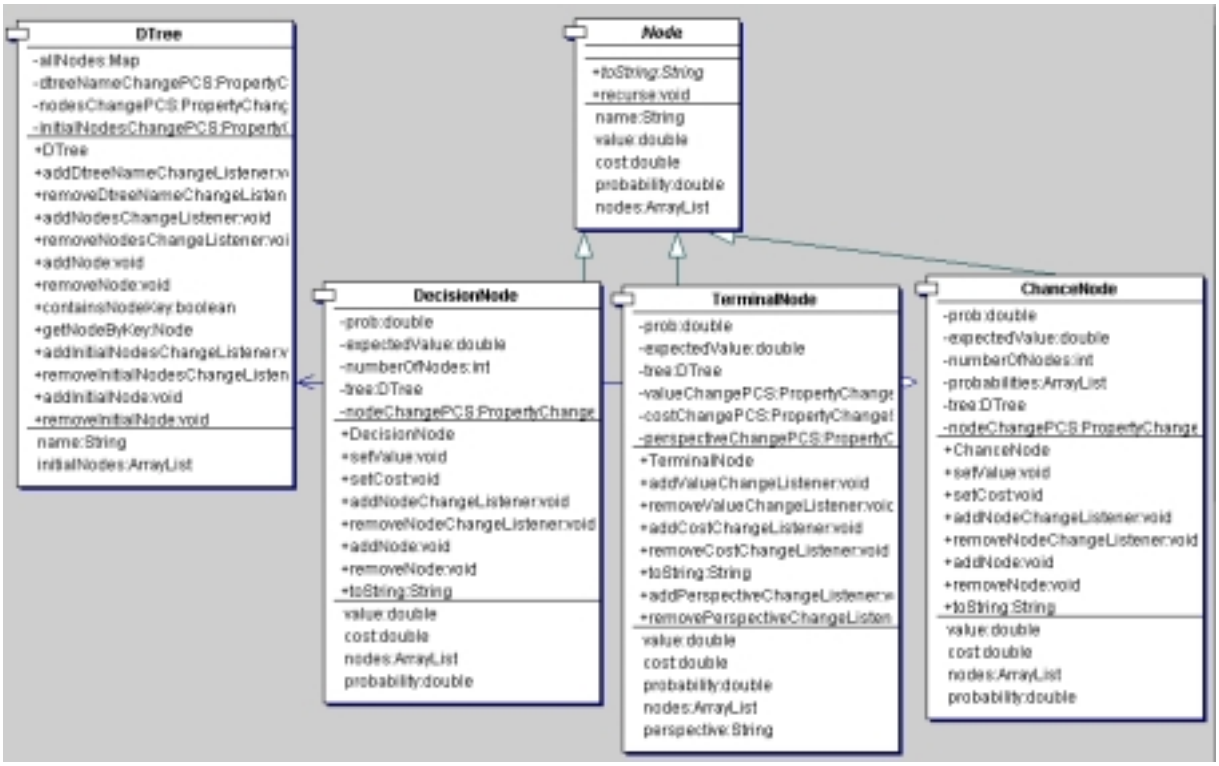


Figure 3. Decision Model. The three main elements of a decision tree are decision nodes, terminal nodes and chance nodes. These nodes all have expected utility values. A decision or choice node typically branch into alternative choices, corresponding to any type of node.

Recall that decision trees are chronologically arranged from left to right. It is therefore logical to assume that the utility measured at a specific choice node is appropriate to rank all the options at that given point in time (but there may be more than one choice node in a tree). Consequently, there is a chance that in a subsequent decision step, a different choice may be taken at execution, which may invalidate the ranking of the branches of the initial choice node. Far from being a weakness, however, this occurrence merely reflects the uncertainty involved in decision-making.

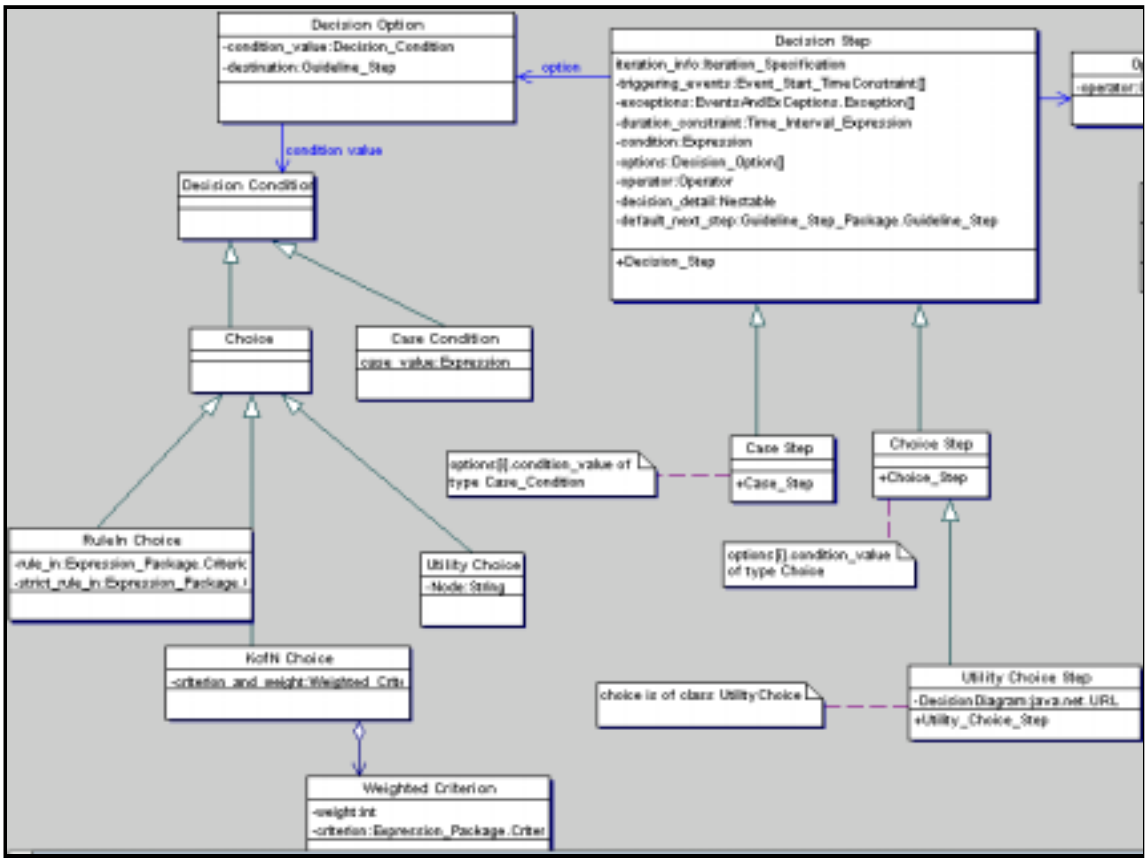


Figure 4. GLIF3 Decision Hierarchy with the Decision Model. As previously noted in Figure 2, the utility choice step inherits from the choice step in GLIF3.

For the purpose of this dissertation, I am using three decision tree models from the medical literature assuming that they are appropriately constructed and validated (2-4). Each of these three decision models contains only one choice node. Furthermore, the branches of these choice nodes are identified as the nodes specifically assigned to the corresponding utility choice of the guideline model. The specifications of the extended GLIF3 model that supports the decision model (see Figure 4) were used to represent three guidelines, which will be discussed in the next paragraphs.

5.3 Authoring Guidelines in the Extended Model

Three guidelines from three different domains (prevention, diagnostic workup and treatment) were represented in the model, and extended to include three domain-specific decision analysis models, respectively (2-8).

