AUTOMATED CREATION OF CLINICAL-PRACTICE GUIDELINES FROM DECISION MODELS

A DISSERTATION
SUBMITTED TO THE PROGRAM IN
MEDICAL INFORMATION SCIENCES
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Gillian D. Sanders
June 1998

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Douglas K. Owens, Principal Adviser
(Department of Medicine)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Robert F. Nease Jr.
(Division of General Medical Sciences,
Washington University School of Medicine)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Edward H. Shortliffe
(Department of Medicine)

Approved for the University Committee on Graduate Studies

Abstract

Rising costs have forced policy makers in the United States to reorganize the finance and delivery of health care. This reorganization, combined with documented variation in practice patterns, has produced widespread interest in methods for improving quality of care, and specifically in dissemination of guidelines for clinical use as one such method. Clinical-practice guidelines (CPGs) provide a systematic means to review patient management and a formal description of appropriate levels of care; their use can enhance the quality, appropriateness, and effectiveness of health care, while also containing costs.

Development of effective CPGs, however, requires input from experts in clinical medicine and in evidence synthesis; at the local level, such resources often are not available, or are prohibitively costly. These requirements thus pose two main problems for developers of CPGs: (1) the prohibitive cost of the required resources conflicts with the desire of local clinical communities to implement their own CPGs that reflect their particular sites or patient populations, and (2) the CPGs are static, placing limitations on the ability of guideline developers to update and maintain CPGs with current clinical findings.

In this thesis work, I developed a new approach that allows developers and users to create, disseminate, and tailor CPGs, using normative decision models (DMs). My approach is designed to improve CPG applicability, relevance, and acceptance by local clinicians and guideline developers, and thus to promote high-quality and cost-effective health care. I propose that guideline developers can use computer-based DMs that reflect known global and site-specific data to generate evidence-based CPGs. Such CPGs are of high quality,

can be custom tailored to specific clinical settings, and can be modified automatically over time as the underlying DM or evidence evolves.

To validate this hypothesis, I defined conceptual models for representing CPGs and DMs, and formalized a method for mapping between these two representations. To make this mapping possible, I designed a DM annotation editor that queries the decision analyst for missing knowledge. The two conceptual models, the DM annotation editor, and the mapping algorithm allow guideline developers to create evidence-based CPGs, and permit local users to do the updating and tailoring needed for the CPGs to succeed. I describe implementation of the ALCHEMIST system that encompasses these two conceptual models, the mapping algorithm, and the resulting tailoring abilities.

I evaluated my thesis work in three ways. First, to show that the DM conceptual model and DM annotation editor provide information necessary for the efficient transformation of a DM into a CPG, I evaluated the design and expressivity of both conceptual models, and demonstrated the accuracy of the DM-to-CPG mapping algorithm. Second, to show that ALCHEMIST produces CPGs that satisfy published criteria for high-quality guidelines, I had guideline users rate the quality of generated CPGs using a guideline-rating key, and rate the performance of ALCHEMIST's CPG browser. Finally, I evaluated ALCHEMIST's dynamic patient and site tailoring abilities.

Because it creates CPGs from normative DMs, ALCHEMIST is able to specify explicitly the alternatives, outcomes, and evidence in a clinical problem. Using the DM best-estimate inputs and sensitivity analyses, the generated CPG quantifies the need for CPG tailoring. ALCHEMIST automates the DM-to-CPG process and distributes the generated CPG over the World Wide Web to allow guideline developers at each local site to apply, tailor, and maintain a globally produced CPG. In my thesis work, I argue that my DM-to-CPG conceptual framework is a method for guideline developers to create and maintain automated evidence-based CPGs, and that it thus promotes high-quality and cost-effective health care.

Acknowledgments

After my freshman year at Princeton, I worked for Dr. Edward Lyons at the Winnipeg Health Sciences Centre on a radiology order-entry expert system. It was during this summer that I was first introduced to the field of Medical Informatics and to the world of medical research. I want to thank Dr. Lyons for all the opportunities that came from the summers I worked with him, but most of all for teaching me to love what I do.

Many people helped me get from that first summer to the point of actually finishing my Ph.D. First and foremost, I would like to acknowledge Doug Owens for being a great adviser and an even better friend. Doug has participated in every phase of my training and has always been there to support and challenge me. I am indebted to my other committee members as well: Bob Nease and Ted Shortliffe. Although Bob lives in St. Louis, he has always been there to help turn my disorganized thoughts into research ideas. Much of the work that I describe in this dissertation builds off of Doug and Bob's previous work and the high quality of their research both inspires and motivates my own efforts. I am grateful to Ted for providing an outstanding Medical Informatics research environment and for helping me write my hypothesis!

Although this is a dissertation in Medical Informatics, many of problems that I address come from the domain of Health Services Research. I am extremely fortunate to be a member of the Cardiac Arrhythmia Patient Outcomes Research Team headed by Mark Hlatky. I am grateful to Mark and members of the PORT team who have provided me with numerous clinical questions, new ideas, helpful criticisms, and exciting opportunities over the past four years — specifically Kathy McDonald, Alan Garber, Nathan Every, Paul

Heidenreich, Carol Cheng, Ryan Harris, and Anne Dembitzer. My research was supported in part by Grant No. HS 08362 to Stanford University from the Agency for Health Care Policy and Research.

Frank Sonnenberg and Greg Hagerty were kind enough to supply me with beta versions of Decision Maker, and Greg speedily made changes to the software to accommodate my never-ending requests. I'd also like to thank Lyn Dupré who took on the daunting task of teaching me to write, kept up with my hectic pace, and told me when to slow down.

It seems like I have always spent my weekends at MSOB and I want to thank my fellow students in Medical Informatics for making my years here so much fun — particularly Ida Sim, Doug Fridsma, Wanda Pratt, and Ramon Felciano for keeping the admission glitch quiet all these years. I want to thank Rhea Pinkertopoulos for liking email, non-sequiturs, and Chevy's almost as much as me — and Brian for understanding our cravings. I am also grateful for the constant support of my friends from outside of MSOB. Marc Tremblay for his Otis Spunkmeyers, as well as Loribee Bassman, Brandice Wrone (!), Mike Sullivan, Keith Ohlfs, Erin Adams, Shannon Feindel, Allison Bracken, Kate Hare, Mark Day, Zia Khan, Dave Degraaff, Matthias Schabel, Scott Stehle, and Andrew Kelly for continuing to be my friends despite the fact that lately they only hear from me through email flurries. Thanks to Annika Andrews for daily phone calls and for always understanding, Eric for once again being a programming God, Laura (aka sbl) Dalston for finally moving to the right state and for knowing how to make me laugh just a little too hard, and Sarah Killien, Fay Hanleybrown, Geoff Wilson, Peter Milne, Christy Richter, Stephanie Weider, Katrina Jensen, Faith Hecht, and Dre Freeman for always making me smile.

I would like to give a special thanks to Scott Schmidler for being my work's harshest critic and my bestest friend (scott — *now* you can read it!). His brilliance, companionship, support, and ability to make even protein-structure prediction exciting never cease to amaze me.

I am grateful to my family (the eternal students) for supporting and encouraging me. Thanks to my sister Christa for always being there, and to my brothers Daeb and Jimmy for reminding me of my Canadian roots. Finally, I dedicate this thesis work to my parents for valuing my education and happiness so highly and for always believing in their little girl. I appreciate all that they have done and sacrificed, and I cannot thank them enough.

Table of Contents

Chapter 1 Automated Creation of Clinical-Practice Guidelines from Decision Models

Chapter 2 Clinical-Practice Guidelines

Chapter 3 Decision Models

Chapter 4 Decision-Model Annotation Editor

Chapter 5 Algorithm for DM-to-CPG Mapping

Chapter 8 Evaluation

Chapter 9 Summary, Contributions, and Future Work

Appendix A Clinical-Practice Guideline Conceptual Model

Appendix B Decision Model Conceptual Model

Appendix C Naming Conventions for Model Variables

Appendix D Global CPG: Lung-Cancer Effectiveness Model

Appendix E Human Subjects: Informed Consent Form

Appendix F Clinical-Practice Guideline Evaluation

List of Tables

List of Figures

Chapter 1

Automated Creation of Clinical-Practice Guidelines from Decision Models

Rising health-care costs have forced policy makers in the United States to reorganize the finance and delivery of health care. This reorganization, combined with documented variation in practice patterns (Chassin et al. 1986, Conway et al. 1995, Fisher et al. 1992, Health Services Research Group 1992a, Iscoe et al. 1994, Keller et al. 1990, Welch et al. 1993, Wennberg and Gittelsohn 1973), has produced widespread interest in methods for improving quality of care, and in dissemination of guidelines for clinical use (Woolf 1990). The Institute of Medicine (IOM) defines **clinical-practice guidelines (CPGs)** as "systematically developed statements to assist physician and patient decisions about appropriate health care for specific clinical circumstances" (Institute of Medicine 1992). CPGs provide a systematic means to review patient management and a formal description of appropriate levels of care; their use can enhance the quality, appropriateness, and effectiveness of health care, while also containing costs (US Department of Health and Human Service 1995).

Development of CPGs, however, requires input from experts in clinical medicine, meta-analyses, decision analyses, clinical epidemiology, cost-effectiveness analyses, and evidence synthesis; such resources often are not available, or are prohibitively costly on a local level (Brook 1989, Eddy 1990e, Fletcher and Fletcher 1990). Because the development of CPGs requires specialized expertise and is costly, CPGs often are developed by national organizations (American College of Physicians 1995, US Preventive Services Task Force 1996) for a population that has "average" characteristics. These CPGs may need to be adapted to local settings in which the characteristics of the patient population or practice may be different (Carter et al. 1995a, Nease and Owens 1994, Owens and Nease 1997). In addition, traditional guidelines are normally static and thus may become out of date (Sanders et al. 1998).

In my thesis work, I explore these problems that impede CPG success. One potential solution is to use **decision models (DMs)** as an aid for developing CPGs. I define DMs as abstract representations of a decision problem that take into account the uncertain, dynamic, and complex consequences of a decision, and the assignment of value to those consequences (Owens and Nease 1993, Owens and Sox 1990). I developed a new approach that allows developers and users to create, disseminate, and tailor CPGs, using normative DMs. My approach is designed to improve CPG applicability, relevance, and acceptance by local clinicians and guideline developers, and thus to promote high-quality and cost-effective health care.

1.1 Hypothesis

I propose that guideline developers can use computer-based DMs that reflect known global and site-specific data to generate evidence-based CPGs. Such CPGs are of high quality, can be custom-tailored to specific clinical settings, and can be modified automatically over time as the underlying DM or evidence evolves.

To validate this hypothesis, I defined **conceptual models** for representing CPGs and DMs, and formalized a method for mapping between these two representations. The two conceptual models and the mapping algorithm allow guideline developers to create evidence-based CPGs in algorithmic form, and permit local users to do the updating and tailoring needed for the CPGs to succeed. I describe implementation of the ALCHEMIST system that encompasses these two conceptual models, the mapping algorithm, and the resulting tailoring abilities. I evaluated three subhypotheses:

1. The DM conceptual model provides information necessary for the efficient transformation of a DM (specifically, of a decision-tree representation of the expected outcomes and available alternatives) into CPGs, thereby allowing the creation of guidelines that are evidence based.
2. ALCHEMIST produces CPGs that satisfy published criteria for good practice guidelines.
3. ALCHEMIST's mapping algorithm allows dynamic patient and site tailoring, and, therefore, produces local CPGs that provide expected health outcomes that are based on the DM and that produce expected outcomes (measured in quality-adjusted life years) that are equal to or better than those expected from static global CPGs for specific patient populations.

My main goal, therefore, has been to develop a method that uses DMs to create CPGs. My method takes advantage of the resources available to large guideline-development organizations, but allows for local tailoring and updating. To accomplish this goal, I answered six questions:

1. What should a user of a CPG be able to do? For example, what tasks should she be able to perform? What questions should she be able to answer? (Section 2.4)

2. What is the minimal set of knowledge required of CPGs for a user to perform all these required tasks? How can we specify this knowledge most efficiently? (Section 2.6)
3. What subset of the required CPG information can be obtained from the input variables and structure of a DM? (Section 3.6)
4. How should we obtain the information required by the CPG that is not currently available in the DM? (Section 4.3)
5. How should we transform the DM representation into a clinical-practice algorithm? (Section 5.4)
6. What subset of CPGs can be based on DMs? (Section 3.3.2)

To answer these questions, I defined five specific aims for my dissertation:

1. To specify the data, structural, procedural, and knowledge requirements for conceptual models for CPGs and for DMs
2. To implement the DM conceptual model; my computer-based implementation requires that decision analysts specify additional information that is necessary to develop a CPG from the DM
3. To design a mapping between the conceptual model of a DM and the conceptual model of a CPG; I show that, given appropriate additional specification of information in the DM, the knowledge represented in either model is available in the other
4. To implement and evaluate the ALCHEMIST system, which uses my mapping between the conceptual models to create an annotated algorithmic CPG directly from the DM; I evaluated the resulting CPG, the updating and tailoring capabilities of ALCHEMIST, and the consistency of the mapping between the two conceptual models

5. To explore the degree to which my conceptual models and implementation can be extended to a larger subset of CPGs

1.2 Clinical-practice guidelines

Current CPGs encompass numerous formats, levels of complexity, and sources. Each has its own methods of development and dissemination, and therefore its own strengths and limitations. CPGs can be represented in several different formats, including text, protocol charts or lists, flowcharts, or any combination thereof. This diversity in CPG formats, styles, graphics, and methods of development limits widespread use and dissemination (Society for Medical Decision Making (SMDM) Committee on Standardization of Clinical Algorithms (CSCA) 1992). I concentrate on CPGs that are represented as clinical algorithms as defined by the SMDM CSCA, and that allow the guideline developer to communicate a complex series of conditional statements in a structured manner. The flowchart representation (Figure 1.1) is integrated with textual output that follows a published structure for CPGs developed by Hayward and colleagues (Hayward et al. 1993). This organization and content promote a consistent structure for reporting CPGs and enhance a user's ability to determine the applicability, importance, and validity of a CPG for her specific population (Hayward et al. 1993).

1.3 Evidence-based guidelines

Many CPGs are developed based on expert opinion, local practice, or consensus. In this dissertation, I emphasize the creation of **evidence-based CPGs (EB-CPGs)**, which I define to be CPGs that developers create using the clinical literature and a decision-analytic framework (Evidence-Based Medicine Working Group 1992, Sox and Woolf 1993). Although EB-CPGs provide more accurate, complete, and accountable information that do most other techniques (Eddy 1992), the creation of such an EB-CPG requires resources

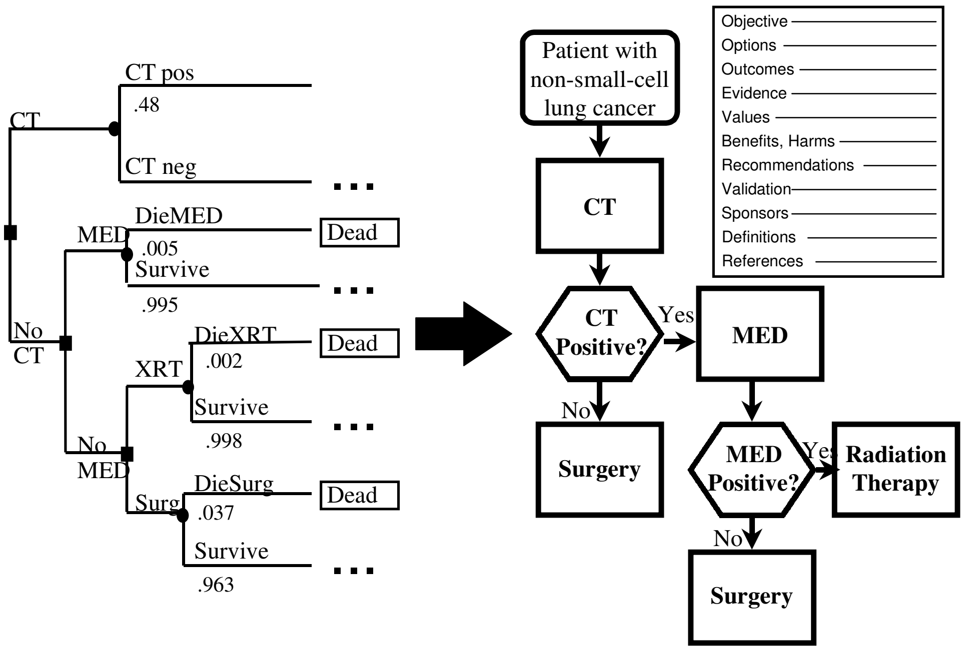


Figure 1.1. Decision-tree (left) and flowchart (right) representations for staging non–small-cell lung cancer. The decision-tree schematic represents the alternatives, outcomes, and preferences in the clinical decision; the annotated flowchart (right) represents the optimal strategy for a specific instantiation of the underlying decision model.

that local guideline developers normally cannot afford. The compromise of implementing locally CPGs that were created by large institutions or funding agencies often leads to a reduction of guideline acceptance and, therefore, of the guidelines' potential effectiveness, because recommendations are not perceived as reflecting the actions appropriate for a specific patient or site (Grimshaw and Russell 1993, Owens 1994). I maintain that my work helps to alleviate this problem.

1.4 Problems in the development of clinical-practice guidelines

I address several problems associated with the creation, dissemination, and updating of CPGs. The creation of EB-CPGs is limited because it requires numerous resources that are not normally available to a local guideline developer. Allowing CPGs to be produced on a global level, yet maintained and updated by the local guideline implementer or clinician, reduces this resource requirement. The assumptions and clinical evidence on which CPGs are based are often hidden from the end user. This weakness causes CPGs to behave like black boxes, limiting their acceptance by clinicians (Abendroth et al. 1988, Shiffman and Greenes 1991). ALCHEMIST attempts to address this problem by making the assumptions an explicit element in the DM conceptual model (and, therefore, in the resulting CPG), and allowing the user to adjust variables and to view the updated CPG. Because developers create DMs using quantitative clinical data, creating a CPG from an existing DM ensures that the guideline is based on evidence, and eliminates many inconsistencies or contradictions commonly found in implemented CPGs (Owens and Nease 1993, Shiffman and Greenes 1994, Wears et al. 1994).

After the CPG is created, problems with dissemination and maintenance often impede its success. Local users who were not involved in the CPG-development process may feel removed from the policy-making process or may not believe that the CPG is applicable to their specific site or patients (Carter et al. 1995b). This lack of local involvement and validation decreases the likelihood of CPG dissemination and implementation. ALCHEMIST allows the user to make changes at the local level while providing access to the clinical evidence on which the CPG is based.

1.5 Decision models

Basing CPG creation on DMs enriches the produced CPGs (Nease and Owens 1991, Oddone et al. 1994, Owens and Nease 1997, Parmigiani et al. (submitted for publication)). DMs provide a normative analytic framework for representing the evidence, outcomes, and preferences involved in a clinical decision. They clearly define the available alternatives and events of interest, and combine these elements in an objective and predictable way to produce a recommendation that is consistent with underlying data and assumptions. In addition, the ability to perform sensitivity analyses allows the guideline user to identify critical variables to focus refinement of the guideline. I assume that guideline developers and clinicians desire such a normative model of a decision, and create CPGs based on such normative models. Current CPGs are not usually based on DMs; even when both representations exist for the same clinical problem, the advice or logic of the two representations may not correlate. I address this lack of correlation and provide a method for mapping between DMs and CPGs. There are several DM representations, including decision trees, influence diagrams, spreadsheet models, and state-transition models. I concentrate on the decision-tree representation initially (Figure 1.1), because it serves as a common model representation used in the medical decision-making community for simple decision analyses, Markov models, and cost-effectiveness studies. Several software packages are also available for building such decision trees on the computer. Such software packages include Decision Maker (Sonnenberg and Pauker 1987), SMLtree (Hollenberg 1984), and Data by TreeAge. I work with decision trees modeled using the Decision Maker software.

1.6 Overview of ALCHEMIST architecture

Figure 1.2 provides an overview of ALCHEMIST's architecture. The following list describes the individual modules of Figure 1.2.

1. *Decision model:* The decision analyst — using the Decision Maker software — creates The DM. Then, using ALCHEMIST's World Wide Web-based interface, the decision analyst loads the DM into the ALCHEMIST system.
2. *DM conceptual model:* ALCHEMIST maps the DM onto the DM conceptual model and automatically produces the results of the model and the flowchart algorithm for the CPG. ALCHEMIST also obtains knowledge explicit in the DM (such as best-estimate values and the available alternatives).

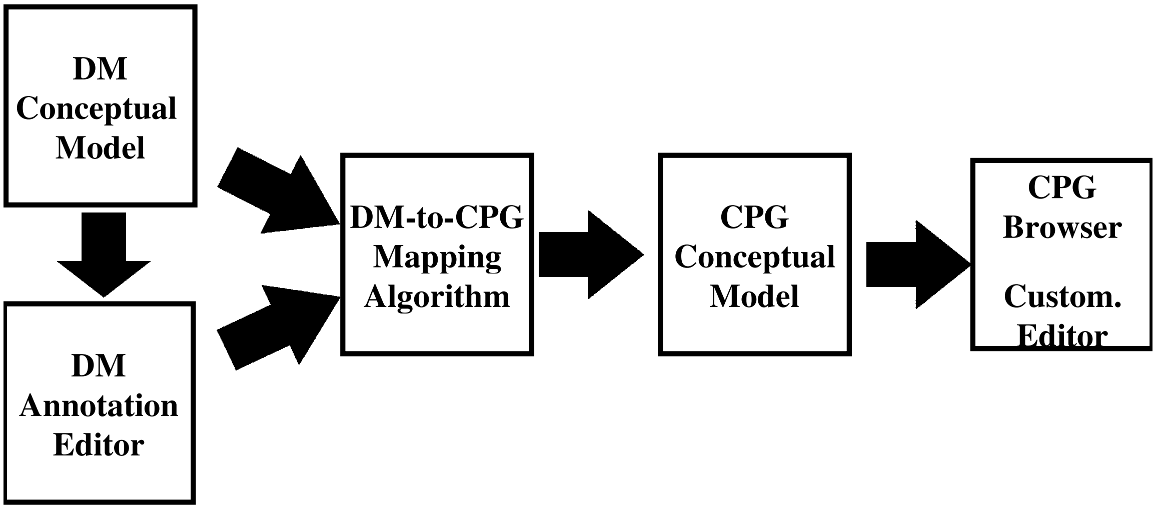


Figure 1.2. Schematic representation of ALCHEMIST's architecture. Each box represents an individual module of this dissertation. DM = decision model, CPG = clinical practice guideline.

3. *DM annotation editor:* ALCHEMIST dynamically creates a web-based DM annotation editor using the information obtained from the DM. The DM annotation editor queries the decision analyst for missing information required by the CPG conceptual model

4. *DM-to-CPG mapping algorithm:* ALCHEMIST's mapping algorithm uses the knowledge that ALCHEMIST extracted automatically from the DM and knowledge that ALCHEMIST obtained through the DM annotation editor to translate formally the DM conceptual model into the CPG conceptual model

5. *CPG conceptual model:* ALCHEMIST instantiates the CPG conceptual model using the information provided through the mapping algorithm.

6. *Guideline browser and custom-tailoring editor:* ALCHEMIST creates a web-based interface using the information from the CPG conceptual model. The interface allows the user to browse the created CPG and to adjust input-variable values. These new values entered by the user on the web page are then fed back by

ALCHEMIST to the DM, and the DM-to-CPG creation process is repeated (without any additional input from the decision analyst or from the DM annotation editor), producing an updated or tailored CPG.

1.7 Conceptual-model framework

Conceptual models are designed to describe a part of the world, the concepts about that part of the world, and the relationships among those concepts. I use conceptual models to represent the different knowledge required and provided by both DMs and CPGs. Figure 1.3 demonstrates the different knowledge representations found in ALCHEMIST. The goal of my thesis research was to determine (1) what knowledge is needed for a CPG (the white ellipse), (2) what knowledge is inherent in a DM (the dark-gray ellipse), (3) what the intersection of these two knowledge sets is (the light-gray section), (4) how to obtain the missing CPG knowledge from the DM (or other sources), and (5) how to link the additional DM knowledge to the CPG such that it can be used for updating and maintaining of the CPG at a later time (i.e., how best to store the union of these sets). It is with this framework in mind that I designed the conceptual models for DMs and CPGs.

1.8 Conceptual model for CPGs

I define a conceptual model for a subset of CPGs (Section 2.6). This subset includes those CPGs that can practically be based on a DM (e.g., I exclude development of guidelines for patients who have numerous comorbid conditions). I discuss the restrictions that this subset places on the applicability of my thesis to other CPG-development work, and describe methods for extending my model to include other CPG formats (Section 2.3). I defined the knowledge that is necessary and sufficient for creating and maintaining a CPG by studying the development and use of existing CPGs. I identified the key questions that a user should be able to answer, and the tasks that she should be able to perform with a

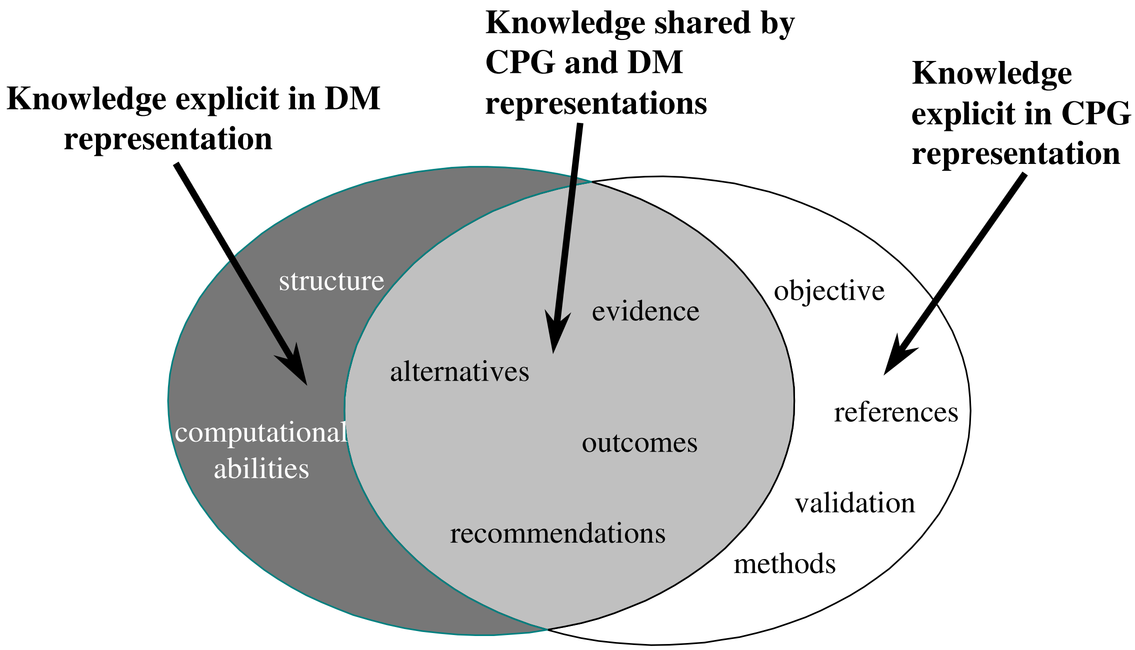


Figure 1.3. Conceptual model framework. The white ellipse represents the knowledge required for a CPG; the dark-gray ellipse represents the knowledge inherent in a DM. The intersection of these two ellipses (light-gray section) represents the knowledge that ALCHEMIST must obtain form the decision analyst to create the CPG. Examples of these different knowledge elements are shown in the respective ellipses.

CPG. I describe the data, functional, and knowledge requirements for a conceptual model to perform these tasks (Section 2.4). I assessed existing conceptual models for CPGs; I describe their advantages and disadvantages (Section 2.5). Of the several existing guideline representations (Cimino 1995, Stoufflet 1995, Stoufllet 1996, Ohno-Machado 1998), not one produces evidence-based CPGs that can be based on DMs.

1.9 Conceptual model for decision models

I studied the insight inherent in a subset of DMs, and determined what knowledge required by the CPG conceptual model is not currently available. I created a conceptual model that details the combination of this knowledge (Section 3.6). I restricted the subset of DMs to which my conceptual model applies, although I included dual utilities (cost-effectiveness models), Markov nodes (to deal with real-time dynamics), and multiple-decision-point models (e.g., sequential testing).

I expanded work by Wellman and associates (BUNYAN) (Wellman et al. 1989), Langlotz (QxQ) (Langlotz 1989), and Sonnenberg (MIDAS) (Sonnenberg et al. 1994). These three works provide a taxonomy for the structure and content of DMs. Both BUNYAN and QXQ, however, are able to support only single-utility models, and none of the three taxonomies can describe Markov processes. I expanded these taxonomies to deal with dual-utility models (for cost-effectiveness studies), and Markov nodes (for management of patients who have chronic diseases) (Section 3.5).

An important addition to these taxonomies is the explicit modeling of the assumptions in the DM. I represent four different types of assumptions in my DM conceptual model:

1. *Modeling perspective* (e.g., societal perspective). The perspective of the DM — especially that of a DM that reflects the costs involved or patient preferences for different strategies — affects the CPG produced.
2. *Best-estimate input values and ranges* (e.g., "the prior probability of mediastinal metastases is 0.46"). The input values and sensitivity-analysis ranges used in the base-case analysis reflect numerous assumptions made by the decision-analysis team. The chosen input values may combine results of clinical trials, meta-analyses of the current literature, and estimates based on expert opinion. Links to evidence tables and sources make these assumptions an explicit part of the generated CPG.
3. *Patient characteristics* (e.g., "the age-specific mortality rates used are for white males"). My DM conceptual model makes explicit assumptions about the patient population for whom the CPG is designed by highlighting the defining characteristics of the population and those variables that depend on the model representing this particular cohort.

4. *Model structure* (e.g. computed tomography and mediastinoscopy are conditionally independent). The structure of the model reflects assumptions regarding both the relationships among variables and the chosen simplifications of the scope of the DM.

Comparison of my DM and CPG conceptual models reveals numerous areas that do not overlap, such as the CPG objective, methods used, alternative and outcome definitions, sensitivity-analysis ranges, levels of evidence, detailed evidence tables, and the data sources used to develop the model. The CPG conceptual model obtains this additional information using the DM annotation editor described in Chapter 4.

1.10 Mapping between DM and CPG representations

I defined a formal mapping between the knowledge in DMs and the knowledge required for CPGs (Section 5.4). This mapping allows representation of an evidence-based CPG as a flowchart algorithm, explicit representation of assumptions in the DM, and maintenance of additional knowledge from the DM for future automated updating and tailoring of the CPG.

Creation of the flowchart algorithm uses a combination of the automated analyses of the DM and additional specifications from the decision analyst. The assumptions represented in the CPG are based on those obtained directly from the structure and variable inputs used in the DM, as well as the assumptions defined explicitly by the decision analyst using the DM annotation editor. The CPG provides information regarding its sensitivity to particular input variables, and the expected outcomes for different strategies. Additional sections of the CPG, such as the definition of the patient population, comprise information obtained directly from the DM (such as the starting age of the population) and from the decision analyst (such as the criteria for identification of patients as survivors of sudden cardiac death).

1.11 Tailoring of CPGs

ALCHEMIST's initial CPG reflects the DM base-case values. To make this CPG applicable to her specific site or patient, the local user must tailor and update the input variables. Such tailoring abilities are beneficial when clinical circumstances vary sufficiently that guideline recommendations differ (Nease and Owens 1994, Owens and Nease 1991a, Owens and Nease 1997). The user of the CPG is able to change the base-case variable values to reflect her specific patient population. After determining that none of the DM modeling assumptions are violated, ALCHEMIST updates the results and recommended CPG flowchart algorithm. Information indicating to which variables the CPG is sensitive, plausible ranges for given variables that ensure that the resulting CPG is still clinically valid, and relationships between variables to guarantee that modeling assumptions are not violated are all represented in the conceptual-model framework and made explicit to the CPG user. *ALCHEMIST does not allow structural updating or custom tailoring of the underlying DM*. Although it is possible to implement changes to the DM from a web interface, the requirements for ensuring that the structurally changed model is complete and that the resulting CPG is valid are complex and are not part of my thesis research.

1.12 Proposed users

In this section, I identify the parties who participate in the CPG process and who may benefit from use of ALCHEMIST. I am proposing that guideline developers adopt a framework and system that changes how certain subsets of guidelines are created, disseminated, and maintained. It is important, therefore, to know who will be affected by use of the ALCHEMIST system — who will benefit and who may be encumbered. I envision two main user groups for ALCHEMIST: decision analysts and guideline developers. Figure 1.4 depicts the flow of information and the users involved as a CPG is created from a DM and custom tailored by a local guideline developer.

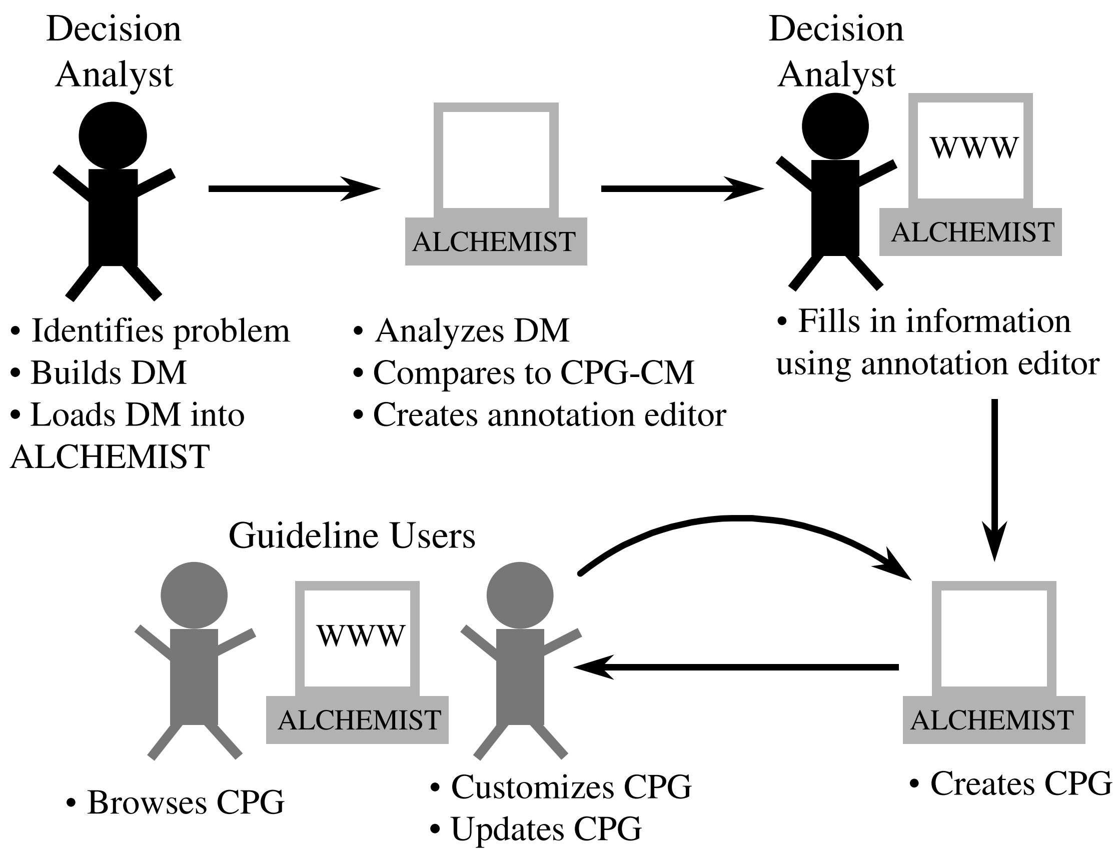


Figure 1.4. Proposed users of the ALCHEMIST system. This schematic demonstrates the flow of information through the ALCHEMIST system and the different intended users. The decision analyst identifies a clinical problem and builds a clinically valid DM that he then loads into ALCHEMIST. ALCHEMIST then analyzes the DM, attempts to instantiate the CPG conceptual model and creates the annotation editor. The decision analyst fills in the needed information in the annotation editor and submits this information to ALCHEMIST. ALCHEMIST then creates the global CPG which is presented to guideline developers who may either browse the CPG, or custom tailor or update the CPG with patient- and site-specific information.

The decision analyst who constructs a DM uses the ALCHEMIST system to upload his previously created DM into the DM conceptual model. ALCHEMIST then creates the DM annotation editor to query the decision analyst for additional information and evidence pertaining to the DM. Completing the DM annotation editor is additional work for the decision analyst, but provides structure to the modeling and evidence-gathering process and helps the analyst to create a DM that has no inconsistencies or structural problems, and that therefore can be transformed into a CPG. The analyst can update the information that he enters into the DM annotation editor if new information becomes available or if he wants to reflect different patient populations. Implementation of the annotation editor on the web gives decision analysts access to the editor from different institutions, and allows decision-analysis teams to share decision-modeling tasks among members located at geographically disparate institutions who are using different computing platforms.

The guideline developer who implements the resulting CPG has access to the structured DM and to that model's various input variables. She therefore can tailor and update the clinical guideline based on patient- and site-specific information, or on new clinical information. The assumptions in the DM are explicit, so that the guideline developer can determine the model's applicability to a specific patient population. For example, the base-case patient population used in model of sudden cardiac death (SCD) described in Section 1.13 comprises survivors of SCD. If a guideline developer wanted to change this model to reflect a different patient population, such as patients who have had a myocardial infarction (MI), she would want to modify several variables that would be affected. A variable such as the yearly cost for treating a patient with an implantable cardioverter defibrillator (ICD) takes into account the number of outpatient visits. A lower-risk population, such as post-MI patients, should have fewer outpatient visits than the original population of survivors of SCD, and therefore the yearly ICD cost should also be reduced. The patient-population description, therefore, is linked to variables such as yearly costs to reflect such dependencies between variables.

The creation of the DM and CPG conceptual models helps decision analysts and guideline developers by providing a detailed specification of the knowledge required for CPG creation. This specification allows these two user groups to create guidelines or DMs that are complete and internally consistent, and that can be updated or tailored, because the DMs specify all the required knowledge for such maintenance.

Although I envision the primary users of the ALCHEMIST system to be decision analysts and local guideline developers, clinicians could also use the generated CPG. ALCHEMIST would provide these clinicians with access to the CPG recommendations, evidence, modeling assumptions and to ALCHEMIST's custom-tailoring capabilities on a patient-specific level (Section 2.3.5).

1.13 Examples of decision problems, mappings, and resulting CPGs

I use three different DMs[1] to demonstrate the abilities of ALCHEMIST. In increasing order of complexity, these models are the following:

1. **Lung-cancer effectiveness model (LC-EM):** This DM represents the optimal strategy for staging the mediastinum of patients with known non–small-cell lung cancer. The only outcome modeled is life expectancy, although there are sequential decisions representing the numerous tests that can be used (Gould et al. 1997, Nease and Owens 1997, Owens et al. 1989).
2. **Lung-cancer cost-effectiveness model (LC-CEM):** This DM represents the same problem as does the LC-EM, but also models the financial costs incurred by the different strategies. It therefore requires representation of a dual-utility model.

1. Note that all three of the DMs used in my dissertation are for example purposes only and the clinical recommendations that ALCHEMIST produces based on these DMs should *not* be used by the reader in a clinical setting.

3. **Sudden cardiac death PORT Markov model (SCD-MM):** This DM represents a subset of the Cardiac Arrhythmia and Risk of Death Patient Outcomes Research Team (CARD PORT) DM that the CARD PORT decision modeling subgroup developed previously (Owens et al. 1997a, Sanders et al. 1996, Sanders et al. 1995). This model incorporates Markov processes; it therefore demonstrates ALCHEMIST's ability to represent time dependencies and recurrent or repeating events.

I use the LC-EM to step through how a user would interact with ALCHEMIST. In this scenario, a guideline-developing organization develops a DM that represents the alternatives, outcomes, evidence, assumptions, and knowledge for a decision problem, such as what mediastinal-staging strategy to use in patients who have non–small-cell lung cancer.

This DM represents a patient who has a known non–small-cell carcinoma of the lung. A chest X-ray examination reveals that the tumor does not abut the chest wall or the mediastinum. If mediastinal metastases are found to be present, then thoracotomy is contraindicated, and the preferred treatment is radiation therapy. However, if mediastinal metastases are absent, then thoracotomy offers a substantial survival advantage. There are several diagnostic tests available to assess any involvement of the mediastinum (Nease and Owens 1997). In this DM, we consider the use of only computed tomography (CT) of the chest and mediastinoscopy. Figure 1.5 shows a schematic representation of the LC-EM decision tree.

The first decision is whether to perform a CT. This decision is then followed by the decision of whether to perform mediastinoscopy (note that, if a CT was performed, the results of this test are available before the decision to perform the mediastinoscopy is made). Mediastinoscopy includes a small (but not insignificant) risk of death. Survival of mediastinoscopy (or the absence of mediastinoscopy altogether) is followed by the treatment decision (thoracotomy versus radiation therapy). The results of both CT and

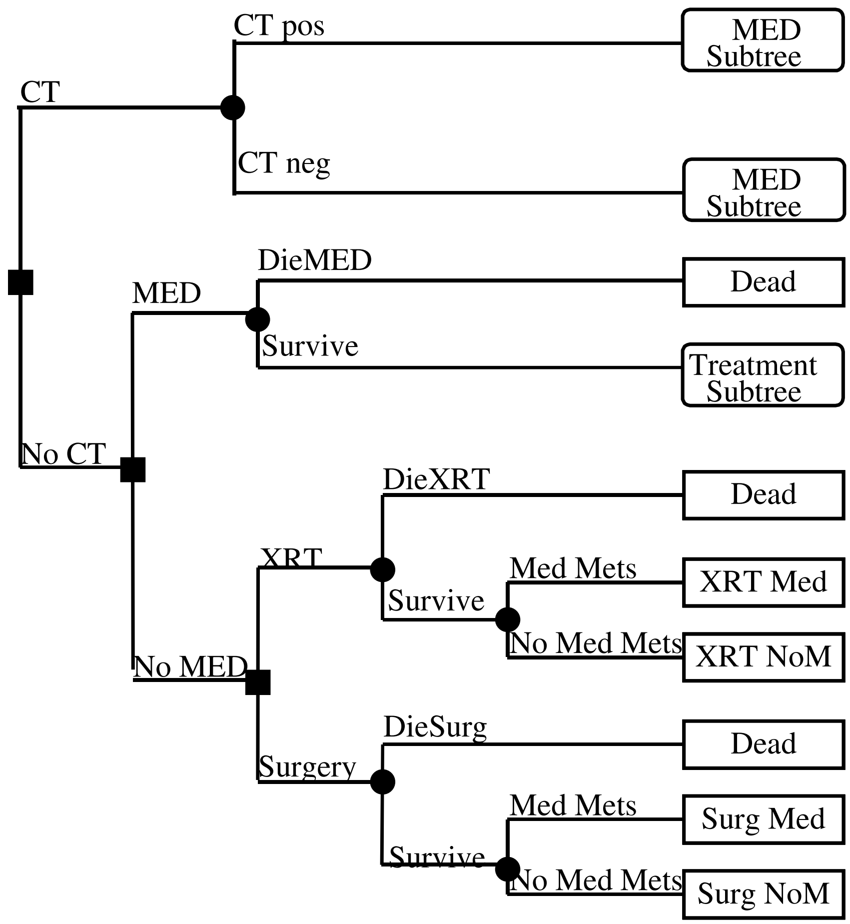


Figure 1.5. Schematic representation of the lung-cancer effectiveness model. Square nodes represent decision nodes, circles represent chance nodes. CT = computed tomography, MED = mediastinoscopy, XRT = radiation therapy, Surgery = thoracotomy, MedMets = mediastinal metastases.

mediastinoscopy are known before the treatment decision is made. Both treatments incur a risk of death and each has a different life expectancy based on the presence or absence of mediastinal metastases.