As stated earlier, the guidelines were encoded manually in Extensible Markup Language (XML) using a text editor. Moreover, I used Protégé, a knowledge-modeling tool (9), tailored to represent the extended GLIF specification, for modeling guideline knowledge. This created a text file in the Resource Description Framework (RDF) format (10). Similarly, the traditional guidelines without the decision-analytic steps involved (used as control) were also encoded using Protégé.

```
<guideline xmlns="x-schema:glSchema.xml">
   <name>Sleep Apnea</name>
   <algorithm>
      <firstStep>S1</firstStep>
      <steps>
         <actionStep ID="S1">
            <stepName>Evaluation</stepName>
            <tasks>
               <actionSpecification>
                  <description>get history and physical
examination</description>
                  <intention>to assess urgency or severity of
symptoms</intention>
               </actionSpecification>
            </tasks>
            <nextStep>S2</nextStep>
         </actionStep>
         <utilityChoiceStep ID="S2">
            <stepName>Diagnostic test for sleep apnea</stepName>
            <dtree ID="D1">
               <dtreeName>Sleep Apnea Decision Tree</dtreeName>
               <nodes>
                  <chanceNode ID="CN1">
                     <chanceNodeName>Polysomnogram</chanceNodeName>
                     <chanceNodeValue> __ </chanceNodeValue>
                     <tree>D1</tree>
                     <nodes>
                        <node>CN2</node>
                        <node>CN3</node>
                     </nodes>
                     <probabilities>
                        <probability>0.85</probability>
                        <probability>0.15</probability>
                     </probabilities>
                  </chanceNode>
```

Figure 5a. XML encoding of a fragment of the Sleep Apnea guideline. This guideline object contains a name and an algorithm. There are two steps above, an action step and a utility choice step. The utility choice step contains a name and a decision tree, which contains several objects of class node.

A fragment of the XML and RDF documents is shown in Figures 5a and 5b to illustrate the decision step model that contains a decision tree class. The algorithm derived from Protégé is shown in Figures 6, 7 and 8 for each of the three guidelines represented.

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<!-- Version Fri Apr 14 14:53:45 EDT 2000 -->
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:rdfs="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#"
 xmlns:rdfutil="http://www.w3.org/rdfutil#"
 xmlns:protege="http://smi-web.stanford.edu/projects/protege/protege-rdf/protege-19992012#"
 xmlns= "http://www.rdfschema.org/mynamespace.rdf#">

 <rdf:Description ID="#Algorithm">
  <rdf:type resource="http://www.rdfschema.org/mynamespace.rdf#Network_Metaclass"/>
<node_slot><rdf:Description ID="#steps"/> </node_slot> </rdf:Description>

 <rdf:Description ID="#sleepApnea_INSTANCE_00175">
  <rdf:type resource="http://www.rdfschema.org/mynamespace.rdf#Local_Material"/>
  <material></material>
  <MIME_type></MIME_type>
  <name>American Sleep Disorders Association's Standards of Practice Committee. Practice
parameters for the use of portable recording in the assessment of obstructive sleep apnea.
Sleep. 1994; 17(4): 372-7.</name>
 </rdf:Description>

 <rdf:Description ID="#sleepApnea_INSTANCE_00064">
  <rdf:type resource="http://www.rdfschema.org/mynamespace.rdf#Local_Material"/>
  <material></material>
  <MIME_type></MIME_type>
  <name>American Sleep Disorders Association's Standards of Practice Committee. Practice
parameters for the use of portable recording in the assessment of obstructive sleep apnea.
Sleep. 1994; 17(4): 372-7.</name>
 </rdf:Description>

 <rdf:Description ID="#Branch_Destination">
  <rdf:type resource="http://www.rdfschema.org/mynamespace.rdf#Connector_Metaclass"/>
<first_object_slot_pointer><rdf:Description ID="#branches"/> </first_object_slot_pointer>
</rdf:Description>

 <rdf:Description ID="#sleepApnea_INSTANCE_00179">
  <rdf:type resource="http://www.rdfschema.org/mynamespace.rdf#Local_Material"/>
  <material></material>
  <MIME_type></MIME_type>
  <name>Chervin RD, Murman DL, Malow BA, Totten V. Cost-utility of three approaches to the
diagnosis of sleep apnea: polysomnography, home testing and empirical therapy. Ann Intern
Med. 1999 Mar 16; 130(6): 496-505.</name>
 </rdf:Description>

 <rdf:Description ID="#sleepApnea_INSTANCE_00184">
  <rdf:type resource="http://www.rdfschema.org/mynamespace.rdf#Decision_Destinations"/>
<first_object><rdf:Description ID="#sleepApnea_INSTANCE_00165"/> </first_object>
<name>choice1</name>
<second_object><rdf:Description ID="#sleepApnea_INSTANCE_00115"/> </second_object>
</rdf:Description>
```

Figure 5b. RDF output from the Protégé representation of the Sleep Apnea guideline.

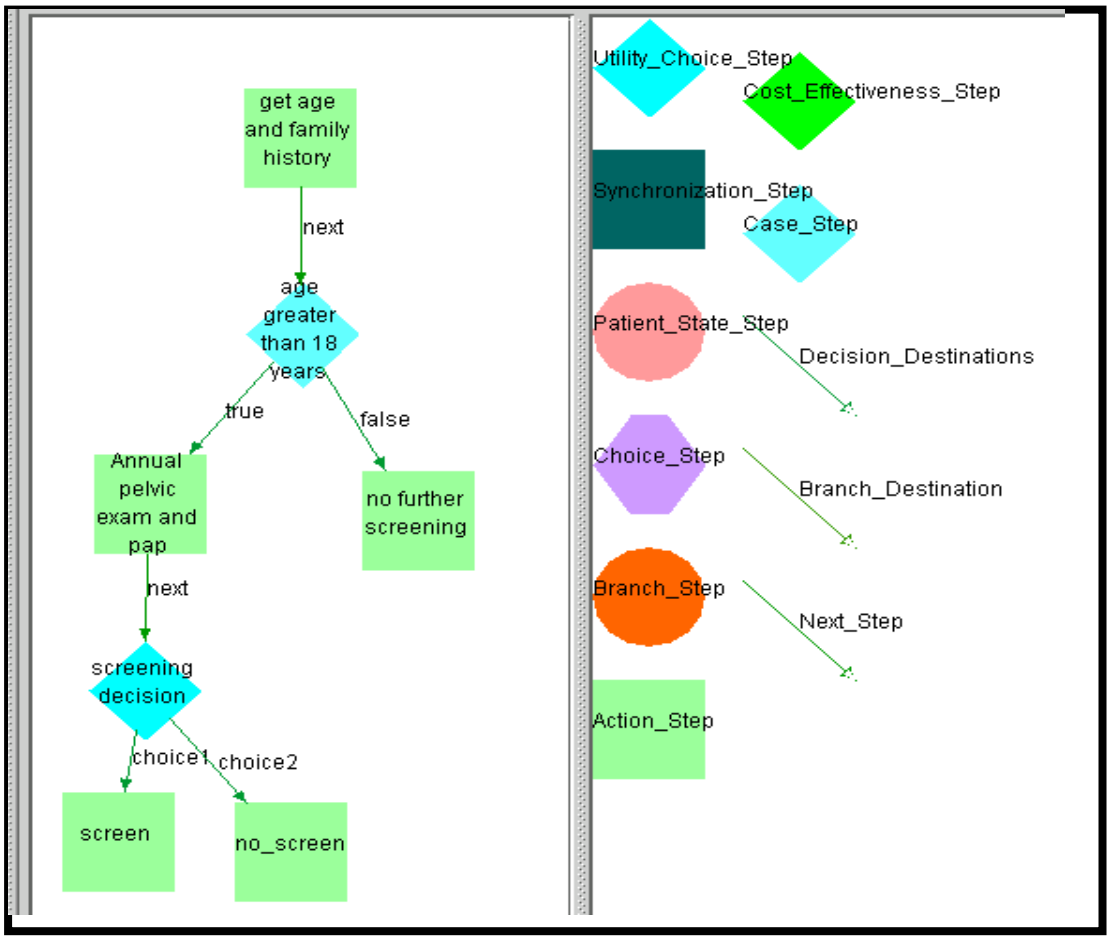


Figure 6. Ovarian Screening Guideline in the Extended Decision Model using Protege. The lower diamond-shaped object is a utility choice step, which has 2 decision options. A decision option would have a decision condition of class utility choice. It also has a corresponding next step. The 2 next steps for each of the decision options are represented by the 2 bottom square boxes, screen and no_screen.

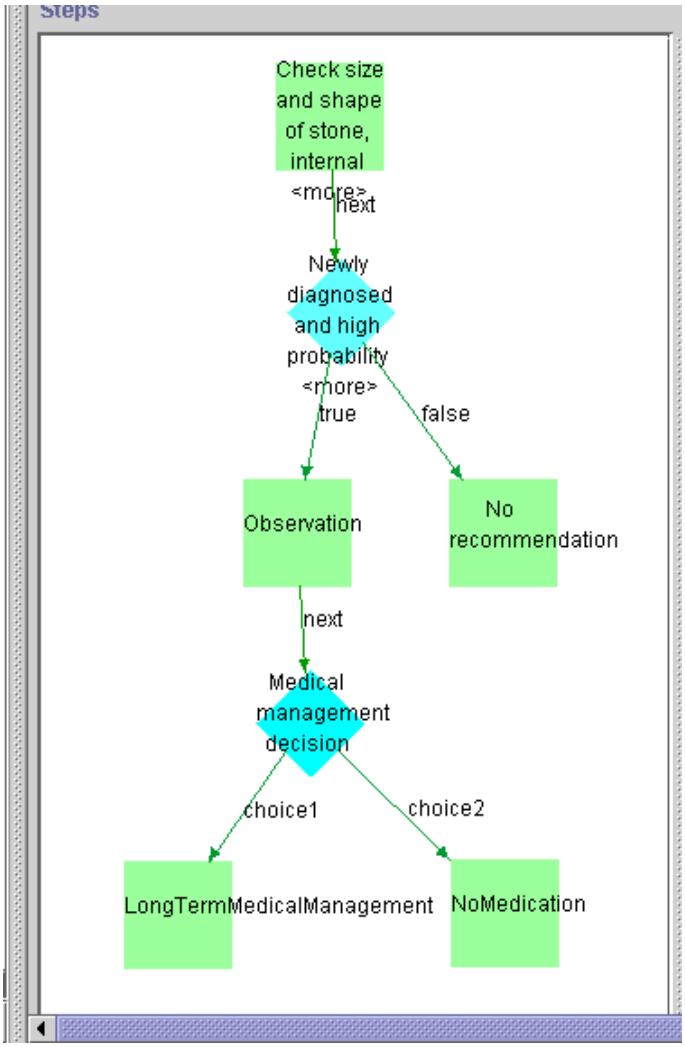


Figure 7. Ureteral Stone Management Guideline in the Extended Decision Model using Protege. The lower diamond-shaped object is a utility choice step (Medical management of stones) with decision options named "long term medical management" and "no medications."

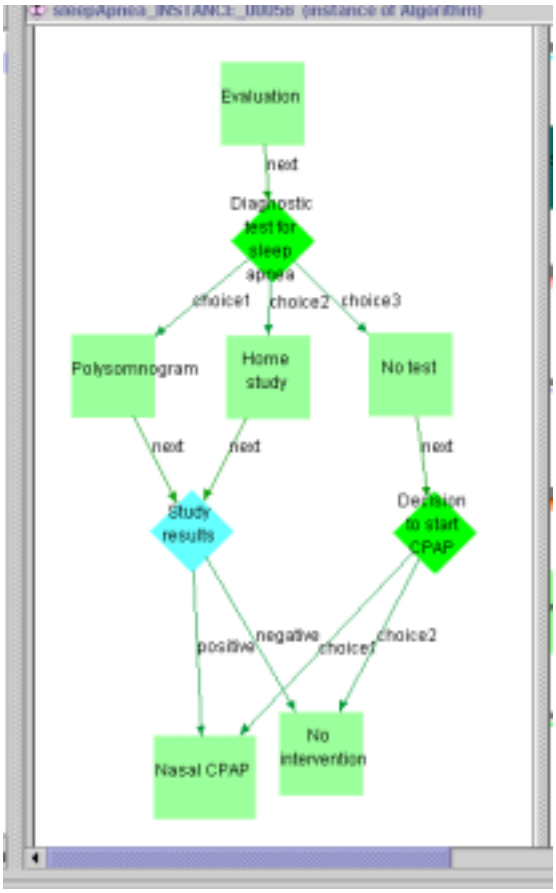


Figure 8. Sleep Apnea Diagnostic Workup Guideline in the Extended Decision Model using Protege. The uppermost diamond-shaped object is a utility choice step with three decision options. These decision options have corresponding next steps, polysomnogram, home study and no test, from which a user can choose from, depending on the expected utility of the node in the decision option's utility choice.

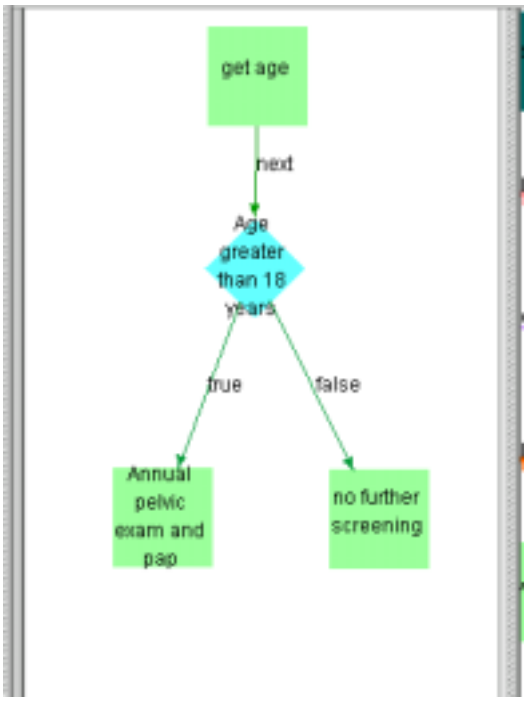


Figure 9. Ovarian Screening Guideline in the Traditional Model using Protege. The guideline does not have a utility choice step.

5.3.1 Preventive Guideline

The guideline for the screening of ovarian cancer represents this domain (5). Typically, a patient's age and family history identify their risk of acquiring this disease. However, no screening is recommended for early detection of ovarian cancer because of the low incidence of the disease and the inadequate specificity of screening techniques. Figure 9 illustrates the traditional guideline for screening ovarian cancer.

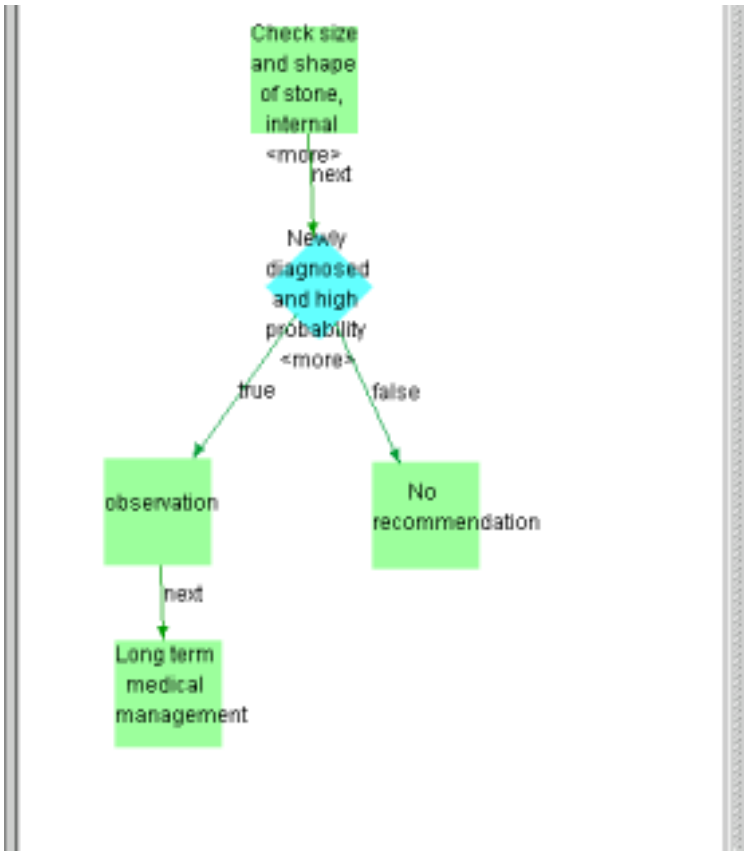


Figure 10. Ureteral Stone Management Guideline in the Traditional Model using Protege. The action step "long term medical management" is the only next step allowed after "observation," when contrasted to the model in Figure 7 where the utility choice step has 2 options.

5.3.2 Management Guideline

The guideline for medical management of recurrent ureteral stones (urolithiasis) represents this domain (6,7). The initial steps are designed to manage the presenting stone. If the stone has a high probability of spontaneous passage and the symptoms are controlled, the recommendation is limited to observing the patient. Subsequently, a decision has to be made whether long-term management is necessary and a decision tree

(3) is modeled to make this decision explicit. Figure 10 illustrates the guideline in the traditional model.

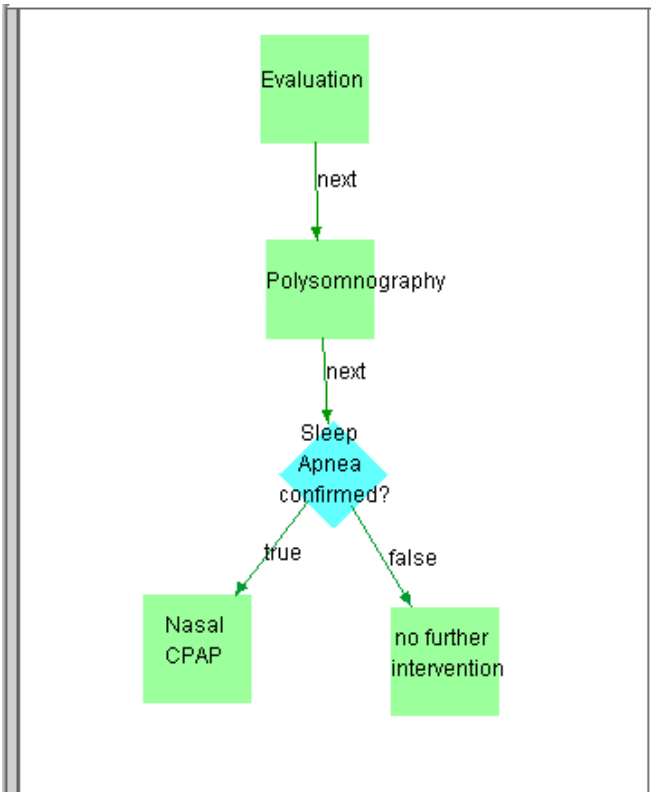


Figure 11. Sleep Apnea Diagnostic Workup Guideline in the Traditional Model using Protege. This model contains an action step "polysomnogram" in place of the utility choice step in Figure 8, which has 3 decision options.

5.3.3 Diagnostic Guideline

The guideline for diagnostic work-up of obstructive sleep apnea syndrome (OSAS) represents this domain (8). OSAS is a medical problem affecting 4% of the middle-aged adults in which complete or partial obstruction of the airway occurs during sleep (11). This disease often results in significant morbidity and even some mortality. Common complaints include loud snoring, disrupted sleep and excessive daytime sleepiness. Polysomnography in a sleep laboratory (sleep study) is the gold standard test for

confirming the diagnosis of sleep apnea. However, the test is expensive and is not widely available. Home sleep studies are less expensive but have lower diagnostic accuracy. A decision tree with a comparison of three options to work-up sleep apnea was recently published (4). The competing choices for the diagnosis of OSAS are shown in Figure 8 and the guideline in the traditional model is shown in Figure 11.

References:

1. Peleg M, Boxwala A, Ogunyemi O, Zeng Q, Tu S, Lacson R, et al. GLIF3 Technical Specification. In: http://smi-web.stanford.edu/projects/intermed-web/guidelines/GLIF1.htm; 2000.
2. Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening: a decision-analysis model. Annals of Internal Medicine, 1993; 118: 838-43.
3. Kuo RL, Aslan P, Abrahamse PH, Matchar DB, Preminger GM. Incorporation of patient preferences in the treatment of upper urinary tract calculi: A decision analytical view. The Journal of Urology, 1999 Dec; 162: 1913-9.
4. Chervin RD, Murman DL, Malow BA, Totten V. Cost-utility of three approaches to the diagnosis of sleep apnea: polysomnography, home testing, and empirical therapy. Annals of Internal Medicine, 1999; 130: 496-505.
5. NIH Consensus Development Panel on Ovarian Cancer. Ovarian cancer: Screeening, treatment and follow-up. JAMA, 1995 Feb; 273(6): 491-7.
6. Segura JW, Preminger GM, Assimos DG, Dretler SP, Kahn RI, Lingeman JE, Macaluso JN. The management of ureteral calculi. American Urological Association, Inc., 1997 Sep; 72(9/97).
7. Goldfarb DS, Coe FL. Prevention of recurrent nephrolithiasis. American Family Physician, 1999 Nov; 60(8): 2269-76.
8. Thorpy M, Chesson A, Ferber R, Kader G, Millman R, Reite M, et.al. Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea. Sleep, 1994; 17(4): 372-7.
9. Musen MA, Gennari JH, Eriksson H, Tu SW, Puerta AR. Protégé II: Computer support for development of intelligent systems from libraries of components. MEDINFO 95, The Eighth World Congress on Medical Informatics. Vancouver, B.C., Canada, 1995: 766-70.
10. Lassila O, Swick RR. Resource Description Framework (RDF) Model and Syntax Specification. W3C Proposed Recommendation, 1999 Jan, PR-rdf-syntax-19990105. In: http://www.w3.org/TR/PR-rdf-syntax/; 2000.
11. Victor LD. Obstructive sleep apnea. American Family Physician, 1999; 60(8): 2279-86.

VI. Evaluation of the Acceptability of the Guidelines in the Extended Decision Model

I conducted an evaluation study to assess the acceptability of the extended decision model in representing variations in recommendations that can be tailored to specific patient preferences. The first part addresses how the utilities were elicited and then used to create patient cases. The latter part of the study addresses how the preferences expressed as utilities were used to generate situation-specific recommendations.

Clinical vignettes were artificially created by the investigator from utilities elicited from volunteers in the study laboratory. Each vignette was further divided into two distinct cases, one with recommendations generated from the decision model and the other one with recommendations that are not based on the decision model. The primary analysis is designed to compare the adequacy of the recommendations for the two groups for the cohort identified.

6.1 Eliciting Utilities using Standard Gamble

Utilities were obtained from 15 adult volunteers at the research laboratory where this study was conducted. Verbal informed consent was obtained from each participant. All the 15 people that were asked to participate in the study agreed to do so. Nine of the 15 participants were female and were asked to give utilities for all three guidelines. The remaining 6 were male and were excluded from the standard gamble procedure to elicit utilities for the ovarian cancer guideline.