I implemented this specific decision tree using Decision Maker. The actual decision-tree implementation has much greater detail than the schematic representation in Figure 1.5 (it includes, for example., probabilities, utilities, and variable bindings). Sufficient knowledge for producing a CPG, however, is not contained in the decision-tree representation.

After the decision analyst has completed the DM, he loads it into ALCHEMIST (Figure 1.6). The ALCHEMIST implementation uses a web-based interface and common-gateway interface (CGI) scripts, so it can be run on any computer platform, and the decision analysts can load any decision tree that is located on his personal-computing environment.

Figure 1.6. Loading the DM into ALCHEMIST. The decision analyst enters her name and electronic-mail address for logging purposes. She then browses her personal files to locate the Decision Maker DM that she will load into the ALCHEMIST system.

The DM is loaded into the ALCHEMIST system, which parses the DM and maps the information obtained from the DM onto the DM conceptual-model framework. ALCHEMIST then creates a web-based DM annotation editor that queries the decision analyst for additional information where needed (Figure 1.7 and Figure 1.8).

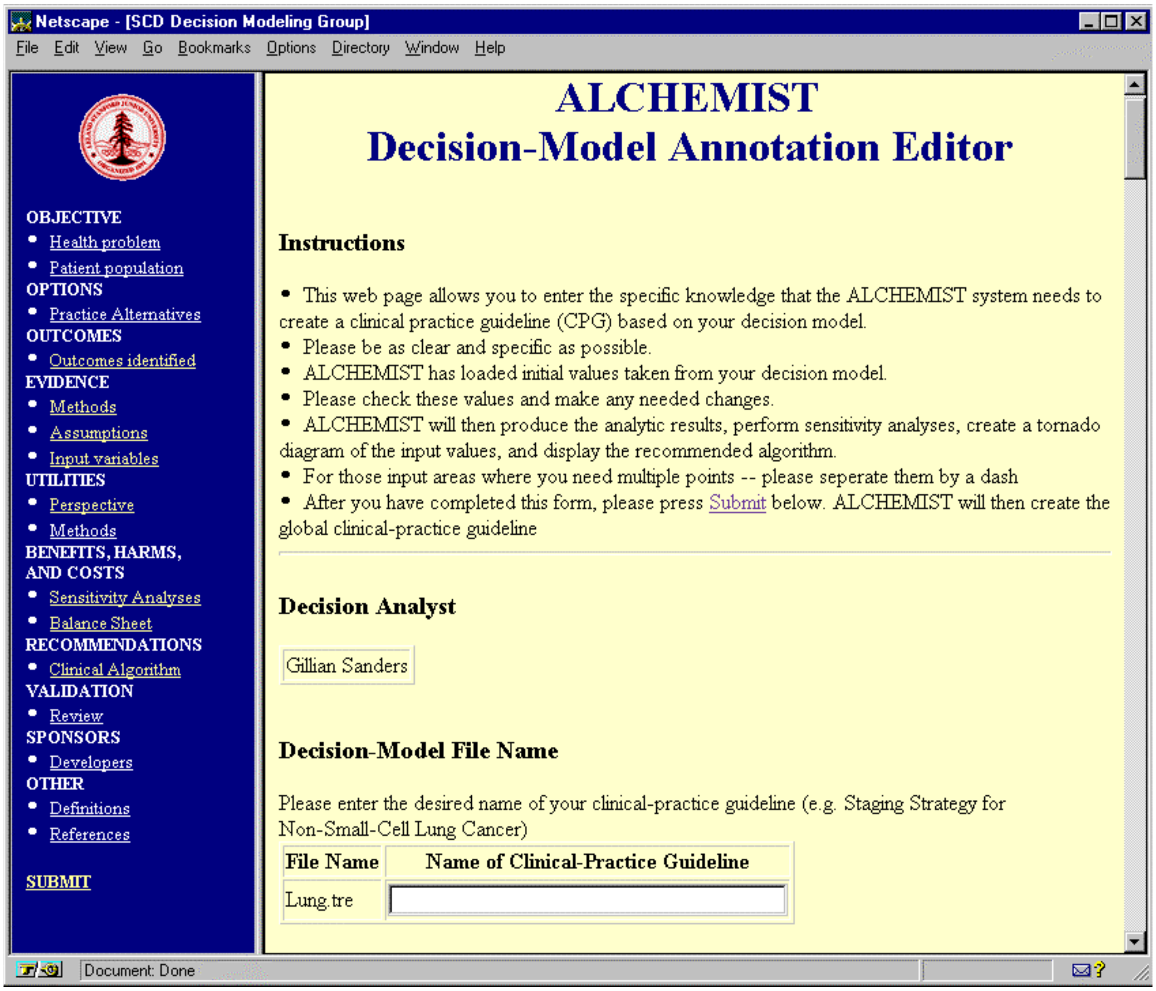


Figure 1.7. Introduction to the DM annotation editor. The menu on the left side of the screen outlines the organization of the CPG and allows the decision analyst to navigate through the numerous areas that require annotation. Note that the results, sensitivity analyses, and clinical algorithm will be generated automatically by ALCHEMIST and therefore do not require annotation.

The annotation editor requires that the decision analyst do additional work. However, the information that he is required to enter should be readily available: It comprises the data and knowledge used in the DM. Once the decision analyst becomes familiar with the format of the annotation editor, he can also use that format to help him to organize the

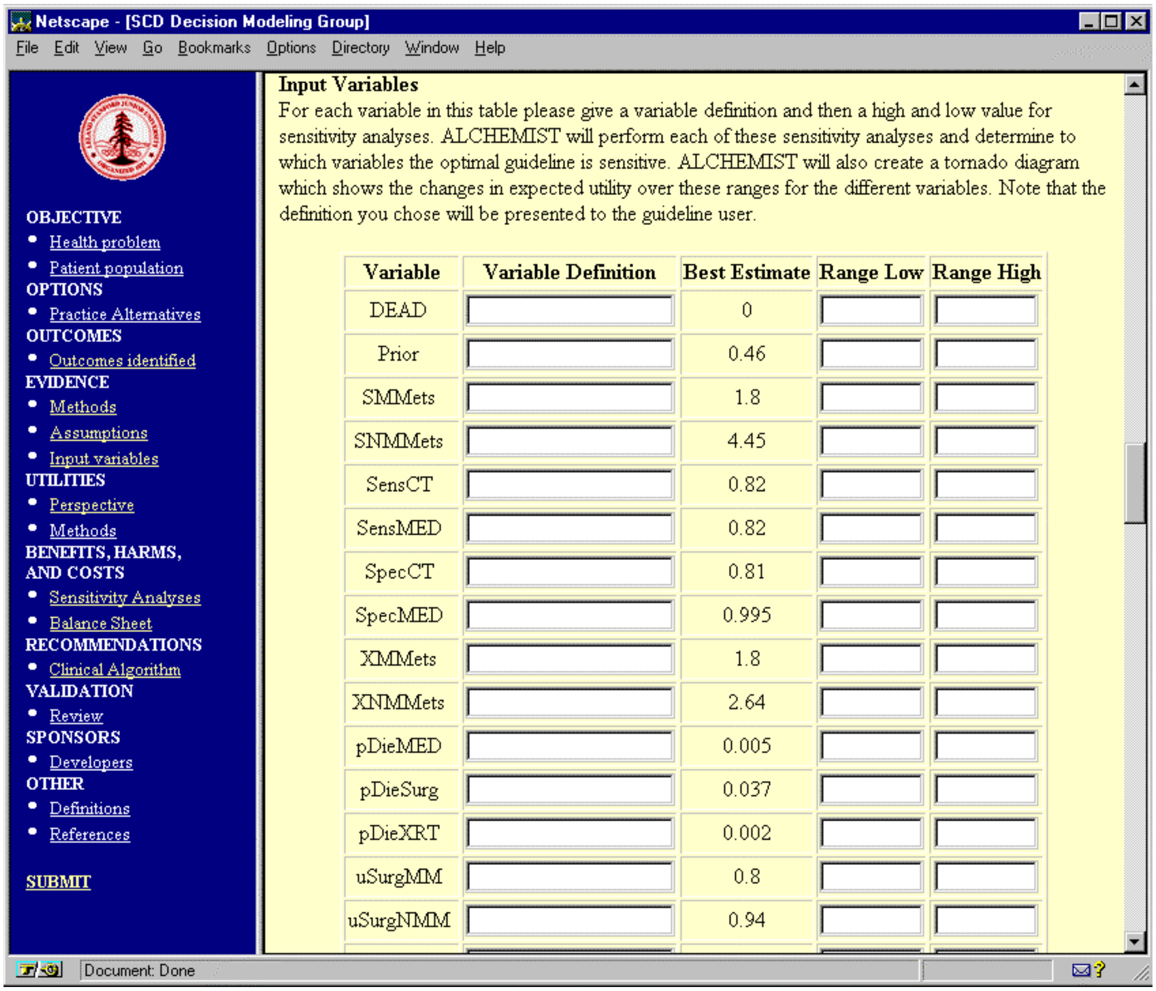


Figure 1.8. Annotation editor. White input boxes indicate where the decision analyst needs to enter input information about the DM input variables, such as definitions, low estimates, and high estimates for the sensitivity analyses. The best estimates are obtained directly from the DM and are listed here.

evidence as he creates the DM. After entering the needed information in the annotation editor, the decision analyst submits this information to ALCHEMIST, which then produces the CPG browser and custom-tailoring editor (Figure 1.9 and 1.10).

This web-based CPG is then subject to internal and peer review, and can be modified. After the guideline is accepted, the developing organization distributes it to the health-care community by adding it to a guideline-repository web page. An example of such a

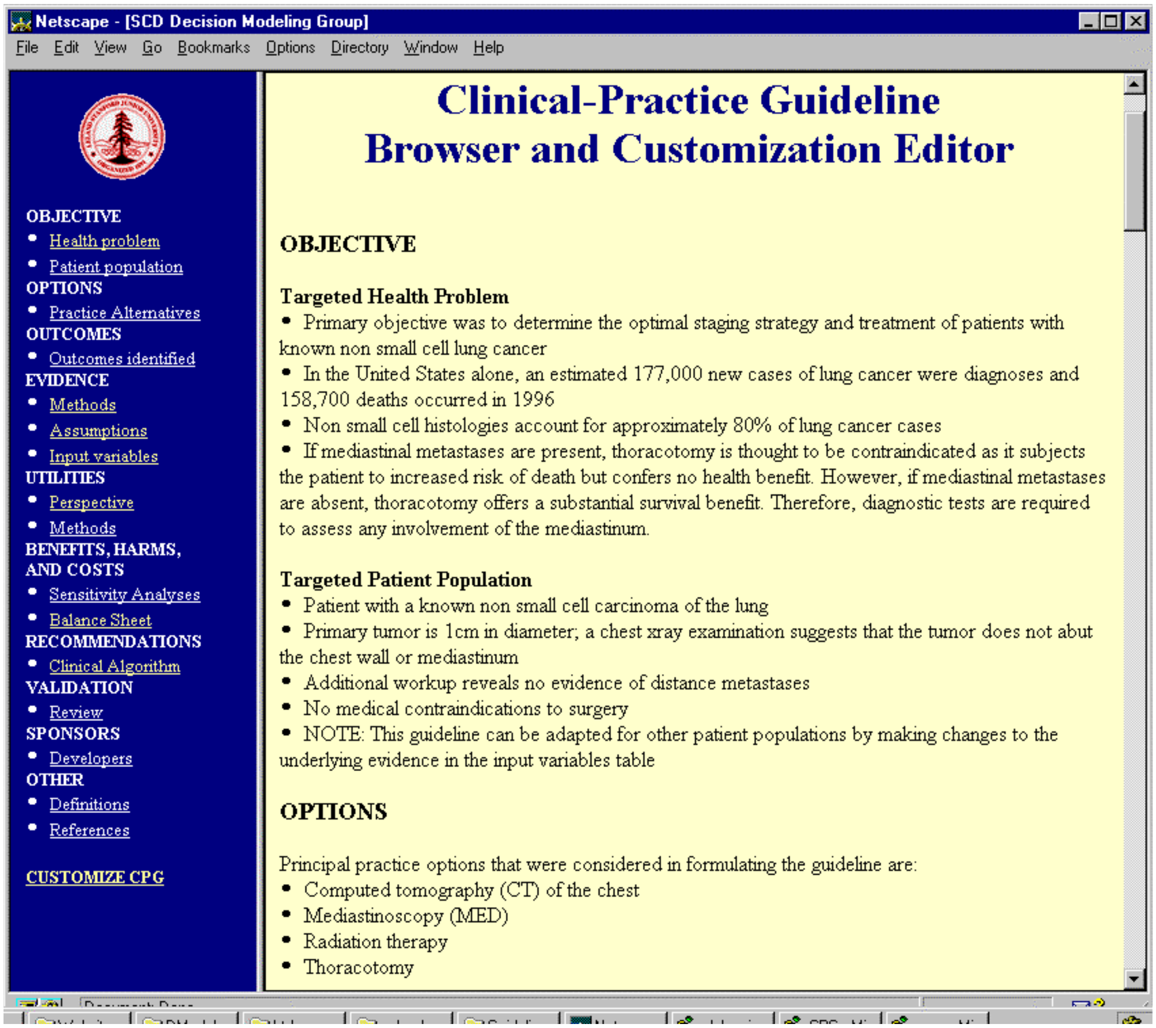


Figure 1.9. CPG that ALCHEMIST produces using information from the annotation editor.

guideline repository is the newly created National Guideline Clearinghouse (NGC) by the Agency for Health Care Policy and Research (AHCPR)) (Agency for Health Care Policy and Research, 1998). The goal of the AHCPR in creating the NGC is to promote widespread access through the development of a comprehensive electronic database for thousands of guidelines. To be included in the NGC, the guidelines must satisfy a set of inclusion criteria. I envision the guidelines created by ALCHEMIST as being eligible for to the NGC; they would then be available for widespread use by local guideline developers. After the global guideline is available over the web, local guideline implementers can

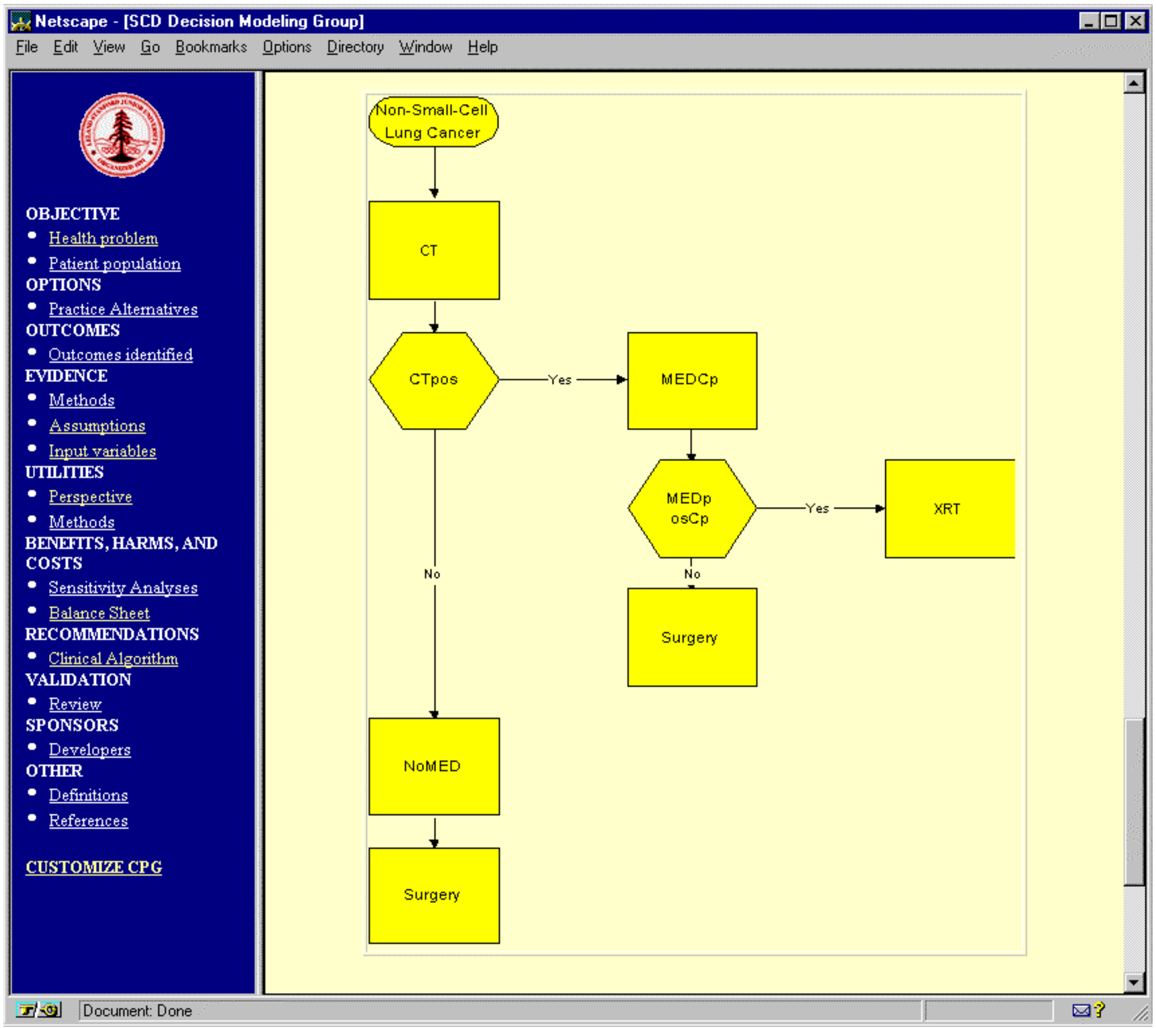
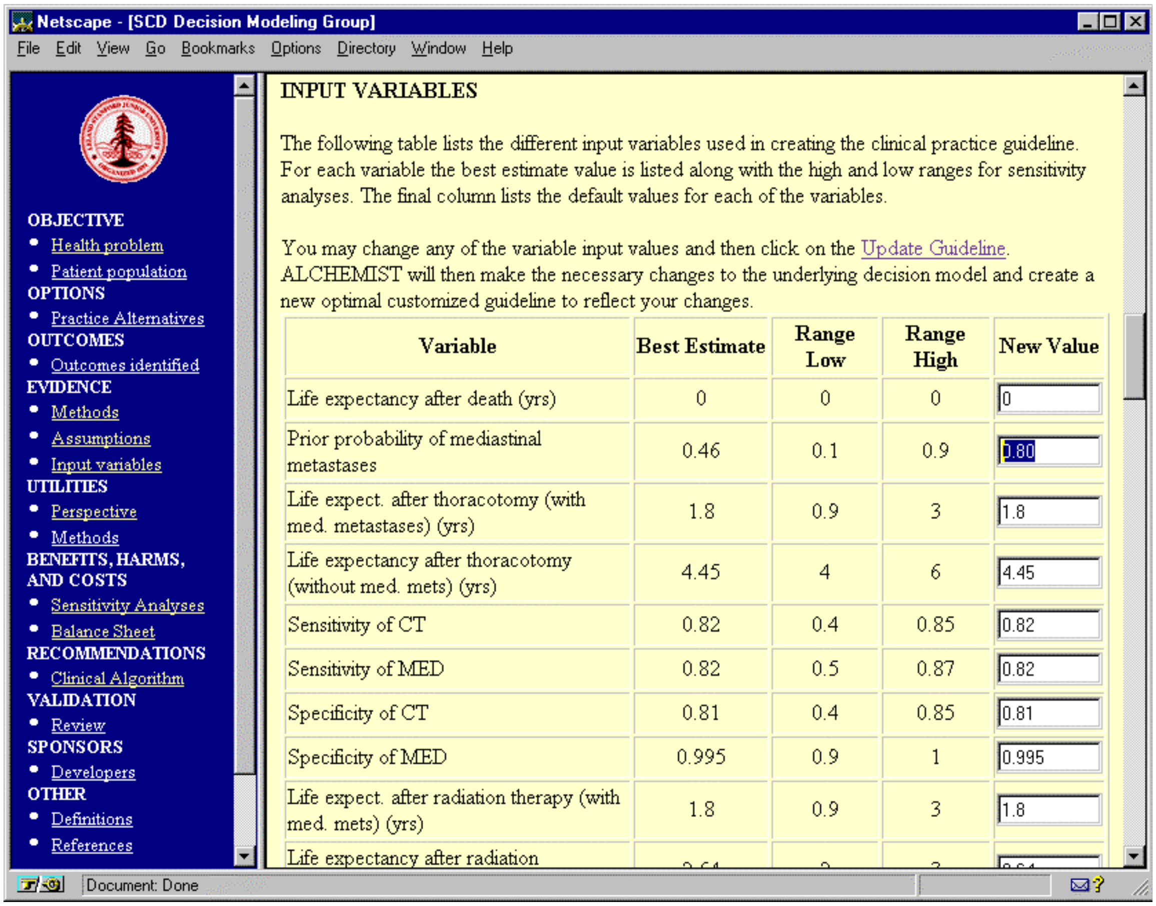


Figure 1.10. CPG flowchart representation generated by ALCHEMIST.

explore it, can examine its evidence and recommendations, and can specify site- or patient- specific input values to produce an updated tailored CPG. Figure 1.11 shows a guideline user changing the prior probability of mediastinal metastases from the base-case value of 0.46 to a new value of 0.80. ALCHEMIST takes this new value, and — after checking that none of the underlying assumptions in the DM are violated — produces a new CPG. Using a base-case value for the prior probability of mediastinal metastases of 0.46, the generated CPG (Figure 1.10) recommended a CT examination followed by mediastinoscopy if the CT was positive, and followed by a thoracotomy if the CT was negative.

Changing the prior-probability value to 0.80 produces the updated CPG (Figure 1.12), which recommends immediate mediastinoscopy.



Variable	Best Estimate	Range Low	Range High	New Value
Life expectancy after death (yrs)	0	0	0	0
Prior probability of mediastinal metastases	0.46	0.1	0.9	0.80
Life expect. after thoracotomy (with med. metastases) (yrs)	1.8	0.9	3	1.8
Life expectancy after thoracotomy (without med. mets) (yrs)	4.45	4	6	4.45
Sensitivity of CT	0.82	0.4	0.85	0.82
Sensitivity of MED	0.82	0.5	0.87	0.82
Specificity of CT	0.81	0.4	0.85	0.81
Specificity of MED	0.995	0.9	1	0.995
Life expect. after radiation therapy (with med. mets) (yrs)	1.8	0.9	3	1.8
Life expectancy after radiation	[illegible]	[illegible]	[illegible]	[illegible]

Figure 1.11. Tailoring of the CPG. The user here has entered 0.80 for the prior probability of mediastinal metastases.

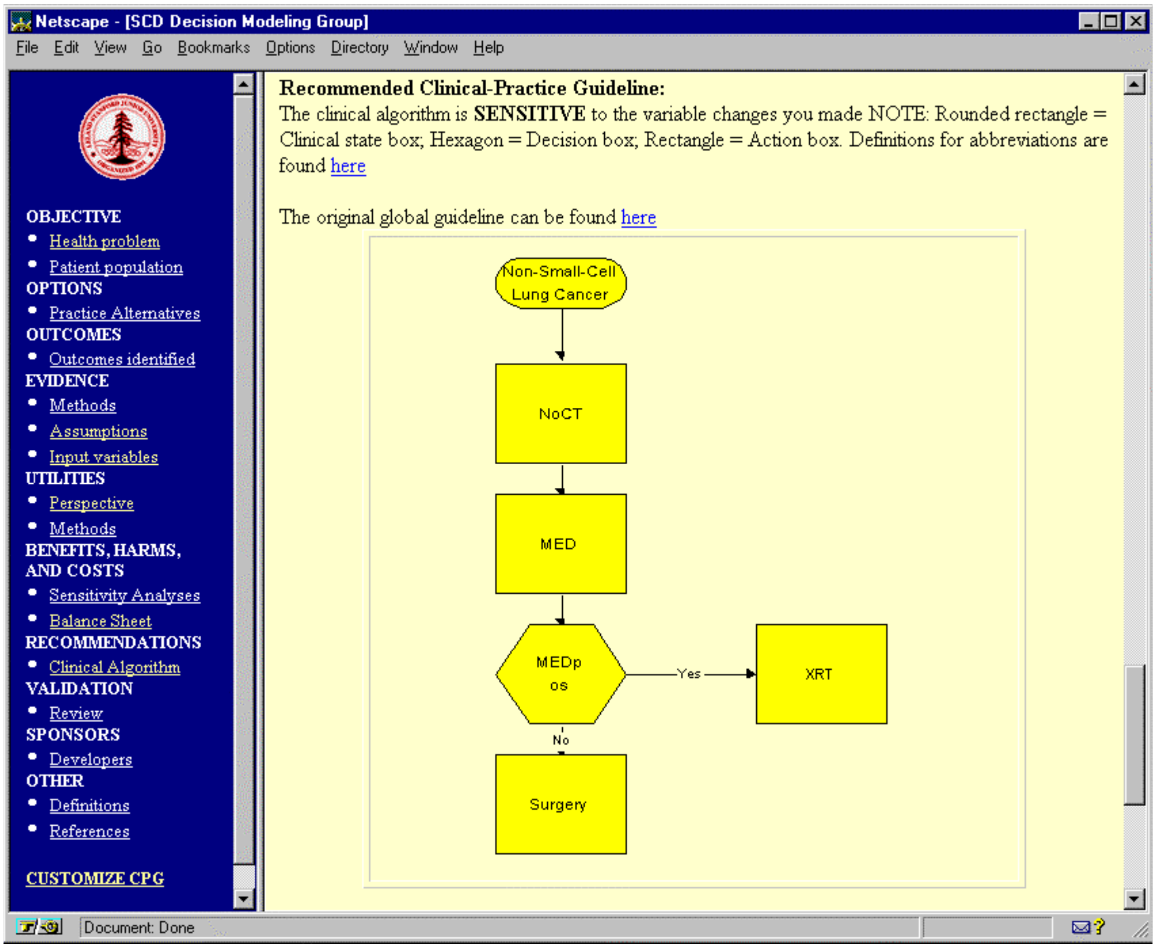


Figure 1.12. Tailored CPG. The flowchart representation has been updated to reflect the 80-percent prior probability of mediastinal metastases that the user entered in Figure 1.11.

1.14 Evaluation

My evaluation had three separate phases: (1) evaluation of the DM and CPG conceptual models and of the mapping algorithm between these conceptual models, (2) evaluation of the generated CPG, and (3) evaluation of the updating and tailoring abilities of ALCHEMIST.

1.14.1 Conceptual-model framework

I evaluated the design and expressivity of the DM and CPG conceptual models. Using the guideline-assessment tool developed by the IOM (Institute of Medicine 1992), I compared the data elements within my conceptual model to those required by the IOM assessment tool. Of the 60 individual data elements, I found 45 (75 percent) in the CPG conceptual model. In Section 8.1.1.3, I explain reasons for exclusion of the remaining 15 elements from my CPG conceptual model. To determine the expressivity of the CPG conceptual model, I retrieved a sample of current guidelines and attempted to express the information within each of these CPGs with my conceptual framework. Overall, the CPG conceptual model was able to capture the major elements of all the guidelines.

1.14.2 Generated CPG

I evaluated ALCHEMIST's resulting CPG in three ways: using a guideline-rating questionnaire, a user-satisfaction questionnaire, and a structured interview. For this part of my evaluation, I solicited 15 subjects who were familiar with guideline use. First, each subject critiqued the LC-EM and a comparison CPG with reference to an established guideline criteria (Basinski 1995, Institute of Medicine 1992, Sonnad et al. 1993). Using the guideline criteria, each subject gave the CPGs a numeric score based on the CPG's usability, accountability, and accuracy. The mean score for the LC-EM was 1.502 (on a scale of 0 to 2), whereas the comparison CPG had a mean score of 0.987. The difference between the two guidelines was statistically significant (p = 0.002). The ALCHEMIST CPG was rated higher than the comparison CPG on all questions except for the one that asked the subject whether the CPG had been peer reviewed.

Second, the subjects completed a user-satisfaction questionnaire to describe their experience with the ALCHEMIST web-based system. The results of this questionnaire produced scores on ALCHEMIST's ease of use, usefulness of the content, and format of presentation.

Using an ordinal scale of 1 to 5 where 5 is ideal; the subjects rated ALCHEMIST's ease of use at 4.76, the usefulness of the content at 3.98, and the format of the presentation at 4.64.

Third, I conducted structured interviews with the subjects to elicit any additional comments or concerns. Overall, the subjects' experience using the ALCHEMIST system was extremely positive. I discuss the results of this structured interview in Section 8.3.1.4.3.

1.14.3 Updating and tailoring abilities

To evaluate the custom-tailoring and updating abilities of ALCHEMIST, the 15 subjects entered between 1 and 3 new clinical scenarios into the ALCHEMIST system. ALCHEMIST generated a new CPG, and I compared the expected health benefit and flowchart algorithm produced by ALCHEMIST to that produced through manual computation. ALCHEMIST's tailoring of the CPG exactly mimicked manual computation of the flowchart algorithm in all patient scenarios. This component of my evaluation demonstrated the feasibility and accuracy of ALCHEMIST to produce tailored CPGs.

1.15 Contributions

My work combines ideas from decision analysis, health policy, and medical informatics to produce a methodology for the automated creation of evidence-based CPGs. The design of the CPG and DM conceptual models combines extensive domain knowledge about the proper structure of CPGs and the knowledge within DMs. The ALCHEMIST system provides a proof of concept that the transformation of DMs into CPGs, and CPG automatic updating and tailoring, can be performed. The evaluation of the produced CPG browser and custom-tailoring editor generated pilot data that will be helpful to people who design future quantitative studies that compare the use of CPGs created automatically from DMs with that of the current existing CPGs. A description of the specific contributions to the domains of medical informatics, decision theory, and health policy follows.

1.15.1 Medicine and medical informatics

- The ALCHEMIST system allows the automated construction of CPGs that represent explicitly the uncertainties and evidence inherent in clinical decision-making problems, allowing users to determine the degree to which the resulting CPG is pertinent to their patient population and thus whether it requires tailoring.
- The mapping algorithm provides an automated, quantitative link between the clinical data from the literature and the CPGs, allowing users to view how changes to the clinical data affect the resulting CPGs and their patient-management strategies.

1.15.2 Decision theory

- The DM conceptual model provides a taxonomy of DMs and the knowledge within those DMs, helping decision analysts to create DMs that are complete and that guideline users can use for creating CPGs.
- The translation of the DM into a CPG demonstrates a method for providing evidence-based guidelines to users who are unfamiliar with the technical and mathematical details of a DM.
- The CPG-tailoring system provides an automated decision-support system to be used for specific sites or patients.
- The mapping of a DM into algorithmic form is domain independent and can be applied easily in fields other than medicine.

1.15.3 Health research and policy

- The CPG conceptual model formalizes the knowledge required for the creation of an algorithmic CPG.
- ALCHEMIST addresses the tension between performing comprehensive high-quality analyses centrally and accommodating legitimate variances in practice patterns and in patient preferences

1.16 Guide to the dissertation

In this dissertation, I define conceptual models of CPGs and DMs, and a method for mapping between these two knowledge representations. If you are to understand my goals, you must first have a general understanding of CPGs, DMs, and the knowledge inherent in these two representations. The next two chapters provide an introduction to these topics, explain the limitations of these formats, provide the background for understanding the contributions of ALCHEMIST, and detail my conceptual models. I describe my mapping algorithm, the implementation of ALCHEMIST, and my evaluation in the remaining chapters.

- Chapter 2 presents an overview of CPGs and the limitations current CPG development have for the three core CPG tasks: creation, dissemination, and maintenance. I define the subset of CPGs that I used, and describe the knowledge necessary to perform the three core tasks for this subset of CPGs. I provide a detailed taxonomy of this knowledge, and define a conceptual model that formalizes this knowledge.
- Chapter 3 introduces DMs and the knowledge explicit in their structure and inputs. I evaluate existing taxonomies of DMs, and describe my conceptual model that formalizes the knowledge already present in the DM, and that requires the decision analyst to enter any missing knowledge.

- Chapter 4 describes the implementation of the DM conceptual model and the DM annotation editor based on the conceptual model formulated in Chapter 3. I list any restrictions on the class of DMs that ALCHEMIST will be able to transform into CPGs, and detail the knowledge required for the creation of the CPG and unavailable in the DM. This chapter concludes by describing extensions that would allow ALCHEMIST to include a greater array of DM types.
- Chapter 5 reviews the justification for basing CPG creation on DMs. I review existing methods for transforming a DM representation into a textual and algorithmic CPG form. I assess these differing methods, and describe my DM-to-CPG algorithm. I detail the restrictions on the translation between DM and CPG representations.
- Chapter 6 describes ALCHEMIST's tailoring and updating abilities. It formalizes the process of making local adjustments to the CPGs that represent changes in the DM, and describes restrictions on ALCHEMIST's abilities (e.g., being able to make adjustments on the level of inputs, such as on local disease prevalence, but not regarding model structure, such as the existence of a new treatment strategy).
- Chapter 7 steps through three extended example translations from DM to CPG. These examples illustrate the performance of ALCHEMIST.
- Chapter 8 details my evaluation and results for validating my hypotheses. I report my evaluation of the two conceptual models, the mapping algorithm, the resulting CPG, and the custom-tailoring abilities of the ALCHEMIST system.
- Chapter 9 concludes by discussing the contributions of my work to health care, to decision theory, and to medical informatics. I point out the limitations of my thesis research, and include future directions for extending my work.

Chapter 2

Clinical-Practice Guidelines

In this chapter, I define **clinical-practice guidelines (CPGs)** and review the rationale and need for CPGs in our health-care system (Section 2.1). I describe current CPG development, implementation, and maintenance projects, and then outline the specific limitations of existing methods (Section 2.2). These current CPG projects encompass a large array of purposes, types, domains, levels of complexity, and formats. In my thesis work, I have placed numerous restrictions on ALCHEMIST's CPG representation. I describe the subset of CPGs that I have addressed, and detail the restrictions that this subset imposes on the generalizability of my work (Section 2.3). I complete this chapter by describing the conceptual model that I developed for representing CPGs (Section 2.6).

2.1 Definition of CPGs

The Institute of Medicine (IOM) defines CPGs as "systematically developed statements to assist physician and patient decisions about appropriate health care for specific clinical circumstances" (Institute of Medicine 1992). This broad definition encompasses numerous guideline formats, types, and purpose. Common CPG formats include prose, protocol lists or charts, and clinical flowchart algorithms. Each of these presentation formats, or any combination thereof, can also be paper or computer based. Similarly, computer-based guidelines can be static or interactive. ALCHEMIST's generated CPG uses a computer-based combination of structured text and a clinical flowchart algorithm. In Section 2.3.4, I detail this CPG format.

There are also numerous CPG types and purposes. CPG purposes range from educational reference for providing explanatory information on approaches to a problem, to care plans or audits to review utilization of resources or to monitor conformance, to CPGs of accepted practices (Eddy 1992, Ohno-Machado 1998). I concentrate on CPGs created to assist clinical decision making by patients and practitioners. These CPGs could be developed to aid practitioners with patient-management, diagnostic, disease-screening, and disease-prevention decisions.

2.2 Current status of CPGs

This section provides a brief background on the justification for CPGs, and describes current CPG projects. I detail the current methods of developing, disseminating, and maintaining CPGs, and also describe current computer-based CPG projects.

2.2.1 Justification

Rising health-care costs and concerns about quality of care have motivated many recent guideline-development projects. Inspiring CPG developers is the belief that, if properly developed and used, CPGs can lead to improved patient outcomes, enhanced patient satisfaction, and reduced health-care costs (Shapiro et al. 1993, Walker et al. 1994, Woolf 1990). These expected benefits derive from the solution of the following health-care problems (Kaegi 1996):

1. *Inappropriate use of services:* CPGs are designed to reduce inappropriate use of health-care services. Studies have documented unnecessary radiologic procedures, laboratory tests, hospital admissions, days of care, and drug use. Ten to 30 percent of diagnostic and therapeutic procedures have been judged to be performed for inappropriate reasons (Kassirer 1993, Leape 1990). By providing a systematic means of management review and a formal guideline for practitioners to follow, developers of CPGs hope to decrease significantly the occurrence of inappropriate care.
2. *Uncertainty about health outcomes from various services:* There is a high degree of uncertainty about health outcomes produced by various medical services. This uncertainty allows physician practice style, peer opinion, tradition, financial incentives, and expectations of patients to affect the chosen therapy strategy and perhaps to cause people to choose an inappropriate or more costly route. CPGs would ideally limit this uncertainty through the explicit documentation of the differing alternative treatments and the latter's respective benefits, harms, and costs.
3. *Variation in practice patterns:* Researchers have documented that there is significant variation in the use of services that cannot be explained consistently by differences in practice location, patient characteristics, or patient preferences (Conway et al. 1995, Fisher et al. 1992, Health Services Research Group 1992a, Iscoe et al. 1994, Keller et al. 1990, Welch et al. 1993). Wide variations in use suggest that patients in certain areas are receiving unnecessary services, whereas others are failing to receive needed services (Leape 1990). These variations may stem from several factors, the most obvious being a lack of consensus within the medical community regarding the appropriate indications for many treatment or screening strategies. The implementation of CPGs would provide a method for reducing this variation by recommending appropriate levels of care given different clinical scenarios.

Physicians are required to remain abreast of a staggering body of medical knowledge (Health Services Research Group 1992b). To make an informed patient-care decision, physicians must identify accurately the available options, possible outcomes, and intermediate consequences, as well as weigh the resulting benefits, harms, and costs according to their patient's, institution's, and society's preferences. Clearly there are many health-care decision problems that are sufficiently straightforward that the corresponding guideline can be developed without a formal method (e.g., whether to use sterile techniques in an operating room) (Eddy 1992). However, many health decisions are too complicated or have too many associated uncertainties for physicians to perform the needed analysis for their individual patients. Indications that a clinical condition falls into this category include the existence of wide variations in practice patterns, ongoing clinical research, and conflicting policies (Eddy 1992). Valid CPGs developed by experts who have the required time and expertise to review all the available evidence, and to weigh all the benefits and harms of each decision, can greatly ease this decision-making task for physicians (Eddy 1990a).

2.2.2 Current CPG projects

Today, estimates of the number of CPGs in the United States range from 1800 (American Medical Association 1996, Physician Payment Review Commission 1995) to greater than 26,000 (ECRI 1997, Woolf 1998). Organizations that develop CPGs include numerous professional societies, such as the American Medical Association and the American College of Physicians; third-party payers; research groups such as the RAND corporation; academic and health maintenance organization (HMO) health centers; commercial precertification and utilization-review programs; and government organizations, such as the National Institutes of Health, the Health Care Financing Administration, Physician Payment Review Commission, U.S. Preventive Services Task Force, and Agency for Health Care Policy and Research. Each institution uses different methods of data collection and synthesis, CPG representation, implementation, maintenance, and evaluation; each has its own objectives for its produced CPG. In Section 2.2.2.1 through 2.2.2.3, I briefly describe the current methods used for development, dissemination, and maintenance of CPGs, indicating areas where ALCHEMIST addresses the current methods' limitations.

2.2.2.1 Development

Guideline-development methods range from informal consensus development (unstructured, subjective group judgment) and formal consensus development (standardized opinion gathering) to evidence-based methods (direct linkage of recommendations with supporting science) and explicit approaches (projections of likely benefit, harms, and costs) (Woolf 1992, Woolf 1998). Methods for evidence-based or explicit CPG development include several important tasks: topic selection, processes for data collection (e.g., expert-panel consensus or meta-analysis of clinical literature), methods for combining the available scientific evidence (e.g., decision analysis), outcomes measurement, techniques for determining and incorporating patient preferences (e.g., time tradeoff or standard gamble), and means for identifying and evaluating inconsistencies and conflicts among CPGs on the same topic (Eddy 1992, Institute of Medicine 1992, Woolf 1991, Woolf 1992). Each guideline-developing organization uses a different combination of these methods. Example differences among CPG development projects involve the emphasis placed on formal literature review and meta-analyses, the reliance on national experts as opposed to local physicians, and the application of rigorous analytical techniques for combining the chosen evidence (Audet et al. 1990, Woolf 1992).

The RAND corporation uses a completely different approach to guideline development. It has designed **appropriateness criteria** for certain medical or surgical procedures. A nine-member multispecialty panel reviews the background material for a given procedure and rates each possible indication on a 9-point appropriateness scale (US Congress Office of Technology Assessment 1994). Using these criteria, the panel rates the appropriateness of up to several thousand separate indications for a given diagnostic or therapeutic procedure. Appropriateness criteria can be distinguished from CPGs, and, therefore, excluded from the subset of CPGs relevant to my thesis work, because the criteria are not primarily designed to assist patient or physician decision making (Institute of Medicine 1992).

As I described in Chapter 1, development of CPGs requires input from experts both in clinical medicine and in evidence synthesis — resources that are not normally available to a local guideline developer. These large resource requirements often limit produced CPGs to global patient-management recommendations. This global perspective may prevent CPGs from being implemented effectively in a local environment, because the guideline user may not believe that the CPG is based an accurate representation of her patient population (National Centre for Reviews and Dissemination 1994, Woolf 1998). ALCHEMIST is able to take advantage of the global expertise used in developing a DM and the corresponding CPG, while providing the local guideline developer the option to tailor the guideline to reflect her specific site or patient population.

2.2.2.2 Dissemination

Although most current guideline-related activity emphasizes the development of CPGs, if CPGs are not properly disseminated, their potential benefit is negligible (Oxman 1993, Shortliffe 1990). The health-care benefit of CPGs is dependent on their widespread dissemination and use, and on physician compliance. Current methods of disseminating CPGs include release and promotion in peer-reviewed journals, mailings, press releases, and presentations at major meetings (Health Services Research Group 1992b). A guideline implementer's desire is to find a method that allows widespread dissemination, emphasizes the validation and support of the guideline by respected institutions and physicians, and allows easy use of the CPGs and understanding of the supporting evidence. The implementation of ALCHEMIST and its resulting CPG on the web will allow widespread distribution of the CPG and will permit easy access to the supporting evidence (See Section 2.2.2.4), but doesn't support currently integration with a patient record at the point of care. I discuss such integration in Section 9.2.

2.2.2.3 Maintenance

After a CPG has been developed and disseminated, the work of the guideline implementer is far from done. Although the guideline may reflect the current state of medical knowledge, periodic reviews, updating, and revisions are required for a CPG to remain valid (Institute of Medicine 1992). The abilities of a guideline developer to use ALCHEMIST to update the evidence used in the DM and to produce an updated version of the CPG alleviate problems with current CPG maintenance. For more major updates (e.g., addition of a new intervention), the underlying decision model (DM) can be changed; then, building off the initial global guideline, ALCHEMIST can create an updated global CPG.

2.2.2.4 Advantages of a computer-based CPG

Guideline developers can address several limitations of current CPG projects by making the CPGs available in a computer-based environment — especially by using a web-based interface. I previously discussed several of these advantages with reference to presenting and explaining medical DMs using a web-base format (Sanders et al. 1996, Sanders et al. 1998). The web-based DM presentation allows authors to present their DM in detail, to link the model inputs to the primary evidence, to disseminate the model to peer investigators for critique and collaborative modeling, and to enable users to analyze interactively the DM at remote sites. Similar advantages are available for a CPG representation. A web-based CPG can incorporate greater complexity than paper-based formats while maintaining a consistent structure and orientation for the user. The level of detail displayed can be varied, allowing the user to view parts of the CPG relevant to her patient or site and thus to tailor the recommendations to her specific population (Abendroth and Greenes 1989, Abendroth et al. 1988). The web allows easy linking to related supporting material, as well as interaction with the underlying DM, and thereby permits updating and tailoring of the generated CPG.