Using the utilities from all the participants for each of the three guidelines, 39 total patient cases were generated artificially. Only the preferences were considered to be dependent variables and these were based on utilities of the volunteers. The rest of the patient scenario presented to the expert physicians for evaluation was artificially generated. Thirty to forty cases were needed in order to have 80-90% power to detect a 0.25 difference in proportion between the percentages of appropriate recommendations generated by the new model that incorporates decision analysis, compared to the traditional model. Expert physicians determined the appropriateness of recommendations on an ordinal scale (discussed below). A thesis from Stanford University that used an ordinal scale to measure quality of automatically generated recommendations from a decision model was used as the basis for the projected 0.25 difference in appropriateness between the guidelines in the traditional and the extended decision models (1). The ordinal scale (0, 1, 2), used to measure appropriateness of guideline recommendations in that study, was converted into a continuous scale (0-1) by dividing the total score by 2. The difference between models was computed to be 0.25. This value was extrapolated for this dissertation to represent the best estimate based on the limited applicable evidence from the literature.

Utilities were elicited in a uniform manner from each volunteer by a single interviewer (the investigator). This assured consistency in generating the utilities. A questionnaire was used for each participant. The initial questionnaire was adapted from the study conducted to incorporate patient preferences in the treatment of upper urinary tract calculi (2). The descriptions of the actual scenario were copied from the same published study.

Similarly, the utilities were subsequently used in the same decision analysis model as was described therein (2). Questionnaires for the ovarian cancer screening and sleep apnea diagnostic workup guidelines were developed de novo by the investigator specifically for this dissertation. The questionnaires were inspected by and discussed with colleagues (knowledgeable about the standard gamble technique) before being used in the study. A copy of the questionnaires can be found in Appendix A.

6.2 Creation of Abstract Patient Cases with Actual Utilities

As previously stated, the patient cases were artificially generated based on typical presentations of patients with urolithiasis, OSAS, and women screened for ovarian cancer. Within each guideline domain, only the patient preferences varied among the cases. To illustrate, the same 38 year-old male with ureteral stones will be presented for all test cases for the urolithiasis guideline. However, the patient preferences for treatment will be varied in each case to create different patient scenarios. Recommendations generated with and without the use of a decision model may vary between similar patient scenarios. To go back to the 38 year-old male patient with stones, if the traditional guideline recommends a particular therapy (regardless of expressed preference for that therapy) such as surgery, the guideline generated with the decision model might avoid a recommendation for which the (hypothetical) patient was noted to be treatment-averse (e.g. medical therapy in lieu of surgery).

The preferences (expressed as utilities) were presented in the questionnaire given to the experts using the Visual Analog Scale. (see Chapter 4) The patient states matching the

terminal nodes of the corresponding decision tree for each guideline were presented as part of the patient abstract. As previously stated, this would be a suitable proxy for patient preferences to represent those preferences a physician might elicit during an actual patient encounter. The preferences were presented as a continuous number between 0 and 100.

6.3 Generation of Recommendations Based on the Extended Decision Model and the corresponding Traditional Guideline Model

Recommendations were generated for each patient case, corresponding to the appropriate guideline being evaluated. In the case of the traditional guideline model, there is no variation in the preferred recommendation for each patient case. For the extended decision model, the recommendations are derived from the expected utilities corresponding to the branch of the choice node that is contained in the decision option.

Results reveal that the agreement rate between recommendations generated when patient preferences are considered, compared to the traditional guideline model vary between 0.60 to 1.0. Table 1 (below) shows the proportion of recommendations that matched the preferred option when using the decision model compared to the traditional guideline model. Results are shows for each of the three guidelines. In two of the three domains (ureteral stone and sleep apnea), the recommendations indeed varied according to patient preferences.

Guideline	**Agreement Rate between Guidelines in the Extended Decision Model and the Traditional Model**
Ovarian Cancer	9/9 = 1.00
Ureteral Stone	9/15 = 0.60
Sleep Apnea	10/15 = 0.67

Table 1: Table showing the proportion of recommendations generated with the decision model that matched the preferred option using the traditional guideline model

6.4 Questionnaires and Expert Evaluation of Recommendations

The goal of this study is to compare the adequacy of the recommendations derived from the decision model compared to the recommendations from the traditional CPG model that did not incorporate decision analysis. Expert physicians evaluated the two sets of recommendations for each of the cases. The experts were provided with the patient cases and the corresponding recommendations generated using the traditional model, as well as patient cases with recommendations generated from the decision model using decision analyses reproduced from published studies (2-4). Their responses were solicited in a questionnaire. These questions were adapted from two studies that compared two models in generating guidelines (1,5). The questionnaire used contains the following items:

1. Would you follow the recommendations? __yes __no
2. Are the patient preferences considered in (generating) the guideline recommendations? __yes __no
3. Comments_____

Appendix B contains the cases and questionnaires that were provided for the experts. The questionnaires were answered by two domain experts for each of the three guidelines. In

selecting the expert physicians, the subspecialties that would be primarily responsible for the three guidelines chosen for this study were identified. Subsequently, board-certified physicians in the corresponding subspecialty from university teaching hospitals who were either known to the investigator or were referred by colleagues were asked to participate in the study. The expert physicians have no previous knowledge of the study or its goals. They were blinded to the study design and the source of model generating the recommendations for each patient case.

The experts received two questionnaires for each patient scenario. However, one questionnaire contained recommendations that were generated using the traditional guideline model and another contained recommendations generated using the decision model. The questionnaires were arranged randomly to minimize easy comparisons for identical cases. In spite of this, an expert may conceivably notice that two questionnaires have identical cases and different recommendations. They may consider it appropriate to agree with one recommendation and not the other for an identical case. This would bias the study towards detecting a difference in acceptability between the two guidelines. A sample questionnaire for each of the three guidelines is attached in Appendix B. An accompanying instruction sheet was given to each of the experts who participated in the study to clarify the patient scenarios, the meaning of the utilities, the visual analog scale and the questions being asked. This can be viewed in Appendix C.

The experts were asked to return the questionnaires after answering the three questions stated above. The answer to the first question measures the acceptability of the

recommendation. The proportion of acceptable recommendations for guidelines that were generated using the traditional model was compared to the proportion of acceptable guidelines that were generated using the decision model.

6.5 Results

The results of the study showed good agreement rate between the two experts in each guideline domain. The agreement rates were 1.0, 0.967 and 1.0 for the ovarian cancer, ureteral stone and OSAS guidelines, respectively. Kappa was not appropriate as a measure of inter-rater agreement because there were columns in the two-by-two table that had 0 values, which automatically yield a kappa of 0, because kappa considers chance agreement.

Using the mean values obtained from the two experts for each patient case, Fisher's Exact test was used for comparing proportions from both guideline models in generating a 2-sided p-value. Table 2 below shows the results of this study.

Guideline	Utility Model	Traditional Model	p-value (2-sided)
Ovarian Cancer	1.000 (9/9)	1.000	1.000
Ureteral Stone	1.000 (15/15)	0.966	1.000
Sleep Apnea	0.400 (6/15)	0.733	0.013
Combined (n=39)	0.769 (30/39)	0.884	0.077

Table 2: Table showing the proportion of cases expert physicians would follow the recommendations using the two models

Other than for the OSAS guideline, where the expert physicians were more likely to follow the recommendations of the traditional guideline, there was no significant difference in the proportion of physicians who would follow guidelines generated from the decision-analytic model, compared to the traditional one.

Guideline	Utility Model	Traditional Model	p-value (2-sided)
Ovarian Cancer	0.444	0.444	1.000
Ureteral Stone	0.867	0.766	0.564
Sleep Apnea	0.533	0.600	0.780
Combined (n=39)	0.641	0.628	1.000

Table 3: Table showing the proportion of cases expert physicians thought patient preferences were used in generating the recommendations obtained from the two models

Table 3 summarizes the answers to Question 2, which inquired whether the experts thought patient preferences were used in generating guideline recommendations. The results indicate no significant differences in the expert physicians' perception of whether patient preferences were considered in generating the recommendations they were given for each patient case using either the utility model or the traditional guideline model.

The next chapter discusses the potential reasons for the results noted in this pilot evaluation of the decision model.

References:

1. Sanders GD. Automated creation of clinical-practice guidelines from decision models. Ph.D. Dissertation, Stanford Medical Informatics, SMI Report No. SMI-98-

0712, Department of Computer Science Report No. STAN-CS-TR-98-1609. Stanford University, Stanford, CA. July 1998.
2. Kuo RL, Aslan P, Abrahamse PH, Matchar DB, Preminger GM. Incorporation of patient preferences in the treatment of upper urinary tract calculi: A decision analytical view. The Journal of Urology, 1999 Dec; 162: 1913-9.
3. Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening: a decision-analysis model. Annals of Internal Medicine, 1993; 118: 838-43.
4. Chervin RD, Murman DL, Malow BA, Totten V. Cost-utility of three approaches to the diagnosis of sleep apnea: polysomnography, home testing, and empirical therapy. Annals of Internal Medicine, 1999; 130: 496-505.
5. Boxwala A, Mehta P, Peleg M, Tu S, Lacson R, Ash N. Modeling guidelines using domain-level knowledge representation components. Submitted to: Proceedings of the 2000 AMIA Fall Symposium, Mar. 2000.

VII. Conclusion and Recommendations

In this chapter, I will discuss the possible explanations for the results of this study. Specifically, I aim to conjecture the reasons why no significant difference was found in the acceptability of guideline recommendations (generated using either of the models discussed). In addition, I will enumerate the contributions of this dissertation and its implications for medical informatics. Lastly, I will recommend future directions for research and areas in need of further elucidation.

7.1 Variations in Patient Preferences are Reflected in the Generated Guidelines

Patient preferences are important and should be factored in the clinical decision-making process. The acceptability of guideline recommendations to patients was identified as one reason affecting the lack of physician adherence to practice guidelines (1). The study showed that certain recommendations varied when patient preferences, in the form of utilities, are taken into consideration. However, this variation in recommendations generated using patient utilities did not influence the acceptability of the recommendations to expert physicians as was expected. These are discussed in greater detail in the succeeding section.

7.2 Physician Experts' Agreement with the Recommendations

The study showed that there was no significant difference in physicians' acceptance of recommendations generated using the decision model, when compared to those generated using the traditional guideline model. There are several reasons that may explain this finding. In a recent review published in JAMA (1999), a discussion on why physicians

do not follow clinical practice guidelines enumerated several reasons, which were later confirmed by clinical studies to be barriers to guideline adherence (1). Three of the barriers enumerated below are especially relevant to the results obtained in this study. One reason for poor adherence to guidelines is the physicians' lack of familiarity with the recommendation. In this case, the recommendation suggested by the traditional model would be more familiar to expert physicians who are expected to be acquainted with the published recommendations. In fact, the study showed that in the case of the OSAS guideline, the traditional recommendation was actually more acceptable to the physicians than the recommendations generated with patient preferences. This may even reflect the physicians' biases that led them to develop these given recommendations in the first place. Second, there is a barrier to overcome the "inertia of previous practice." This can very much apply to the experts' decisions on the recommendations from the OSAS guideline. Third, the physician may not agree with the guideline recommendations. This reason can explain why a physician may go for more aggressive or more conservative strategies. Continuing with the example on OSAS, it is conceivable that physicians felt that the patient preference(s) to avoid polysomnography (because of aversion to CPAP therapy), is unreasonable. They may strongly believe in polysomnography and think that avoiding the test and the corresponding treatment will put the patient at greater risk for death. Reasons for non-acceptance in this case would include physicians' perception that the benefits of polysomnography are so great compared to any other diagnostic modality Other extreme reasons for non-acceptance of guidelines could include doubts on the credibility of the guideline authors.

More disturbing is the fact that physicians would agree with conflicting recommendations in cases where the recommendations differ for each model. Recall that in the same case with the same patient scenario, two different recommendations may have been presented to the expert physicians depending on whether patient preferences were incorporated. The study showed that our panel of experts had a similar rate of agreement for both guideline models' recommendation. It suggests that the options given are probably not unacceptable nor mutually exclusive. Thus, the degree of support for choosing one over the other may not be based on clinical evidence, but rather on consensus. In reality, the options that are available in a decision analysis model are really choices that have to be made given some uncertainty. There are probably other factors not considered in this study that may influence a decision-maker to go one way as opposed to another. Perhaps the context in which the decision is being made (e.g. political geography, religion, social undertones, etc.) needs to be addressed. Patient preferences alone may not be strong enough to push certain biases aside and direct the physician to agree or disagree with a recommendation.