Prior work by Liem and colleagues describes a process for implementing algorithmic CPGs in a graphical format on the web. That work however, assumes the prior existence of these algorithms and does not allow creation of tailored or updated CPGs; rather, it allows only interactive browsing of existing flowchart algorithms (Liem et al. 1995).

2.2.2.5 Current web-based CPGs

During the past few years, the World Wide Web has become an alternate source for guideline developers to disseminate CPGs. There are substantial guideline resources available currently on the web, and use of the web by guideline developers continues to evolve rapidly. Table 2.1 lists several existing websites that provide full-text CPGs, serve as directories of CPGs, or describe ongoing research in the development of CPGs.

In contrast to CPGs that appear in peer-reviewed journals, a web-based CPG does not need to undergo the same level of internal or external review before its dissemination. AHCPR's creation of a national guideline repository (the National Guideline Clearinghouse discussed in Section 1.13) may help to ensure that web-based CPGs are of high quality, while providing guideline developers a flexible medium through which to disseminate and implement CPGs.

Table 2.1. Selected guideline resources available on the World Wide Web. (Reprinted by permission of *The Western Journal of Medicine. Source:* Owens DK. Use of Medical Informatics to Implement and Develop Clinical Practice Guidelines. West J Med 1998; 168:166-175).

Organization or website	Description of site	URL
American College of Cardiology (ACC)	ACC/AHA guidelines on management of heart disease	http://www.acc.org/clinical/guidelines/index.html
American College of Physicians	Search capability for guidelines published in the *Annals of Internal Medicine*	http://www.acponline.org/journals/annals/pastiss.htm
Canadian Medical Association	Links to over 200 guidelines, indexed by discipline, title, and developer	http://www.cma.ca/cpgs/

Table 2.1. Selected guideline resources available on the World Wide Web. (Reprinted by permission of *The Western Journal of Medicine. Source:* Owens DK. Use of Medical Informatics to Implement and Develop Clinical Practice Guidelines. West J Med 1998; 168:166-175).

Organization or website	Description of site	URL
Centers for Disease Control (CDC) Prevention Guidelines Database	Comprehensive summary of all the official guidelines and recommendations published by the CDC	http://aepo-xdv-www.epo.cdc.gov/wonder/prevguid/prevguid.htm
Decision Sciences Group, Harvard University	Interactive implementation of National Cholesterol Education Project guidelines	http://dsg.harvard.edu/public/guidelines/cholesterol/chlintun.html
McMaster University, Guidelines Appraisal Project (GAP) Homepage	Appraisals, summaries, and dissemination of information about CPGs	http://hiru.mcmaster.ca/cpg/default.htm
Medical Matrix, Healthtel Corporation	Links to a database of resources that can assist clinicians in patient care; the linked websites are reviewed by an editorial board composed of members of the Internet Working Group of the American Medical Informatics Association	http://www.medmatrix.org/SPages/Practice_Guidelines.asp
Radiologic Society of North America	Links to many guidelines indexed by the developer	http://www.rsna.org/practice/guidelin/guidelin.html
Stanford EON Project Homepage	Description of the EON project to support automated reasoning about protocol-based care	http://www-smi.stanford.edu/projects/eon/
U.S. National Library of Medicine, Health Services Technology Assessment Text	Access to guidelines by AHCPR, NIH, and the U.S. Preventive Services Task Force	http://text.nlm.nih.gov/

2.2.3 The role of cost-effectiveness analyses in CPG development

CPGs are often promoted as being a means for cost containment, yet many CPGs are currently developed on the basis of clinical effectiveness alone. If costs are examined, they usually play only a secondary role (Gold et al. 1996). CPGs that are constructed by decision-analysis teams that consider both costs and efficacy can ideally aid in slowing the rise in health-care costs by identifying three kinds of treatment: ineffective treatment that can be withheld, the most cost-effective treatment among equally effective alternatives, and alternatives that are both more effective and more expensive (Shapiro et al. 1993). The inclusion of a dual-utility function in my CPG conceptual model enables cost-effectiveness analyses and thus allows ALCHEMIST to produce CPGs based on the cost effectiveness of given treatment strategies.

2.2.4 Efficacy of CPGs

Although policy makers have expressed great interest in using CPGs and have recognized CPGs' potential to increase quality of care while containing costs, they are still uncertain about CPG efficacy (Walker et al. 1994, Woolf et al. 1990). Although several studies have demonstrated changes and improvement in physician behavior after guidelines have been disseminated (Weingarten 1995, Wachtel 1990, Eagle 1990, Weingarten 1993, Weingarten 1994, Weinstein 1997), other studies document that clinicians are often unaware of existing guidelines or, if they are aware, they often fail to change their behavior based on the CPGs (Cohen et al. 1985, Grilli et al. 1991, Kosecoff et al. 1987, Lomas et al. 1989, Lomas and Haynes 1988, Maiman et al. 1991, Pierre et al. 1991, Romm et al. 1981, Weingarten et al. 1994). In 1993, Grimshaw and Russell studied 59 published evaluations of CPGs with rigorous evaluations (Grimshaw and Russell 1993); they found that implementation of CPGs significantly improved the process and outcome of care. The degree of improvement, however, varied considerably. They did not study the costs associated with the development and implementation of the CPG. CPGs developed at a national level and published in journals were found to have a low probability of affecting the process of care. Internal development strategies combined with a patient-specific reminder at the point of care produced the highest probability of efficacy.

In two related studies, Hayward and colleagues designed and performed a survey of practitioners to determine the importance of CPG features in helping these physicians to decide whether to adopt a guideline, and to assess the usefulness of dissemination strategies in making guidelines accessible and effective (Hayward et al. 1996, Hayward et al. 1997). Table 2.2 lists those authors' recommendations for increasing CPG effectiveness, and gives the corresponding features of ALCHEMIST.

Table 2.2. Adherence of ALCHEMIST approach to published suggestions for a successful CPG. (Hayward et al. 1996, Hayward et al. 1997, Tunis et al. 1994).

Hayward internist survey	ALCHEMIST approach
Endorsement by respected colleagues and major organizations	ALCHEMIST CPG will be peer reviewed before distribution
Short pamphlets and manuals summarizing a number of guidelines	Main webpage provides overview of entire CPG and main recommendations
Guideline presentation requires concise recommendations, synopsis of supporting evidence, and quantification of benefit	Recommendations are provided in a compact flowchart algorithm form; evidence table and result tables provide concise summaries
Comparison between this guideline and that of other competing organizations	Currently not addressed
Strength of recommendations	Levels of evidence are provided for supporting evidence; sensitivity analyses describe uncertainty in produced recommendations
How the guideline can be applied to individual patients	Sensitivity analyses and identification of sensitive variables helps the user in determining applicability to a given patient; ALCHEMIST's CPG updating and tailoring ability allows the applicability to be determined explicitly.

2.3 My subset of CPGs

The IOM categorizes guidelines along five dimensions: clinical orientation, clinical purpose, complexity, format, and intended users (Institute of Medicine 1992). I use these five areas to define the subset of CPGs that I address in my thesis work, as well as to detail the restrictions that this subset places on the applicability of ALCHEMIST. I also use the underlying requirement that the CPG must be able to be based on a DM.

2.3.1 Clinical orientation

Clinical orientation reflects whether the chief focus of the guideline is a clinical condition, (e.g., lung cancer) or a technological device or process (e.g. implantable cardioverter defibrillator). My CPG subset is not restricted to a particular clinical orientation; the CPG problem needs only to be represented by a DM (Section 3.3.2). CPGs such as clinical pathways that specify what to do at each point in a patient's hospitalization cannot be feasibly represented by a valid DM.

2.3.2 Clinical purpose

The **clinical purpose** of a CPG reflects whether the guideline advises the user about screening and prevention, evaluation or diagnosis, aspects of treatment, or other aspects of health care. The example DMs that I have chosen for my research describe patient-treatment decisions. The CPG conceptual model is not limited to representing CPGs whose clinical purpose is treatment; it can also represent CPGs that have a wide variety of clinical purposes. Screening or prevention guidelines that require long sequences of events (e.g., the proper workup of an abnormal PAP smear) may go beyond the complexity limits of DMs and, therefore, may not be applicable to my subset.

2.3.3 Complexity

The IOM defines **complexity** as a measure of whether the guidelines are straightforward in presentation and discussion or are marked by considerable detail, complicated logic, or lengthy narrative and documentation (Institute of Medicine 1992). I restrict the textual complexity of the CPG by organizing the evidence and knowledge of the CPG in a structured-abstracted format.

2.3.4 Format

There are numerous methods for **formatting** and representing the knowledge within a CPG, including free text, tables, IF-THEN statements, critical pathways, decision tables, protocol lists and charts, and flowchart algorithms (Gottlieb et al. 1992). Figure 2.1 through 2.3 show three different possible representations of subsets of the same CPG. Figure 2.1 shows the flowchart algorithm for managing a patient with acute low-back pain; Figure 2.2 uses a table to list possible red flags for serious conditions that can cause low-back pain, and supplemental textual guideline information is given in Figure 2.3.

I combine a structured textual and clinical flowchart-algorithm representation for depicting CPGs produced by ALCHEMIST. The flowchart representation enhances the accompanying prose by depicting the logic and conditional statements behind the recommended clinical decisions, and by using the branching structure of the flowchart to capture relationships among the elements of the decision that are difficult to characterize with textual output (Abendroth et al. 1988). It also provides a quick visual summary of the guideline for users, and depicts clearly any changes made to the guideline when the evidence is updated or tailored.

The belief that clinical algorithms are a beneficial format for representing guideline recommendations is not unanimous among guideline developers. Possible oversimplifications in a clinical algorithm format include reducing a complex medical DM to a binary yes-no choice, omitting important options at decision nodes, presenting decisions in a linear fashion that are properly approached in conjugate because of their complex interrelationships, and failing to recognize feedback loops that require the repetition of tests and treatments (Woolf 1998). These possible problems with clinical algorithms are all important; however, several of them are eliminated if we restrict our CPG subset to those that can be based on DMs (as described in Section 3.3.2). By expanding the available choice options, a decision analyst can reduce a DM to all binary choices, the possibility of omitting important options at decision nodes should be addressed in the decision modeling when the important outcomes are identified for the given intervention. Similarly, the problems of linear representations and feedback loops both will be addressed by the decision-analysis team when it designs the underlying DM. Finally, guideline users may worry that the use of clinical algorithms may eliminate an individualized approach (Woolf 1998). ALCHEMIST attempts to alleviate this concern about clinical algorithms by allowing the guideline user to make changes to the underlying DM and to custom-tailor the generated CPG to represent more accurately than the global CPG her patient population.

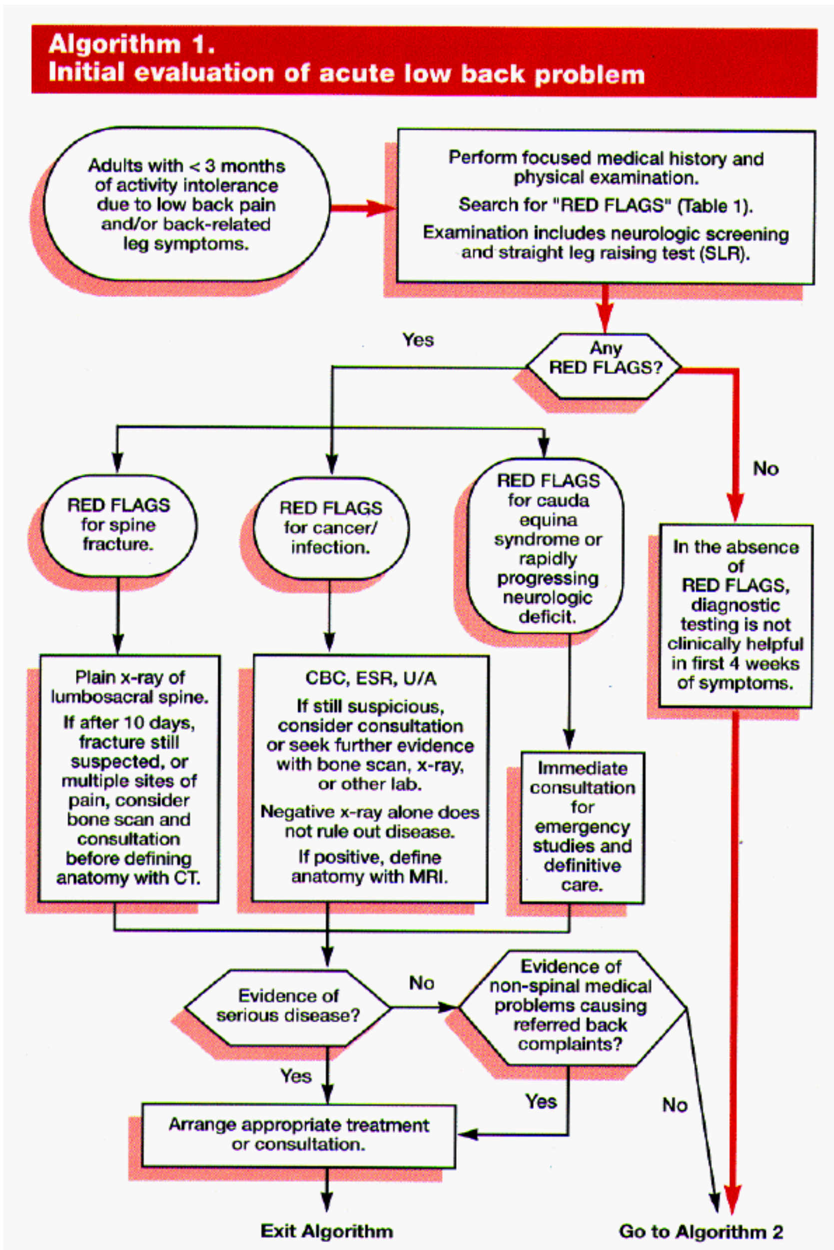


Figure 2.1. Clinical-flowchart algorithm representation. This flowchart algorithm demonstrates AHCPR's recommended initial evaluation of acute low-back problem (*Source:* Acute Low Back Problems in Adults: Assessment and Treatment. Quick Reference Guide for Clinicians, Number 14. 27 pp. (AHCPR 95-0643)).

Netscape - [Page Contents]

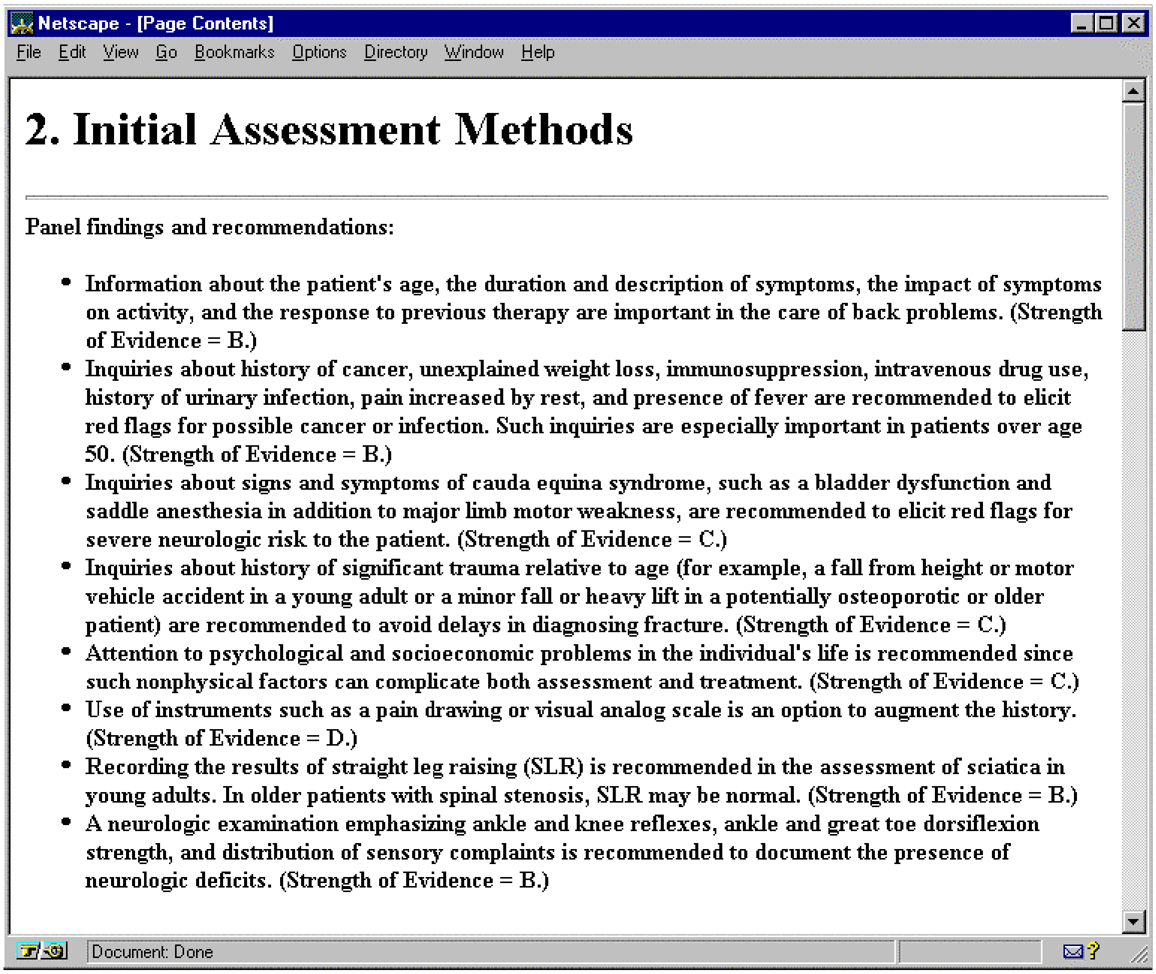
File Edit View Go Bookmarks Options Directory Window Help

Table 1. Red flags for potentially serious conditions

Possible fracture	Possible tumor or infection	Possible cauda equina syndrome
From medical history		
Major trauma, such as vehicle accident or fall from height. Minor trauma or even strenuous lifting (in older or potentially osteoporotic patient).	Age over 50 or under 20. History of cancer. Constitutional symptoms, such as recent fever or chills or unexplained weight loss. Risk factors for spinal infection: recent bacterial infection (e.g., urinary tract infection); IV drug abuse; or immune suppression (from steroids, transplant, or HIV). Pain that worsens when supine; severe nighttime pain.	Saddle anesthesia. Recent onset of bladder dysfunction, such as urinary retention, increased frequency, or overflow incontinence. Severe or progressive neurologic deficit in the lower extremity.
From physical examination		
		Unexpected laxity of the anal sphincter. Perianal/perineal sensory loss. Major motor weakness: quadriceps (knee extension weakness); ankle plantar flexors, evertors, and dorsiflexors (foot drop).

Document: Done

Figure 2.2. Protocol chart representation. Corresponding chart to recommendations described in the algorithm in Figure 2.1 for the AHCPR low-back pain guideline (*Source:* Acute Low Back Problems in Adults: Assessment and Treatment. Quick Reference Guide for Clinicians, Number 14. 27 pp. (AHCPR 95-0643)).



2. Initial Assessment Methods

Panel findings and recommendations:

- **Information about the patient's age, the duration and description of symptoms, the impact of symptoms on activity, and the response to previous therapy are important in the care of back problems. (Strength of Evidence = B.)**
- **Inquiries about history of cancer, unexplained weight loss, immunosuppression, intravenous drug use, history of urinary infection, pain increased by rest, and presence of fever are recommended to elicit red flags for possible cancer or infection. Such inquiries are especially important in patients over age 50. (Strength of Evidence = B.)**
- **Inquiries about signs and symptoms of cauda equina syndrome, such as a bladder dysfunction and saddle anesthesia in addition to major limb motor weakness, are recommended to elicit red flags for severe neurologic risk to the patient. (Strength of Evidence = C.)**
- **Inquiries about history of significant trauma relative to age (for example, a fall from height or motor vehicle accident in a young adult or a minor fall or heavy lift in a potentially osteoporotic or older patient) are recommended to avoid delays in diagnosing fracture. (Strength of Evidence = C.)**
- **Attention to psychological and socioeconomic problems in the individual's life is recommended since such nonphysical factors can complicate both assessment and treatment. (Strength of Evidence = C.)**
- **Use of instruments such as a pain drawing or visual analog scale is an option to augment the history. (Strength of Evidence = D.)**
- **Recording the results of straight leg raising (SLR) is recommended in the assessment of sciatica in young adults. In older patients with spinal stenosis, SLR may be normal. (Strength of Evidence = B.)**
- **A neurologic examination emphasizing ankle and knee reflexes, ankle and great toe dorsiflexion strength, and distribution of sensory complaints is recommended to document the presence of neurologic deficits. (Strength of Evidence = B.)**

Figure 2.3. Prose representation. Textual representation of the same clinical information of the initial assessment recommendations for patients who present with acute low-back pain (*Source:* Acute Low Back Problems in Adults: Assessment and Treatment. Quick Reference Guide for Clinicians, Number 14. 27 pp. (AHCPR 95-0643)).

2.3.4.1 SMDM clinical-algorithm standard

My CPG's flowchart representation is based on the standard described by the Society for Medical Decision Making (Society for Medical Decision Making (SMDM) Committee on Standardization of Clinical Algorithms (CSCA) 1992). This standard has the following components: boxes, arrows, numbering scheme, and links.

- **Boxes.** The standard uses several types of boxes to represent the different states of a CPG.

1. **Clinical-state box** (rounded rectangle). This box defines the clinical state or problem. It has only one exit path and may or may not have an entry path. This box always appears at the beginning of an algorithm
2. **Decision box** (hexagon). This box represents a branching decision whose response will lead to one of two alternative paths. It always has one entry and two exit paths
3. **Action box** (rectangle). This box represents an action that usually is either therapeutic or diagnostic.
4. **Link box** (oval). This box is used in place of an arrow to clarify page breaks or to connect separated nodes to maintain path continuity.

- **Arrows.** Arrows flow top to bottom of the CPG, and usually flow from left to right. Arrows should never intersect one another. Arrows originating from decision boxes should be labeled "yes" or "no," with the "yes" arrow pointing to the right and the "no" arrow pointing downward, whenever possible.
- **Numbering scheme.** Boxes are numbered sequentially from left to right and top to bottom.
- **Title.** The title of the algorithm should clearly define the clinical topic and the intended users. Authors of the CPG should be listed under the title with their degrees and institutional affiliations.

Figure 2.4 shows an example CPG that conforms to the SMDM standard.

The main limitation of representing CPGs using the SMDM standard is the requirement that all choice boxes have only "yes" and "no" exit points. This restriction implies that all chance nodes within the corresponding DM should be binary (i.e., have the format of "Result-Yes" and "Result-No"). This restriction does not cause problems for a decision tree such as the LC-EM, where the chance nodes have children such as "CTpos" and

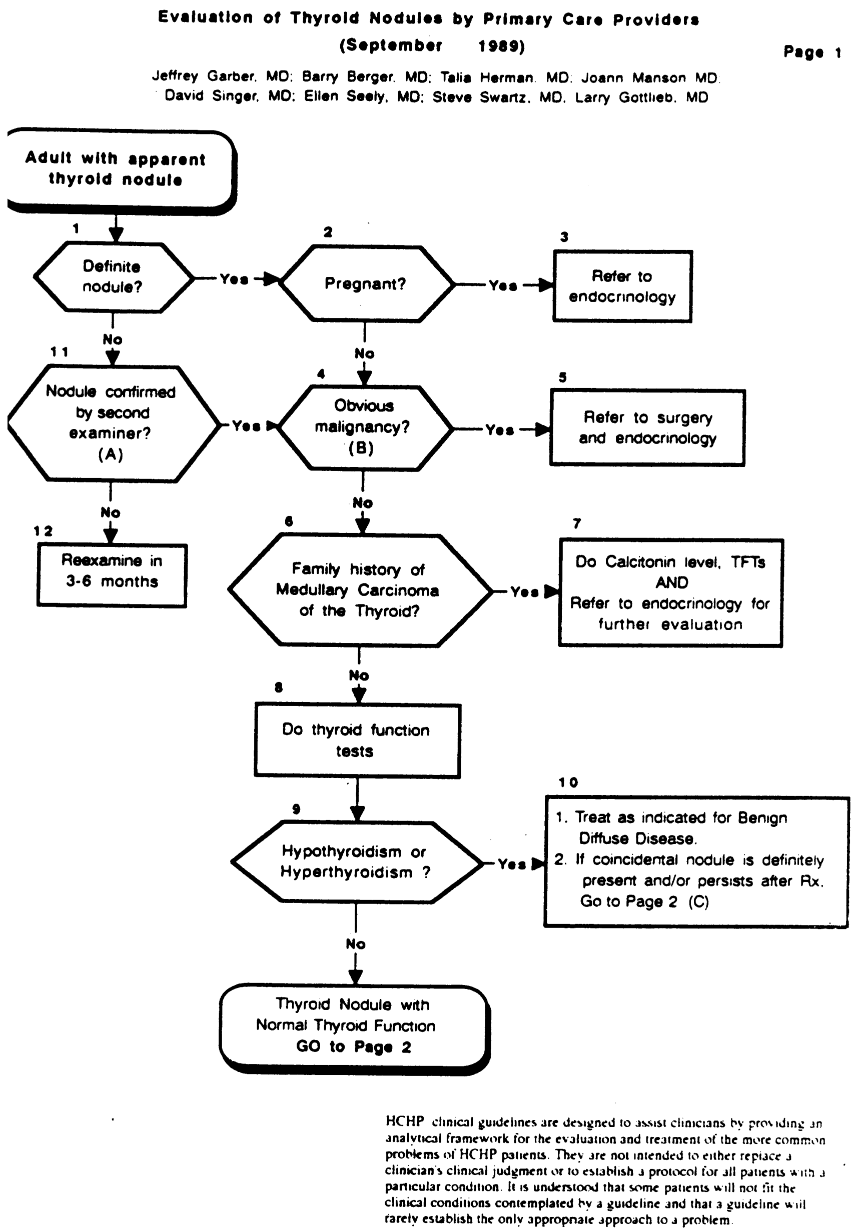


Figure 2.4. Example CPG that conforms to the SMDM standard. (Reprinted with permission. *Source*: Society for Medical Decision Making Committee on Standardization of Clinical Algorithms. Proposal for Clinical Algorithm Standards.Medical Decision Making, 1992, 12(2): 149-154).

"CTneg." However, decision analysts may need to restructure more complicated trees in order to comply with the binary-node restriction.

2.3.4.2 Guideline abstracts

I combine the SMDM flowchart representation with a structured abstract of the CPG. This abstract is based on a proposal by Hayward and colleagues, for writing informative abstracts describing CPGs (Hayward 1995, Hayward and Laupacis 1993, Hayward et al. 1993). Hayward recommends use of the following classes of knowledge for organizing a CPG: objective, options, outcomes, evidence, values, benefits, harms, and costs, recommendations, validation, and sponsors. I detail these classes in Section 2.6, when I describe my CPG conceptual model.

2.3.5 Intended users

CPGs are directed typically at physicians, nurse practitioners, or physician assistants, though their use has broadened to include payer, patients, and other health-care workers. The guideline's **intended users** is an important determinant of the CPG’s scope. The intended users of the CPGs produced by ALCHEMIST, as described in Chapter 1, are local guideline developers and implementers. These guideline implementers could browse the generated CPG and its primary evidence, and then tailor the CPG to reflect their patient population. In the future, the intended users of my CPG subset could be expanded to include individual physicians; however, to ensure the validity of the CPG for a particular patient, ALCHEMIST would need to allow proper assessment of the patient's utilities, to allow extraction and integration of specific patient probabilities, and to represent existence of comorbid conditions.

2.3.6 Dimensions of practice policies

Eddy has described three dimensions of practice policies (Eddy 1992). In this section, I describe these three dimensions and discuss where my CPG subset falls within this space of practice policies. These dimensions are the **intended use**, the **intended type of guidance**, and the **intended degree of flexibility**. The cube defined by these three dimensions is shown in Figure 2.5. Those categories that are included in my subset of CPGs are shaded in gray.

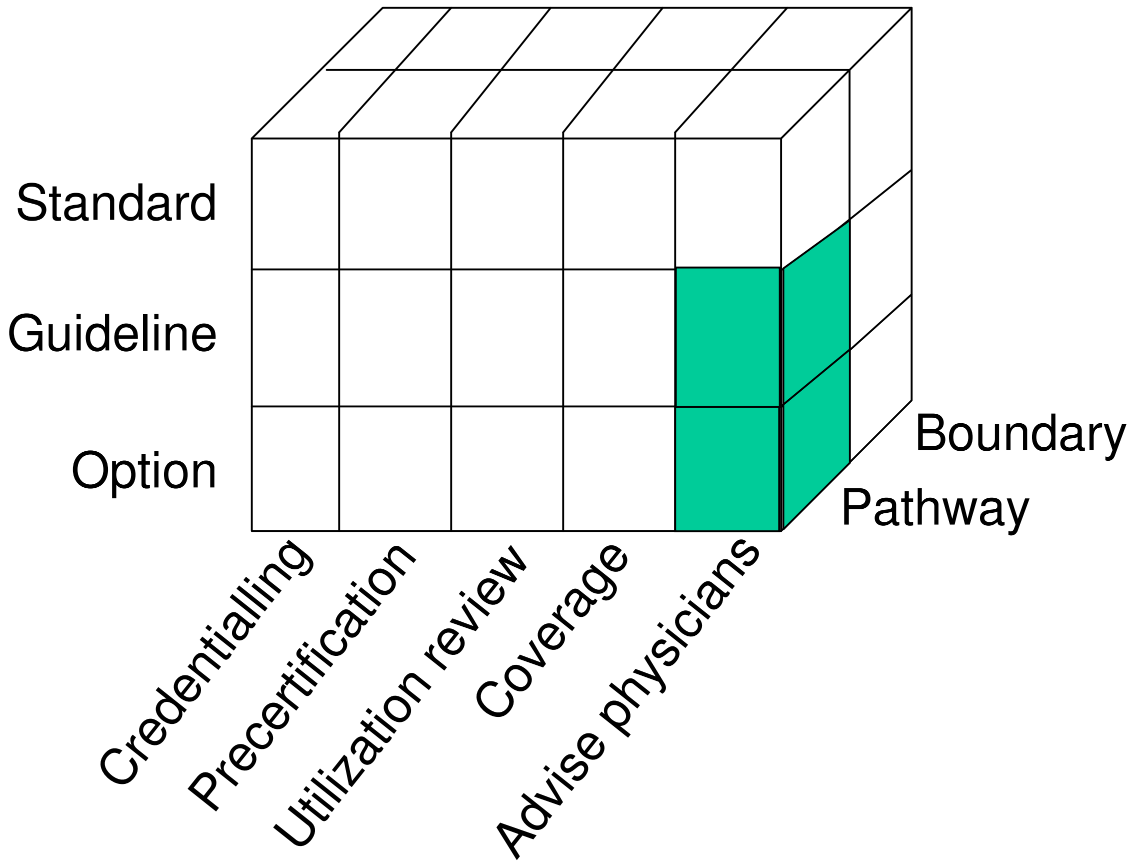


Figure 2.5. ALCHEMIST dimensions of practice policies. ALCHEMIST is currently able to represent option and guideline pathways to provide advise to physicians. The non-shaded boxes indicate those types of CPGs that ALCHEMIST can be extended to cover in future work (Reprinted with permission. *Source:* Eddy DM. *A manual for assessment of health practices and designing practice policies: the explicit approach*. The American College of Physicians, Philadelphia. 1992; 7).

Within the domain of CPGs, there are five main categories of intended use. They are (1) provision of advice to practitioners, (2) determination of the appropriate use before an intervention is performed (precertification), (3) determination of the appropriate use after an intervention has been used (utilization review), (4) determination of whether the cost of an intervention should be paid (coverage), and (5) determination of who should be allowed to perform an intervention (credentialling). My CPG subset concentrates on giving advice to practitioners about the correct patient-management plan. The second dimension involves whether a policy is intended to specify a preferred practice or management path (pathway) or to describe the limits of acceptable practices (boundary). My CPG subset is restricted to pathway guidelines. The third dimension refers to the intended flexibility. Such flexibility is determined by the degree of certainty among experts about the outcomes of the intervention and the degree of agreement among patients about the desirability of the outcomes. Depending on the intended flexibility, the CPG is classified as a standard, a guideline, or an option. The flexibility that the guideline user has to make changes to the underlying DM classifies my CPGs as guidelines or options.

2.4 CPG conceptual model

A **conceptual model** is a description of a part of the world, the concepts about that part of the world, and the relationships among these concepts. To specify the CPG conceptual model, I describe the required knowledge for creating and using a CPG.

CPG tasks: What tasks should CPG users be able to perform? What questions should they be able to answer? A series of published articles describes how users should use published CPGs (Hayward et al. 1995, Wilson et al. 1995). Design of my CPG conceptual model reflects the knowledge that guideline users must have to answer these identified questions and to perform these tasks. Hawyard and colleagues explore the required CPG tasks in a series of articles (Hayward et al. 1995, Wilson et al. 1995). Based on these articles, I determined that a user of a CPG developed with my conceptual model should be able to answer each of these questions for the generated CPG.

1. **Are the recommendations valid?**
   - Were all important options and outcomes clearly specified? Was an explicit and sensible process used to identify, select, and combine evidence?

- Was an explicit and sensible process used to consider the relative value of different outcomes? Is the guideline likely to account for important recent developments? Has the guideline been subject to peer review and testing?

2. **What are the recommendations?**
   - Are practical, clinically important, recommendations made?
   - How strong are the recommendations? What is the quality of the investigators that provide the evidence for the recommendations? What is the magnitude and consistency of positive outcomes relative to negative outcomes? What is the relative value placed on different outcomes?
   - What is the effect of uncertainty associated with the evidence and values used in the guidelines? The weaker the evidence linking intervention to outcome, and the greater the possible range of competing values, the greater the need for a sensitivity analysis.
3. **Will the recommendations help you in caring for your patients?**
   - Is the primary objective of the guideline consistent with your objective?
   - Are the recommendations applicable to your patients?

To provide answer to these questions, a CPG must be based on solid evidence; have an explicit method of synthesizing and analyzing this evidence; and allow the user to review, understand, and apply the recommendations to her patient population (Wilson et al. 1995).

2.5 Prior work in CPG modeling

Although several groups have done research on CPG representations, these previous studies differ from my representation in two ways: (1) they do not model CPGs based on DMs; and (2) they produce representations of the algorithmic flow of the guideline, but often do not represent explicitly additional knowledge, such as the CPG objective or patient population. This knowledge is often integral to the understanding and effective implementation of a CPG. The CPG conceptual model developed as part of my thesis work combines methods of representing CPGs from these previous research projects, while adding additional functionality through the use of DMs and annotation.

Many of the current approaches to modeling CPG knowledge are based on the desire of developers to automate the execution of these CPGs. Current work includes that by Stoufflet and colleagues on the Guided Entry of Data Elements for Clinical Management (GEODE-CM), by Cimino and colleagues on the Arden syntax using Medical Logic Modules, by Barnes and colleagues on the Modeling Better Treatment Advice (MBTA) practice guideline system, and by Musen and colleagues on the automation of protocol-based care using the EON architecture. Each of these research projects concentrates on a different subset of CPGs, and provides a different method for representing the logic and steps in CPGs and the execution of this logic. (Barnes 1995, Cimino 1995, Stoufflet 1995, Stoufflet 1996, Musen 1996).

Researchers from these projects have formed the **InterMed collaboratory**. InterMed, an interdisciplinary project that promotes collaborative medical informatics research, involves six institutions: Stanford University, Brigham and Women's Hospital, Massachusetts General Hospital, Columbia University, McGill University, and the University of Utah. As part of the collaboration, the researchers have produced a common language for describing the sequential nature of CPGs. The **GuideLine Interchange Format (GLIF)** language supports the description of the relationships among the different steps in the guideline; researchers plan to extend the GLIF language to represent a description of the guideline knowledge (Ohno-Machado 1998). As an example, the GLIF language represents the Guideline class as shown in Figure 2.6 (Deibel 1996).

Although ALCHEMIST's current CPG does not have an explicit intention element, it does have an element describing the clinical objective and target population. The eligibility

```
interface Guideline {
      attribute string name;
      attribute sequence <string> authors;
      attribute string intention;
      attribute Criterion eligibility_criteria;
      attribute sequence <Guideline_Step> steps;
      attribute Guideline_Step first_step;
      attribute sequence <Supplemental_Material> didactics;
};
```

Figure 2.6. GLIF representation of the Guideline Class. This figure shows how each Guideline class has a name, list of authors, CPG intention, eligibility criteria, a list of the steps in the CPG, the starting step of the CPG, and a list of supporting didactic material.

criteria of the CPG are represented in my conceptual model as the patient population and this population's defining characteristics. The steps of the CPG are represented as the clinical-flowchart algorithm; the evidence, definitions, and sources elements of my CPG conceptual model provide the needed supporting material. One possible extension of my thesis is work that would ensure that the GLIF language can represent the knowledge that ALCHEMIST requires to produce CPGs based on DMs. If GLIF had the necessary expressivity, then ALCHEMIST could produce output using GLIF, and could allow incorporation of my generated guidelines into the EON protocol-based decision-support system.

2.6 My CPG conceptual model

I designed the CPG conceptual model by studying current representations of CPGs and literature describing the needed components, development methods, and shortcomings of current CPGs. I modeled the knowledge needed to perform tasks related to the creation, dissemination, use, and maintenance of a CPG (Section 2.4). The resulting CPG conceptual model identifies knowledge required to produce a flowchart algorithm for the CPG recommendations, as well as the following essential CPG elements: objective, alternatives considered, outcomes measured, supporting evidence and sources, values used, summary of expected results, recommendations, key analysis information, clinical definitions, and guideline validation (Eddy 1990b, Hayward et al. 1993). Although the recommended global CPG flowchart algorithm can be produced without much of this information, successful local implementation and tailoring is hindered without this supporting knowledge. Figure 2.7 is a graphical depiction of the top layers of the CPG conceptual model and a more detailed description of the conceptual model and how it is used to perform the needed CPG tasks is included in Appendix A.

I describe each of these elements, and give examples from a generated CPG. Using the knowledge in the CPG conceptual model, the guideline user can answer the questions enumerated in Section 2.4. Of course, the CPG is based on a DM, so it will be only as valid as the underlying DM and corresponding DM conceptual model. In Chapter 3, I discuss the development of valid DMs and the knowledge explicit in their representation.

Each CPG has an **objective**. This objective includes knowledge regarding what the targeted health problem is, who the patient population is, who the intended users of the CPG are, the burden of the existing clinical problem, and why there is a need for a new recommendation. For example, in the LC-EM, the targeted health problem is the optimal mediastinal-staging strategy in patients who have non–small-cell lung cancer, and one characteristic of the targeted patient population is that the patients have known non–small-cell lung cancer. Because often there are numerous guideline-developing projects for a given clinical problem it is important for a guideline developer to establish precisely the clinical problem, population, and use of the generated CPG.

Each CPG has a list of practice **options**. Options include diagnostic and therapeutic alternatives used in the guideline (e.g., CT, mediastinoscopy, thoracotomy, and radiation therapy). A CPG should define these alternatives clearly and establish the evidence for including — or excluding — an alternative from the CPG recommendations. Each option may also have associated with it acute or chronic complications — or mortality.

Each CPG must also identify the guideline's **outcomes**, which are what health outcomes were identified (e.g., quality-adjusted life years, or life expectancy), what economic

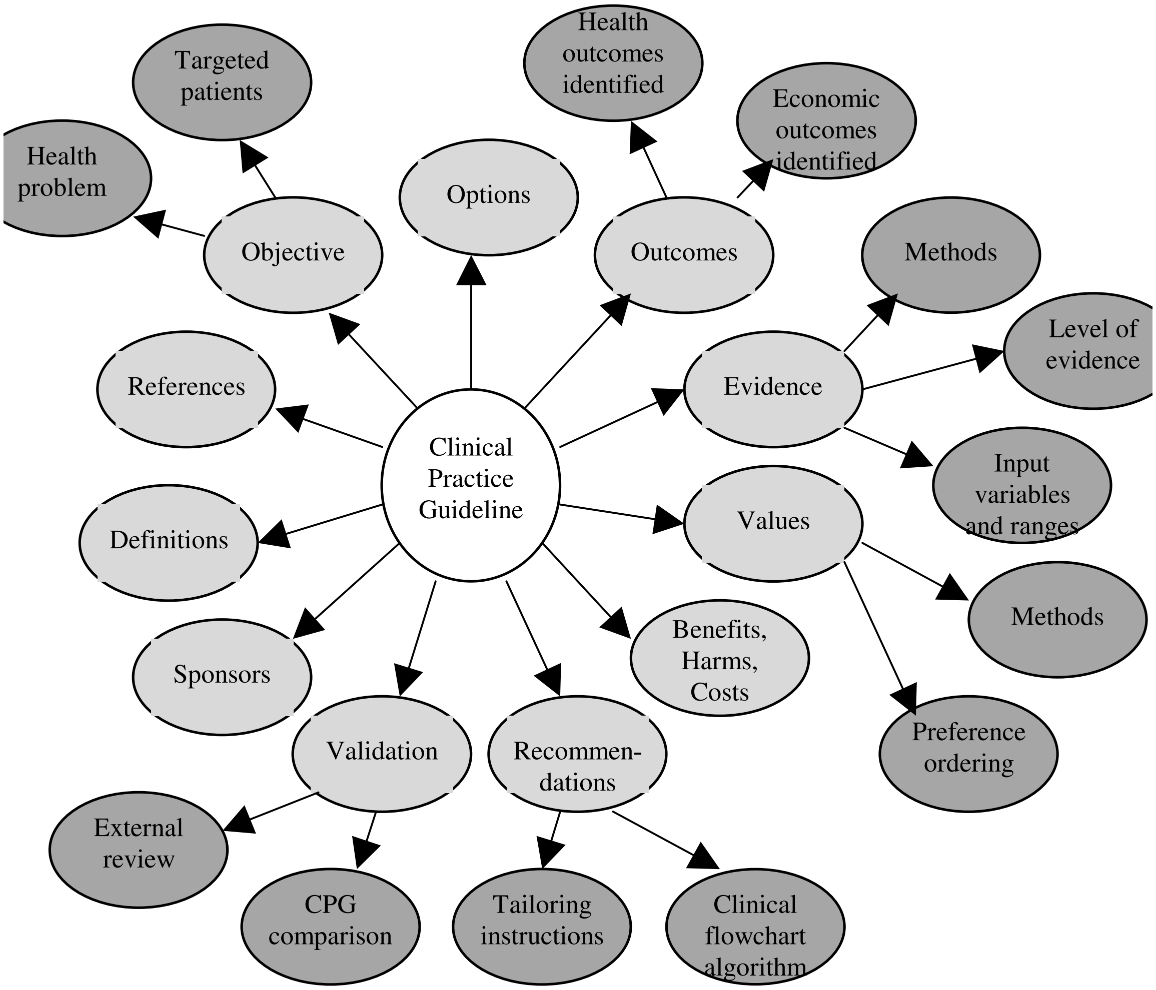


Figure 2.7. CPG conceptual model. This schematic of the CPG conceptual model shows the 11 main elements of ALCHEMIST's CPGs and several of the subclasses. For example, each CPG objective has a targeted health problem and patient population.

outcomes were identified (e.g., dollar costs), and whether comorbid conditions were modeled. Information about how these outcomes are evaluated is also important. For example, for a CPG that incorporates cost analyses, the CPG should indicate the currency and year of all costs used, as well as the method for inflating past costs to present dollars. Also the CPG should indicate if the costs reported reflect actual costs or hospital charges. Finally, the CPG should list the discount rate the analysis used to discount future costs. If the guideline developers identify an intermediate outcome (e.g., arrhythmic events), the CPG must make clear the definition of this outcome for the future guideline users (e.g., whether arrhythmic events include ventricular fibrillation and tachycardia).