7.3 Importance of Patient Preferences

Another key finding is the apparent inability of physicians to discern whether the recommendations were generated using patient preferences or not. One major barrier to the use of preferences in the design of CPGs is the inability to fully describe preferences in a meaningful manner (2). In this case, it may explain why the patient preferences, expressed in the form of utilities on a visual analog scale, were inconsistently interpreted and found not to influence recommendations. Another reason would be the physicians'

lack of familiarity with decision analysis and utility theory. In this case, there is difficulty reconciling patient utilities (which are for health states at the terminal end of a decision tree) with the choice to be made at a given time. This "disconnect" may be an artifact produced by the use of the decision analytic method.

It is important to emphasize that the decision-making process using decision analysis and utility theory does not necessarily reflect how physicians actually make decisions (3). It is modeled to reflect how decisions should be made if someone was thinking rationally and applying all criteria uniformly in each and every case. Given a limited amount of information and without the non-verbal clues and stimulation, it may be difficult for the expert physician to integrate the patient's preferences when expressed as distinct and permanent criteria as used in the decision analytic model. Thus, it may be difficult for anyone unfamiliar with the technique to understand how to apply these patient utilities.

Lastly, physicians often think that they know what is best for their patients. This traditional paternalistic view drives the physician to consider a multitude of factors and use his best judgement and clinical experience. It is, therefore, very difficult to explicitly factor in patient preferences alone in deciding appropriate recommendations. To illustrate, a patient who adamantly requests cosmetic surgery may be considered by the physician as an inadequate candidate for several reasons and so he/she refuses to accept this patient's preference, acting in the physician's perceived notion of what is the patient's best interest.

7.4 Contributions

The major goal of this thesis is to create a model for explicit representation of the decision-making process in CPGs. The decision-making process is not specified in current CPG models. However, the actual decisions or recommendations that a traditional CPG provides usually include this decision-making process, albeit implicitly. Thus, it will not be an added burden to tease out and express these preferences during the guideline development process. If a decision analytic approach is used, this decision model is then applied and explicitly stated during the guideline representation phase. Furthermore, I believe the object-oriented approach used for incorporating the decision model in GLIF3 will allow easy adaptation of the model to other representations for computable guidelines.

In addition, the "explicit" decision model will enable supporting information about CPGs to be stated specifically. The users will have a better understanding of reasons why specific options are better than others. The creation of a situation-specific and patient-specific model such as the decision model is needed for medical decision-making. The evidence-based medicine working group supports this approach in evaluating CPGs (1). Local adaptation of a "centralized guideline" may then be more feasible in this context because it gives the users more flexibility in following local practices when the recommendations support these. Furthermore, the flexibility also encourages evidence-

based practice because users would be forced to justify reasons explicitly for specific choices.

Another major advantage of using this model is the built-in mechanism that allows for easy representation for changes, as new information or new technology become available, when updating CPGs. To illustrate, when there are changes in disease prevalence, it would be easier to update information in the model because the prevalence is explicitly factored in the guideline at a specific point. This will also allow for a direct path to trace the consequences downstream to the resulting CPG recommendations.

Another contribution of this thesis is the extension of an existing CPG representation (GLIF3) to include decision analytic approach (4). Using this model, recommendations could be derived and compared to the previous model. This can be a paradigm for future research when evaluating innovations in guideline representation for performance, and acceptability.

Finally, an added contribution of this study was the comprehensive review and comparison of current CPG models available. The variations in the features of these models may help the developers identify areas where future work can be directed.

7.5 Future Direction

7.5.1 Improved User Interfaces

The use of a context-specific model for CPGs may make the CPGs appear more complex, especially for clinicians that prefer to view traditional text-based guidelines. However, this may be addressed, at least in part, by creating well designed, user-friendly interfaces. The actual decision analysis may be hidden from users until called-up by request, or alternatively, restricted to specific authors or implementers of a CPG in an institution. As clinicians slowly adapt to evidence-based medicine, an increasing demand for the decision analysis basis of recommendations may develop, especially when faced with a more informed patient.

7.5.2 Measuring Patient Preferences

An important area for future work is the development of better methods for defining patient preferences in a meaningful manner. The use of utilities is the most widespread approach currently. However, it is still not universally accepted nor understood; and the process of eliciting utilities is tedious.

Where this method may be beneficial, practitioners may not even be familiar with their use. One potential remedy would be the inclusion of tutorials illustrating specific probability and utility concepts phrased in a practical and concise format. Still, this does not preclude the need for more intuitive and perhaps, more representative means of capturing true and direct patient preference(s).

7.5.3 Need for Specific Patient Cases that take other Patient Factors into Account

The study design did not include nor address the importance of other patient and physician variables that are influential in clinical decision-making. Further studies using this decision model is warranted to compare recommendations that are generated with other situation-specific factors.

Another area for development of this model is the addition of cost-effectiveness analysis. The inclusion of cost-effectiveness information in the model is important when several options are equally preferable, but differ significantly in cost. This added dimension can be easily incorporated in this decision model.

7.5.4 Tracking Variations in Choices or Deviations from Recommendations

The study design was able to document how variability in practice may fall within the same evidence base. Variation in recommendations may be inevitable depending on local situations and patient factors. Specifying these factors which cause disparity will serve to illustrate in greater detail areas of potential disagreement within the CPG decision-making process. These disagreements are often not explicitly described in the CPG documents. These areas of disagreement can, in turn, stimulate further study to clarify and further strengthen the evidence base leading to refinement of CPG recommendations.

It would be useful to understand areas where physician and patient preferences may play a major role in variations in clinical decision-making. The role of experience and physician heuristics may be assessed from the variety of choices that are made for

specific clinical situations. Areas where there is disagreement due to lack of evidence can spur hypothesis driven research to drive practice towards one or the other decision path.

7.5.5 Evaluation of Guideline Adherence and Patient Outcomes

Finally, a feedback loop to the CPG developers may be created. This will allow for tracking of local probabilities or utilities that are then specified when adapting CPG to local clinical practice. These may then be used to identify areas where major deviations from the recommendations of the core CPG are common. Furthermore, this feedback will provide the CPG developer with adequate information from different locations about the areas of the CPG that are generalizable nationally or internationally. Additionally, it can potentially identify areas in the CPG where future users can be advised to look into the local situation prior to clinical application. Most importantly, the impact of any or all of these local changes in the CPG on actual patient outcomes,\ needs to be monitored and evaluated. Similar to post-marketing surveillance of new drugs, the evaluation of the performance of a CPG will necessarily entail a carefully monitored application and implementation to the patient population. After all, the final goal is for the improvement of the quality of care provided to patients.

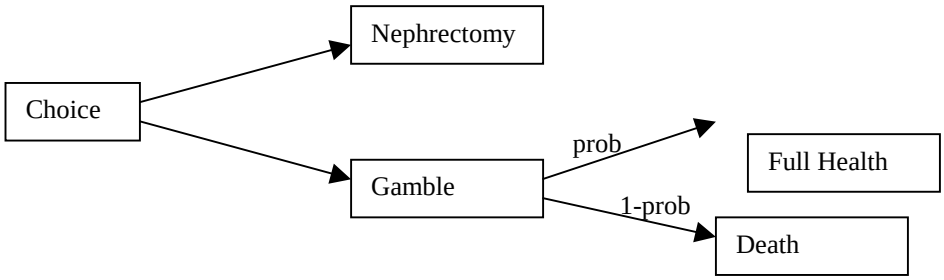
References:

1. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PC, et. al. Why don't physicians follow clinical practice guidelines? A Framework for improvement. JAMA, 1999; 282(15): 1458-65.
2. Kuo RL, Aslan P, Abrahamse PH, Matchar DB, Preminger GM. Incorporation of patient preferences in the treatment of upper urinary tract calculi: A decision analytical view. The Journal of Urology, 1999 Dec; 162: 1913-9.
3. Torrance GW, D Feeny. Utilities and quality-adjusted life years. Intl J of Technology Assessment in Health Care, 1989; 5: 559-75.
4. Peleg M, Boxwala A, Ogunyemi O, Zeng Q, Tu S, Lacson R, et al. GLIF3 Technical Specification. In: http://smi-web.stanford.edu/projects/intermed-web/guidelines/GLIF1.htm; 2000.

Appendix A: Questionnaire Used to Elicit Preferences Using Standard Gamble

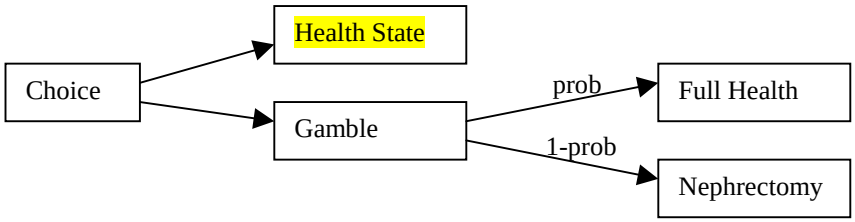
Age: _____
Sex: _____
Race: _____
Experience with Stones: _____
Experience with Sleep Apnea: _____
Experience with Ovarian CA: _____

Recurrent Urolithiasis:



Nephrectomy: You experience an intense, continuous pain in the groin as the result of a kidney stone, making it difficult to carry out most daily activities. After about 2 days in pain, you will be operated on in the hospital. The operation will remove not only the kidney stone, but also one of your kidneys. As a result, you will have a large scar in your side, and must stay in the hospital for 1 week after the operation. For 3 weeks after you leave the hospital, you must refrain from physical exertion, and go through rehabilitation therapy at home. Following this, you will return to your normal health and have no difficulties in carrying out any daily activities. This event is called nephrectomy.

Prob: _____



Severe Pain: Imagine that, at some point in the next month, you will experience an intense, continuous pain in the groin as the result of a kidney stone, making it difficult to carry out most daily activities. After about 2 days in pain, you will be operated on in the hospital. During the operation, you will be placed under a general anesthetic, and a tube will be inserted into your kidney to suck out the stone. There is a small chance of bleeding occurring during the operation, in which case you will require a transfusion, but in any case, you will be able to leave the hospital in perfect health after about 3 days.

Prob: _____

Moderate Pain: In this state, you will feel a chronic back ache that is painful but does not keep you from carrying out daily activities. After about 2 weeks, you will undergo an operation at which you will be placed under a general anesthetic, and a tube will be inserted into your kidney to suck out the stone. There is a small chance of bleeding occurring during the operation, in which case you will require a transfusion, but in any case, you will be able to leave the hospital in perfect health after about 3 days.

Prob: _____

Mild Pain: In this state, you will feel a dull ache in the back, which causes discomfort, but does not keep you from carrying out daily activities. After 2-4 weeks, you will undergo an operation at which you will be placed under a general anesthetic, and a tube will be inserted into your kidney to suck out the stone. There is a small chance of bleeding occurring during the operation, in which case you will require a transfusion, but in any case, you will be able to leave the hospital in perfect health after about 3 days.

Prob: _____

Mild pain with lithotripsy: In this state, you will feel a dull ache in the back, which causes discomfort, but does not keep you from carrying out daily activities. After 2-4 weeks, you will undergo a non-surgical procedure during which you will be sedated and sound waves will be used to break up the stone. The procedure will take about 3 hours and there is a small chance that it will cause an arrhythmia or irregular heart beat. In any case, there will be no long-term effects, and you will be able to go home later that day.

Prob: _____

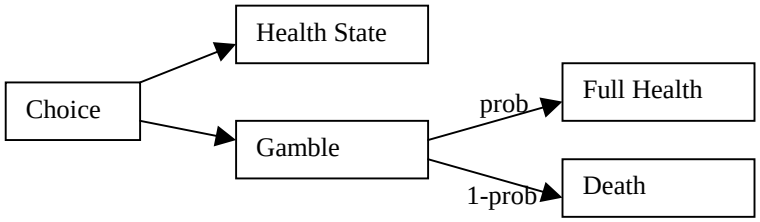
Long-term medication: In this state, you will have to take 2 tablets twice a day. In addition, every 6 months you will need to visit the doctor for some tests. The tablets may also occasionally cause mild nausea. Other than this, you will be in perfect health with no other problems.

Prob: _____

You can choose to go on long-term medication, as described in the last question. If you do go on medication, there is still a 10% chance that you will have a kidney stone in the next year. If you do have a stone, 20% of the time it will be a mild stone, 30% of the time it will be moderate, and half of the time it will be a kidney colic. If you do have a stone, it can be treated with lithotripsy half of the time, and half of the time it must be treated with surgery. If it is treated with surgery, there is a 5% chance that you will also have a nephrectomy and lose a kidney. Your other choice is not to use long-term medication. In this case, there is a 50% chance that you will have a stone in the next year. If you have a stone, the chances of it being mild, moderate or severe; and the chances of having surgery or lithotripsy and of losing a kidney are exactly the same as before, with medication. Would you choose to be on medication or not?

Choice: _____

Ovarian CA Screening



Survive laparotomy after early disease: You are diagnosed with ovarian cancer at an early stage. There is about 10% chance that you have involvement of 2 ovaries. If confined to only one ovary and the tumor grade is 1 or 2, you may undergo unilateral ovary removal if you still want to have children. Otherwise, you may have complete removal of the uterus, 2 ovaries and fallopian tubes. If some residual tumor is left, one may undergo oral chemotherapy or intraperitoneal radioactive colloids. Chances are the tumor is removed and there is no further need for treatment after discharge.

Prob: _____

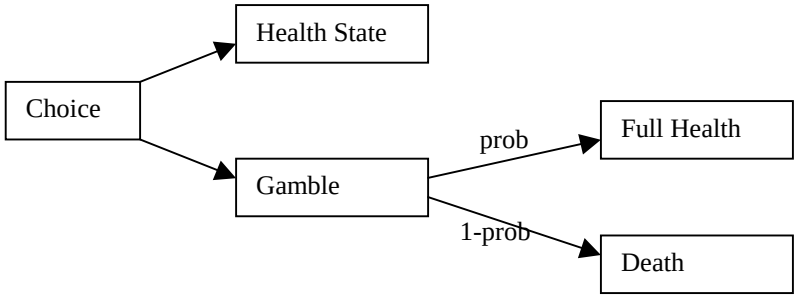
Survive laparotomy after late disease: You are diagnosed with ovarian cancer at a late stage. You undergo laparotomy where removal of the uterus and both ovaries plus removal of as much tumor as possible is performed. Most of the time, extensive residual disease is left behind. You undergo either radiation therapy or combination chemotherapy to debulk those tumors. You may then have to undergo a second surgery. After this, you may then undergo total abdominal irradiation. If the tumor reduction surgery necessitates removal of the intestines, you may have to live with a colostomy. There are major side effects associated with the radiation therapy such as nausea and vomiting, loss of appetite, diarrhea, and weight loss. Furthermore, the radiation may lead to bowel obstruction or GI bleeding. In most patients, these subside in a few weeks but may persist for months to years in 30% of patients.