An important part of every CPG should be the supporting **evidence**. The evidence of a CPG includes the methods used to gather data (e.g., literature search, expert opinion, clinical trial), to determine what evidence was used in the CPG (e.g., inclusion and exclusion criteria, levels of evidence), and to synthesize this evidence (e.g., decision analysis, meta-analysis, consensus panel). Note that developers of CPGs need to do more than just to cite references if they are to claim that the recommendations are evidence based. My CPG conceptual models, therefore, include descriptions of how the evidence was collected, evaluated, and translated into recommendations (Woolf 1998). For the variables used in the CPG, the generated CPG should include the best-estimate value, description of the variables, sensitivity-analysis range, level or quality of evidence, assumptions, and a corresponding evidence table. The evidence table describes the study design, number of patients involved, outcomes measured, patient population, interventions compared, biases, observed outcomes, reported effects of the intervention, as well as the methods the guideline developers used to synthesize the numerous evidence sources (Eddy 1992). Uncertainty normally is associated with a portion of the evidence in any CPG; therefore, the CPG should acknowledge the uncertainties in the scientific evidence, and should make explicit any weaknesses in corresponding recommendations.

Another element of a CPG is the **values**, which are the methods that the developers used to obtain any utilities used in the CPG that are used for outcomes (e.g., standard gamble, time-tradeoff analysis), a preference ordering of the outcomes (e.g., thoracotomy or radiation therapy with mediastinal metastases < radiation therapy without mediastinal metastases < thoracotomy without mediastinal metastases), and a declaration of the modeling perspective (e.g., societal, patient, third-party payer). The CPG should highlight any patient utilities that affect the recommendation and provide a simplified method for the guideline user to assess these utilities in their patient population.

The **benefits, harms, and costs** of a CPG are the type, magnitude, and level of uncertainty of these outcomes, and any related sensitivity-analysis results. I use a balance sheet similar to that described by Eddy (Eddy 1990a) to represent these outcomes. This balance sheet includes the type and magnitude of each outcome for the best-estimate strategy and any additional custom-tailored strategies. These results should also include information about the variables to which the CPG is sensitive, and how the expected utility is affected by the different variables over their given clinically valid ranges. Therefore, I include in my conceptual model a list of the sensitive variables and a tornado diagram[1] for the numerous variables.

A significant part of every CPG is the **recommendations**. In my CPG conceptual model, the recommendations include a clinical flowchart algorithm in the SMDM format (Figure 2.4), and instructions for tailoring the CPGs. The CPG should identify decisions in which patient preferences or patient-specific attributes are important — the important variables in the CPG are listed as being “sensitive variables”. If possible, the recommendations should also compare the clinical-flowchart recommended policies to those of earlier guideline projects, or to current practice.

Before a CPG is disseminated to the medical community, it should undergo internal and external peer review to ensure that the CPG is based on a valid synthesis of the available evidence. Therefore, another element of a CPG is declaration of CPG **validation**. This element includes an explicit statement of any results from external review (e.g., publication of the underlying DM in a peer-review journal) or internal validation. The CPG sponsors element includes definition of the key persons or groups that developed, funded, or endorsed the guideline. The date of publication of the CPG and the sources used in creating its recommendations are listed explicitly for the guideline user. Finally, any **definitions** used in the CPG (e.g., positive CT = greater than or equal to one lymph node with

---

1. A tornado diagram is a graphical representation of the change in the expected utility of a given strategy as each variable is varied along its sensitivity-analysis range.

short-axis diameter greater than 10mm on CT examination) and sources used for the evidence should be listed as part of the CPG.

ALCHEMIST produces CPGs using the elements described in this section, combined with the knowledge obtained from the DM. As part of my thesis work, I evaluated the quality of these evidence-based CPGs (See Chapter 8).

2.7 Summary

In this chapter, I reviewed the rationale for CPG development. I documented current development methods and detailed the limitations of those methods. I then described the subset of CPGs with which I worked. I explained how to formalize this CPG representation, and described a CPG conceptual model that identifies the knowledge required for creating and maintaining a CPG. In Chapter 3, I examine DMs and the knowledge explicit within their representation, delineating the required CPG information not provided by the DM, and establishing the need for annotation of the DM by the decision analyst.

Chapter 3

Decision Models

In this chapter, I discuss **decision models (DMs)** and explain the clinical problems that they can represent. I describe briefly several possible representations of DMs, including decision trees and influence diagrams (Section 3.2). In my work, I place several restrictions on the DMs that ALCHEMIST can use. I describe the subset of DMs that I address in relation to my DM conceptual model and to my current implementation (Section 3.4). I detail the limitations that this subset impose on the generalizability of my research. I then describe the conceptual model that I developed for representing DMs (Section 3.6). I complete this chapter by comparing my DM conceptual-model representation to other work in this field, and by listing the advantages and limitations of my approach.

3.1 Introduction to DMs

An abstract representation of a decision problem, a DM takes into account the uncertain, dynamic, and complex consequences of a decision, and assigns values to those consequences (Owens and Nease 1993, Owens and Sox 1990). A DM for clinical medicine is a simplification of the real clinical situation; therefore, the DM reflects the decision-analysis team's conception of how a treatment or screening intervention is used and the way in which that intervention affects the natural course of the disease, and the health status of the target patient population (Gold et al. 1996). In my research, I concentrate on DM representations that include not only the structural and mathematical relationships between variables, but also the underlying assumptions, evidence, and perspectives that decision analysts use in creating them.

3.2 DM representations

Once a decision-analysis team identifies a clinically appropriate decision problem, there are several ways for the analyst to represent the problem, the alternatives, and the expected outcomes. DM-representation methods differ in computational, graphical, and analytic complexity. Example representations of the underlying mathematical structure of a DM include a decision tree, influence diagram, spreadsheet model, or, for certain problems, a state-transition model.

Influence diagrams are compact graphical representations of the probabilistic relationships and influences among variables in a DM (Nease and Owens 1997, Owens et al. 1997b). **Decision trees** represent these same relationships structured according to the variables' observation ordering (Nease and Owens 1997, Owens and Nease 1993). Figure 3.1 displays decision-tree and influence-diagram graphical representations for a single HIV-screening decision problem. Both representations depict whether to screen a patient for HIV, what the possible test results are, what the infection status of the patient is, and whether there is an observed behavior change in the screened population. The simple decision tree consists of decision nodes, chance nodes, probabilities, utilities, and an ordering of events. The corresponding influence diagram also has decision nodes, chance nodes, utilities, and probabilities; in addition, it has arcs indicating influence between variables. Both representations have advantages and limitations. Conditional independence between variables is represented numerically in a decision tree and is depicted graphically through an absence of arcs in an influence diagram. Structural asymmetry is represented numerically in an influence diagram, and is shown clearly through structural differences between branches in a decision tree.

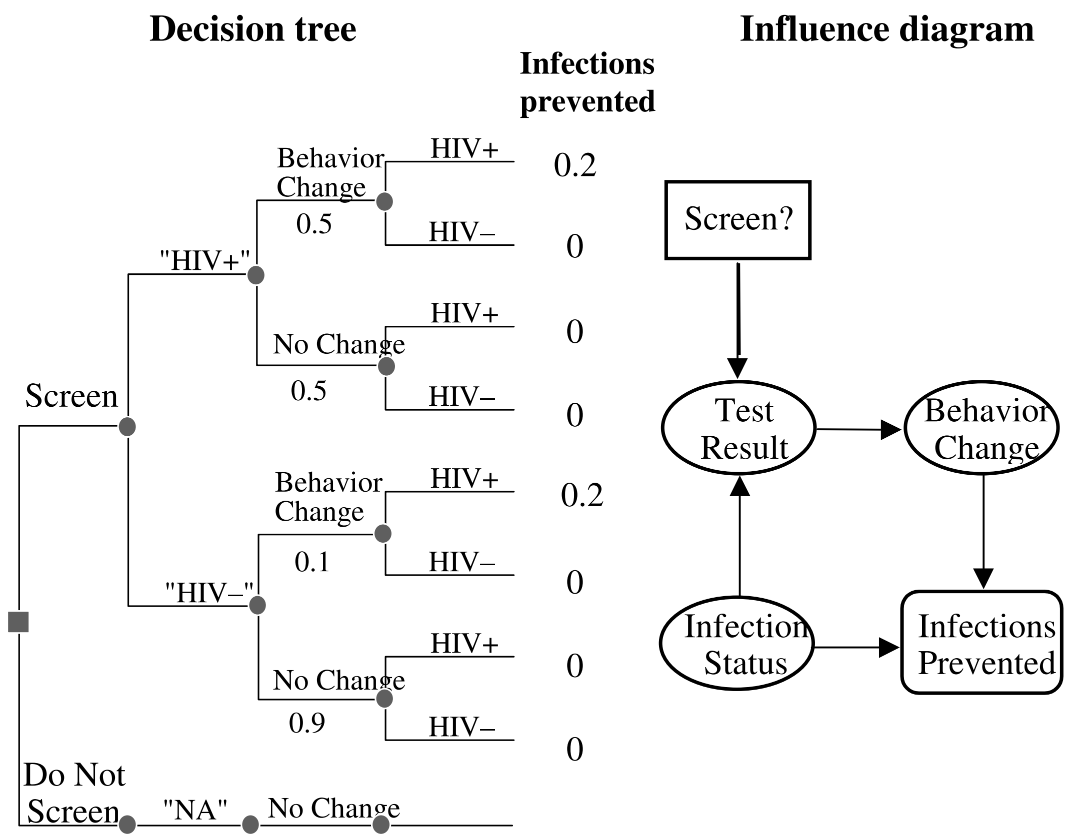


Figure 3.1. Schematic representation of a decision tree (left) and influence diagram (right) for the HIV-screening decision problem. Both DM representations depict whether to screen a patient for HIV, what the possible test results are, what the infection status of the patient is, and, whether there is an observed behavior change in the screened population. (Reprinted with permission. *Source*: DK Owens, RD Schacter, RF Nease. Representation and analysis of medical decision problems with influence diagrams. *Med Decis Making*. 1997 17(3): 241-62).

DMs are not complete representations of real clinical situations, but rather are much simplified and highly structured models of what the decision analyst considers to be the most important elements (Detsky et al. 1997a, Detsky et al. 1997b). These simplifying assumptions affect the DM's recommendations. ALCHEMIST makes these assumptions explicit, to allow the guideline user to establish whether they are reasonable when she is considering whether to use the generated clinical-practice guideline (CPG), and whether updating or tailoring it will produce valid recommendations.

Analysis of DMs is restricted by software availability, computational complexity, and the validity of existing data. These limitations also restrict the ease with which a corresponding CPG can be produced. In Section 3.4, I describe the DM subset and the corresponding limitations on the generalizability of ALCHEMIST's approach.

3.3 Brief background of DM development and use

In this section, I describe the process of creating DMs and the restrictions on the clinical problems that DMs can represent accurately.

3.3.1 Development methods for DMs

My thesis work assumes that the decision analyst has created a valid DM. A valid DM is based on strong clinical evidence and is structurally sound (i.e., it is syntactically coherent and does not contain impossible strategies, dominated strategies, or symmetry violations) (Wellman et al. 1989).

To create a DM, the decision analyst must perform the following sequence of tasks (Sox et al. 1988). Knowledge provided by completion of these tasks is reflected in the design of my DM conceptual model (Section 3.6).

1. Define the decision problem.
2. Identify the decision alternatives.
3. List the possible clinical outcomes of each decision alternative.
4. Represent the sequence of events leading to the clinical outcomes by a series of chance nodes and decision nodes.
5. Choose a time horizon and discount rate for the problem.
6. Determine the probability of each chance outcome.
7. Assign a value to each clinical outcome.

After these tasks are completed, the DM contains sufficient information for the decision analyst to calculate the expected value (the probability-weighted average value of the potential outcomes) of each decision alternative or strategy, and to perform sensitivity analyses. I assume that an experienced decision analyst has created the DM. I do not evaluate the quality of the initial DM; there is extensive literature on the proper development of DMs (Detsky et al. 1997a, Detsky et al. 1997b, Krahn et al. 1997, Naglie et al. 1997, Naimark et al. 1997).

3.3.2 Clinical problems that can be described by DMs

Although various existing guidelines have been based in part on DMs (American College of Physicians 1994, Carlson et al. 1994, Eddy 1991a, Eddy 1991b, Eddy 1991c, Eddy 1991d, Fahs et al. 1992, Grady et al. 1992, Littenberg et al. 1991, Melton et al. 1991, Schapira et al. 1993, Singer et al. 1991, Sox et al. 1991a, Sox et al. 1991b), creating a high-quality CPG does not always require a DM. Decision-analytic techniques are useful for unfamiliar problems when there is uncertainty about the appropriate clinical strategy for patients who are in a given health state (Detsky et al. 1997a). There are certain clinical problems for which there is little uncertainty and, therefore, for which a CPG is more appropriately produced without a decision analysis. There are also clinical problems that are so complicated that a DM cannot represent accurately all the required information. For example, the decision of how best to work up a routine PAP-smear examination in a given population requires multiple steps with numerous sequential decisions, including obtaining specimens at the appropriate intervals from the entire population at risk, obtaining specimens that are most suitable for reading, handling them so that they can get to the laboratory promptly in good condition, reading them accurately and expeditiously, notifying the clinician of the results appropriately, notifying the patient, and performing suitable follow-up care for women who have an abnormal result (Schoenbaum and Gottlieb 1990). A DM that represented all these steps would be extremely complex.

The subset of all CPGs that ALCHEMIST can create is limited only in that the clinical problems must be ones that DMs can represent. In general, if direct evidence is available for the effect of the intervention and the outcomes of interest — and if the recommendations are not sensitive to patient preferences — it is neither necessary nor appropriate to use DMs. However, if there is no direct evidence, guideline developers can use DMs to estimate the outcomes explicitly. DMs may be beneficial in clinical domains in which problems are dynamic, with importance factors changing over the period of interest, if there are many options and events, or if the period of interest is long. Example clinical domains that may require DMs include repeated screening, repeated or continuous monitoring, compound diagnostic problems with many options, and diagnostic or treatment problems with many possible events separating the intervention and a long-term outcome of interest. The most common methods of addressing these types of problems are decision trees and Markov chains. The DM conceptual model can represent both these types of DMs, and, therefore, ALCHEMIST can create automatically the corresponding CPGs.

When modeling most decision problems, it is neither efficient nor feasible for the decision analysts to gather evidence for, and to model, every possible health benefit, harm, cost, alternative, event, or outcome. The decision whether to include a particular DM element should take into account both the expense and difficulty of inclusion, and the potential importance of these elements in the analysis. For example, in the SCD-MM, the marginal cost effectiveness of implanting an ICD in patients at high risk for SCD changes from approximately $75,000 to about $50,000 per quality-adjusted life year gained, as the frequency of battery replacement varies between the base-case value of every 4 years and the manufacturer's predicted 8 years (Owens et al. 1997a). The generated CPG will be sensitive to this variable, so the latter should be included in the DM.

There are also limitations imposed by the implementation software. There are several software packages designed for creating and analyzing DMs. These packages vary in computational abilities, analysis techniques, and user interface. Influence-diagram packages include Analytica, Ergo, DATA, IDEAL, Microsoft Bayes Network (MSBN), and Netica. Decision-tree packages include Decision Maker (Sonnenberg and Pauker 1987, Sonnenberg and Pauker 1997), SMLtree (Hollenberg 1984), and DATA by TreeAge. My current implementation of ALCHEMIST uses DMs that experienced decision analysts have modeled using the Decision Maker software for Windows. I chose to use the Decision Maker software because it is able to provide a range of analytical functions, including cost-effectiveness analyses and Markov processes. In addition, Sonnenberg and colleagues maintain the software actively, and continue to improve its functionality. Decision Maker's Object Linking and Embedding (OLE) interface allows remote calls from other applications and therefore permits ALCHEMIST to control Decision Maker from its web interface. SMLtree files also can be converted to Decision Maker format and analyzed by ALCHEMIST.

3.4 My subset of DMs

In this section, I describe the conceptual and implementation subset of DMs with which I worked, and the future methods that I will use in extensions of my thesis to expand this subset to include a greater variety of DMs.

3.4.1 Decision-tree representation

In the medical decision-making community, decision trees are a common method of representing simple decision analyses, Markov processes, and cost-effectiveness studies. Therefore, I concentrated on the decision-tree representation of DMs. The restrictions that I have placed on the decision-tree representation were mainly for implementation purposes and do not set limitations on the DM conceptual model.

Decision Maker uses ASCII text files with the following syntax:

[NODES]
BLANK LINE
[GLOBAL VARIABLES]
BLANK LINE
[TABLES]
BLANK LINE

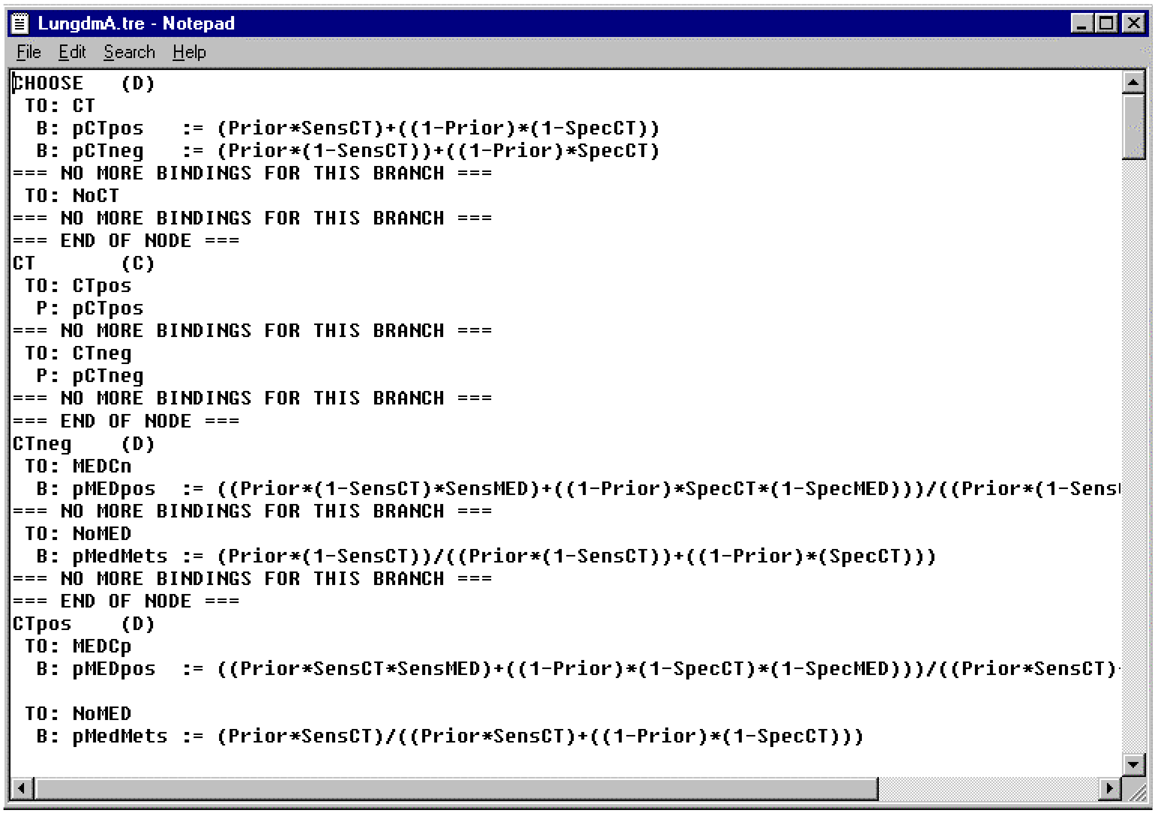
where each element enclosed in brackets has an additional specific format. For example, the format of a node [NODES] is

NODENAME
[BRANCHES]

where NODENAME is

{NodeName} (nodetype)

Figure 3.2 shows an excerpt of the LC-EM ASCII text file. Currently, I do not place any restrictions on variable naming; if standard conventions were used by the decision analyst, these structured variable names would enable ALCHEMIST to do additional automated parsing of the decision tree (e.g., if "InitAge" were always used for the starting age of the population in time-dependent models, then ALCHEMIST could infer automatically that this age information should be part of the patient-population definition for the generated CPG).




```
CHOOSE    (D)
 TO: CT
  B: pCTpos    := (Prior*SensCT)+((1-Prior)*(1-SpecCT))
  B: pCTneg    := (Prior*(1-SensCT))+((1-Prior)*SpecCT)
=== NO MORE BINDINGS FOR THIS BRANCH ===
 TO: NoCT
=== NO MORE BINDINGS FOR THIS BRANCH ===
=== END OF NODE ===
CT        (C)
 TO: CTpos
  P: pCTpos
=== NO MORE BINDINGS FOR THIS BRANCH ===
 TO: CTneg
  P: pCTneg
=== NO MORE BINDINGS FOR THIS BRANCH ===
=== END OF NODE ===
CTneg     (D)
 TO: MEDCn
  B: pMEDpos   := ((Prior*(1-SensCT)*SensMED)+((1-Prior)*SpecCT*(1-SpecMED)))/((Prior*(1-Sens
=== NO MORE BINDINGS FOR THIS BRANCH ===
 TO: NoMED
  B: pMedMets := (Prior*(1-SensCT))/((Prior*(1-SensCT))+((1-Prior)*(SpecCT)))
=== NO MORE BINDINGS FOR THIS BRANCH ===
=== END OF NODE ===
CTpos     (D)
 TO: MEDCp
  B: pMEDpos   := ((Prior*SensCT*SensMED)+((1-Prior)*(1-SpecCT)*(1-SpecMED)))/((Prior*SensCT)

 TO: NoMED
  B: pMedMets := (Prior*SensCT)/((Prior*SensCT)+((1-Prior)*(1-SpecCT)))
```

Figure 3.2. Decision Maker ASCII file format for LC-EM. This file shows the ASCII format of Decision Maker that ALCHEMIST uses to get information about the DM. For example, it lists the node CT and a chance (C) node with two branches (TO: CTpos and TO: CTneg) and corresponding probabilities (P: pCTpos and P: pCTneg). CT = computed tomography; CTpos = positive CT; CTneg = negative CT; pCTpos = probability of a positive CT; pCTneg = probability of a negative CT.

Table 3.1 summarizes the implementation-specific restrictions that I imposed on my DM subset.

Table 3.1. ALCHEMIST's implementation DM subset. This table shows the restrictions on the DM subset imposed by Decision Maker and by my current implementation of ALCHEMIST.

Decision-tree concept	Limit/ restriction
Decision Maker implementation	
Maximum number of nodes	140

Table 3.1. ALCHEMIST's implementation DM subset. This table shows the restrictions on the DM subset imposed by Decision Maker and by my current implementation of ALCHEMIST.

Decision-tree concept	Limit/ restriction
Maximum number of variables	255
Maximum number of bindings per branch	255
Maximum number of tables	255
Maximum length of expressions	255 characters
Maximum number of Markov states	64
ALCHEMIST implementation	
Decision-tree format	Decision Maker compatible (.tre) file
Maximum number of children per chance node	2
Incorporation of tunnel states[a]	Not allowed

a. Incorporation of tunnel states is a method of allowing certain decision-modeling software packages use to allow memory in Markov processes. This modeling solution creates a set of *tunnel* states that are strung together using the residual probability. Tunnel states are not allowed currently in my DM subset though they could be a future extension of my thesis work.

3.4.2 DM attribute subset

My DM conceptual model is able to express a simplified subset of all DMs. This section describes three DM attributes that my DM conceptual model represents. In Chapter 7, I demonstrate ALCHEMIST's ability to represent these attributes using my three example DMs.

1. **Sequential decisions.** Many CPGs produce recommendations for clinical problems where there are several sequential decisions involved. The LC-EM and the LC-CEM both have sequential decisions.

2. **Dual utilities (cost-effectiveness studies).** As I described in Chapter 2, most CPGs are based on effectiveness. For CPGs to promote cost-effective health care, the DM must model both the costs and benefits of treatments. A cost-effectiveness analysis uses conflicting cost and effectiveness attributes to value outcomes (Sonnenberg and Pauker 1997). To evaluate a cost-effectiveness analysis, the

decision analyst must maintain two independent utility scales (the first using a measure of cost such as dollars, and the second using a measure of effectiveness, such as life expectancy). Therefore, my DM conceptual model incorporates dual utilities. If I am to create a CPG, a tradeoff between these two utilities is required. In ALCHEMIST, the decision analyst establishes a marginal cost-effectiveness threshold for determining the allowed cost-effectiveness ratio among alternate strategies. This default threshold value is currently set at $50,000 per quality adjusted life year saved[1]; it can be changed by the decision analyst or by the guideline developer to reflect the perspective and outcomes of the implementing institution, changing conventions, changing financial situations, or other new data. The LC-CEM and SCD-MM models incorporate both costs and utilities and, therefore, demonstrate the dual-utility abilities of my DM conceptual model.

3. **Markov model (time-dependent studies).** Certain clinical conditions require modeling of repetitive events or modeling of patients at continuous risk. Modeling a chronic disease (e.g., breast cancer) differs from modeling an acute disease (e.g., appendicitis) in that the risks and benefits for chronic diseases occur over an extended time period and the disease behaves as a process, rather than as an instantaneous event (Manton and Stallard 1988). A valid DM for chronic diseases requires a model that takes into account the time dependence of the disease-hazard rates, and the patient's age-specific mortality rates, discounting of both future costs and benefits, as well as provides a structure that reflects continuous risk. Markov models are one method with which we can model such time dependence. A Markov model (in the medical domain) is a type of state-transition model in which the transition probabilities depend on only the current

1. Because cost effectiveness is relative, it is difficult to state whether a given treatment is “cost effective”. Therefore, specific interventions that society has chosen to implement of not implement are often used as benchmarks (e.g., cost-effectiveness ratios for dialysis for end-stage renal disease). ALCHEMIST’s default cost-effectiveness threshold is $50,000/QALY, however, the decision analyst or guideline developer may change this threshold.

patient state (Beck and Pauker 1983, Sonnenberg and Beck 1993). Use of Markov models has become standard, and the common decision-analysis software packages (SMLtree, DATA, and Decision Maker) can evaluate such models (Gold et al. 1996). My DM conceptual model includes Markov nodes (branching points within the decision tree that lead into a Markov process) (Gold et al. 1996). The SCD-MM is a 13-state Markov model; I use it to demonstrate the ability of ALCHEMIST to transform time-dependent DMs into CPGs.

3.4.3 Expansion of the DM subset

Although my decision-model subset currently is restricted to the decision-tree representation, this restriction is based on implementation requirements. The DM conceptual model is designed to represent the knowledge in other DM representations, such as influence diagrams. However, allowing ALCHEMIST to create a CPG automatically from an influence-diagram DM would require further implementation. Although such work was not part of my thesis, it is an obvious future extension of this work (see Section 9.2.1)

3.5 Past DM conceptual models

Several groups have done research on DM taxonomies. My DM conceptual model expands their work to include dual utilities, Markov processes, and explicit modeling of DM assumptions and evidence.

Wellman and colleagues developed the BUNYAN system for automated critiquing of medical decision trees (Wellman et al. 1989). BUNYAN uses an underlying taxonomy of node and branch types to represent decision trees, and uses this taxonomy to determine the essential structural features of a problem without a detailed understanding of specific medical concepts. Node types include decision, terminal, and chance (Figure 3.3); branch types include action (either treatment or test) and event (either cost or physiologic state).

BUNYAN detects potential problems in a DM by matching general pattern expressions that refer to branch and node types. BUNYAN provides information about decision, terminal, and chance nodes, although it does not define representation of Markov nodes or have the ability to deal with dual-utility DMs.

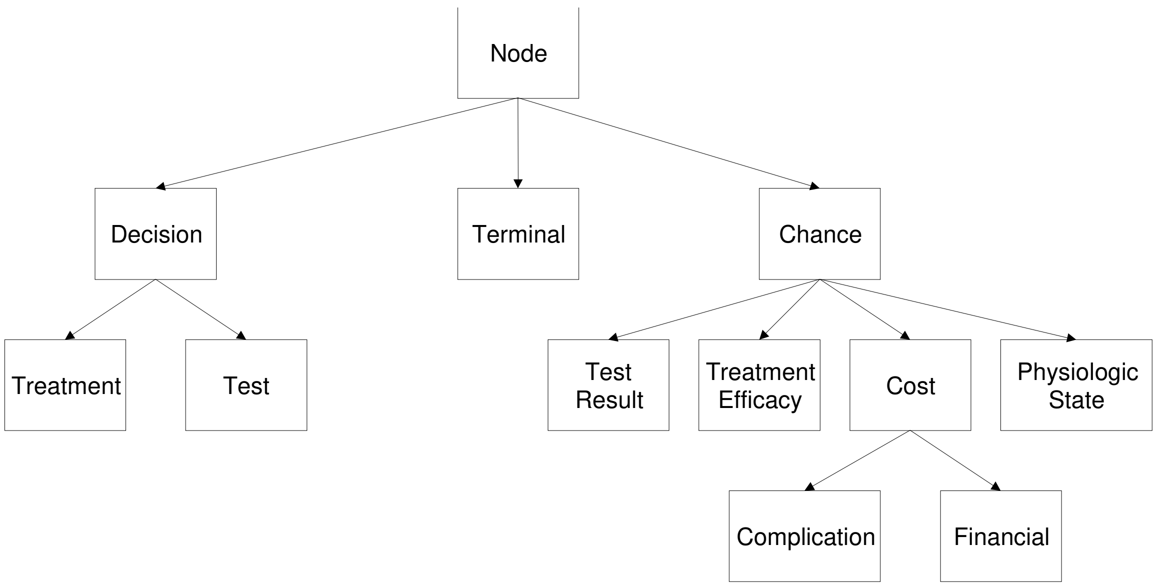


Figure 3.3. The BUNYAN taxonomy of node types. BUNYAN categorizes node and branch types. Nodes are partitioned into decision, chance, and terminal types. Further distinctions are made based on the nature of the decision or treatment represented. (Reprinted with permission. *Source*: From MP Wellman, MH Eckman, C Fleming, SL Marshall, FA Sonnenberg, and SG Pauker. Automated critiquing of medical decision trees. *Med Decis Making* 1989;9:272-284).

As part of his dissertation work, Langlotz developed the QXQ system, which used symbolic-reasoning techniques to generate text and graphics to explain and interpret the results of formal DMs (Langlotz 1989). QXQ uses frames, slots, and values to describe the concepts, important characteristics, and relationships between nodes in a decision tree. For example, in Figure 3.4, the box labeled B describes a value node. The slot-value pairs in box B indicate such characteristics of the node as its offspring, probability, probability expression, utility, utility expression, and situation. Box A shows similar information for a chance node. Langlotz's ongoing work on automated explanation for DMs will provide a valuable extension to my thesis research.

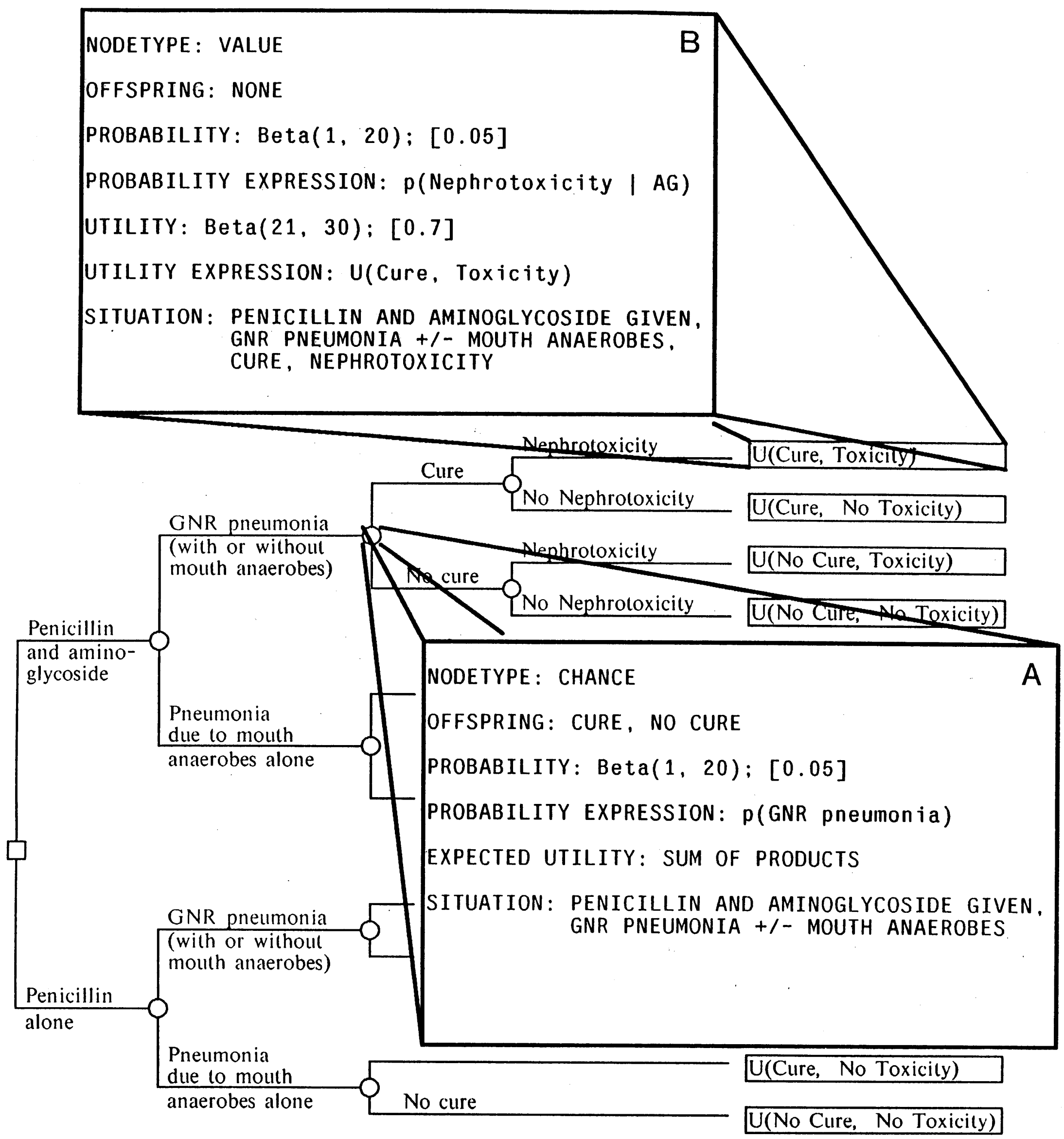


Figure 3.4. The internal structure of two frames representing nodes in QXQ. The box labeled A contains a frame describing a chance node. The box labeled B describes a value parameter. The node frame has slots for nodetype, offspring, probability, probability expression, expected utility, and situation. (Reprinted with permission. *Source:* CP Langltoz, EH Shortliffe, LM Fagan. A methodology for generating computer-based explanations of decision-theoretic advice. *Med Decis Making*. 1988, 8:290-303).

Most recently, Sonnenberg and colleagues developed MIDAS, an architecture for knowledge-based construction of DMs (Sonnenberg et al. 1994). This system is based largely on the BUNYAN DM taxonomy, but it adds a knowledge base for a specific clinical domain and allows cost-effectiveness analyses using a dual-utility model. None of these existing taxonomies model Markov processes. Comparison of these taxonomies to the knowledge required for my CPG conceptual model indicates that the knowledge found in these taxonomies is not sufficient to create the corresponding CPG representation. For example, although QXQ lists the probability and utility expressions for each node, it does not detail the evidence behind these expressions. Also, none of these conceptual models incorporates such elements as the model perspective or patient population into their framework.

Table 3.2 summarizes past DM conceptual-modeling work; it also indicates what additional abilities ALCHEMIST has.

Table 3.2. Summary of previous DM taxonomy work. The table presents a brief summary of the main work from which ALCHEMIST builds its conceptual model, and the extensions to this work that ALCHEMIST provides. DM = decision model; KB = knowledge base; CPG = clinical-practice guidelines.

System	Description	Model types	Single utility	Dual utility	Markov process	Other
BUNYAN Wellman 1989	Decision tree critiquing program; taxonomy	Decision tree	X			
QXQ Langlotz 1989	Representation of DMs; generation and interpretation of results	Decision tree; influence diagram	X			Can do dual utility using two models

Table 3.2. Summary of previous DM taxonomy work. The table presents a brief summary of the main work from which ALCHEMIST builds its conceptual model, and the extensions to this work that ALCHEMIST provides. DM = decision model; KB = knowledge base; CPG = clinical-practice guidelines.

System	Description	Model types	Single utility	Dual utility	Markov process	Other
MIDAS Sonnenberg 1994	Automated construction of DMs; clinical KB	Decision tree; influence diagram	X	X		Based on BUNYAN
ALCHEMIST	Automated creation of CPGs from DMs	Decision tree; influence diagram	X	X	X	Incorporates assumptions, evidence tables, and perspective

3.6 My conceptual model

Using previously developed decision-tree and influence-diagram DMs, combined with existing DM taxonomies, I developed a DM conceptual model that identifies the knowledge inherent in simple DMs, as well as in decision analyses that require sequential decisions, dual utilities, or Markov processes. I also concentrated on the knowledge inherent in DMs that can later be used for creating evidence-based CPGs.

I identified seven main elements of a DM: variables, assumptions, analysis-specific information, alternatives, events, outcomes, and analytic results. Each of these elements has further subclasses of information. Figure 3.5 is a graphical depiction of the top layers of the DM conceptual model and I include the a description of the DM conceptual model in Appendix B.

A DM has numerous globally and locally defined **variables**. For each variable, there exist corresponding assumptions about why this variable was important enough to include in the analysis, as well as assumptions about the best-estimate input value and range

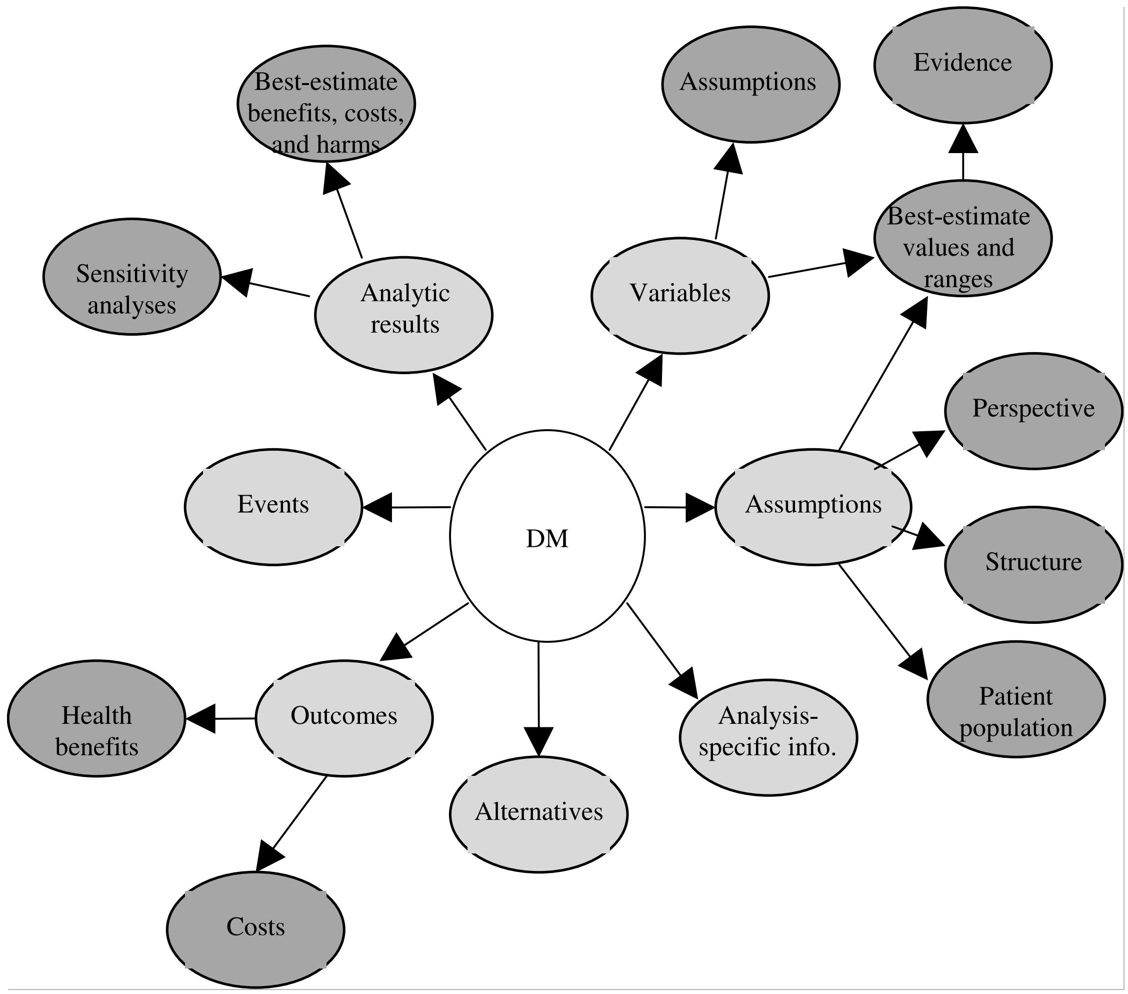


Figure 3.5. Schematic representation of DM conceptual model. This figure shows the seven main elements of each DM with some of their subclasses of information depicted. For example, every DM has numerous variables that have associated with them assumptions about why they are included in the analysis, best-estimate input values and ranges, and evidence for these values and ranges.

(described further when I discuss the assumptions elements). In addition, there is evidence to support this input value and range. For example, for the prior probability of mediastinal metastases, there is the assumption that this probability may vary depending on histology, location, and size of the primary tumor, or on the appearance of the mediastinum on a chest X-ray image; such variables therefore should be included in the DM. The best-estimate input value is 0.46 and ranges from 0.10 to 0.90. Evidence for this input value and range was obtained from a meta-analysis of the literature that is represented in a corresponding evidence table.

Each DM also has numerous simplifying **assumptions**. Some of these assumptions are explicit within the DM structure or variable definitions; other assumptions must be obtained directly from the decision analyst (Chapter 4). The assumptions can be divided into four categories:

1. **Modeling perspective.** The perspective of the DM — especially that of a DM that reflects the costs involved or patient preferences for different strategies — affects the CPG produced. Therefore, I model it explicitly in my DM conceptual model. For example, the LC-CEM uses diagnostic and treatment costs that reflect the cost incurred directly by the patient; therefore, the LC-CEM is modeled from the patient perspective. The perspective of the DM usually affects the costs and utilities used in a DM, as well as the ordering of the resulting outcomes. If a guideline user wants to adjust the modeling perspective, ALCHEMIST identifies the affected variables to allow the user to make the necessary modifications.
2. **Best-estimate input values and ranges.** The input values and the sensitivity-analysis ranges used in the DM analysis reflect numerous assumptions made by the decision-analysis team. The chosen values may combine results of clinical trials, meta-analyses of the current literature, or estimates based on expert opinion. For example, the 0.005 probability of death from mediastinoscopy combines results from two main analyses (Fishman and Bronstein 1975, Larsson 1976). Links to evidence tables and sources make these assumptions an explicit part of the generated CPG, allowing a guideline user to explore the evidence behind particular data assumptions and to change these input values to reflect data more suitable to her patient population.

3. **Patient characteristics.** My DM conceptual model makes explicit the assumptions about the patient population for whom the CPG is designed by describing in detail the patient population and any corresponding assumptions, and by indicating to the user when they change a patient characteristic.
4. **Model structure.** The structure of the model reflects assumptions regarding both the relationships among variables and the chosen simplifications of the scope of the DM. For example, in the LC-EM, positron-emission tomography scans are not an available diagnostic testing alternative. If a guideline developer did not agree with this simplification, structural reorganization of the model would be required. The facility to make structural changes to the DM currently is not implemented in ALCHEMIST. Instead, these structural assumptions are merely listed explicitly for CPG users, so that they can determine whether the assumptions are valid in the clinical situation.

Analysis-specific information, such as the most recent modification of the model structure or data, the decision modeling software used, and the decision analyst's or analysis team's name, is an explicit part of the DM conceptual model. For example, noting that a DM on HIV treatment was updated most recently in 1985 conveys information about the model's applicability to today's patients.

Because DMs are built to model decision problems, every DM has at least two decision **alternatives**. These alternatives reflect the possible treatment or screening options available to the decision maker; they may be available to the decision maker all at once or sequentially over time. For example, in the LC-EM, available alternatives are diagnosis by CT, mediastinoscopy, or both; and treatment by thoracotomy or radiation therapy.

Each alternative has associated with it a number of possible resulting **events**. For example, in the LC-EM, the CT alternative results in a chance that the CT will return a positive result, and a chance that the CT will return a negative result. Instead of having only simple chance nodes, alternatives can also have associated with them Markov nodes. For example, in the SCD-MM, a patient following the amiodarone strategy enters a Markov node that corresponds to his being in one of the following states at the end of each 1-month cycle: arrhythmic death, nonarrhythmic cardiac death, noncardiac death, well, or amiodarone toxicity but otherwise well.

Each path through a DM, composed of alternatives and events, ends with an **outcome**, which is valued in terms of a utility (or a dual utility in cost-effectiveness models). For example, in the LC-EM, a patient with mediastinal metastases who has undergone a thoracotomy is given a life expectancy of 1.8 life years. As I mentioned in Section 2.2, for ALCHEMIST to create a cost-effectiveness CPG, the DM requires a threshold value for dealing with tradeoffs in a dual-utility model.

Finally, the DM conceptual model has information regarding the expected **analytic results**. A user of a DM can calculate the estimated costs, effectiveness (using a measure such as life expectancy or quality-adjusted life expectancy), or cost effectiveness of the differing strategies using the best-estimate input values. For dual-utility models, ALCHEMIST can obtain marginal cost-effectiveness ratios between the available strategies. Using the DM, ALCHEMIST can also perform sensitivity and threshold analyses, and can present the results to the guideline developer in graphical format (e.g., as a tornado diagram showing the effect of chosen variables on the expected results (Figure 5.5)). A user also can calculate all these results using discounted or undiscounted costs and life years.

At present, my DM conceptual model is based on the decision-tree representation and is implemented for only decision trees created by a decision analyst using the Decision Maker software. The restriction of my subset to decision trees is imposed for implementation purposes; the DM conceptual model is designed to describe the knowledge within an influence diagram and other DM representations as well.

3.7 Interactive decision models

Although DMs can provide a formal foundation for CPG development, their widespread use is often limited by the lack of platform-independent software that geographically dispersed users can access and use easily without extensive training. As a predecessor to my thesis work on automated guideline creation from DMs, I developed PORTAL, a web-based interface for previously developed DMs (Sanders et al. 1998). PORTAL allows a decision analyst to use a web browser to interact with a DM and to change the value of input variables within prespecified ranges, to specify sensitivity or threshold analyses, to evaluate the DM, and to view the results generated dynamically. The web site (http://alchemist.stanford.edu/wbds/) also provides linkages to an explanation of the model, and evidence tables for input variables. This system has the potential to increase the usefulness of DMs by enabling a broader audience to incorporate systematic analyses into both policy and clinical decisions. A second interactive web site was developed by Kattan and Fearn at Baylor College of Medicine. Their system allows a user to load a previously developed DM, to specify which variables should be interactive, and to publish this interactive DM on their website (http://utility.urol.bcm.tmc.edu). Similarly, users of the DATA decision-analytic software by TreeAge can create a "Custom Interface" to a developed DM. This interface (currently not web-based) allows a decision analyst to designate which variables may be changed by a future user, and which analyses this user will be able to perform manually. All three of these interactive systems allow users who are not familiar with the mathematical details of DMs to interact with and explore the analytic results of such analyses. Unlike ALCHEMIST however, these systems do not produce the corresponding CPG recommendation or algorithm, and they do not contain sufficient information to produce a high-quality CPG.

3.8 Summary

My development of the DM and CPG conceptual models identified the essential information needed for the ALCHEMIST system to create CPGs directly from DMs. Comparison of these conceptual models identifies the crucial elements that a decision analyst has to add to the DM representation so that sufficient information is present for ALCHEMIST to use in creating a CPG. In Chapter 4, I discuss these missing elements and my method for obtaining this additional knowledge from a decision analyst. In Chapter 5, after detailing the expected benefits from basing CPG creation on DMs, I outline the mapping between the DM and the CPG conceptual model.

Chapter 4

Decision-Model Annotation Editor

In this chapter, I describe ALCHEMIST's **decision-model (DM) annotation editor**. ALCHEMIST uses this annotation editor to retrieve from the decision analyst additional information that is required for the creation and automatic updating of a clinical-practice guideline (CPG). Although the DM conceptual model is able to represent the knowledge required for several key elements of a CPG, additional knowledge is essential to the production of a high-quality CPG that is not available directly from the DM. I identify this missing information (Section 4.2), explain ALCHEMIST's method for obtaining it, and detail the DM annotation editor's current web interface (Section 4.3).

4.1 Comparison of the DM and CPG conceptual models

Although the DM conceptual model provides a great deal of information regarding CPG elements — such as available alternatives, possible outcomes, and best-estimate input values used in the initial analysis — a comparison of even the top level of the DM and CPG conceptual models identifies elements that do not have an apparent mapping. For example, there is no explicit information in the DM conceptual model to correspond to the targeted health problem element of the CPG. This information is not provided by the DM, yet clearly is required for the creation of the corresponding CPG. Fortunately, it is information to which the decision analyst should have ready access.

In Table 4.1, I align the elements in the two conceptual models, to the extent possible, and highlight those areas where the DM conceptual model does not provide a direct mapping onto the CPG conceptual model. For example, there is nothing in the DM representation that helps ALCHEMIST to provide the guideline user with a definition of a positive CT examination, yet this information is required by the guideline user to determine what branch of the clinical flowchart algorithm she should follow. Although there is information in the DM that refers to the targeted patient population (i.e., ALCHEMIST can provide the starting age of the cohort), there is no information about the targeted health problem, about the intended user of the CPG, or about the need for recommendations, yet all these elements are parts of the CPG's objective.

Table 4.1. Comparison of CPG and DM conceptual models. An asterisk (*) indicates CPG elements that are not addressed at all by the DM representation. Note that those elements, such as the CPG objective, that indicate a corresponding DM conceptual-model element may not be fully described and may require additional annotation.

CPG conceptual-model element	DM conceptual-model element
Objective	Assumptions
Options	Alternatives
Outcomes	Outcomes
Evidence	Variables; assumptions; analysis information
Values	Variables (utilities)
Benefits, Harms, and Costs	Analytic output
Recommendations	Analytic output
Validation	*
Sponsors	*
Definitions	*
Sources	*

The DM conceptual model also contains probabilities and utilities that ALCHEMIST uses to calculate the recommendations and benefits, harms, and costs that are not normally represented explicitly in the CPG. ALCHEMIST retains this information to use for future modifications of the DM when a guideline user wants to update or tailor the existing CPG.

4.2 Identification of missing knowledge

Before it acquires the needed additional knowledge from the decision analyst, ALCHEMIST must determine what knowledge is omitted in the DM representation.

After the decision analyst has loaded the decision tree into ALCHEMIST, ALCHEMIST completes the DM conceptual model and attempts to instantiate the CPG conceptual model. ALCHEMIST can produce automatically certain CPG elements, such as the available alternatives, by locating the unique branches of decision nodes. Additional information is implicit in the DM and can be inferred through analytic techniques. For example, ALCHEMIST can determine the logic of the CPG flowchart algorithm by analyzing the decision tree and pruning the tree to reflect the optimal path at each decision point. Other knowledge is not available either explicitly or through computation and must be obtained from the decision analyst. ALCHEMIST identifies those CPG elements that it is unable to complete, and creates dynamically a web-based DM annotation editor to query the decision analyst for the needed information.

The DM annotation editor uses the same organization and format that the final CPG interface will use. This format combines attributes of a structured CPG abstract, a clinical flowchart algorithm, and the information provided in my DM and CPG conceptual models. ALCHEMIST automatically displays for the decision analyst's verification information that is obtained directly from the DM. ALCHEMIST assesses the DM, identifies missing information, and creates input areas so that the decision analyst can provide this needed information.

After the decision analyst has entered the required information in the DM annotation editor, he submits the information to ALCHEMIST, which completes the DM-to-CPG mapping and creates the CPG.

4.3 Acquisition of the missing knowledge

After ALCHEMIST identifies the missing knowledge, it creates the DM annotation editor and displays for the decision analyst a structured CPG abstract — which will also be used in the CPG browser and custom-tailoring editor. The menu on the left side of the web interface is consistent throughout ALCHEMIST; it will help the decision analyst to navigate through the DM annotation editor, and will help the guideline user to navigate through a detailed CPG. Figure 4.1 shows an example DM annotation-editor page for the LC-EM.

Elements of the CPG that can be extracted directly from the DM by ALCHEMIST are displayed for the decision analyst. Figure 4.2 shows the DM annotation editor's primary-variables table, which incorporates the variable names and values from the DM. The decision analyst will be required to provide ALCHEMIST with the variable definitions and sensitivity-analysis ranges.

ALCHEMIST creates input areas for those CPG elements that it cannot infer directly from the DM. The decision analyst enters the required information in these input areas. Much of the requisite information is entered as free text (e.g., the variable definitions shown in Figure 4.2). Certain information uses radio buttons (one option button in the group can be selected at a time), checkboxes, or scrolling lists (multiple options may be selected). The DM annotation editor always provides the decision analyst with an option of other (paired with a text area) for those cases where the DM does not conform to conventional modeling techniques. Figure 4.3 illustrates how the DM annotation editor uses checkboxes to indicate the outcome measures that are part of the decision analysis.

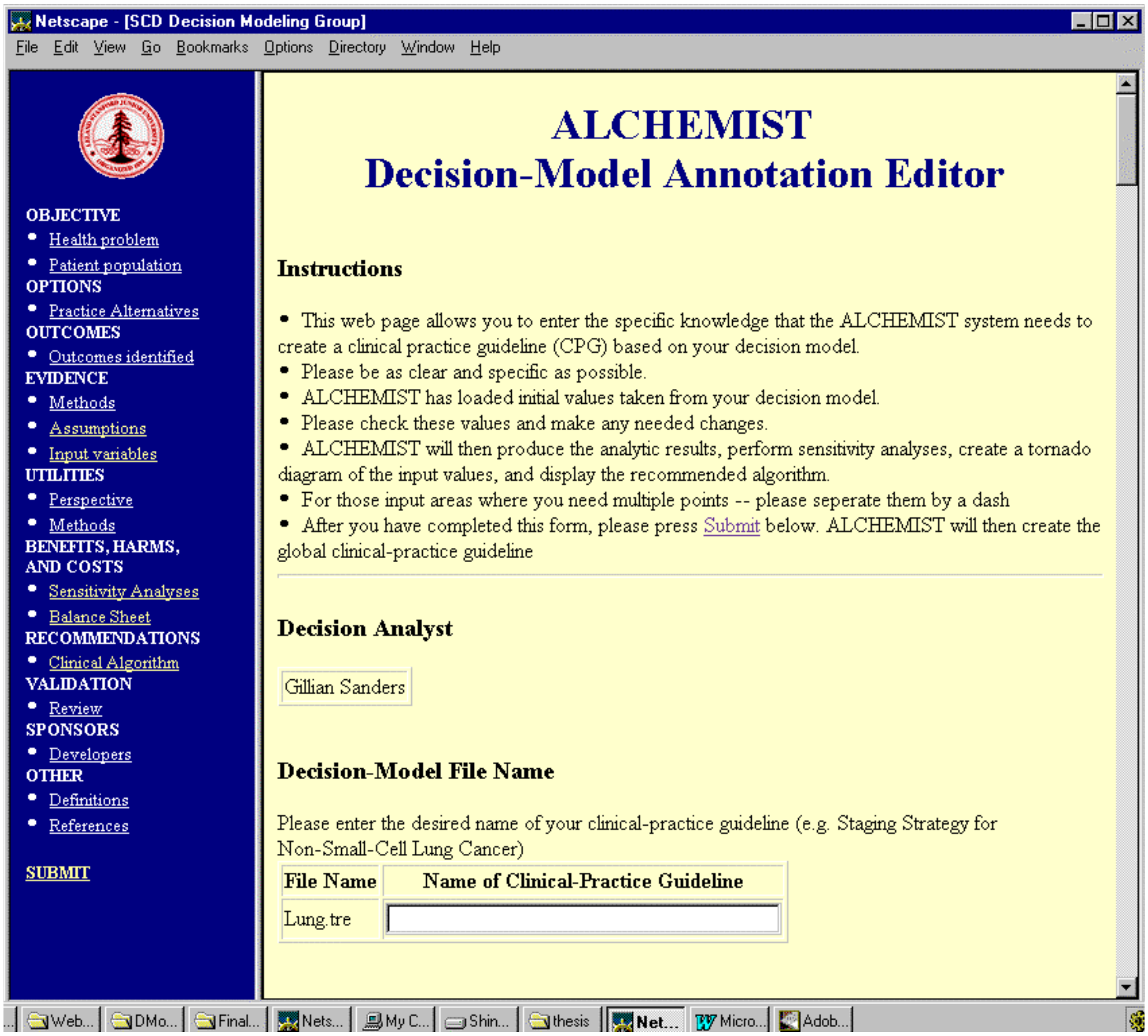


Figure 4.1. Overall structure of the DM annotation editor. The menu on the left details the overall structure of the DM annotation editor with links to the CPG elements that require decision-analyst annotation.

Other input areas, such as the evidence tables, are created dynamically based on the needs of the decision analyst. Using a scrolling list, the decision analyst indicates which of the input variables require an evidence table. ALCHEMIST then creates for each of those variables evidence tables with blanks for the following possible elements: study design, number of patients involved, outcomes measured, patient population, interventions compared, biases, observed outcomes, level of evidence, and reported effects of the intervention (Eddy 1992). The decision analyst then enters into these evidence tables information for

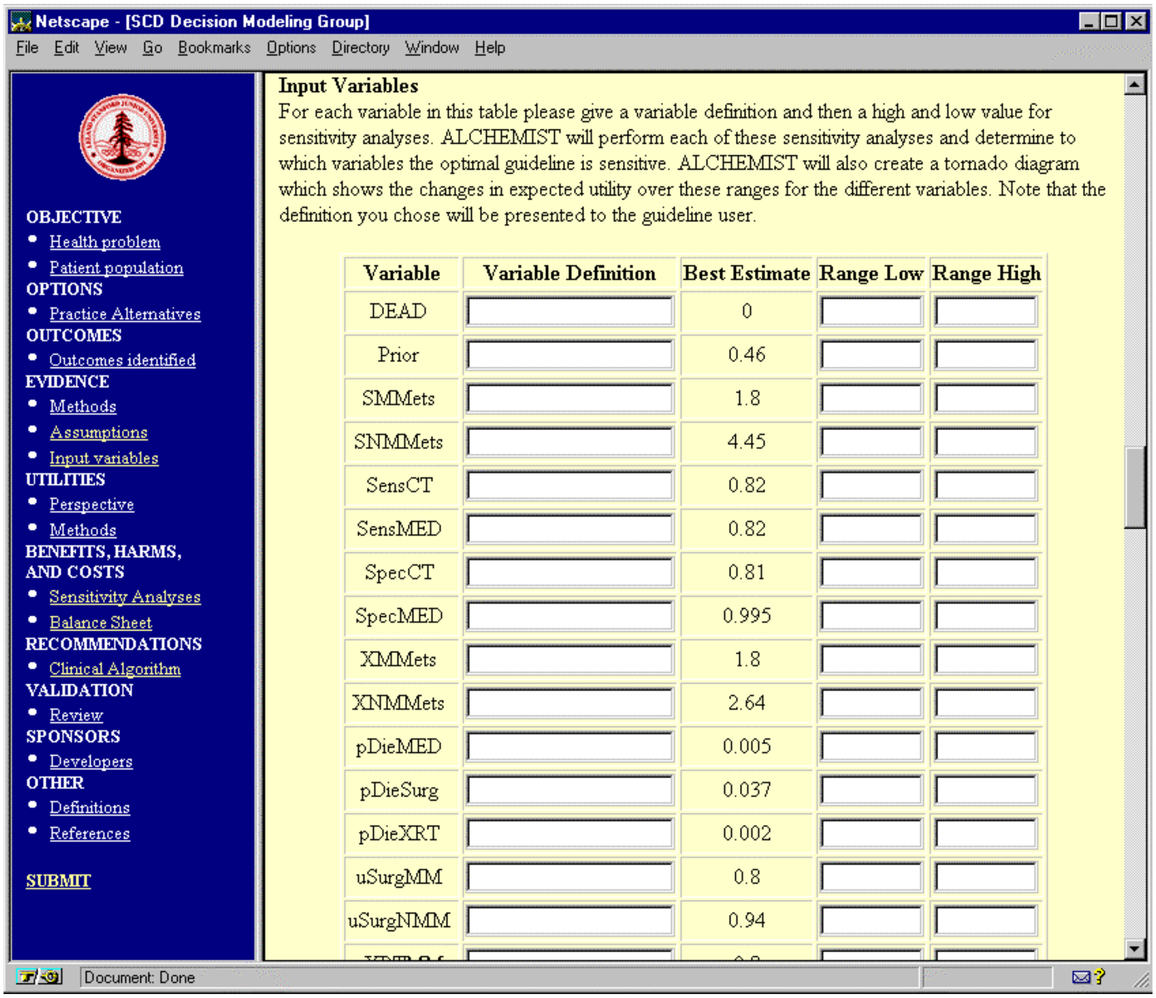


Figure 4.2. Automated creation of the DM input table. The primary-variables table is created dynamically from the DM; it lists the variable names and base-input values.

the relevant categories. Those elements that are not used will be discarded in the generated CPG. For example, a decision analyst who wants to create an evidence table for the amiodarone withdrawal rate due to toxicity used in the SCD-MM may choose to include the following evidence table elements: study name, study design, level of evidence, patient population, description of intervention, outcome, number of patients. Figure 4.4 shows a sample evidence table from the LC-EM. The decision analyst uses this evidence table to enter information regarding the prior probability of mediastinal metastases.

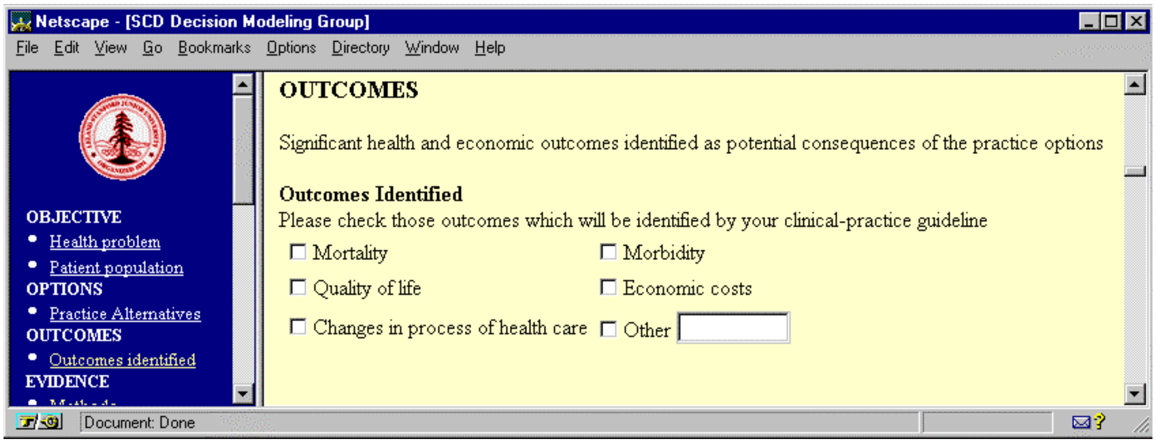


Figure 4.3. Checkbox input in the DM annotation editor. The DM modeling outcome measures are indicated as possible checkboxes, allowing the decision analyst to choose among mortality, morbidity, quality of life, economic costs, changes in process of health care, or other outcomes.

A future extension of my thesis work, I plan to incorporate certain aspects of research performed by Sim regarding sharable databases of randomized clinical trails (Section 9.2.7). Sim designed Ocelot-CCM, a core-conceptual model of clinical trials; and is developing a Trial-Bank open architecture-programming interface (API) that will allow other software packages to communicate with her Trial-Bank system via the Internet and take advantage of its underlying evidence (Sim 1996, Sim 1997). This API will be available to ALCHEMIST and could help my system to create and maintain the evidence table components used in my DM annotation editor.

Most of the information that is required from the decision analyst should be straightforward to provide. An important part of the DM representation required for CPG creation — and more specifically for future tailoring of the CPG — is the explicit definition of the model's underlying assumptions. All modeling costs and utilities are linked initially to the model perspective. When a future guideline user changes the model perspective, these variables are flagged, and ALCHEMIST will query that guideline user for any needed modifications. Several transition probabilities will depend on the patient population, and the decision analyst is required to identify these key variables. For example, if a guideline

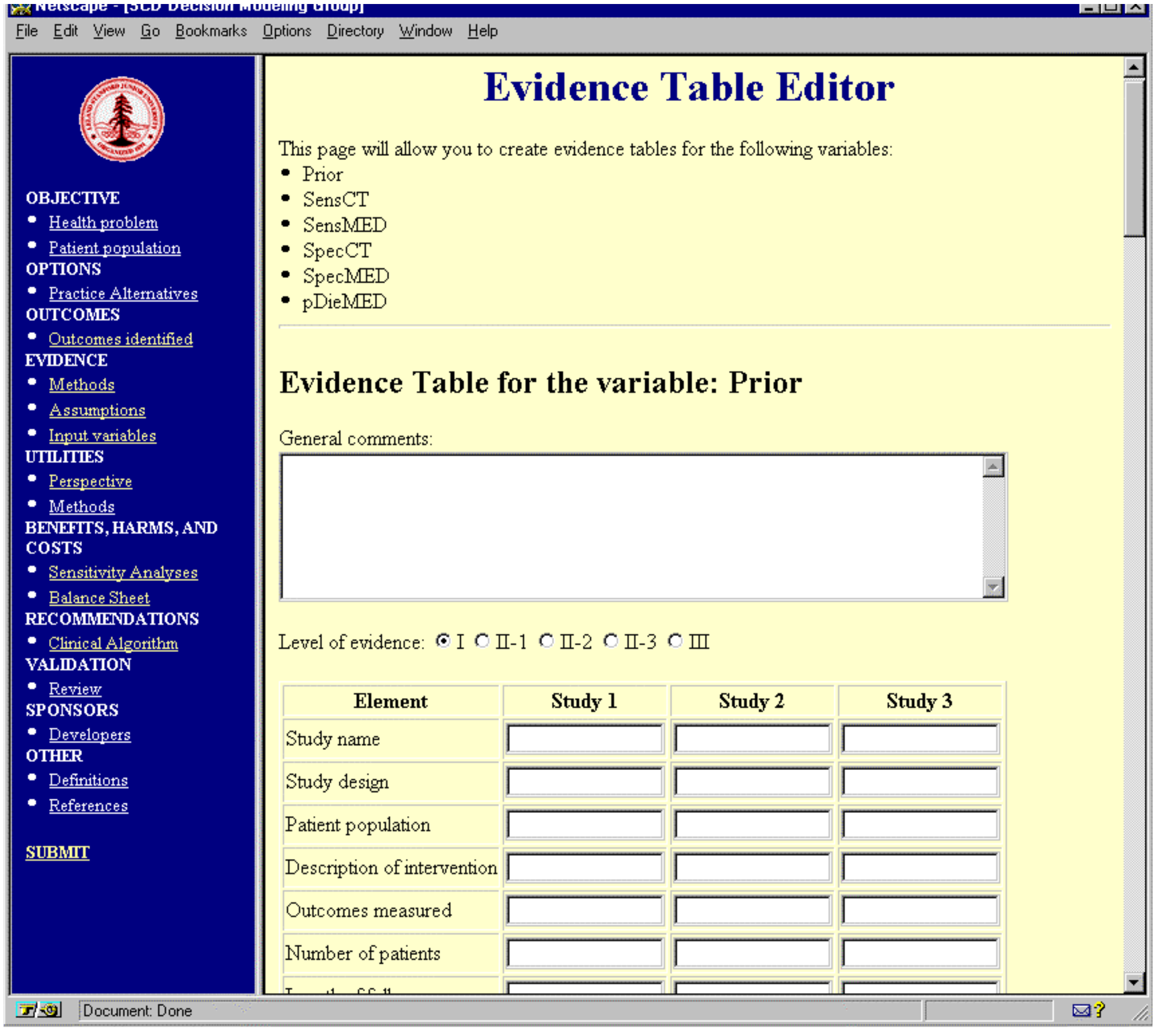


Figure 4.4. Dynamic creation of evidence tables with the DM annotation editor. The decision analyst is able to choose which variables and which evidence-table elements to include for the DM data.

developer of the SCD-MM wants to change the patient population to describe a post–myocardial infarction low-risk population, rather then the initial survivors of SCD, variables such as the initial hospitalization cost, the probability of arrhythmic events, and even the probability of noncardiac death could change. Using the DM annotation editor, the decision analyst will identify the key variables that depend on patient population or on model perspective.

4.4 Users of the DM annotation editor

The users of the DM annotation editor are the members of the decision-analysis team that created the DM. Interaction with the DM annotation editor requires that the decision analyst perform additional work, but provides structure to the modeling and evidence-gathering process and helps the decision analyst in the creation of a DM that has no inconsistencies or structural problems, and that can be transformed into a CPG.

When developing a DM, a decision analyst gathers a collection of evidence from numerous sources of differing quality. The annotation editor allows the decision analyst to enter and store this information in a dynamically created, interactive format. The implementation of the annotation editor on the web also allows other members of the DM development team — even those who are located at remote sites — to peruse the DM representation and to share in the evidence-gathering and DM-creation processes. The information entered in the DM annotation editor is stored in an Excel file; therefore, a decision analyst can complete the annotation task over several web sessions. Error messages produced by ALCHEMIST regarding the structural soundness or logical inconsistencies of the DM may also help the less-experienced decision analyst to produce a valid DM.

The completion of the DM annotation editor is a one-time task that the decision analyst must perform before ALCHEMIST can produce the initial CPG. Any future tailoring and updating of the CPG will use the annotation editor's information and will not require additional input from the decision analyst.

After becoming familiar with the ALCHEMIST CPG-creation system, a decision analyst may become aware of the important elements for CPG creation and may use this information to develop more complete future DMs.

4.5 Need for a structured vocabulary

Many of the required inputs in the DM annotation editor are definitions of variable names used in the DM. Without this information, mapping of a DM into a CPG would result in a CPG that could not be understood or used by guideline users. For example, in Figure 4.2, the variable pDieXRT refers to the probability of death given radiation therapy. Although this interpretation may seem obvious to the decision-analysis team, a future guideline developer might not understand what the variable denotes and therefore might interpret incorrectly the base values or recommendations. If a decision analyst created a DM model using a structured vocabulary, then much of this annotation would not be required. In my research, I did not look at the use of a structured vocabulary, although I did use certain variable-naming conventions (e.g., CycLen for the cycle length used in Markov DMs, InitAge for the age of the patient at the start of the analysis in years) to ease the DM analytical and annotation process (Appendix C). An extension of my thesis work would be to explore whether the use of a structured vocabulary for decision-model variables would allow the automation of the DM-CPG process to be more efficient, or would enable the DM — and the CPG based on this DM — to be integrated with a computer-based patient record. I discuss the feasibility and ramifications of this extension in Chapter 9.

4.6 Summary

In this chapter, I compared the knowledge provided by a DM with that required for the creation and use of a CPG. I also described the DM annotation editor, which is a tool for obtaining this missing information from the decision analyst. The DM annotation editor eases the process of providing this needed information by presenting to the decision analyst a structured format, dynamically created interface, and potential values. In Chapter 5, I discuss how ALCHEMIST combines the information from this DM annotation editor with the DM and CPG conceptual-model framework to produce a formal mapping algorithm between a DM and a corresponding CPG.

Chapter 5

Algorithm for DM-to-CPG Mapping

Although several guideline developers have based their work on clinical-practice guidelines (CPGs) in part on a decision model (DM) (American College of Physicians 1992, American College of Physicians 1994, Carlson et al. 1994, Eddy 1991a, Eddy 1991b, Eddy 1991c, Eddy 1991d, Fahs et al. 1992, Grady et al. 1992, Littenberg et al. 1991, Melton et al. 1991, Schapira et al. 1993, Singer et al. 1991, Sox et al. 1991a), this practice is not the norm. In this chapter, I discuss my rationale for using DMs to create CPGs (Section 5.1) and identify the main conceptual and practical differences between these two representations (Section 5.2). After examining previous work on mapping between DMs and CPGs (Section 5.3), I conclude by describing ALCHEMIST's DM-to-CPG mapping algorithm, its limitations, and possible future extensions (Section 5.4 and 5.5).

5.1 Creation of CPGs from DMs

I base my research on the assumption that it is beneficial to the guideline user and to the patient to create a CPG from a normative DM. In this section, I provide justification for this assumption.

In recent articles, several authors support the use of DMs for the creation of evidence-based CPGs (Goldman 1996, Kulikowski 1996, Mutnick and Szymusiak-Mutnick 1996, Nease and Owens 1991, Nease and Owens 1994, Oddone et al. 1994, Owens 1998, Owens and Nease 1991b, Owens et al. 1993, Owens et al. 1997b, Shiffman and Greenes 1994, Shiffman et al. 1993). These authors explain that basing CPGs on DMs allows guideline developers to use an evidence-based method of linking explicitly the CPG to the underlying probabilities and utilities of the DM, thereby allowing the generated CPG to specify recommended patient management for common situations, and well as allowing guideline users to apply the analytic techniques of the DM in more difficult clinical problems. DMs also emphasize the critical evaluation of evidence, rather than expert opinion, to define proper patient care (Sox and Woolf 1993). Although the authors promote the use of decision analyses for CPG creation, they do not provide the necessary framework for such a DM-to-CPG mapping. ALCHEMIST attempts to fill this void by providing an automated method of creating CPGs from DMs.

A DM defines clearly the alternatives, outcomes, and patient preferences in a given decision problem. Such definition helps the decision analyst to ensure that relevant factors are considered and it enables others to review and check the reasoning behind the decisions. ALCHEMIST combines these elements in an objective and predictable way to produce a recommendation. ALCHEMIST performs sensitivity analyses to identify critical variables to focus any future refinement of the CPG, and to alert the guideline user that these variables heavily affect the CPG recommendations. In addition, as I discuss in Chapter 6, ALCHEMIST's DM-to-CPG mapping allows easy updating of the CPG when new clinical information becomes available.

Although there are many benefits of basing CPGs on DM, one limitation is that the DM must represent the decision problem with sufficient fidelity. Developing such a DM is a nontrivial task — yet it is a task that a decision-analysis team can perform using the available literature and clinical trial results to reduce costs.

5.2 Differences between CPGs and DMs

A DM specifies the probability that a specific clinical situation exists, and quantifies the value of the outcome of a decision. A CPG ideally has this information inherent in its recommendations, but does not represent the information explicitly for the guideline user. Although the DM and the CPG representations of a clinical problem can be seen as two representations for the same set of clinical recommendations, there is information contained in each representation that is not available in the other. For example, in the LC-EM, the sensitivity of the CT examination for mediastinal metastases is 82 percent. Although a corresponding CPG may use this information, the latter may not be provided explicitly for the guideline user. Similarly, the CPG for the lung-cancer decision problem described in the LC-EM may have the objective of determining the optimal staging strategy for patients with non–small-cell lung cancer. Clearly, the decision analyst would also have this objective in mind when she develops the DM, yet it is not contained explicitly in the DM representation. ALCHEMIST's DM-to-CPG mapping algorithm translates between the two representations while maintaining a link to the underlying DM to permit future CPG updating or tailoring.

Table 5.1 summarizes the main differences between DMs and CPGs

Table 5.1. Comparison of DMs and CPG representations. (Hayward 1995, Kamae and Greenes 1991, Margolis 1983, US Congress Office of Technology Assessment 1994).

Decision models	Clinical-practice guidelines
Specifies explicitly the probability that a particular clinical state exists	Reflects only implicitly the underlying utilities and probabilities
Quantifies the value of the outcome of a decision	May rely on qualitative reasoning
Answers: (1) Is it more desirable to do A or do B? (2) With what probability is A the most desirable action?	Algorithm prescribes that, given X, do Y
Focuses on pivotal decisions at a local stage	Emphasizes a particular clinical context
	Often deals with multistage workup and management

5.3 Previous work on transforming DMs into CPGs

Although researchers have promoted basing CPGs on DMs, they have not done a large amount of work in the clinical domain on transforming decision-tree representations of DMs into CPGs. Two relevant research projects are work by Shiffman and colleagues on the use of decision tables to improve clinical guidelines (Shiffman 1991, Shiffman 1995, Shiffman and Greenes 1991, Shiffman and Greenes 1992, Shiffman and Greenes 1994, Shiffman et al. 1993), and work by Kamae and Greenes on the use of a computational model of approximate Bayesian inference for associating clinical algorithms with decision analyses (Kamae and Greenes 1991).

Shiffman and colleagues used decision-table techniques to ensure logical completeness, to eliminate ambiguity, and to translate the clinical logic of a decision problem into a flowchart representation (Reinwald and Soland 1966, Reinwald and Soland 1967, Shiffman 1991, Shiffman 1995, Shiffman and Greenes 1991, Shiffman and Greenes 1992, Shiffman and Greenes 1994, Shiffman et al. 1993, Wears et al. 1994). **Decision tables** are matrices that relate a set of conditions to a set of actions. Each column in the decision table is considered to be a rule of the form: IF <conditions> THEN <actions>. Using reduction rules, a user can reduce the size of the decision table by combining those rules that result in the same diagnostic conclusion but differ in one finding; this strategy eliminates any tests from a condition set whose results are not necessary for a conclusion to be reached. In their research, Shiffman and colleagues defined an augmented decision-table format that allows the incorporation of probability and utility data. For example, Figure 5.1 is a reduced and sorted augmented decision table that represents a strategy for the diagnosis of appendicitis. The authors use this augmented decision table to create a corresponding clinical algorithm (Figure 5.2) (Shiffman and Greenes 1991).

	1	2	3	4	5	6	7	8	9
Location	**RLQ**	**RLQ**	**RLQ**	**RLQ**	**LH**	**LH**	**LH**	**Other**	**Other**
$P(\text{finding} \mid \text{app})$	.74	.74	.74	.74	.13	.13	.13	.13	.13
$P(\text{finding} \mid \text{NSAP})$	.29	.29	.29	.29	.09	.09	.09	.62	.62
Rebound	Y	N	N	–	Y	N	N	Y	N
$P(\text{finding} \mid \text{app})$	.95	.05	.05		.95	.05	.05	.95	.05
$P(\text{finding} \mid \text{NSAP})$	.26	.74	.74		.26	.74	.74	.26	.74
Severe	N	N	N	Y	–	–	–	–	–
$P(\text{finding} \mid \text{app})$	.61	61	61	.39					
$P(\text{finding} \mid \text{NSAP})$	.81	.81	.81	.19					
Rectal tender	–	Y	N	–	–	Y	N	–	–
$P(\text{finding} \mid \text{app})$		.43	.57			.43	.57		
$P(\text{finding} \mid \text{NSAP})$		.16	.84			.16	.84		
$P(\text{comb} \mid \text{App})$	.429	.010	.013	.289	.124	.003	.004	.124	.007
$P(\text{comb} \mid \text{NSAP})$	.061	.028	.146	.055	.023	.011	.056	.161	.459
Likelihood Ratio	7.021	0.349	.088	5.238	5.278	2.62	.066	.766	.014
posttest odds	3.932	.195	.049	2.933	2.956	.147	.037	.429	.008
$P(\text{app} \mid \text{comb})$	.80	.16	.05	.75	.75	.13	.04	.30	.01
Operate	X	X		X	X	X		X	
Observe			X				X		X

Figure 5.1. Example reduced and sorted augmented decision table. Conditional probabilities of individual findings in appendicitis and nonspecific abdominal pain (NSAP) are indicated, as are probabilities of combinations of findings, and the likelihood ratios. RLQ = right lower quadrant; LH = lower half of the abdomen; Other = other location; app = appendicitis; Y = yes; N = No. (Reprinted with permission from the Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care 1992 American Medical Informatics Association Published by Hanley & Belfus, Inc., Philadelphia, PA. *Source:* RN Shiffman and RA Greenes, Use of augmented decision tables to convert probabilistic data into clinical algorithms for the diagnosis of appendicitis. Proceedings of SCAMC, 686-690, 1991).

Using a different approach, Kamae and Greenes define a finite-state automaton model that simulates the discrete sensitivity analyses in a decision tree. An automaton model is a machine whose only devices are those for input and output, and a control. In Kamae's work, the finite-state-automaton model receives clinical findings as input, and returns as output the maximal expected utility for the best action among alternative therapeutic alternatives. The authors detail the relations among this automaton model, decision analysis, and clinical-flowchart algorithm. Using their approach, the automaton model links the clinical algorithm with the underlying probabilities and utilities of the decision analysis,

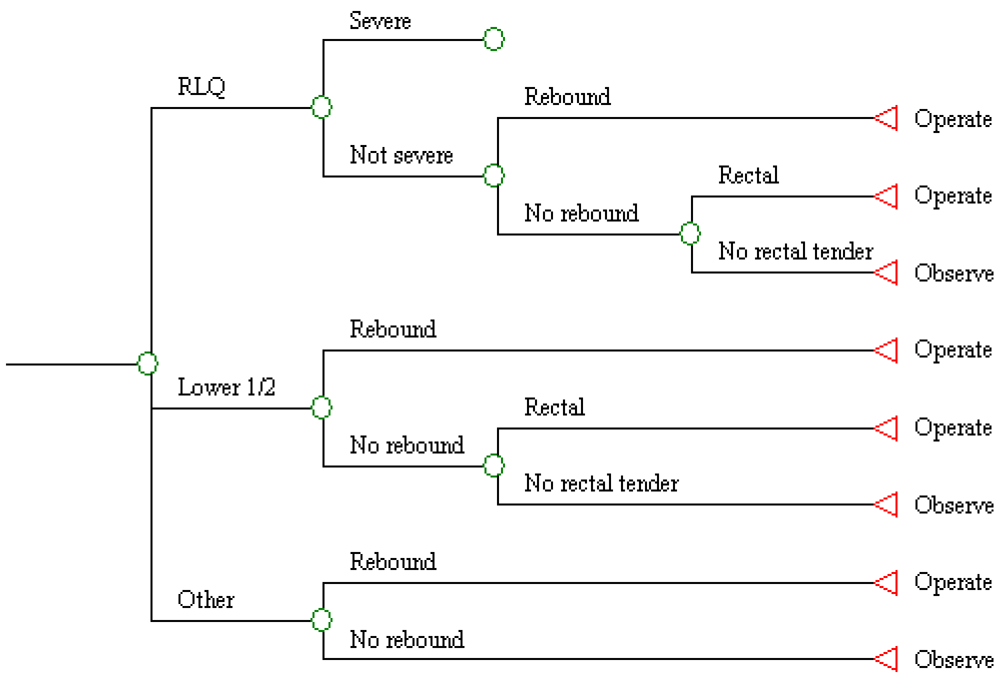


Figure 5.2. Clinical algorithm derived from the decision table in Figure 5.1. *(*Reprinted with permission from the Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care 1992 American Medical Informatics Association Published by Hanley & Belfus, Inc., Philadelphia, PA. *Source:* Adapted from RN Shiffman and RA Greenes, Use of augmented decision tables to convert probabilistic data into clinical algorithms for the diagnosis of appendicitis. Proceedings of SCAMC, 686-690, 1991).

and, through the recognition of equivalent clinical algorithms, permits the user to discover other equivalent or more optimal sequences of the clinical algorithm (Kamae and Greenes 1991).

Both of these approaches produce clinical-flowchart algorithms, although these flowcharts are not substantiated with the additional knowledge needed to make a CPG successful (e.g., CPG objective or evidence tables). Shiffman and colleagues do preserve the probability data in an augmented decision table to provide a means of documentation for the produced CPG. This documentation, however, is not sufficient to produce a high-quality CPG. Decision tables display the factors to be considered and the actions to be taken, yet the evidence on which the numbers are based and the clinical significance of these findings are not readily available to the user (Holland 1975). ALCHEMIST provides a corresponding flowchart representation of the recommendations, using the SMDM standard format, and includes additional information (e.g., targeted health problem, modeling perspective) that is required for a successful CPG.

5.4 ALCHEMIST mapping algorithm

In this section, I describe ALCHEMIST's algorithm for DM-to-CPG mapping. This mapping allows ALCHEMIST to create a CPG from a DM, and to maintain the underlying DM for future tailoring and updating. Figure 5.3 is an overview of the architecture of ALCHEMIST's algorithm for DM-to-CPG mapping.

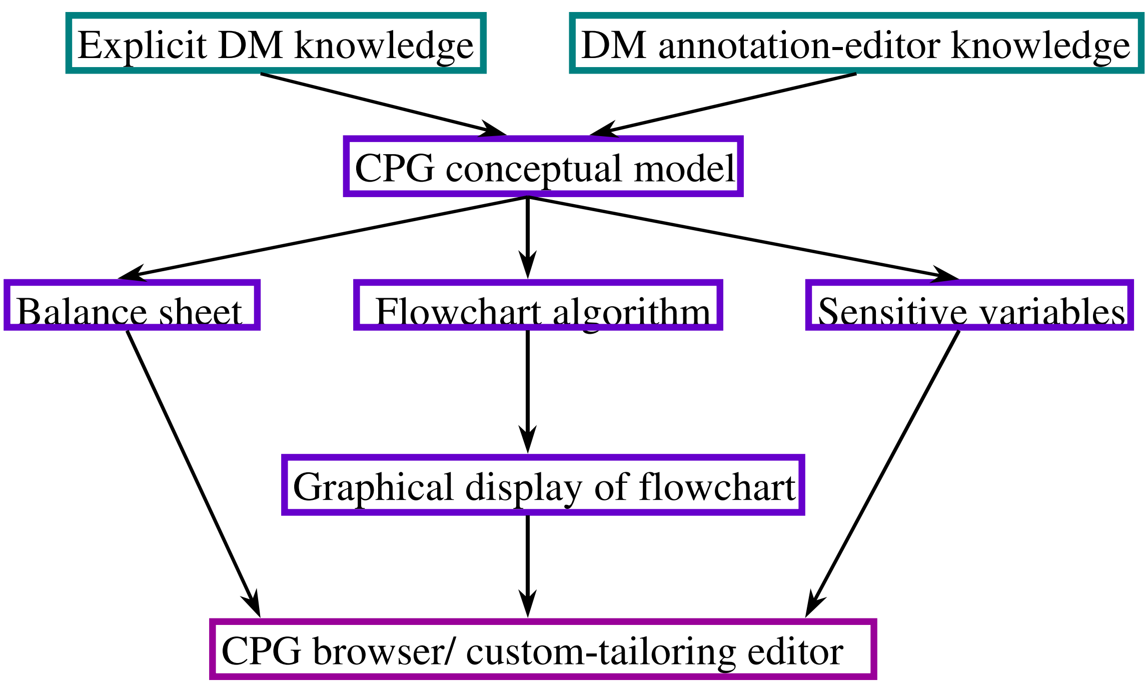


Figure 5.3. Overview of the mapping-algorithm architecture. ALCHEMIST obtains information from the DM and from the DM annotation editor, and uses the mapping algorithm to instantiate the CPG conceptual model. ALCHEMIST then creates a balance sheet, flowchart algorithm, and determines the sensitive variables. This information is combined and used to produce the CPG browser and custom-tailoring editor.