Prob: _____

Survive laparotomy with no disease: You are diagnosed with ovarian cancer. You are naturally distressed by this diagnosis. You undergo surgery and are told afterwards that you do not have cancer after all. You are discharged after 3 days in the hospital with a small abdominal scar.

Prob: _____

Sleep Apnea:



OSAS with CPAP: You are diagnosed with sleep apnea. The symptoms associated with this disease include daytime sleepiness, decreased concentration, memory loss, irritability, moodiness, depression, psychosis, decreased libido, impotence, nocturnal bedwetting, snoring and marital problems. In addition, it is associated with high blood pressure, heart failure, and arrythmias. There is inconclusive evidence if the disease reduces survival. You decide to have treatment with CPAP. The untoward effects include dryness of the nasal mucosa, awkwardness of the nasal mask (that has to be worn all night, every night) and the noise that the machine makes. There is improvement of the oxygen saturation in the blood during sleep, the snoring, moodiness, daytime sleepiness and decreased concentration. There is also possible improvement in 5-year survival in terms of heart disease but this remains unproven.

Prob: _____

No OSAS with CPAP: You are diagnosed with sleep apnea. The symptoms associated with this disease include daytime sleepiness, decreased concentration, memory loss, irritability, moodiness, depression, psychosis, decreased libido, impotence, nocturnal bedwetting, snoring and marital problems. In addition, it is associated with high blood pressure, heart failure, and arrythmias. There is inconclusive evidence if the disease reduces survival. You decide to have treatment with CPAP. The untoward effects include dryness of the nasal mucosa, awkwardness of the nasal mask (that has to be worn all night, every night) and the noise that the machine makes. You get some improvement in snoring and sleepiness, but no other effect because you did not have sleep apnea after all, but you do not know about it.

Prob: _____

OSAS with no CPAP: You are misdiagnosed as not having sleep apnea. You have symptoms associated with this disease including daytime sleepiness, decreased concentration, memory loss, irritability, moodiness, depression, psychosis, decreased libido, impotence, nocturnal bedwetting, snoring and marital problems. In addition, it is associated with high blood pressure, heart failure, and arrythmias. There is inconclusive evidence if the disease reduces survival. You may receive some other diagnosis and given other less efficacious treatment. Or you may not get any treatment.

Prob: _____

No OSAS and no CPAP: You do not have sleep apnea although you have some of its symptoms. The associated morbidities may be the same. You may get another diagnosis. You do not get CPAP.

Prob: _____

Appendix B: Sample Questionnaire for Each of the Three Guidelines Sent to Expert Physicians to Assess Acceptability of the Recommendations

1. Ovarian Cancer Screening

39 year old female presents for possible ovarian cancer screening. She is asymptomatic and has no family history of ovarian cancer. Her preferences for some health states are shown below.

From 0 to 100 where 0=death and 100=perfect health, participants were asked to indicate how much they value being in a certain health condition.

Survives laparotomy after early disease _____|_
0 99 100

Survives laparotomy after late disease _____|_____
0 60 100

Survives laparotomy but has no disease _____|_
0 98 100

Guideline recommended for this particular patient:

1. Get patient's age and family history regarding breast, ovarian and other cancers.
2. Pelvic exams may or may not be performed for diagnostic purposes.
3. No further screening (versus screening with CA-125 and transvaginal sonography which would gain 2.6 extra hours of the patient's life)

Questions:

1. Would you follow the recommendations? ___yes ___no
2. Are the patient preferences considered in the guideline recommendations? ___yes ___no
3. Comments _____

2. Urolithiasis

29 year old male presents with a ureteral stone in situ.

From 0 to 100 where 0=death and 100=perfect health, participants were asked to indicate how much they value being in a certain health condition.

Nephrectomy
0 70 100

Nephrolithotomy after severe pain
0 91 100

Nephrolithotomy after moderate pain
0 94 100

Nephrolithotomy after mild pain
0 97 100

Lithotripsy after mild pain
0 98 100

Long-term medication
0 88 100

Guideline recommended for this particular patient:

1. Check size and shape of stone, internal anatomy and history of stone passage.
2. If newly diagnosed ureteral stone with a high probability of spontaneous passage and whose symptoms are controlled, observation is recommended for initial treatment.
3. Choices: a. Advise no long-term medication - preferred
   b. Long-term medication for medical management of nephrolithiasis

Questions:

1. Would you follow the recommendations? ___yes ___no
2. Are the patient preferences considered in the guideline recommendations? ___yes ___no
3. Comments _____

3. Sleep Apnea

39 year old female presents with daytime sleepiness, decreased concentration, memory loss, irritability, moodiness, depression, psychosis, decreased libido, impotence, nocturnal incontinence, snoring and marital problems.

From 0 to 100 where 0=death and 100=perfect health, participants were asked to indicate how much they value being in a certain health condition.

O.S.A.S. : Obstructive Sleep Apnea Syndrome
C.P.A.P.: Continuous Positive Airway Pressure
No O.S.A.S. : Patients with the same symptoms but no documented OSAS

OSAS with CPAP
0 72 100

OSAS with no CPAP
0 71 100

no OSAS with CPAP
0 60 100

no OSAS and no CPAP
0 90 100

Guideline recommended for this particular patient:

Choices for the diagnosis of OSAS:
- Standard polysomnography is the preferred test for the diagnosis of OSAS.
- Home Study – 2nd choice
- No test – 3rd choice

Questions:

1. Would you follow the recommendations? ___yes ___no
2. Are the patient preferences considered in the guideline recommendations? ___yes ___no
3. Comments _____

Appendix C: Instructions Given to Expert Physicians for Answering the Questionnaire

Instructions:

1. Read the patient scenarios, which is almost the same for each case.
2. The preferences are actual preferences of these patients for certain therapies such as being on long-term medications or having a nephrectomy. (The idea being to delineate when medical opinion for best management may conflict with personal patient preference and if you would still recommend that they get it.)
3. There is a preferred recommendation from clinical practice guidelines for each patient indicated at the bottom of each case. You need to indicate if you agree with it or not. You may indicate your rationale.
4. For the second question, you are asked your opinion on whether or not you thought that patient preferences were taken into consideration when making the guideline recommendation.
5. There is no right or wrong answer. Just be objective.

Note:

1. When a patient indicates a score of 40 out of 100, that means that if they had a hundred years to live after the treatment, they would trade that for 40 years in perfect health than to have to live 100 years in that current state. So the lower the score, the less they prefer the state. Another case would be a score of 1 for a State X, for example. This means that a patient would rather live 1 year in good health than 100 in a specific state. You know that they hate that state in this case.

2. The recommendations were specific for the patients in some cases and not in others in this sample of patients. For question 2, just answer whether (or not) you think that the patient's preferences were considered in generating these specific guideline recommendations for each patient (since you have a particular patient's preferences for treatment).