This algorithm is implemented in a Windows environment using Windows common gateway interface (WinCGI) scripts written in Visual Basic (Microsoft Visual Basic 4.0). ALCHEMIST uses object linking and embedding (OLE) commands to interact with the underlying decision model, and other existing software packages (e.g., Excel spreadsheets, and Visio flowcharts).

After ALCHEMIST obtains the knowledge explicit in the DM[1], and the decision analyst has entered any additional information using the DM annotation editor, ALCHEMIST performs the following steps to produce a CPG:

1. Map explicit DM knowledge onto the CPG conceptual model
2. Map DM annotation-editor knowledge onto the CPG conceptual model
3. Prepare a balance sheet of the benefits, harms, and costs
4. Calculate the recommended flowchart algorithm using the best-estimate input values
5. Create a tornado diagram of the change in expected utility
6. Determine to which variables the CPG is sensitive
7. Create the CPG browser and custom-tailoring editor

I describe these six steps in Section 5.4.1 to Section 5.4.7. In Chapter 7, I step through this process with the three example DMs

1. The DM is implemented using a **forward-star** representation. An array FirstLink holds the index of the first branch leaving each node. Another array, ToNode, tells to which other node the branch points. A sentinel entry at the end of the FirstLink array points just beyond the last entry in the ToNode array. The forward star representation allows the algorithm to perform operations quickly and easily on the links that leave a particular node

5.4.1 Map explicit DM knowledge onto the CPG conceptual model

ALCHEMIST obtains a large portion of the needed information directly from the DM. This information includes knowledge regarding the available options (the unique children of all the DM decision nodes), the possible outcomes (the terminal nodes), and the evidence (the variables and best-estimate values).

5.4.2 Map the DM annotation-editor knowledge onto the CPG conceptual model

ALCHEMIST obtains many of the remaining required CPG elements directly from the DM annotation editor. These elements include such information as the CPG objective, validation methods, sponsors, definitions, sources, evidence tables, and sensitivity-analysis ranges.

5.4.3 Prepare a balance sheet of the benefits, harms, and costs

Eddy proposes that guideline developers use a **balance sheet** to represent the benefits, harms, and costs of a decision problem. A balance sheet lists the beneficial and harmful outcomes and their magnitudes. The balance sheet should contain all the information that a patient or practitioner would need to make an informed decision about the intervention (Figure 5.4). Through the use of the balance sheet, a guideline developer attempts to provide the CPG user with an accurate understanding of the important consequences of the different options (Eddy 1990a). ALCHEMIST prepares a balance sheet by producing a table with a column for the expected outcome (i.e., benefits, harms, and financial costs) and additional columns for possible strategies. The strategies that ALCHEMIST represents in its balance sheet are the optimal base-case strategy, and any other competing strategies that the guideline user identifies. ALCHEMIST represents three outcomes: costs, life expectancy, and quality-adjusted life expectancy. ALCHEMIST updates the outcomes of the strategies in the balance sheet whenever a guideline user submits new input-variable values. Currently, ALCHEMIST is not able to calculate automatically intermediate outcomes (e.g., lung tumors averted, reduced cases of tuberculosis). Such calculations would require additional information from the decision analyst and are not part of this work.

TABLE 7.1 Benefits, Harms, and Costs of One Colorectal Cancer Screening Strategy*

Outcomes	No Screening	FOBT Every Year and Scope Every 3 y	Difference Caused by Screening	Range of Uncertainty About Difference
Benefits				
Probability of getting colorectal cancer	10.3% (103/1000)	7.3% (73/1000)	−3.0% (30/1000)	0%-6%
Probability of dying of colorectal cancer	5.3% (53/1000)	2.9% (29/1000)	−2.4% (24/1000)	0%-5%
Probability of harboring a hidden cancer†	0.1% (10/10 000)	0.03% (3/10 000)	−0.07% (7/10 000)	0.05%-0.09%
Harms				
Probability of a false-positive FOBT	0%	40% (400/1000)	+40% (400/1000)	20%-60%
Probability of perforation‡	0%	0.3% (3/1000)	+0.3% (3/1000)	0.1%-1.0%
Inconvenience/anxiety/discomfort				
FOBT, No. of tests	0	26	26	. . .
Scope, No. of tests	0	9	9	. . .
Financial costs, $§				
Screening	0	643	643	. . .
Treatment	1155	1106	−49	. . .
Net	1155	1749	594	. . .

*FOBT indicates fecal occult blood test; and scope, 60-cm flexible sigmoidoscope.
†This is the probability that the patient has a cancer that will develop signs or symptoms in the coming year.
‡This is the probability of perforation due to workup of false-positive fecal occult blood test or sigmoidoscopy.
§These are present values, discounted at 5%. A $4 fecal occult blood test and $135 60-cm flexible sigmoidoscopy are assumed. Costs of screening and treatment vary widely in different settings.

Figure 5.4. Example balance sheet of the benefits, harms, and costs for a colectoral cancer screening strategy. This table indicates the benefits, harms, and costs of one colectoral cancer screening strategy. In the left hand column is the different outcomes that are computed for the selected strategies. The following two columns list these outcomes for two strategies. The next column indicates the difference between the two strategies and the final column quantifies the uncertainty (Reprinted with permission *Source*: Eddy DM. Clinical Decision Making: From Theory to Practice. Comparing Benefits and Harms: The Balance Sheet. *JAMA* 263(18):2493-2505 (Copyright 1990, American Medical Association)).

5.4.4 Calculate the recommended flowchart algorithm using the best-estimate input values

Using the algorithm described in Section 5.5, ALCHEMIST creates the recommended flowchart representation. ALCHEMIST also determines the expected health outcome of this strategy. When a guideline user custom tailors a CPG for her specific patient population, ALCHEMIST compares this new CPG to the base-case CPG; if the CPG has changed, ALCHEMIST notes the sensitivity of the guideline to this new population. ALCHEMIST then displays the flowchart graphically for the user. The flowchart representation adheres to the SMDM standard for clinical algorithms that is described in Section 2.3.4.

5.4.5 Create a tornado diagram of the expected utility

A **tornado diagram** is graphical representation of the change in the expected utility of a given strategy as each variable value is varied along its sensitivity-analysis range (Howard 1988). A tornado diagram allows the user to compare one-way sensitivity analysis for many input variables at once (Clemen 1996). An example tornado diagram is shown in Figure 5.5. To create the needed tornado diagram, ALCHEMIST evaluates the DM for the low and high value of each variable (defined by the decision analyst in the DM annotation editor). Then, ALCHEMIST graphs the change that these new variable values produced in the base-case strategy's expected utility and sorts these changes to identify those variables that have the greatest effect on the expected utility. ALCHEMIST displays this tornado diagram for the guideline user as part of the CPG browser and custom-tailoring editor. Note, however, that the tornado diagram documents the change in the expected utility; it does not document the change in the recommended algorithm. Therefore, although a variable may cause a large change in the expected utility, the corresponding CPG algorithm can remain unchanged.

5.4.6 Determine to which variables the CPG is sensitive

Using the values of the high and low sensitivity-analysis ranges defined by the decision analyst, ALCHEMIST performs a one-way sensitivity analysis for each of the variables. ALCHEMIST then calculates the flowchart representation and expected health benefit (as described in Section 5.5). ALCHEMIST compares the resulting flowchart representation to the best-estimate flowchart representation. If the flowcharts differ, then ALCHEMIST marks this variable as sensitive. When ALCHEMIST creates the CPG browser, it lists all sensitive

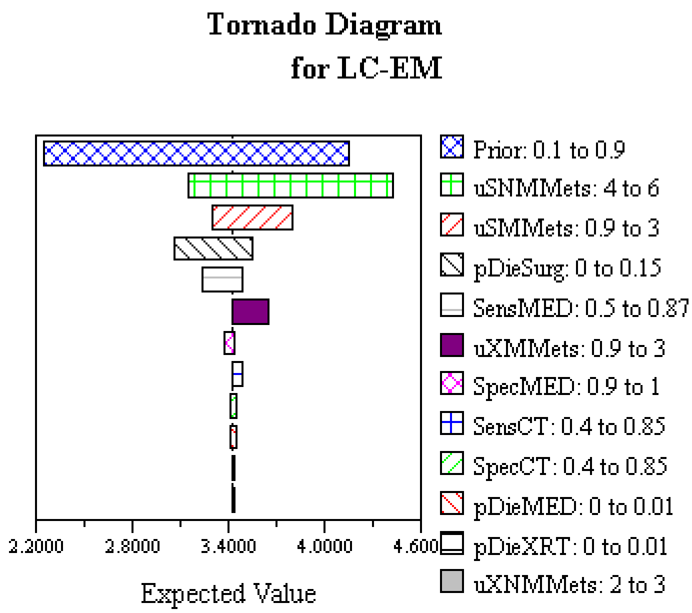


Figure 5.5. Example tornado diagram for the LC-EM decision model. A tornado diagram compares one-way sensitivity analyses for many input variables at once. Those variables that — within a given sensitivity analysis range — affect the expected outcome the greatest are listed at the top of the diagram (e.g., prior probability of mediastinal metastases), while those with little effect on the model's expected utility are at the bottom (e.g., life expectancy after radiation therapy).

variables to indicate to the guideline user that the resulting flowchart algorithm is sensitive to them within the given ranges.

5.4.7 Create the CPG browser and custom-tailoring editor

After ALCHEMIST has completed the DM-to-CPG knowledge mapping, created the best-estimate flowchart representation and tornado diagram, and performed the necessary sensitivity analyses, it produces the HTML code that it needs to create a web interface for the CPG browser and custom-tailoring editor. I discuss the CPG custom-tailoring editor in Chapter 6 and the CPG browser in Chapter 7.

5.5 Creation of the CPG flowchart algorithm

In this section, I describe the algorithm that ALCHEMIST uses to create the CPG flowchart representation of the DM recommendations, detailing how ALCHEMIST creates the flowchart using the best-estimate values. ALCHEMIST must also produce flowcharts for the purpose of sensitivity analyses, and later for a guideline user who wants to update or custom tailor the CPG. In all three scenarios, ALCHEMIST uses this same algorithm; in the latter two, it merely changes — in the DM — the variable values chosen for the sensitivity analysis or by the guideline user.

This algorithm assumes that the decision analyst has created a valid DM (defined in Section 3.3.1) and has entered the decision tree into the ALCHEMIST system. ALCHEMIST has created the DM annotation editor and has obtained from the decision analyst the missing DM knowledge. ALCHEMIST then can start the DM-to-CPG mapping, and can create the accompanying clinical-flowchart representation.

ALCHEMIST initializes an empty flowchart. In my current implementation, I restrict the DM to decision trees that have binary chance nodes. I maintain this restriction to enable the resulting CPG flowchart to adhere to the SMDM standard that requires each decision point to have only yes and no options, thereby requiring all chance nodes in the DM to be binary. After initializing the flowchart, ALCHEMIST adds to it the decision-tree root node. In ALCHEMIST's flowchart representation, this step corresponds to creating a clinical-state box (a rounded rectangle) that has as its label the defined CPG clinical state listed. The root node is labeled the **CurrentNode**. ALCHEMIST creates the flowchart representation by determining the node type of the CurrentNode and performs the actions I detail in Section 5.5.1 through 5.5.5 until there are no more nodes in the DM for ALCHEMIST to traverse.

5.5.1 Decision node

If the CurrentNode is a **decision node**, then ALCHEMIST evaluates the node to determine the expected utility of the node's children. ALCHEMIST saves the results of this evaluation, and parses these results to determine the optimal path to follow. The child with the greatest expected utility is designated the **OptimalNode**. This OptimalNode is labeled as an action box (i.e., a rectangle in the graphical display) and is added to the flowchart. The nonoptimal children of the decision node are pruned, and thus are no longer evaluated as part of the possible flowchart. ALCHEMIST maintains a path string that keeps track of the current path through the DM. This path string enables accurate evaluation of the decision nodes (because node names are not required to be unique, although paths to nodes are so required), and also allows a final record of the paths in the flowchart. ALCHEMIST then advances to the next node, which it names CurrentNode. For example, in the LC-EM, when the CurrentNode is the CTpos, reflecting that the computed tomography (CT) examination showed at least one lymph node with a short-axis diameter greater than 10 mm, ALCHEMIST determines the names of CTpos's children (i.e., MED, and NoMED), evaluates the expected utility of these child nodes, and determines the optimal path to follow (i.e., MED). Figure 5.6 shows the part of the CPG that results from this evaluation.

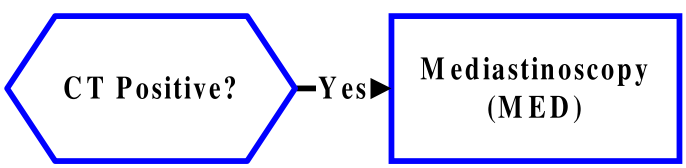


Figure 5.6. Flowchart representation that results from the evaluation of a decision node.

In the LC-EM, when ALCHEMIST encounters the CTpos decision node, the optimal strategy is to perform mediastinoscopy. ALCHEMIST represents this logic in the flowchart by the flow of the CTpos decision box into the mediastinoscopy action box.

5.5.2 Chance node

If the CurrentNode is a **chance node**, ALCHEMIST determines the children of this node and adds these new nodes to the flowchart for evaluation. The graphical flowchart display requires that ALCHEMIST designate one of these nodes as the hypothesis and the other child as the null hypothesis. For example, in the LC-EM, when it visits the CT chance node, ALCHEMIST determines that CT's children are CTpos and CTneg. To create a SMDM-acceptable flowchart, ALCHEMIST must define the CTpos variable, defined as positive CT by the decision analyst, to be the label used in the hexagon decision box, and to correspond to the yes arrow to the right of this box. The CTneg outcome would follow the no arrow straight down from the decision box. ALCHEMIST then advances to the next node and names this new node the CurrentNode. Figure 5.7 shows the corresponding flowchart representation that results from ALCHEMIST's evaluation of the CT chance node.

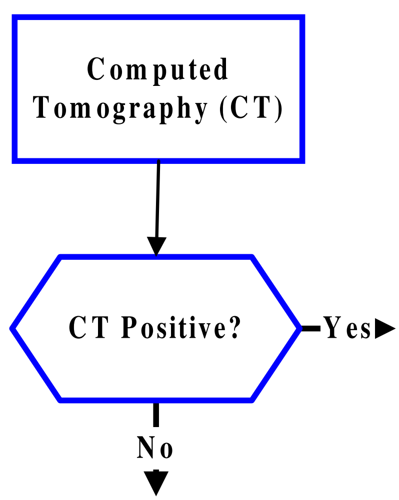


Figure 5.7. Flowchart representation that results from the evaluation of a chance node. In the LC-EM, when ALCHEMIST encounters the CT chance node, the possible events are a positive CT or a negative CT. ALCHEMIST represents these possibilities in the flowchart by creating the CTpositive decision box with yes and no arrows.

5.5.3 Terminal node

A **terminal node** represents DM outcomes, such as death or treatment with or without mediastinal metastases. If CurrentNode is a terminal node, then ALCHEMIST advances to the next untraversed path and names the first node the CurrentNode.

5.5.4 Nadir node

A **nadir node** is similar to a decision node, except that, when it is evaluated, it takes the minimum, rather than the maximum, value of its branches. If CurrentNode is a nadir node, then ALCHEMIST uses the same logic as that it used for a decision node, although the branch with the lowest expected utility, rather than that with the highest expected utility, is chosen.

5.5.5 Markov node

If the CurrentNode is a **Markov node**, then ALCHEMIST considers this node analogous to a chance node in that the node has several branches, each with an associated probability. Yet this node is different from a normal chance node in that the expected utility of a Markov node is equal to the value returned when a Markov simulation specified by the structure of the node is run. The Decision Maker implementation of Markov nodes requires that there be no decision, nadir, or other Markov nodes downstream from the Markov node, and that all terminal nodes downstream from a Markov node have the names of Markov states as their utilities. These restrictions on Markov nodes simplify the required mapping algorithm, allowing ALCHEMIST to consider the entire subtree represented by the Markov node and the associated branches as a special kind of terminal node, with the value of the Markov process analogous to a utility (Sonnenberg and Pauker 1997). Relaxation of these restrictions produces DMs that common decision-modeling software cannot evaluate; therefore, this extension is not within the scope of my thesis.

Once ALCHEMIST has traversed the entire DM tree, it has created the necessary flowchart logic and graphical requirements. Figure 5.8 provides a psuedocode summary of the flowchart-representation algorithm.

```
Initialize empty flowchart
Add root node as clinical state box
Do While (there are available nodes)
      Look at nodetype
            If decision node Then
                  Evaluate node
                  Designate child with greatest expected outcome to be
                  optimal child
                  Add optimal child to flowchart as an action box
                  Prune nonoptimal child and descendants from tree
                  Go to next node
            Else

            If chance node Then
                  Determine children of node
                  Designate one as label of decision box with yes arrow
                  Designate other as no arrow
                  Go to next node
            Else

            If terminal node Then
                  Go to next node
            Else

            If Nadir node Then
                  Evaluate node
                  Designate child with lowest expected outcome to be
                  optimal child
                  Add optimal child to flowchart as an action box
                  Prune nonoptimal child and descendants from tree
                  Go to next node
            Else

            If Markov node Then
                  Evaluate Markov process
                  Label Markov node as Terminal node with results of
                    Markov process as its expected utility
                  Go to next node
            Else
            End If
End Loop
Display flowchart representation
```

Figure 5.8. Psuedocode for ALCHEMIST to create the CPG flowchart representation.

5.6 Creation of the CPG flowchart algorithm for dual-utility models

If the decision analyst loads into the ALCHEMIST system a dual-utility model, ALCHEMIST assumes (unless the decision analyst explicitly states otherwise in the DM annotation editor) that the first argument in a utility expression represents some measure of cost, and that the second argument represents some measure of effectiveness. Because a dual-utility model requires a tradeoff between the two utility measure, the decision analyst enters a cost-effectiveness threshold that ALCHEMIST will use to determine the optimal strategy. Following the most cost-effective branch at each individual decision node may not lead to the overall most cost-effective strategy. Therefore, I use a different method to create the CPG flowchart for dual-utility models. Given a dual-utility model, ALCHEMIST determines all possible paths through the DM; it then presents these paths to the decision analyst in the DM annotation editor. The decision analyst has the opportunity to prune from this set of paths any strategies that are clinically irrelevant. ALCHEMIST converts this reduced set of strategies to normal form (i.e., a single large decision node at the beginning of the tree). It then calculates the cost and effectiveness of each of these strategies, ranks the strategies from least to most effective, and calculates the marginal cost-effectiveness (MCE) ratios (checking for dominance and extended dominance of strategies). ALCHEMIST selects as the optimal strategy the overall strategy that has the lowest MCE ratio (and that does not exhibit extended dominance); it displays the optimal strategy's corresponding flowchart for the guideline user.

5.7 Limitations of my approach

Currently, ALCHEMIST performs only one-way sensitivity analyses of each variable in the DM. Perhaps two- or three-way sensitivity analyses, or another method of measuring the uncertainty in DMs (e.g., probabilistic or bootstrap sensitivity analyses), could be more appropriate. Use of these other methods however, is beyond the scope of my thesis. As I mentioned in Section 5.4.5, the tornado diagram that ALCHEMIST creates as part of the CPG reflects the effects of the variables on the expected utility of the base-case strategy, rather than on the algorithm itself. An extension of my work would be to represent graphically the amount of change in the guideline over the range of variable values. Such as analysis would require a method for quantifying the difference between two algorithms.

The computation of the initial CPG takes substantial time because ALCHEMIST needs to retrieve information from the DM conceptual model and the DM annotation editor, and to perform all the required sensitivity analyses and flowchart comparisons. This initial work, however, occurs as part of the CPG development and thus does not affect the guideline end user. Any future custom tailoring or updating of the model requires only one recalculation, and, therefore, accounts for a fraction of the computation.

5.8 Summary

In this chapter, I justified my assumption that it is beneficial to create CPGs from normative DMs. I then described the DM-to-CPG mapping algorithm, detailing ALCHEMIST's current approach and limitations. In Chapter 6, I discuss why it is important that a guideline user be able to update and custom tailor a CPG, and describe ALCHEMIST's method for addressing this need.

Chapter 6

Custom Tailoring and Updating

In this chapter, I discuss why ALCHEMIST's abilities to update and custom tailor a CPG allows a guideline user to create a successful CPG (Section 6.1 through 6.3). In Section 6.4, I discuss current work on the custom tailoring of CPGs. In Section 6.5, I discuss ALCHEMIST's custom-tailoring technique: its abilities and limitations. I conclude this chapter with an examination of the potential legal issues involved when a guideline user is allowed to update a CPG dynamically (Section 6.6).

6.1 Definition of CPG custom tailoring and updating

In my research, I identify two main problems with current CPG implementation projects: their lack of ability to **custom-tailor** a CPG, and the absence of a method to **update** the produced CPGs. Immense resources and work go into CPG development, yet methods of maintaining the validity of these guidelines over time have not been well developed. Medical knowledge and possible treatments are not static; therefore, guideline developers need to be able to update CPGs to take into account medical advances. Furthermore, guideline developers produce valid evidence-based CPGs using data from sources such as randomized clinical trials or meta-analyses; therefore, these CPGs then reflect global patient preferences, costs, and transition probabilities. A guideline user needs to be able custom tailor such a global CPG to her specific patient population or site. The problem, therefore, is how to produce a valid guideline that can be adopted widely, but also can to be tailored such that it describes important differences in patient populations and practice environments (Owens and Nease 1997)?

In my research, instead of developers producing individualized guidelines for each subpopulation or site, ALCHEMIST uses DMs and their supporting evidence to produce a global guideline, and then offers site-specific custom tailoring and updating through its CPG custom-tailoring editor. I distinguish between ALCHEMIST's custom-tailoring and updating abilities in several ways. ALCHEMIST treats CPG custom tailoring as a temporary change to a produced CPG. If a guideline user wants to custom tailor a given CPG to reflect her patient population, then she can enter populations-specific information into the CPG custom-tailoring editor, and ALCHEMIST will produce a tailored CPG. This tailoring of the CPG, however, does not produce a permanent change in the underlying DM. To update a CPG, ALCHEMIST makes a permanent change to the underlying DM; therefore, CPG updating requires a peer-review process.

6.2 Importance of custom tailoring and updating

Guideline developers hope that their CPGs, after implementation, will reduce inappropriate practices and improve the efficiency and quality of health care. In this section, I discuss two problems that impede the efficacy of current CPGs in the clinical environment: the need for custom tailoring (Section 6.2.1), and the need for updating (Section 6.2.2).

6.2.1 Need for custom tailoring

Guideline users are concerned that globally developed CPGs may not reflect important characteristics of local practice settings and may be insensitive to the needs and preferences of an individual patient (Fletcher and Fletcher 1990, Woolf 1998). The benefits of global CPGs are limited because the latter recommend actions for a large collection of heterogeneous patients, and each patient will have slightly different preferences and outcomes (Eddy 1990d). For example, Owens and Nease studied the effect of custom tailoring a guideline for HIV screening for specific patient populations (Owens and Nease 1997). The authors assumed a cost-effectiveness threshold of $50,000 per quality-adjusted life year (QALY). They produced a global guideline using the average seroprevalence among 26 hospitals that participated in the CDC Sentinel Hospital Study (St. Louis et al. 1990). The authors used the average seroprevalence to compute the cost-effectiveness of HIV screening as compared to no screening. Because this cost effectiveness was greater than $50,000 per QALY, the corresponding CPG recommended that clinicians not screen for HIV. However, the seroprevalence of the different patient populations studied varied from 0.1 to 7.8 percent. When the authors tailored the DM to reflect these site-specific prevalences, the cost effectiveness of screening ranged from $182,400 to $42,300 per QALY. Figure 6.1 shows how the recommended screening policy changes when a guideline user tailors the CPG to a specific site (Owens and Nease 1997).

The example of HIV screening demonstrates how the optimal CPG can differ across specific sites. This CPG can also differ within a given site's population if this population is heterogeneous and the guideline is sensitive to this heterogeneity. For example, McNeil and colleagues interviewed patients who had operable lung cancer about attitudes toward varying periods and quality of survival. McNeil and colleagues combined the results of their interviews with published survival data, and determined that the patient preferences influenced the preferred treatment plan (i.e., thoracotomy or radiation therapy) (McNeil et

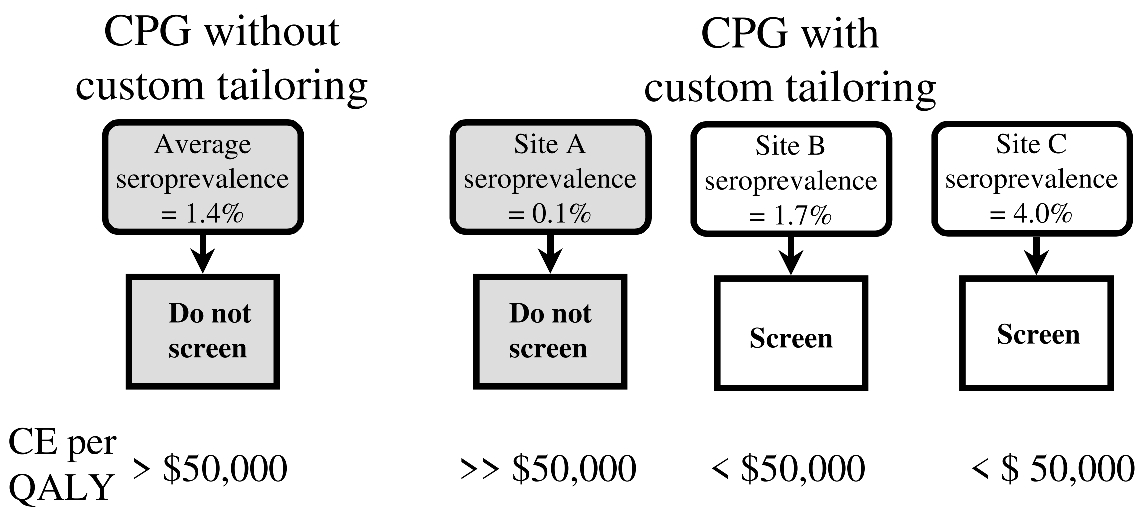


Figure 6.1. Effect of custom tailoring HIV screening guidelines for site-specific seroprevalence. The single CPG on the left is the recommended HIV screening policy if guideline developers used the average seroprevalence. The three CPGs on the right show how the recommended strategy can change with the seroprevalence (Owens and Nease 1997).

al. 1978). These studies emphasize the importance of developers combining patient- or site-specific evidence with objective clinical data when they create CPGs.

6.2.2 Need for updating

Because of the clinical complexity of and uncertainty associated with many of the topics that CPGs address, development of a new CPG can take up to 3.5 years to complete and can cost more than $1 million (US Congress Office of Technology Assessment 1994). It is therefore extremely important that a produced CPG be used and maintained. An example of the need for CPG updating is the development and subsequent retraction of the AHCPR guideline for the evaluation and management of early HIV infection (Figure 6.2). This CPG took over 2 years and $1 million to produce; it was published in 1994, and, by 1995, its recommendations were out of date. It is no longer distributed by the AHCPR (El-Sadr et al. January 1994). This example accentuates the clear need for guideline developers to produce CPGs that can be maintained as the clinical evidence evolves.

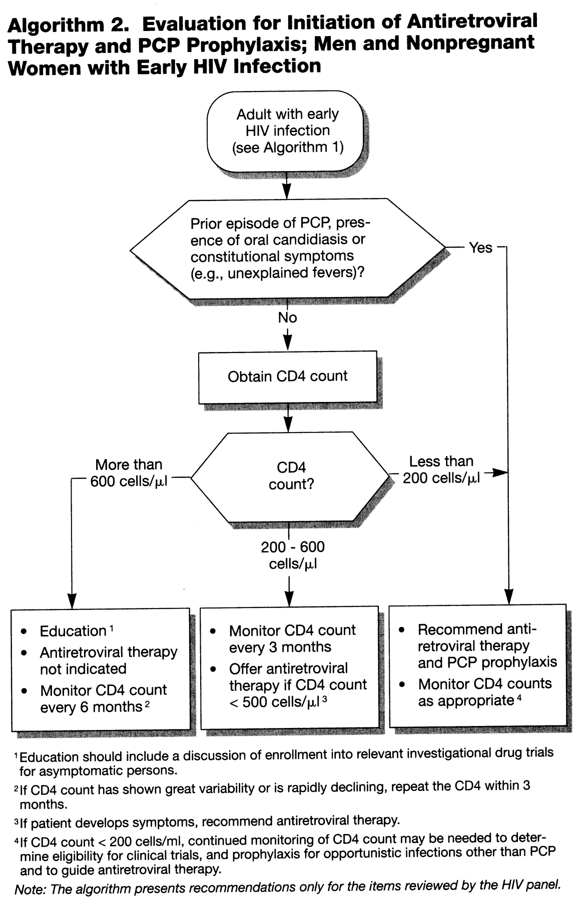


Figure 6.2. CPG for the AHCPR guideline for the evaluation and management of early HIV infection. (*Source:* El-Sadr W, Oleske JM, Agins BD, et al. Managing Early HIV Infection: Quick Reference Guide for Clinicians. AHCPR Publication No 94-0573. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, January 1994).

ALCHEMIST is able to produce a global CPG that is applicable broadly, yet it allows users to tailor and update the CPG — given certain restrictions — to describe the relevant and current patient population or local institution costs and practices.

6.3 Need for a global CPG

If ALCHEMIST is able to create a custom-tailored CPG given specific patient data, why not create an individualized CPG for each patient who requires care? Although this proposition may seem attractive, accepting it requires that guideline users invest huge resources. Gathering the necessary data for each patient to create a new CPG is a time-consuming, expensive, and often-infeasible task. Although a custom CPG may produce a greater expected health benefit, the underlying DM — and, therefore, the resulting CPG — normally is not sensitive to every input variable. Therefore, obtaining patient-specific information on each variable would not be cost-effective methods for guideline development or for use of clinical resources. Instead, using the highest-quality evidence available, ALCHEMIST creates a global CPG, calculates the sensitivity (within a given range) of the DM variables, and displays for the guideline user a tornado diagram and list of the variables to which the CPG is most sensitive; thus, it tells the user which variables in the CPG would be most valuable to custom tailor.

A global CPG allows ALCHEMIST to custom tailor the generated CPG, and to indicate to the user those variables that may influence the generated CPG and, therefore, may be worth tailoring to a specific patient population. ALCHEMIST however, reuses the underlying DM and much of the CPG and previously computed flowchart algorithm, thereby conserving resources.

6.4 Previous work on custom tailoring of CPGs

Despite the frequent criticism of CPGs that they do not capture important differences in clinical situations, there is currently no formal method for guidelines developers or users to assess the need for custom tailoring, or to incorporate such custom-tailoring abilities into CPG development (Audet et al. 1990, Eddy 1990b, Eddy 1990c, Eddy 1990d, Fletcher and Fletcher 1990, Granata and Hillman 1998, Lomas et al. 1989, Mutnick and Szymusiak-Mutnick 1996, Owens and Nease 1997).

Owens and Nease developed a framework to allow a guideline developer to determine whether the cost of creating a site-specific guideline is justified based on the incremental health benefits obtained. They used cost-effectiveness value-of-information analyses to custom tailor guidelines for specific clinical practice settings (Nease and Owens 1994, Owens and Nease 1991a, Owens and Nease 1991b, Owens and Nease 1997, Owens et al. 1993).

Fridsma performed related work on making generic guidelines site specific (Fridmsa et al. 1996). He described a prototype system, CAMINO, that helps developers to create site-specific guidelines based on existing annotated generic guidelines. This generic CPG is annotated with information about the assumptions and intentions of the guideline developers, so that changes to the CPG can occur while the validity of the recommendations is maintained. Fridsma plans to incorporate into his work a site model that contains activities supported by the organization and a workflow manager that helps protocol users to organize and to coordinate the care of patients (Fridmsa et al. 1996). ALCHEMIST, by maintaining links to the underlying DM and to the information obtained from the DM annotation editor, also attempts to maintain the validity of the CPG while producing tailored evidence-based CPGs for specific populations.

As I described in Section 1.13, as a member of the CARD PORT decision-modeling group, I developed PORTAL, a web interface to the Decision Maker decision-modeling software. This interface allows the user to browse the base-case-variable input values used in the DM. The user can change any of the variable inputs, or can perform a cost-effectiveness analysis, sensitivity analysis, or simple foldback of the decision tree. The results of the chosen analysis, which Decision Maker performs using the new input values, are then presented to the user. This interface allows a decision analyst to perform decision analyses and to interrogate a DM from a remote site using diverse computing platforms.

6.5 ALCHEMIST's approach to custom tailoring and updating

When ALCHEMIST creates its initial CPG, it maintains all of the information in the DM or in the DM annotation editor. By maintaining this link to the DM framework, and by preserving the DM's information, ALCHEMIST is able to make changes to the underlying DM and to use this new DM to update or custom tailor the CPG. In this section, I describe ALCHEMIST's approach to custom tailoring and updating. I detail what attributes of the guideline can be tailored or updated, and how ALCHEMIST conserves the validity of the new CPG. I then give examples from the implementation of the CPG custom-tailoring editor.

6.5.1 Custom-tailoring options

ALCHEMIST allows the guideline user to custom tailor or update any of the DM input variables. These changes can be made in two ways: (1) the guideline user can change any or all of the best-estimate values given in the input-variables table, or (2) the guideline user can change a global CPG element (e.g., model perspective) that is linked to the model variables that define it. For example, if a guideline user changes the model perspective of a generated CPG from patient to societal, the new analysis will consider all significant health outcomes (e.g. disease transmission) and costs to everyone potentially affected by the intervention being given to a specific population. Therefore, ALCHEMIST presents the input-variables table with the cost and utility variables highlighted, and queries the user for these new input values.

After the guideline user has changed the input-variable values, ALCHEMIST verifies that these values are possible (i.e., that all probabilities still sum to 1) and do not violate any of the DM modeling assumptions. This process is described in Section 6.5.2

Currently, ALCHEMIST treats utility variables the same as any other input variable. A future extension of my work could be to incorporate research on computer-based utility assessment to aid the guideline user in determining her patient's utilities (Lenert et al. 1995, Nease et al. 1996, Sanders et al. 1994, Sumner et al. 1991). I discuss this extension of my work in Section 9.2.4.

My current implementation does not allow structural updating or custom tailoring of the underlying DM. Although it is possible to implement changes to the DM from a web interface, the requirements for ensuring that the structurally changed model is complete and that the resulting CPG is valid are complex and are beyond the scope of my thesis research.

6.5.2 Maintenance of modeling-assumption validity

ALCHEMIST custom tailors CPGs by allowing guideline developers to tailor and update input variables to describe the specific sites or patient populations. ALCHEMIST changes the underlying DM to these new variable inputs, and then produces a tailored CPG to present to the guideline user. This added flexibility, however, incurs the danger that ALCHEMIST will update or modify the original evidence-based global guideline in a manner that violates the underlying assumptions of the original DM or CPG, and, therefore, will produce CPG recommendations that are no longer clinically justified. In this section, I describe ALCHEMIST's approach to maintaining the four types of DM modeling assumptions and to avoiding their potential violation during CPG updating or tailoring.

6.5.2.1 Modeling perspective

DM modeling normally uses a societal, patient, health-care institution, or third-party-payer perspective. The chosen perspective influences the DM costs and utilities, thereby affecting the recommendation of the generated CPG. If a guideline user wants to tailor a CPG to reflect a different modeling perspective, she checks the modify perspective radio button (Figure 6.3). ALCHEMIST highlights those variables in the input table that may be affected (i.e., the costs and utilities); the user can change these variable inputs. Similarly, if a guideline user changes a cost or utility in the variables table, her reason may be simply that her institution uses a different cost, or may be that she wants to use an alternate perspective. ALCHEMIST attempts to differentiate these two possibilities by indicating to the guideline user that she has changed a cost or utility and asking her whether she wants to change the modeling perspective. If the guideline user answers yes, ALCHEMIST highlights those potentially influenced variables for further modification. Note that, when a guideline user changes the modeling perspective, the structure might change as well. ALCHEMIST does not allow such structural changes.

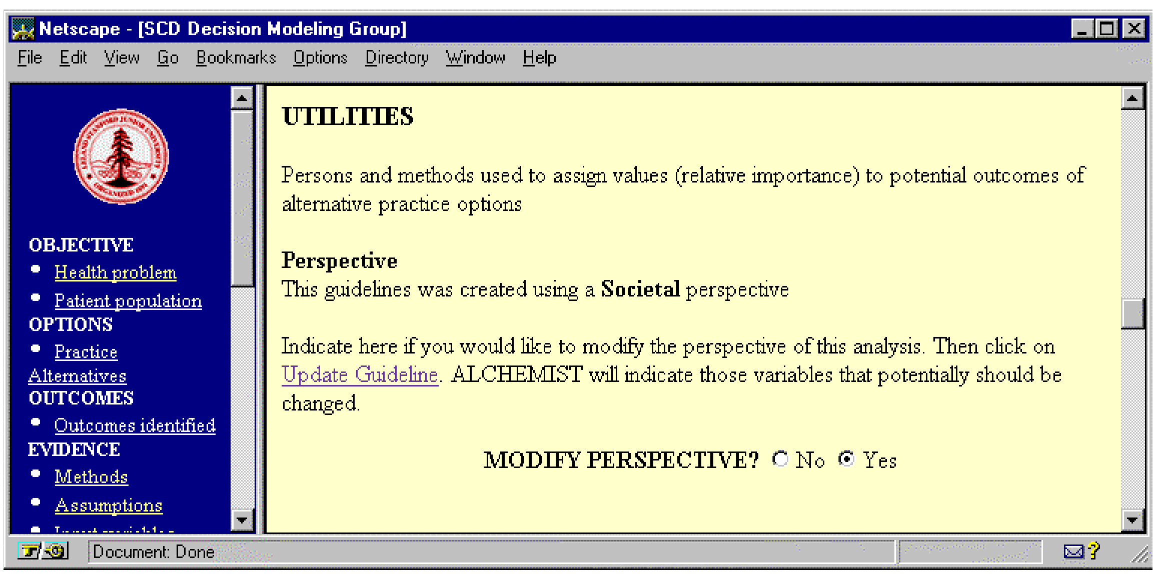


Figure 6.3. The CPG custom-tailoring editor. The guideline user has chosen to modify the modeling perspective of the guideline. ALCHEMIST then indicates to the guideline user which variables she may need to modify.

6.5.2.2 Best-estimate input values

ALCHEMIST creates the original CPG using the best-estimate input values and sensitivity ranges defined in the DM or in the DM annotation editor. A guideline user can change any of the input variables to a new value within these given ranges (Figure 1.11), and can then view the updated CPG. If a guideline developer changes a variable to a value outside of this range, ALCHEMIST produces a warning message by highlighting the variable in red and telling the guideline user that the variable value is outside the sensitivity-analysis range defined by the available evidence. If the value is a legal DM value (i.e., the model can still be analyzed), then the guideline user can override this warning, and ALCHEMIST will produce the tailored CPG — although it will continue to warn the user that the variable values are outside the evidence-based ranges.

6.5.2.3 Patient characteristics

Just as she can change the DM modeling perspective, the guideline user can change the patient characteristics of the target population of a CPG. The decision analyst indicated which variables should be linked to the patient population, therefore, the guideline user who wants to change the patient population simply clicks on the modify patient characteristics button. ALCHEMIST then highlights the appropriate variables that the decision analyst defined using the DM annotation editor.

6.5.2.4 Model structure

The final type of DM modeling assumption involves the model structure. These assumptions are listed explicitly for the guideline user; however, she is not able to change these assumptions, because doing so would require tailoring the DM structure — a process that is beyond the scope of my thesis work. For example, the LC-EM guideline assumes that computed tomography and mediastinoscopy are the only options for diagnosis. A guideline user, however, may want to make a structural change by adding a positron-emission tomography (PET) scan as an available diagnostic test. Such a change would require adding branches to the DM and adjusting existing probabilities (depending on the conditional dependence or independence of the existing diagnostic-test results with the PET-scan results). These changes are nontrivial and would require that an experienced decision-analytic team incorporate the needed changes correctly. Ideally, any structural change that a user wants to perform could be relayed to the guideline organization; if, as in this example, there is a new diagnostic technique available that should be incorporated into the CPG, the organization can change the underlying DM, after which the global CPG can be updated.

6.6 Legal considerations

The CPG custom-tailoring editor gives guideline users the ability to change the input values in the DM (and its corresponding CPG) to reflect their particular patient populations. This ability, however, carries the drawback that a user may violate the underlying assumptions of the CPG, or may perform analyses that are at odds with the best available evidence. To maintain the validity of the tailored or updated CPG, ALCHEMIST attempts not to violate any of the DM modeling assumptions. However, there is always the possibility that, even if all the assumptions defined explicitly by the DM and through the DM annotation editor are maintained, the generated CPG will violate an undefined CPG intention or modeling assumption. If a guideline user is to lessen the effect of this eventuality, all CPG updating will need to be validated by the developing organization, or by a peer-review process, before it can be included in the default CPG. Any temporary custom tailoring of the CPG, however, would not be subject to such peer review. ALCHEMIST reminds the guideline user that the generated CPG is to be used as a guideline, rather than as a standard, for patient management, and that she should always use her clinical judgment in accepting or rejecting its recommendations.

6.7 Summary

A successful CPG should be capable of being both tailored by guideline users to reflect their patient population, and updated by a guideline developer to take into account medical advances and new costs, clinical trials results or patient preferences. In this chapter, I discussed the importance of a guideline user being able to custom-tailor and update a CPG. I described ALCHEMIST's approach to custom tailoring and its method for maintaining the validity of the resulting CPG. In Chapter 7, I describe in detail my example DMs, and follow each DM from its decision-tree representation through to the generated CPG and to its tailored version.

Chapter 7

ALCHEMIST Examples

In this chapter, I describe the three example decision models (DMs) that I used throughout my thesis work (Section 7.1). In Section 7.2, I use one of these models to step through an example translation of a DM to a clinical-practice guideline (CPG), emphasizing the user interaction with the ALCHEMIST system. I end this chapter by documenting any differences in the DM-to-CPG translations for a cost-effectiveness DM (LC-CEM) and for a Markov model (SCD-MM) (Section 7.3 and 7.4 respectively).

7.1 Example DMs

I developed three example DMs for my research. These DMs are from two different medical domains, and represent a variety of DM and CPG attributes (sequential decisions, cost-effectiveness studies using dual-utility models, and time-dependent studies using Markov processes).

In this section, I describe the clinical rationale, structure, and distinguishing attributes of each model. I implemented all three models using the Decision Maker modeling software.

7.1.1 Lung-cancer effectiveness model (LC-EM)

The **lung-cancer effectiveness model (LC-EM)** is the base-case DM for my thesis work, and is the DM that I use to step through the DM-to-CPG translation in Section 7.2. I picked this DM as my base-case model because, although it is a simplified model and has only binary-branching points and a single-utility model, it contains sequential decisions and several underlying assumptions that allow ALCHEMIST to demonstrate the conceptual framework, the DM-to-CPG mapping, and the main attributes of the ALCHEMIST implementation. Figure 7.1 shows a schematic representation of the LC-EM.

This DM represents the optimal staging strategy for the mediastinum of patients who have non--small-cell lung cancer. The only outcome modeled is quality-adjusted life expectancy, although there are sequential decisions representing the numerous tests that can be used.

I adapted the LC-EM from a DM created by Owens and colleagues (Gould et al. 1997, Nease and Owens 1997, Owens and Nease 1991a, Owens et al. 1989). The DM portrays the case of a patient who has a known non–small-cell carcinoma of the lung. Preliminary workup reveals no evidence of distant metastases. If mediastinal metastases are absent, then the preferred treatment is thoracotomy, followed by lobectomy or pneumononectomy, which offers the patient a substantial survival advantage over alternative treatments. However, if mediastinal metastases are found to be present, thoracotomy is contraindicated, and radiation therapy is the preferred treatment, because thoracotomy subjects the patient to a risk of death but confers no known health benefit. There are several diagnostic tests available to assess involvement of the mediastinum. The version of the LC-EM that I used in this work considers the use of only computed tomography (CT) of the chest and mediastinoscopy.

The LC-EM assumes that clinicians use noninvasive tests before invasive tests; therefore, the first decision is whether to perform a CT. This decision is then followed by the decision of whether to perform mediastinoscopy (note that, if a CT is performed, the results of this test are available before the decision to perform the mediastinoscopy is made). The mortality rate for mediastinoscopy depends on the characteristics of the individual patient, but in general is much less than 1 percent; therefore, in the LC-EM, mediastinoscopy includes a small risk of death. The decision maker knows the results of both the CT and mediastinoscopy examinations before a treatment decision is made. Both treatments incur a risk of death, and the two have different associated quality-adjusted life expectancies based on the presence or absence of mediastinal metastases (Nease and Owens 1997).

7.1.2 Lung-cancer cost-effectiveness model (LC-CEM)

The **lung-cancer cost-effectiveness model (LC-CEM)** adds to the LC-EM the relevant costs associated with the tests (i.e., the cost of a CT examination or of mediastinoscopy) and with the treatments (i.e., the cost of a thoracotomy or of radiation therapy). ALCHEMIST evaluates this model on the basis of cost effectiveness, with effectiveness measured in length of quality-adjusted life. The model structure is the same as that shown in Figure 7.1, with the costs added to the appropriate nodes. The LC-CEM thus requires that ALCHEMIST represent a dual-utility model.

7.1.3 Sudden-cardiac-death PORT Markov model (SCD-MM)

The **sudden-cardiac-death PORT Markov model (SCD-MM)** is based on work I have done previously as member of the Cardiac Arrhythmia and Risk of Death Patient Outcomes Research Team (CARD PORT) (Owens et al. 1997a, Sanders et al. 1996, Sanders et al. 1995). The SCD-MM models the cost effectiveness of using an implantable cardioverter defibrillator (ICD) as compared to that of using amiodarone (a leading antiarrhythmic drug) for patients who are at high risk of sudden cardiac death (SCD). For the patient populations modeled, risk of SCD is a chronic condition; therefore, a patient is at risk

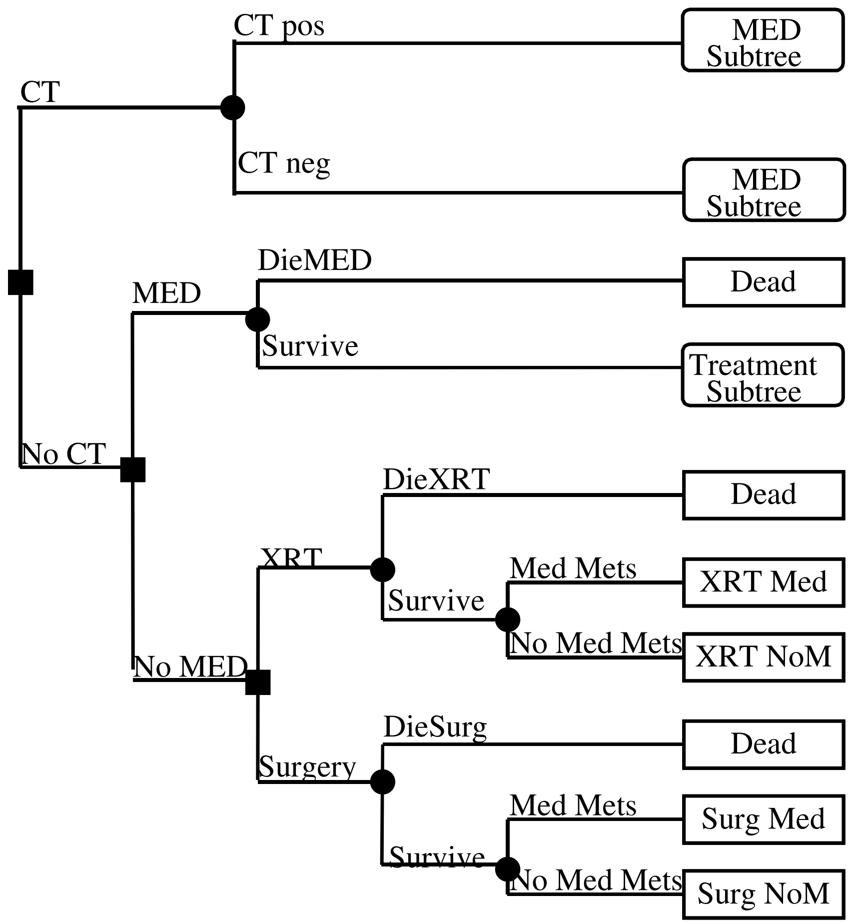


Figure 7.1. Schematic representation of the LC-EM. The square nodes represent decision nodes, the circles represent chance nodes, and the blackened rectangles represent terminal nodes. CT = computed tomography, MED = mediastinoscopy, XRT = radiation therapy, Surgery = thoracotomy, MedMets = mediastinal metastases.

each month for antiarrhythmic death, nonarrhythmic death, or noncardiac death, or, if the patient receives amiodarone, for death from drug toxicity. Patients who have an ICD are also at risk for perioperative death. This continuous risk is represented in the SCD-MM by Markov nodes. The SCD-MM, therefore, allows me to demonstrate ALCHEMIST's ability to represent time dependencies and recurrent or repeating events.

The SCD-MM tracks a hypothetical cohort of patients over time until eventual death. Each patient receives either an ICD or amiodarone. If a patient has ventricular fibrillation or ventricular tachycardia, the patient dies, survives with neurological impairment, or survives without neurologic impairment. Patients who are treated with amiodarone are at risk for acute drug toxicity. The model includes a decrement in quality of life for patients who survived an arrhythmic event but have neurologic sequelae (Figure 7.2) (Owens et al. 1997a).

7.2 LC-EM example

In this section, I step through the DM-to-CPG translation for the LC-EM described in Section 7.1.1. I emphasize how the user would interact with the ALCHEMIST system.

In this example, an organization identifies as a clinical-decision problem worthy of guideline development the staging strategy for mediastinal metastases in patients who have non–small-cell lung cancer. The organization employs experts in decision analysis and meta-analysis to create a DM that represents the relevant alternatives, outcomes, evidence, assumptions, and knowledge. After the decision-analysis team has built the DM, a decision analyst enters the DM into the ALCHEMIST system. ALCHEMIST attempts to map the DM to the DM-to-CPG conceptual framework. ALCHEMIST creates dynamically the DM annotation editor to query the decision analyst for the missing information that it needs to produce the corresponding CPG representation. After the decision analyst has completed the DM annotation editor, he submits this information, and ALCHEMIST creates the initial CPG, the flowchart representation, and the accompanying CPG browser and custom-tailoring editor.

The CPG browser, with its initial CPG, is subject to internal and external peer review. The developing organization then distributes the accepted CPG to the health-care community and adds it to a guideline-repository web page. Local guideline developers can explore the global guideline, examine its evidence and recommendations, and specify site- or patient-specific input values to produce an updated or tailored CPG. In Section 7.2.1 through Section 7.2.5, I detail this translation process using the LC-EM.

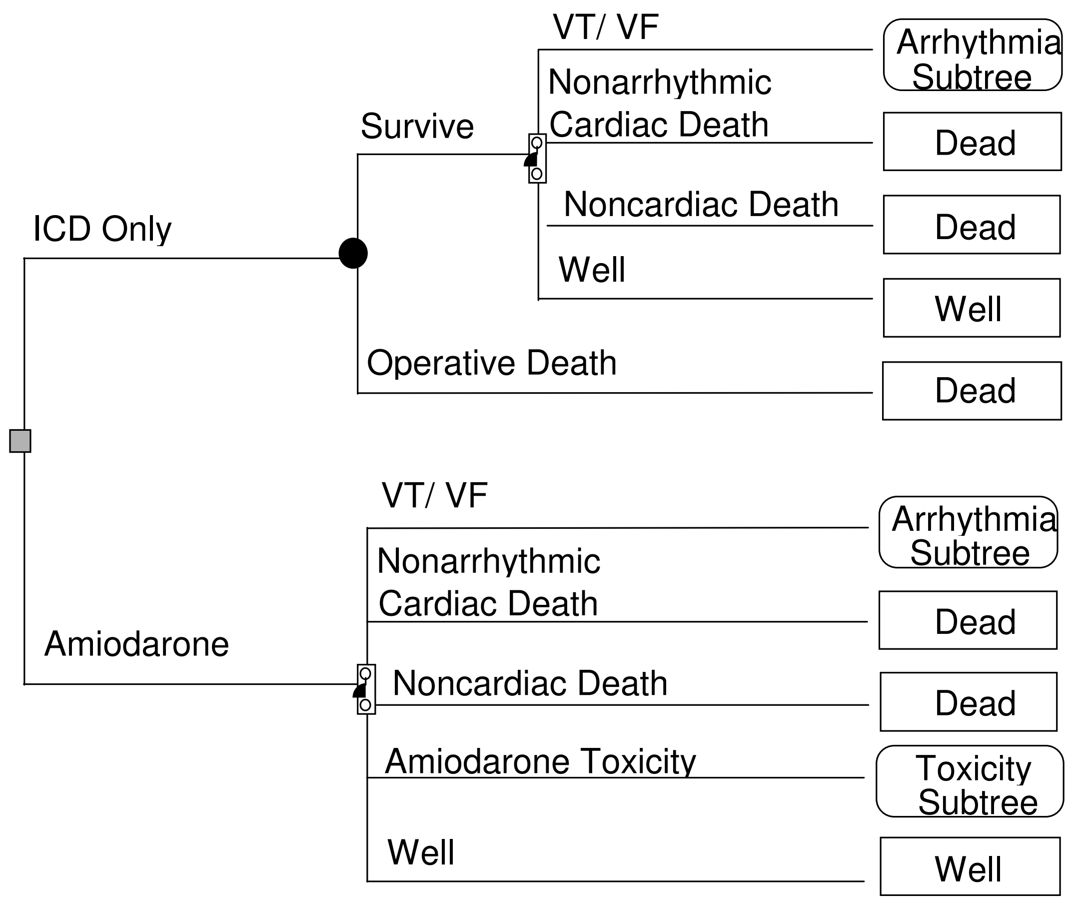


Figure 7.2. Schematic of the SCD-MM decision model. The square node represents a decision to use either an implantable cardioverter defibrillator (ICD) or amiodarone. The circle represents a chance node. After a regimen is chosen, the patient enters a Markov tree (denoted by rectangles containing circles and an arrow). A patient who enters the ICD regimen enters the Markov tree only if he survives ICD implantation. The Markov trees represent the clinical events that can occur during each 1-month period as a patient is followed until his death. For example, a patient who receives the ICD is at risk each month for ventricular tachycardia (VT), ventricular fibrillation (VF), nonarrhythmic cardiac death, and noncardiac death. If none of these events occur, the patient remains well for the 1-month period. A patient who is well at the end of that 1-month period reenters the Markov tree. Subtrees are denoted by rounded rectangles. Patients who receive amiodarone are at risk for VT, VF, nonarrhythmic cardiac death, noncardiac death, and toxicity from amiodarone. If toxicity from amiodarone occurs, the patient enters the amiodarone toxicity subtree (Owens et al. 1997a).

7.2.1 Decision Maker implementation of the DM

I implemented the LC-EM using the Decision Maker modeling software. Figure 7.3 shows the Decision Maker interface displaying a portion of the LC-EM. Figure 7.4 shows an excerpt from the corresponding Decision Maker ASCII file. Although the DM representation can draw information from this representation, sufficient knowledge for producing a CPG is not contained in the decision-tree representation.

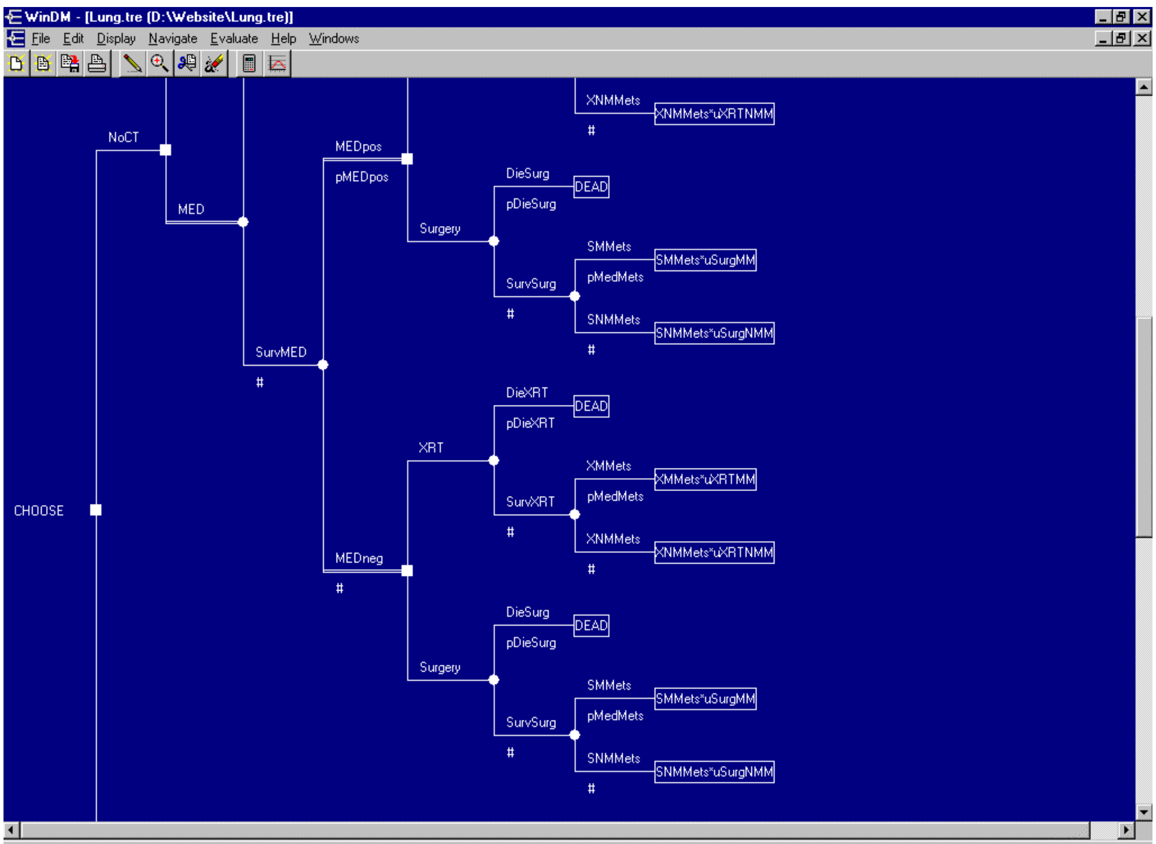
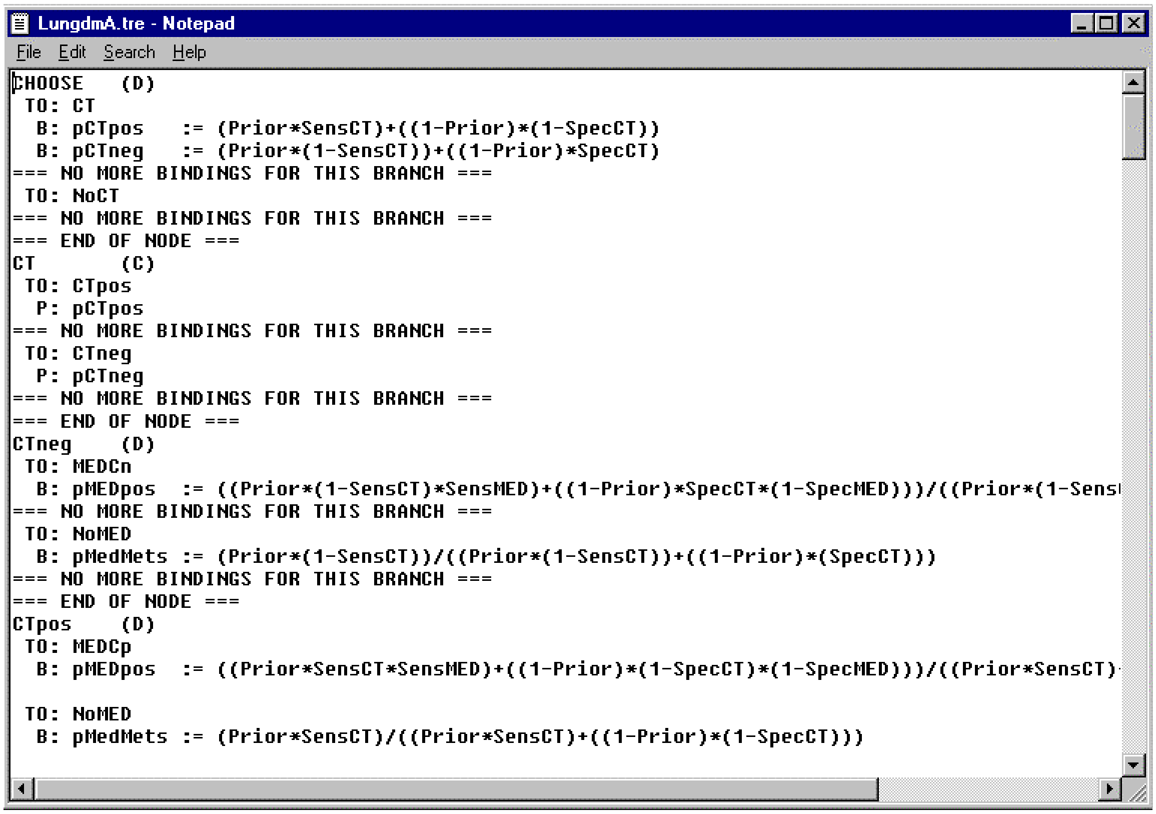


Figure 7.3. Decision Maker interface. Using this graphical interface, the decision analyst navigates the decision-tree representation and makes changes to the model's structure, probabilities, and utilities.

After the decision analyst has completed the DM, he loads it into ALCHEMIST (Figure 7.5). The ALCHEMIST implementation uses a web-based interface and common gateway interface (CGI) scripts, so it can be run on any computer platform, and the decision analyst can



LungdmA.tre - Notepad

File Edit Search Help

```
CHOOSE    (D)
 TO: CT
  B: pCTpos   := (Prior*SensCT)+((1-Prior)*(1-SpecCT))
  B: pCTneg   := (Prior*(1-SensCT))+((1-Prior)*SpecCT)
=== NO MORE BINDINGS FOR THIS BRANCH ===
 TO: NoCT
=== NO MORE BINDINGS FOR THIS BRANCH ===
=== END OF NODE ===
CT        (C)
 TO: CTpos
  P: pCTpos
=== NO MORE BINDINGS FOR THIS BRANCH ===
 TO: CTneg
  P: pCTneg
=== NO MORE BINDINGS FOR THIS BRANCH ===
=== END OF NODE ===
CTneg     (D)
 TO: MEDCn
  B: pMEDpos  := ((Prior*(1-SensCT)*SensMED)+((1-Prior)*SpecCT*(1-SpecMED)))/((Prior*(1-Sens
=== NO MORE BINDINGS FOR THIS BRANCH ===
 TO: NoMED
  B: pMedMets := (Prior*(1-SensCT))/((Prior*(1-SensCT))+((1-Prior)*(SpecCT)))
=== NO MORE BINDINGS FOR THIS BRANCH ===
=== END OF NODE ===
CTpos     (D)
 TO: MEDCp
  B: pMEDpos  := ((Prior*SensCT*SensMED)+((1-Prior)*(1-SpecCT)*(1-SpecMED)))/((Prior*SensCT)

 TO: NoMED
  B: pMedMets := (Prior*SensCT)/((Prior*SensCT)+((1-Prior)*(1-SpecCT)))
```

Figure 7.4. Decision Maker ASCII file format. This file shows the ASCII format of Decision Maker file that ALCHEMIST uses to get information about the DM. For example, it lists the node CT and a chance (C) node with two branched (TO: CTpos and TO: CTneg) and corresponding probabilities (P: pCTpos and P: pCTneg). CT = computed tomography; CTpos = positive CT; CTneg = negative CT; pCTpos = probability of a positive CT; pCTneg = probability of a negative CT.

load any decision tree that is located in his personal-computing environment. ALCHEMIST uses simple error checking to ensure that the file being loaded is in Decision Maker format, and that the user has provided the proper identifying information (i.e., name and electronic-mail address). The decision analyst also notes whether the DM is based on a cost-effectiveness analysis (Section 7.3).

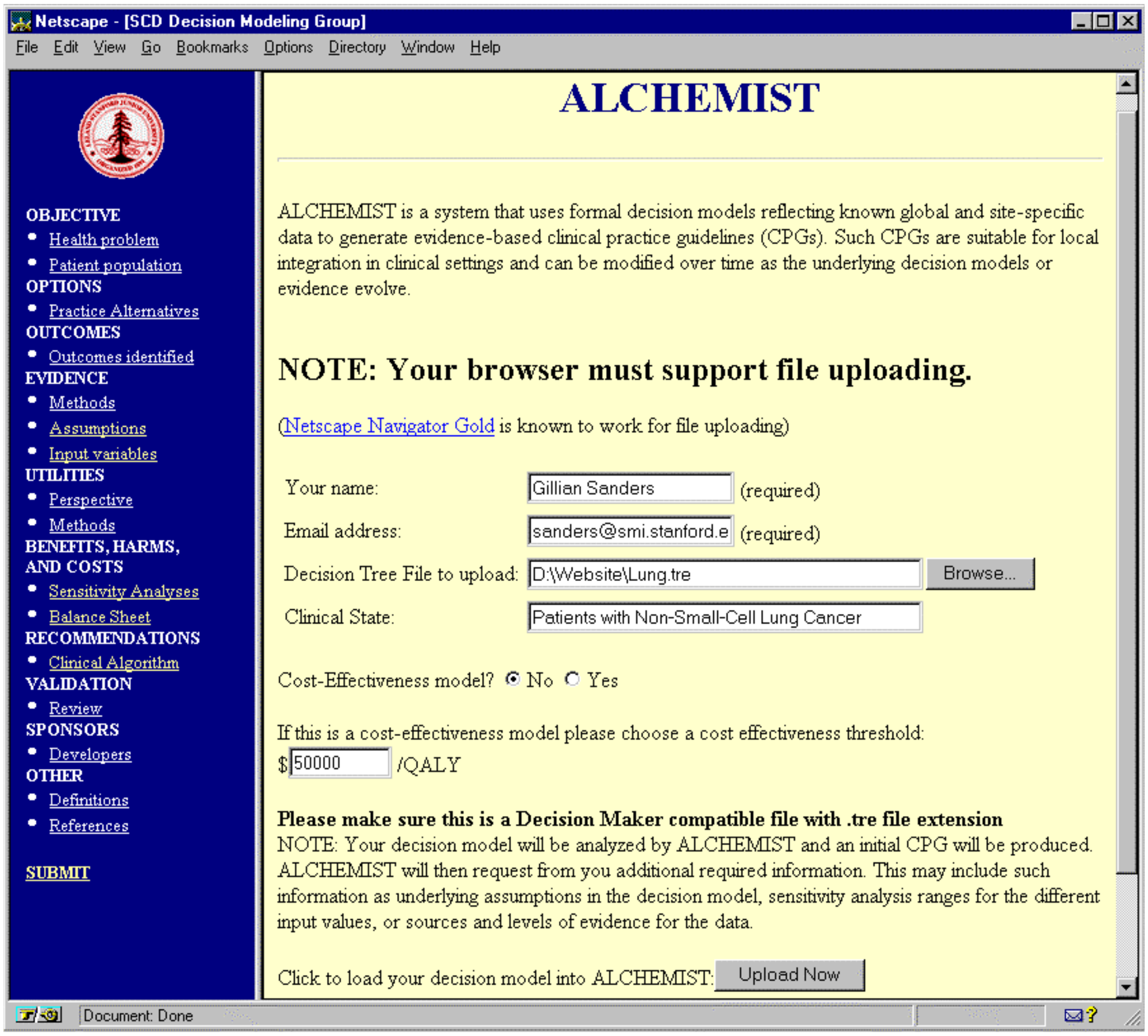


Figure 7.5. Loading the DM into ALCHEMIST. The decision analyst enters her name and electronic-mail address for logging purposes. She then browses her personal files to locate the Decision Maker DM that she will load into the ALCHEMIST system. The decision analyst can also indicate whether the DM represents a cost-effectiveness analysis.

7.2.2 Representation of the DM in the conceptual-model framework

The DM is loaded into the ALCHEMIST system, which parses the DM, and maps the information obtained directly from the DM onto the DM conceptual model. For example, ALCHEMIST determines the possible practice alternatives by looking at each decision node in the DM, determining that node's children, removing repetitive alternatives, and creating a table of the remaining practice alternatives. ALCHEMIST creates a web-based DM annotation editor that displays the explicit DM information, and that queries the decision analyst for additional information where needed. Figure 7.6 shows part of the DM annotation editor that ALCHEMIST completes automatically and displays for the decision analyst's confirmation.

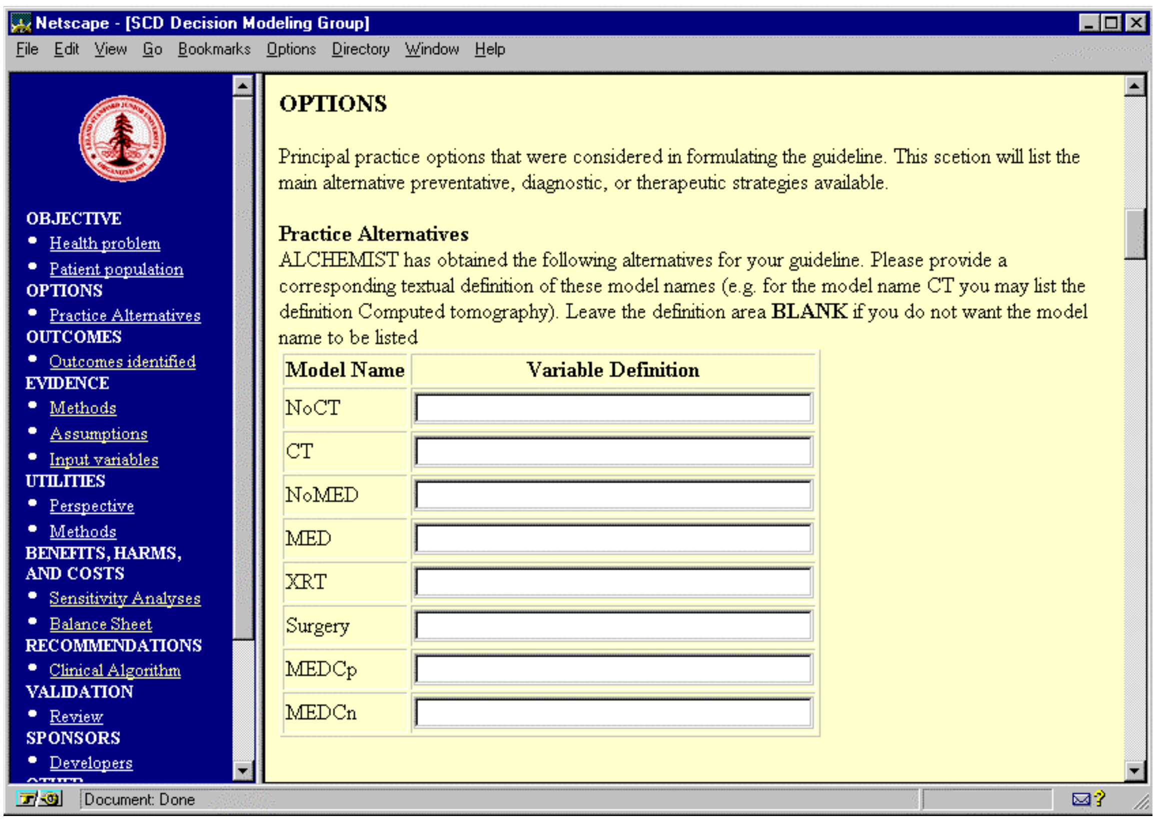


Figure 7.6. Knowledge obtained directly from the DM by ALCHEMIST. ALCHEMIST creates a table with these variables names and queries the decision analyst for the associated free-text definitions.

7.2.3 DM annotation editor

Although a portion of the information needed for CPG creation is located explicitly in the DM, ALCHEMIST must obtain other information directly from the decision analyst. The DM annotation editor asks the decision analyst to provide three types of input. For certain

CPG elements (e.g., health problem, patient population, and definitions of variable names), ALCHEMIST asks the decision analyst simply to provide free-text descriptions (Figure 7.6). For other elements, ALCHEMIST asks the decision analyst to choose among potential values for a given CPG element (e.g., model perspective, and method of obtaining values) (Figure 7.7). In specific situations, ALCHEMIST asks the decision analyst to

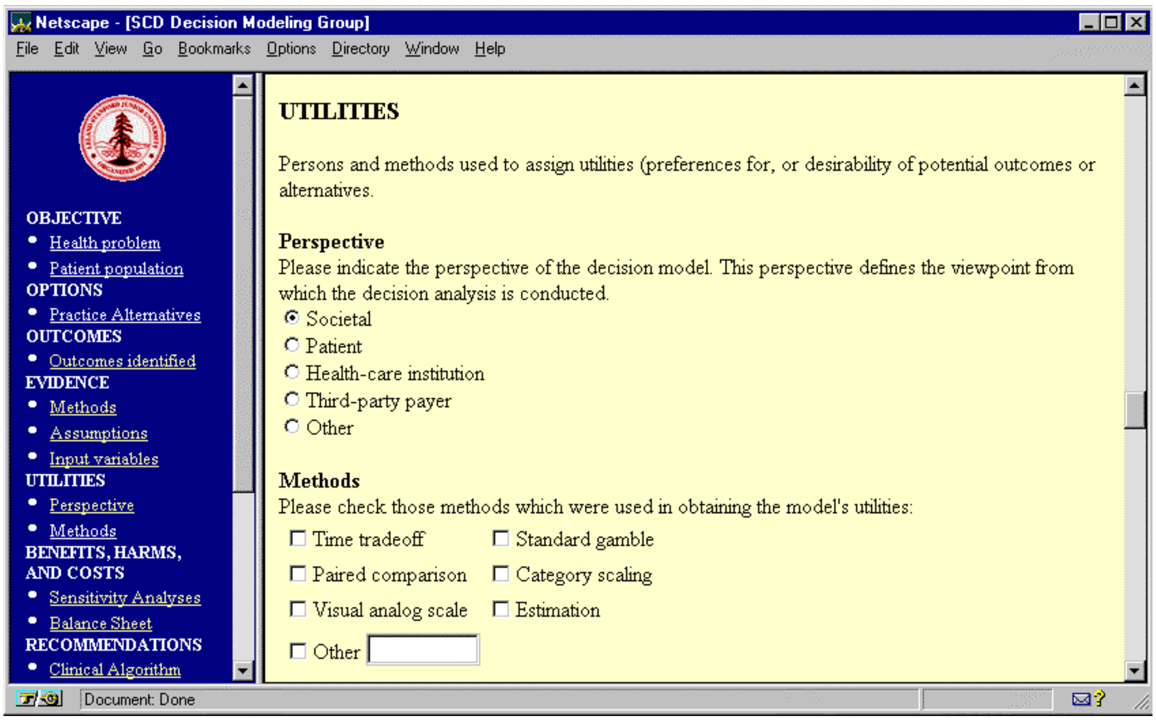


Figure 7.7. Radio-button input in the DM annotation editor. . The DM annotation editor uses radio-button input from which to provide the decision analyst with five possible modeling perspectives to choose.

describe characteristics of the DM or of its evidence to help the DM annotation editor to create an appropriate input form (e.g., creation of evidence table format) and then to obtain the needed information from the decision analyst. For example, in Figure 7.8, the decision analyst has chosen to create an evidence table for the studies that he used to calculate the prior probability of mediastinal metastases. He has provided evidence for this variable based on the following elements: study name, author, study design, level of evidence, patient population, description of intervention, outcome, and number of patients. ALCHEMIST creates the corresponding evidence table for the decision analyst to complete.

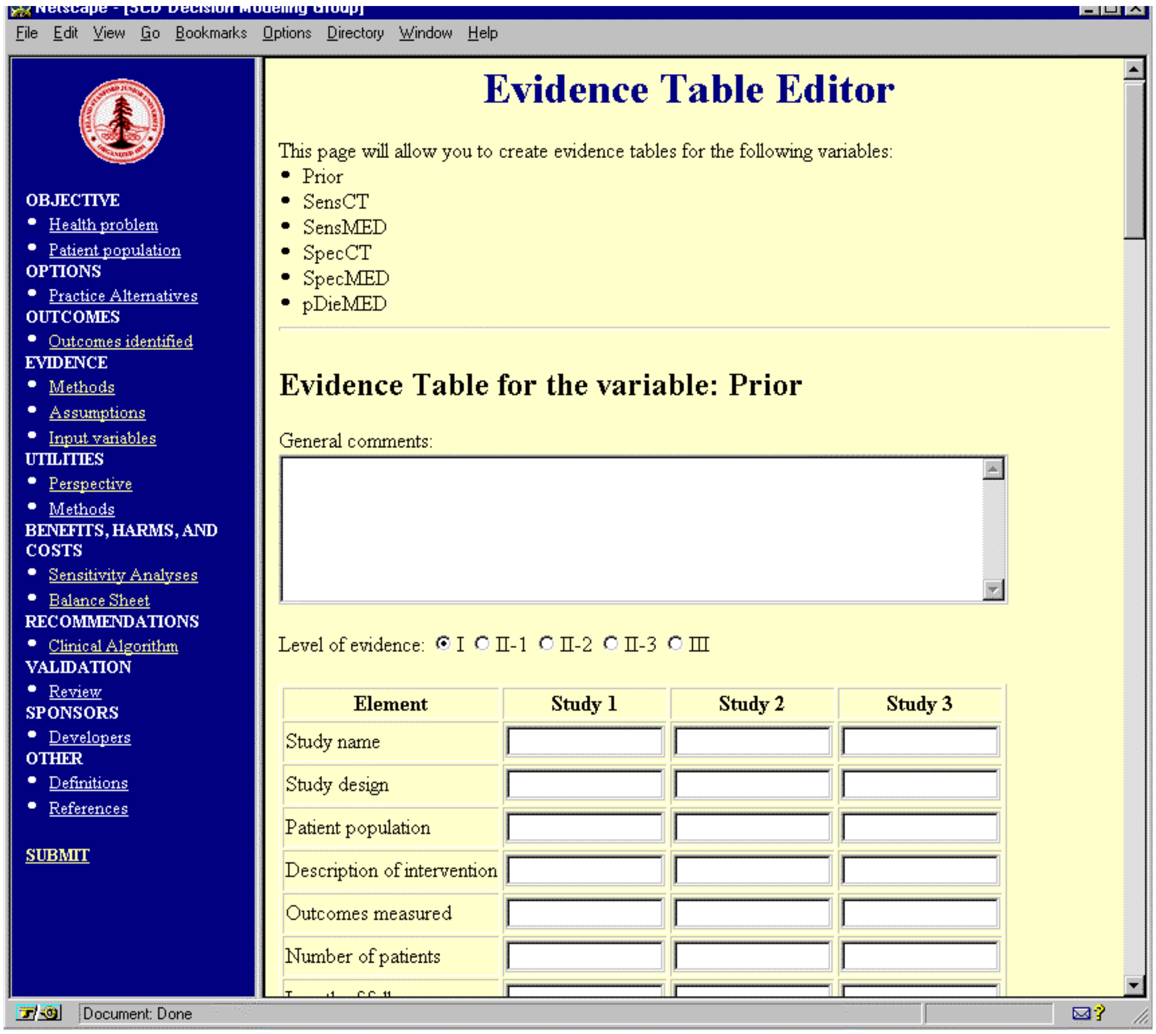


Figure 7.8. Creation of evidence tables by the decision analyst through use of the DM annotation editor. ALCHEMIST creates blank evidence tables for each variable designated by the decision analyst. Using the DM annotation editor, the decision analyst can then fill in the needed evidence for these specific variables.

After entering the needed information in the DM annotation editor, the decision analyst submits this information to ALCHEMIST, which then produces the CPG browser and custom-tailoring editor. This web-based CPG is subject to internal and peer review before it is broadly distributed.

7.2.4 CPG browser

Using the flowchart-creation algorithm that I described in Chapter 5, ALCHEMIST creates a graphical display of the CPG's recommendations. Figure 7.9 displays the generated algorithm for the best-estimate values of the LC-EM.

The CPG browser allows the guideline user to browse the CPG and to examine the evidence and recommendations (Figure 7.10). A high-quality CPG provides a large amount of information for the guideline user; without a consistent organization of this information, the guideline user could become lost. The CPG browser maintains a consistent structure to orient the guideline user. Example tasks that the guideline user can perform include the following: browse the guideline objective, view the related evidence for input variables, view the best-estimate recommendations displayed as a flowchart (Figure 7.9), view a tornado diagram of the input variables (Figure 7.11), and examine results of sensitivity analyses (Figure 7.12).

7.2.5 CPG custom-tailoring editor

The CPG browser includes a CPG custom-tailoring editor. This editor allows a local guideline user to specify site- or patient-specific input values to produce an updated or tailored CPG. The guideline user is able to tailor the CPG in two ways: (1) by making changes to specific input variables (Figure 7.13), or (2) by changing specific CPG elements, such as the modeling perspective or patient population, and having ALCHEMIST highlight the potentially affected variables (Figure 7.14). For example, if a guideline user wants to determine how changing the probability of death from thoracotomy affects the

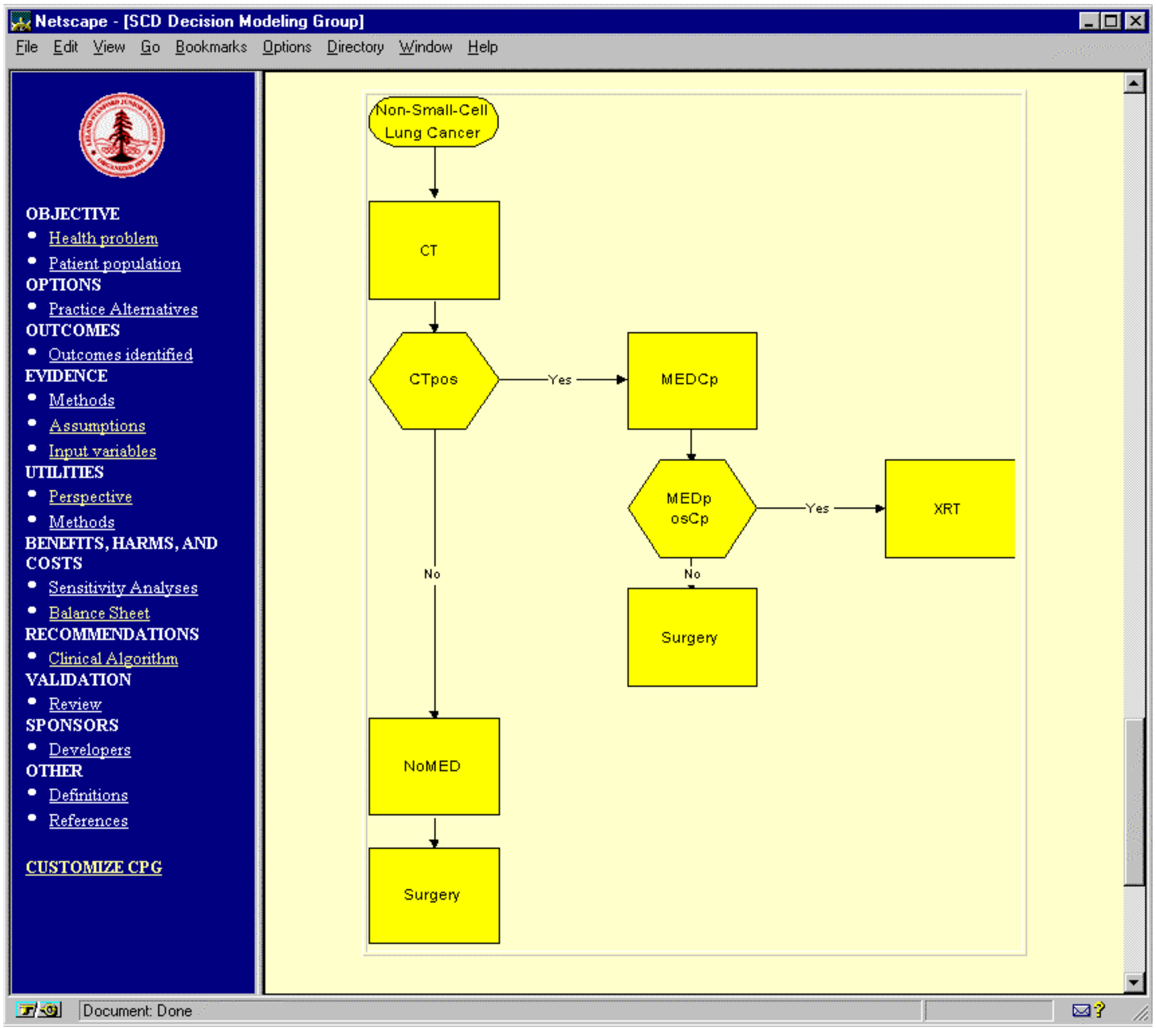


Figure 7.9. ALCHEMIST's LC-EM flowchart. Clinical-state boxes are rounded rectangles, action boxes are rectangles, and decision boxes are hexagons. This flowchart shows that, for the best-estimate values of the LC-EM, the optimal strategy is to perform a CT exam. If this CT exam is negative, then surgery should be performed. If the CT exam is positive, the physician should order a mediastinoscopy. If this second diagnostic test is positive, then radiation therapy should be administered; otherwise, surgery is the treatment of choice.

CPG, this reduced probability increases the health benefit provided by a thoracotomy; thus, the clinical recommendation is sensitive to this variable.

When a guideline user specifies a new input value, ALCHEMIST checks that none of the underlying assumptions in the DM are violated, and creates a new CPG that highlights the

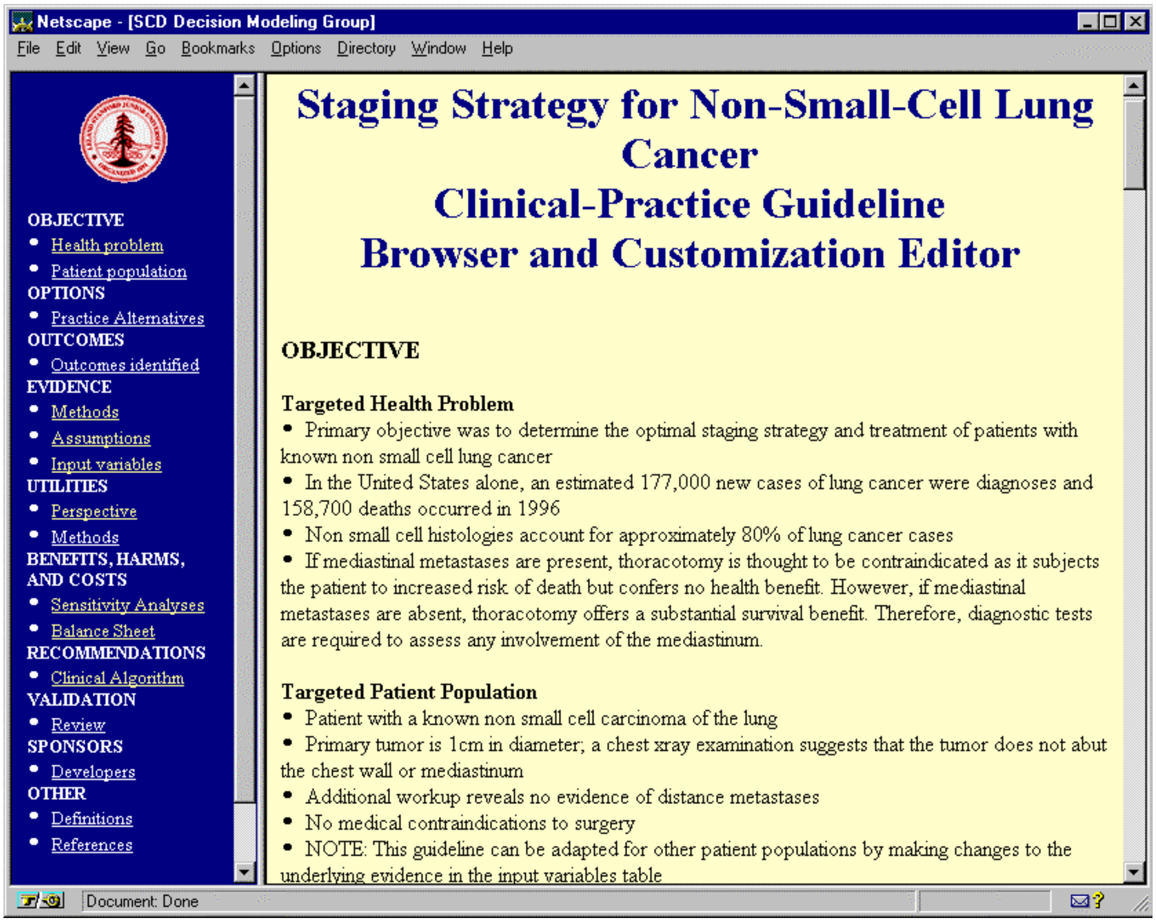


Figure 7.10. **ALCHEMIST's CPG browser.** The menu on the left side of the screen outlines the structure of the CPG and allows the user to move through the elements of the CPG.

changes to the CPG based on the information specified by the guideline user (Figure 7.16). ALCHEMIST also creates an updated balance sheet (Figure 7.15).

In Section 7.3 and 7.4 I step through the DM-to-CPG translations for the LC-CEM and SCD-MM. I accentuate any differences from the LC-EM process.

7.3 LC-CEM example

In addition to being able to create CPGs based on pure-effectiveness DMs, ALCHEMIST can also translate **cost-effectiveness DMs (CE-DMs)** into the corresponding CPGs.

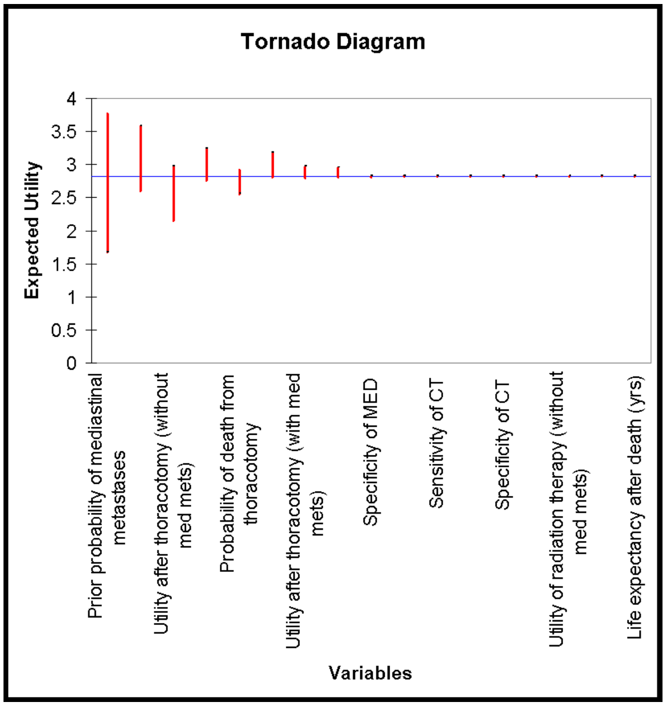


Figure 7.11. Tornado diagram produced for the LC-EM. This diagram displays for the user graphically the change in expected utility as a function of the different variables. The variables are ranked according to the magnitude of the change from left to right.

When a decision analyst loads a CE-DM into ALCHEMIST, he indicates the dual-utility nature of the model by checking the cost-effectiveness box on ALCHEMIST's file-loading screen. ALCHEMIST also asks the decision analyst for his desired cost-effectiveness threshold (Figure 7.17). Establishing that the DM is a cost-effectiveness model causes ALCHEMIST to add several elements to the DM annotation editor. ALCHEMIST (1) requires that the decision analyst indicate which variables are cost variables, through the use of a scrolling list (Figure 7.18); (2) asks the decision analyst to indicate the year of the costs used, and the method used to inflate costs to the given year (e.g., medical consumer price index or gross domestic product price deflator); and (3) computes the best-estimate flowchart algorithm using the chosen cost-effectiveness threshold and following the algorithm detailed in Section 5.6 (Figure 7.19).

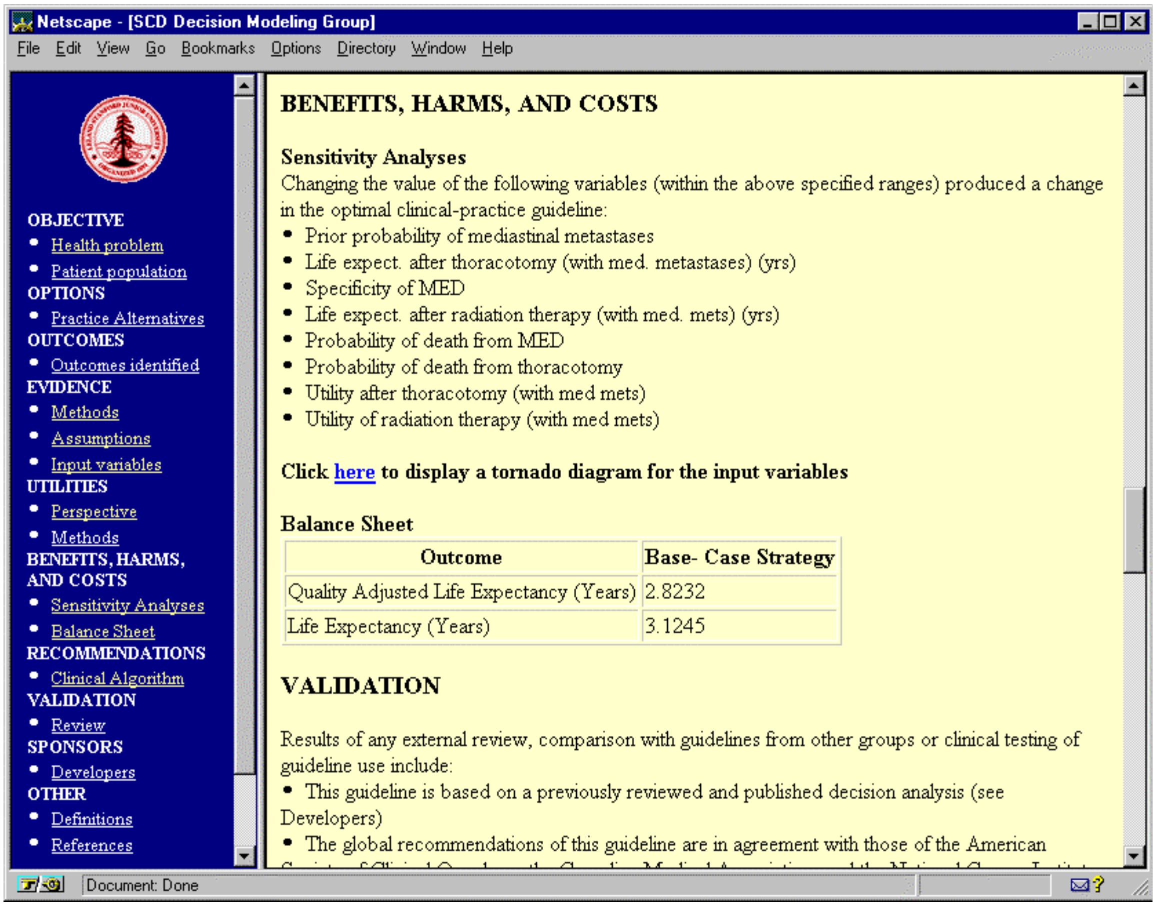
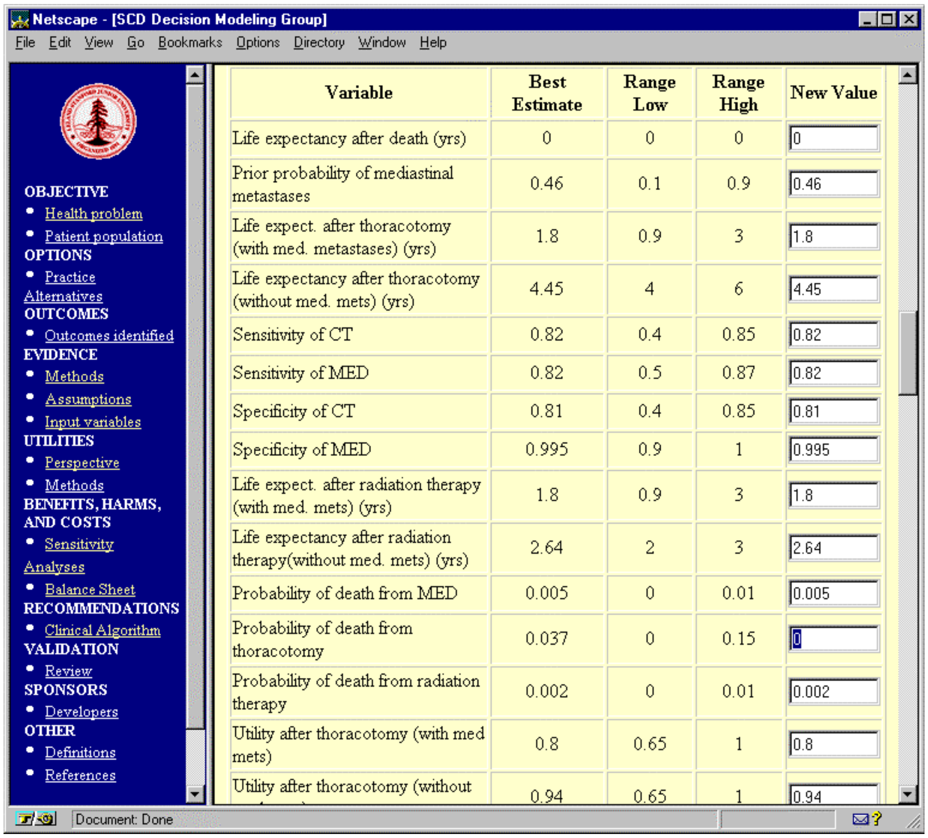


Figure 7.12. ALCHEMIST's list of sensitive variables. ALCHEMIST computes the optimal algorithm for each variable using the variables high- and low-range values. If the algorithm is sensitive to this change, then the corresponding variable is marked as "sensitive" and is listed here for the guideline user.

After the decision analyst has completed the required information in the DM annotation editor, ALCHEMIST creates the CPG browser and custom-tailoring editor. ALCHEMIST adds functionality to the CPG browser and custom-tailoring editor for CE-DMs. ALCHEMIST modifies the balance sheet to include cost outcomes, and the user can modify the cost-effectiveness threshold and view the updated CPG (Figure 7.20). If a guideline user wants to modify the perspective used in a CPG, as described in Section 7.2.5, ALCHEMIST indicates which variables are utility variables, as well as which ones are costs.



Variable	Best Estimate	Range Low	Range High	New Value
Life expectancy after death (yrs)	0	0	0	0
Prior probability of mediastinal metastases	0.46	0.1	0.9	0.46
Life expect. after thoracotomy (with med. metastases) (yrs)	1.8	0.9	3	1.8
Life expectancy after thoracotomy (without med. mets) (yrs)	4.45	4	6	4.45
Sensitivity of CT	0.82	0.4	0.85	0.82
Sensitivity of MED	0.82	0.5	0.87	0.82
Specificity of CT	0.81	0.4	0.85	0.81
Specificity of MED	0.995	0.9	1	0.995
Life expect. after radiation therapy (with med. mets) (yrs)	1.8	0.9	3	1.8
Life expectancy after radiation therapy(without med. mets) (yrs)	2.64	2	3	2.64
Probability of death from MED	0.005	0	0.01	0.005
Probability of death from thoracotomy	0.037	0	0.15	0
Probability of death from radiation therapy	0.002	0	0.01	0.002
Utility after thoracotomy (with med mets)	0.8	0.65	1	0.8
Utility after thoracotomy (without	0.94	0.65	1	0.94

Figure 7.13. **ALCHEMIST's CPG custom-tailoring editor.** The guideline user is able to change the input value for any or all of the variables in the input-variables table. In this example, the user has changed the probability of death due to thoracotomy to zero.

7.4 SCD-MM example

To the decision analyst and guideline user, interaction with the SCD-MM CPG does not appear any different from interaction with simpler DMs. As I described in Section 5.5.5, ALCHEMIST treats Markov nodes as a special type of chance node. Figure 7.21 shows the flowchart algorithm produced from the SCD-MM; note that there are no sequential decisions in the SCD-MM, and, therefore the algorithm reflects just one decision. One special feature of ALCHEMIST for analyzing a Markov model is its treatment of the CycLen variable. If a DM has a variable labeled CycLen, then ALCHEMIST recognizes this variable and lists the variable in the input variables tables, but does not allow a guideline user to change

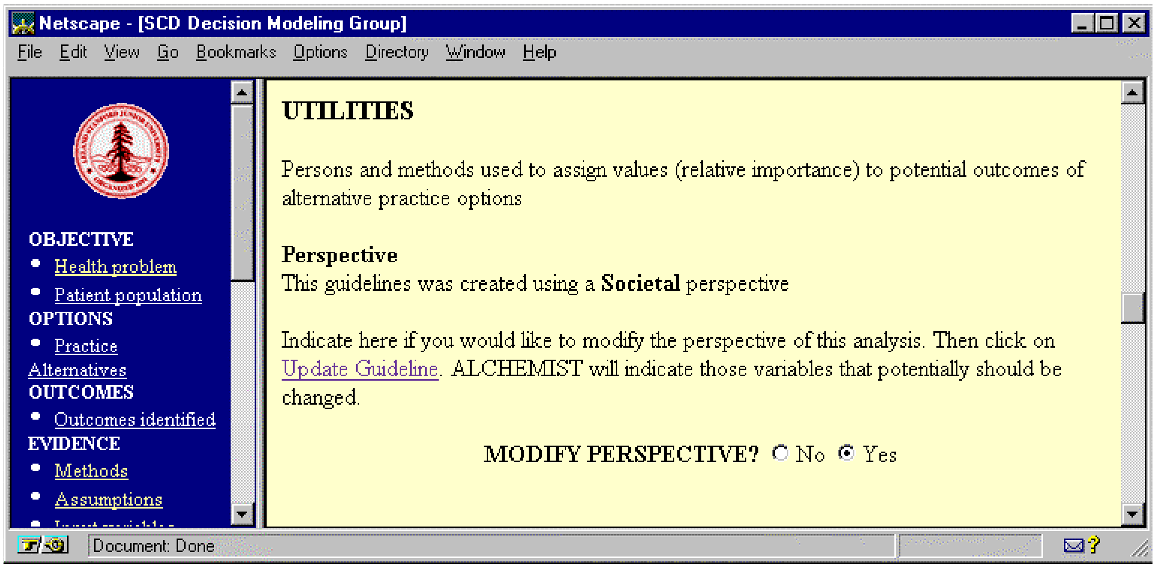


Figure 7.14. Alternate method of CPG custom tailoring. The guideline user indicates that she would like to change the modeling perspective of the model. ALCHEMIST then displays for the user the input-variables table, with the potentially relevant variables highlighted.

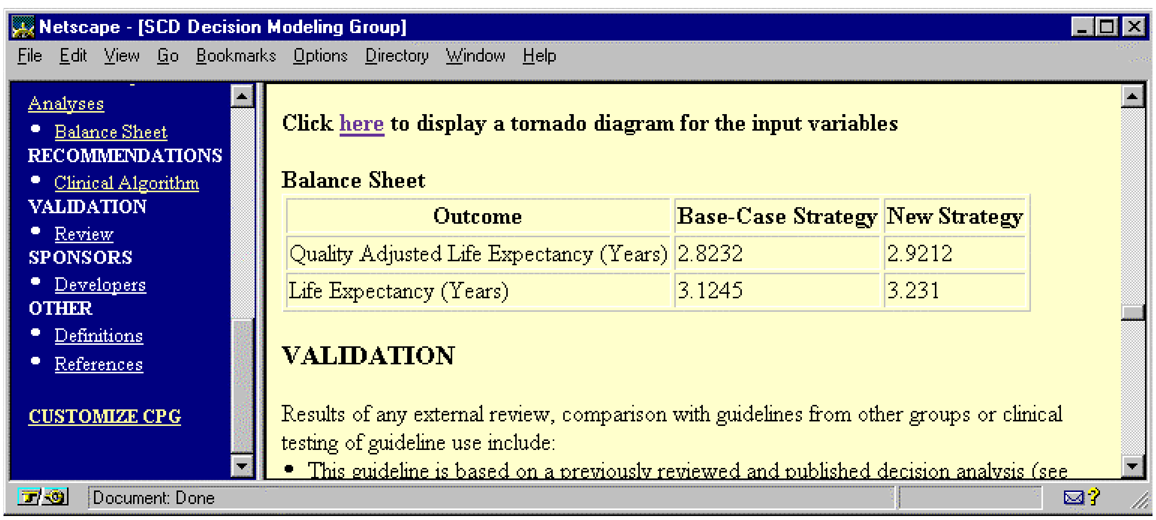


Figure 7.15. Updated balance sheet. The custom tailored CPG's balance sheet lists the various outcomes for the base-case strategy and the new strategy being assessed.

the variable's value. I place this restriction because of the numerous variables within a Markov model that depend on the models' cycle length. If ALCHEMIST allowed a user to change this cycle length, then, although the model could still run, maintaining the validity

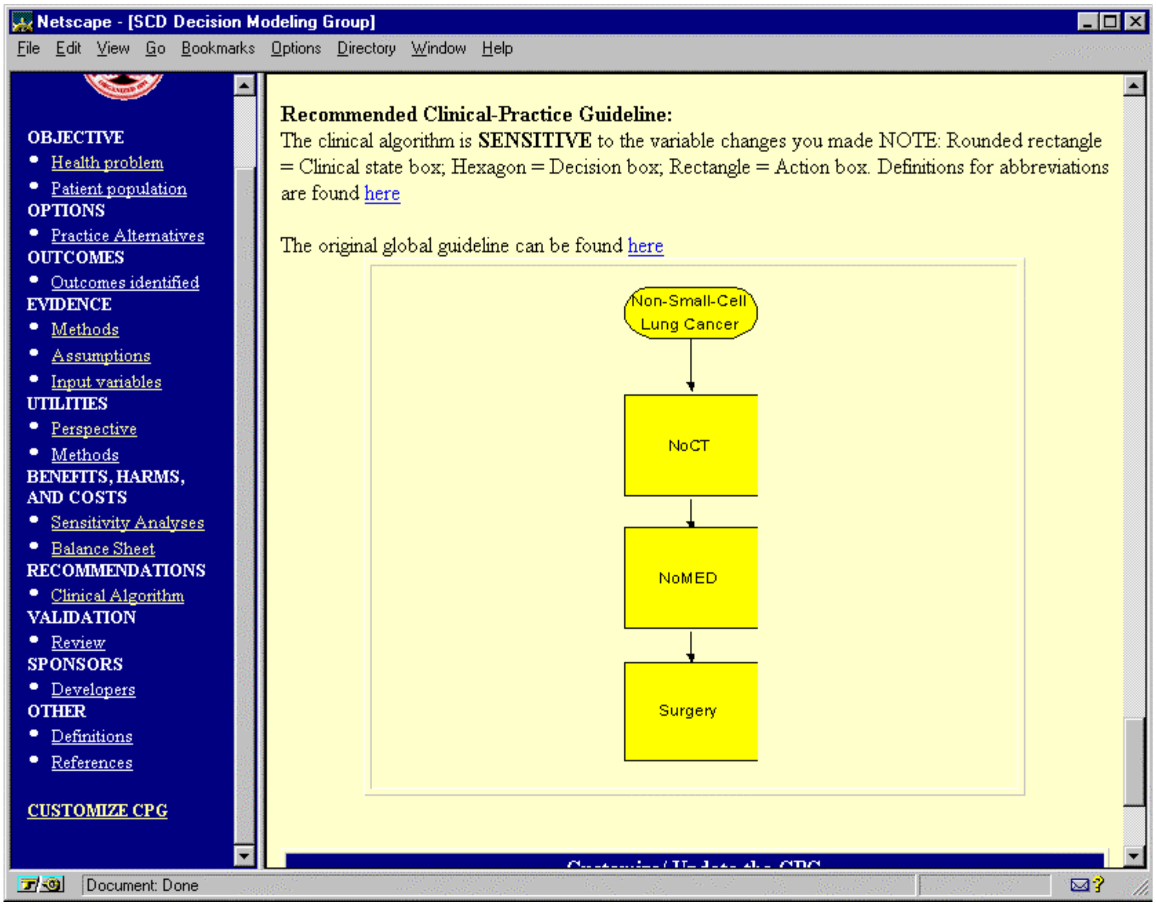


Figure 7.16. Tailored CPG browser. ALCHEMIST has updated the flowchart representation to reflect the probability of of death from thoracotomy that the user entered in Figure 7.13.

of the model would entail requiring the user to change numerous other probabilities from the current to the new time frame (e.g., changing the CycLen in the SCD-MM from 12 to 1 would require changing the probability of having an arrhythmic event from a monthly to a yearly value). Keeping track of such required modifications is beyond the scope of my thesis. The guideline user is able to change all the other input-variable values.

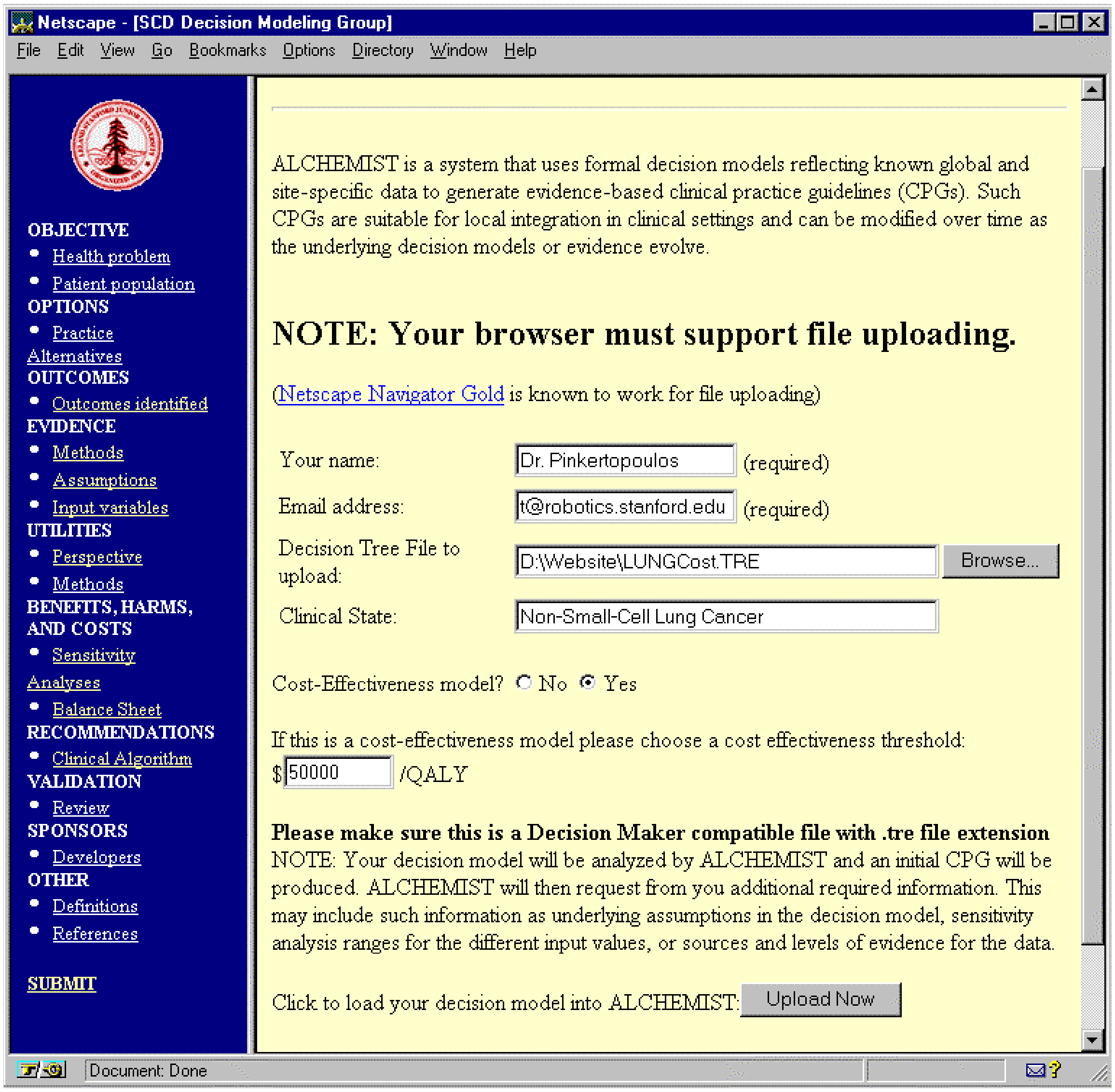


Figure 7.17. Loading of the a cost-effectiveness DM into ALCHEMIST. The decision analyst enters her name and electronic-mail address as before, but also indicates that his DM is a cost-effectiveness analysis, and specifies his desired cost-effectiveness threshold for future analyses.

7.5 Summary

In this chapter, I presented the three example DMs that I used for my thesis work and for my evaluation of ALCHEMIST. I then stepped through example DM-to-CPG translations,

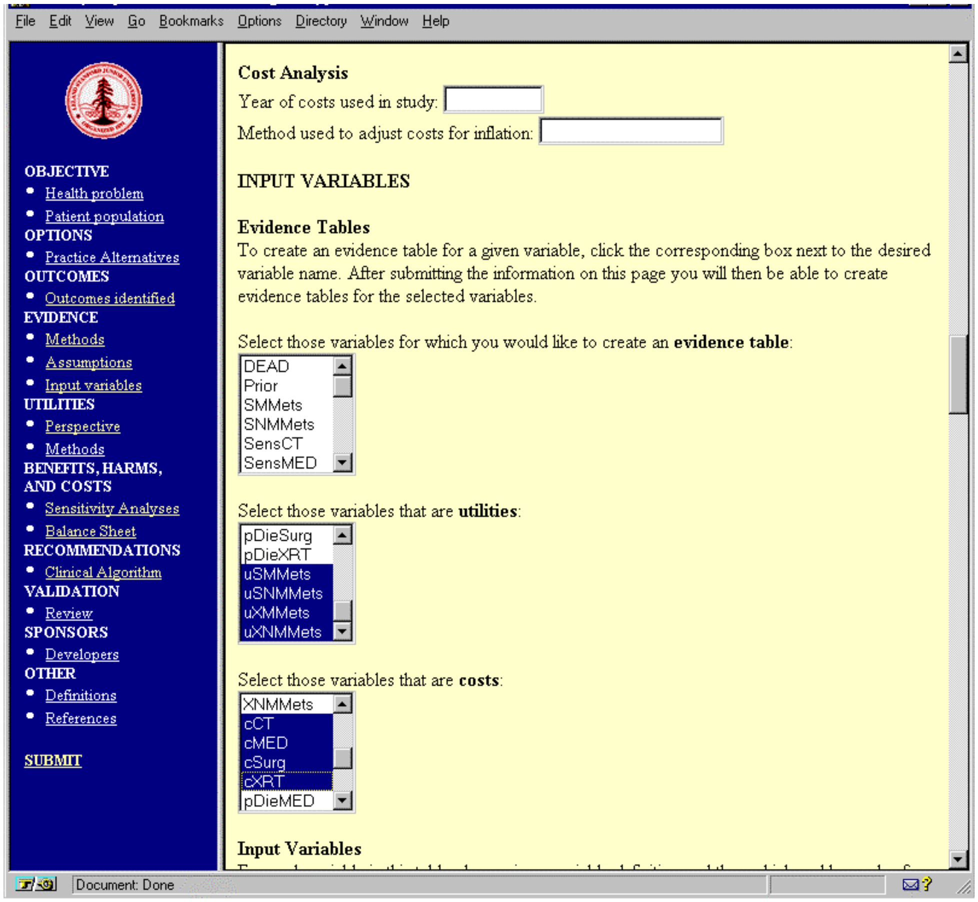


Figure 7.18. Choice of the cost variables. Using the scrolling lists, the decision analyst must indicate which of the input variables are cost variables.

demonstrating the web interface and functionality of the ALCHEMIST system. In Chapter 8, I describe my evaluation of the conceptual framework, the generated CPG, and ALCHEMIST's updating and tailoring abilities.

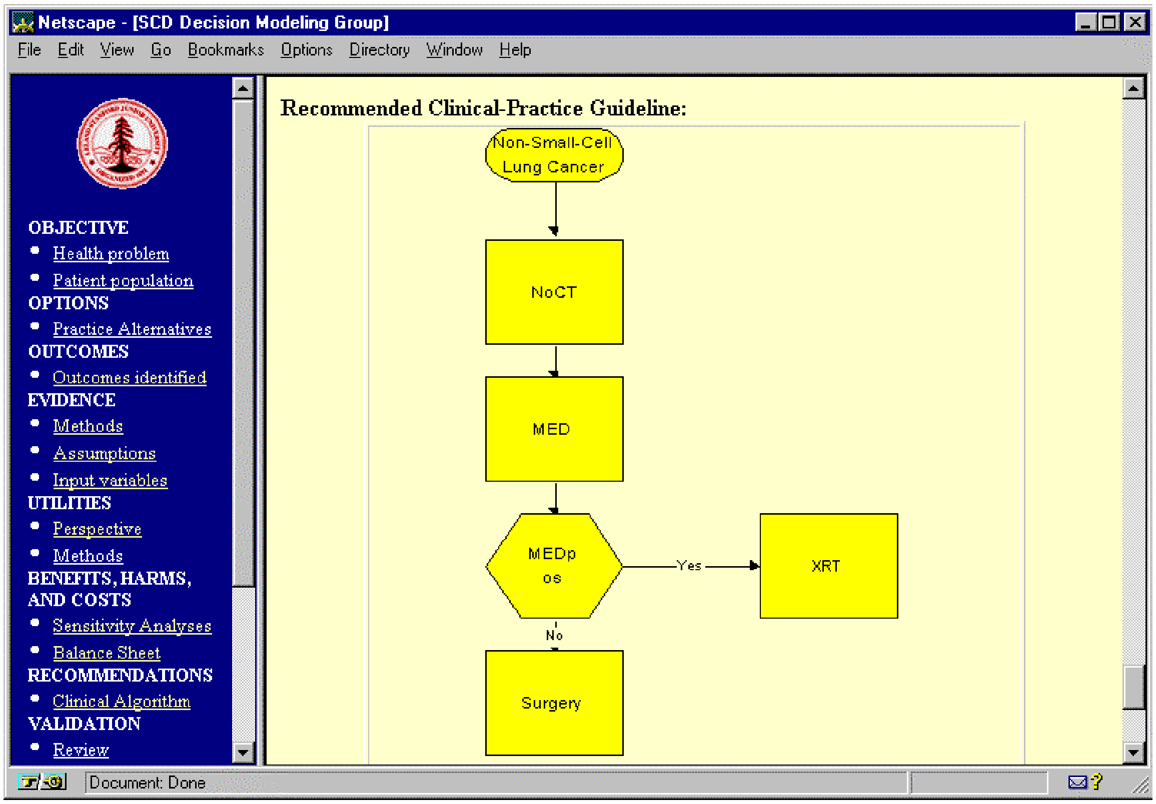


Figure 7.19. CPG flowchart algorithm produced for the LC-CEM. ALCHEMIST uses the chosen CE threshold to determine the optimal strategy at each decision point. Note that the structure of this flowchart is different from that of the flowchart produced in the LC-EM CPG (Figure 7.9), which evaluated the decisions based on only the effectiveness.

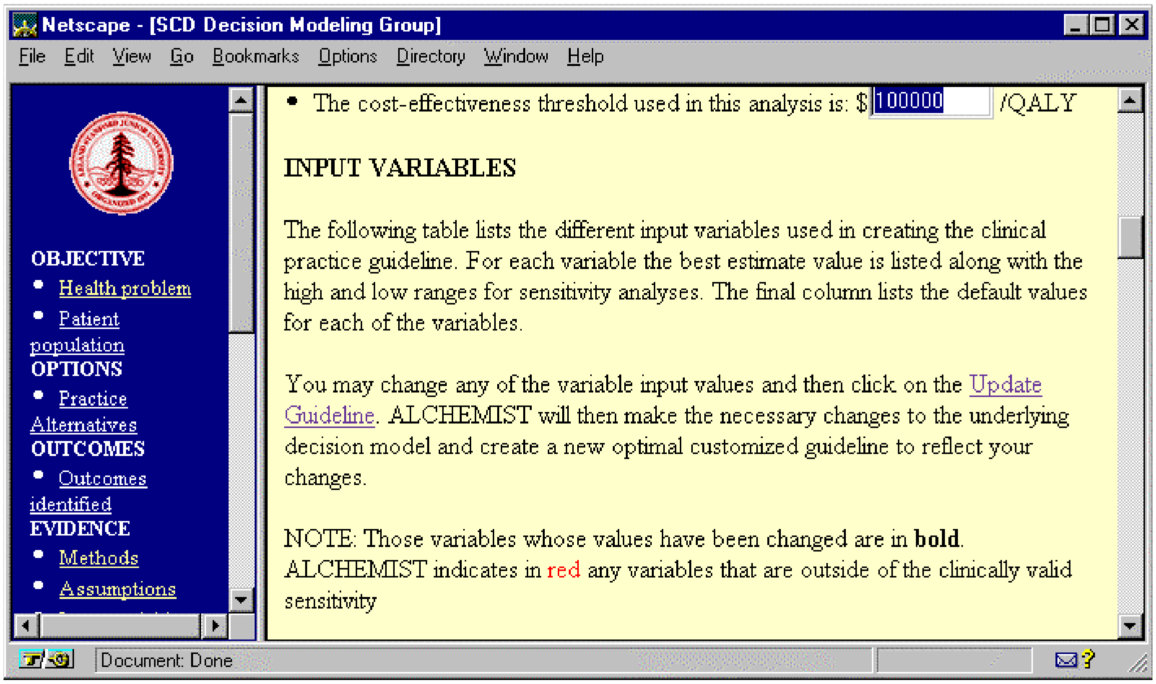


Figure 7.20. Modification of the cost-effectiveness threshold. The guideline user can change the cost-effectiveness threshold and can then view the updated CPG. This figure shows the guideline user changing the cost-effectiveness threshold from $50,000 per QALY to $100,000 per QALY.

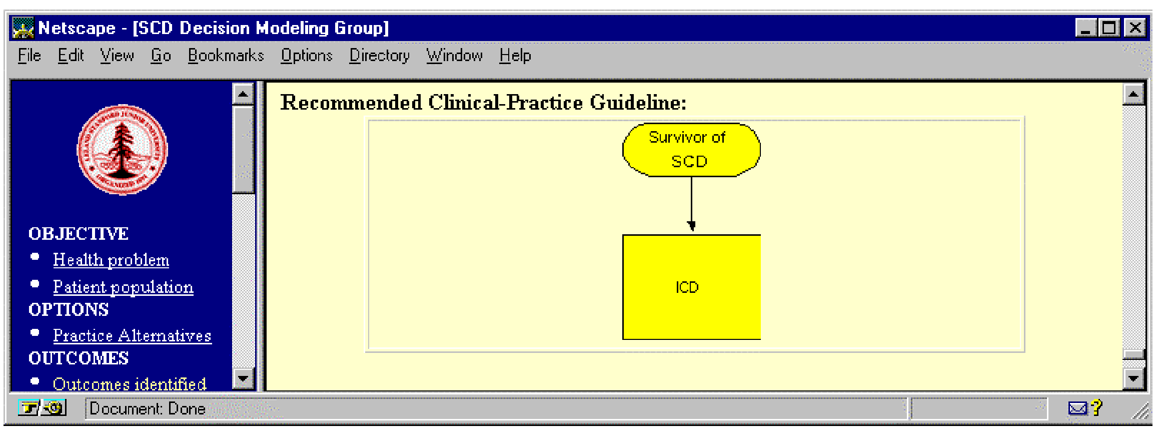


Figure 7.21. Best-estimate flowchart for the SCD-MM. This algorithm shows that, with a cost-effectiveness threshold of $100,000 per QALY, ALCHEMIST's recommended treatment strategy for survivors of sudden cardiac death is implantation of an ICD.

C h a p t e r 8

Evaluation

My research has produced ALCHEMIST, a web-based system that creates clinical-practice guidelines (CPGs) automatically from decision models (DMs). ALCHEMIST uses conceptual models of DMs and CPGs to represent the knowledge within these two representations. In this chapter, I report my evaluation of the DM and CPG conceptual models (Section 8.1), the quality of the resulting CPG (Section 8.3.1) and the custom-tailoring abilities of the ALCHEMIST system (Section 8.3.2).

As I described in Chapter 1, my hypothesis is that guideline developers can use computer-based DMs that reflect known global and site-specific data, to generate evidence-based CPGs. Such CPGs are of high quality, can be custom-tailored to specific clinical settings, and can be modified automatically over time as the underlying DM or evidence evolves. I consider my work to have validated this hypothesis if I can show the following:

1. The DM conceptual model provides information necessary for the efficient transformation of a DM (specifically, of a decision-tree representation of the expected outcomes and available alternatives) into CPGs, thereby allowing the creation of guidelines that are evidence based.
2. ALCHEMIST produces CPGs that satisfy published criteria for good practice guidelines.
3. ALCHEMIST's mapping algorithm allows dynamic patient and site tailoring, and, therefore, produces local CPGs that provide expected health outcomes that are based on the DM and that produce expected outcomes (measured in quality adjusted life years) that are equal to or better than those expected from static global guidelines for specific patient populations.

Section 8.1 through 8.3 describe the experiments that I performed to evaluate my thesis formally.

8.1 Evaluation of the DM-to-CPG conceptual framework

The evaluation of my DM and CPG modeling work had two critical goals: (1) to demonstrate the reasonableness of the conceptual-model design, and (2) to define the types of DMs and CPGs for which the core CPG tasks (defined in 2.4) can be accomplished given the two conceptual models as defined in Appendices A and B.

8.1.1 Evaluation of the conceptual models

To determine whether the design specification of my DM and CPG conceptual models is reasonable, I must determine whether the resulting CPGs satisfy established criteria for a high-quality guideline. As a test of this assertion, I compare the design specification of my DM and CPG conceptual models to a composite of available CPG-assessment tools. To evaluate the mapping algorithm among the DM, the DM annotation editor, and the generated CPG, I confirmed that each element in the CPG conceptual model could be instantiated by an element in the DM conceptual model or in the accompanying DM annotation editor. Because ALCHEMIST's generated CPG is a combination of the knowledge inherent in the DM and the additional information obtained from the DM annotation editor, I assume that the existence of a data element in either the DM or the DM annotation editor indicates that this element is available in the resulting CPG.

8.1.1.1 Objective

My objective was to validate the reasonableness of the DM and CPG conceptual models.

8.1.1.2 Methods

In 1992, the Institute of Medicine published *Guidelines for Clinical Practice: From Development to Use* (Institute of Medicine 1992). In this seminal work, the authors described an assessment instrument for critiquing CPGs. Since then, several other organizations have developed similar instruments, although the latter have all been based in part on the IOM tool (Cluzeau et al. 1997, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993)]. Therefore, I assume that the criteria assessed by the IOM instrument reflect the current gold standard for CPG assessment. Note that all the instruments used in my evaluation assess the *quality* of the CPG, rather than the *effectiveness* of the CPG in changing clinician behavior or improving patient quality of care. It is assumed that the quality of the CPG will correlate with the clinical effectiveness of the CPG (Eddy

1982, Eddy 1990a, Eddy 1990b, Eddy 1990c, Eddy 1990d, Eddy 1990e, Eddy 1990f, Basinski 1995).

I identified all the types of CPG information — the **data elements** — requested by the IOM and by three additional CPG-assessment tools (Cluzeau et al. 1997, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993)]. Data elements also had to be relevant to the critiquing of CPGs, and to all clinical domains. After I identified all the data elements in the instruments, I attempted to match each data element to an equivalent data requirement in the DM or CPG conceptual model. For example, the data element *Description of patient population* is satisfied in the DM annotation editor when ALCHEMIST requests that the decision analyst to describe the targeted patient population.

8.1.1.3 Results and discussion

Table 8.1 shows the data elements obtained from the set of CPG-assessment tools. I indicate whether the CPG conceptual model (which is instantiated from information in the underlying DM or DM annotation editor) satisfies the data element. I explain in further detail those data elements that the conceptual framework does not satisfy explicitly. Of the 60 individual data elements, I found 45 (75 percent) in the CPG conceptual model. In Table 8.1, I explain reasons for exclusion of the remaining 15 elements from my CPG conceptual model.

Table 8.1. Data requirements found in CPG-critiquing instruments. This table lists data elements that current CPG developers believe indicate a high-quality CPG. A checkmark in the CPG-CM column means that the corresponding data element is available in the CPG conceptual model (from the DM or the DM annotation editor) (Cluzeau et al. 1997, Institute of Medicine 1992, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993).

CPG data element	CPG CM	Explanation for those data elements not in design specification
Clinical applicability		
Statement of guideline objective	✔	

Table 8.1. Data requirements found in CPG-critiquing instruments. This table lists data elements that current CPG developers believe indicate a high-quality CPG. A checkmark in the CPG-CM column means that the corresponding data element is available in the CPG conceptual model (from the DM or the DM annotation editor) (Cluzeau et al. 1997, Institute of Medicine 1992, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993).

CPG data element	CPG CM	Explanation for those data elements not in design specification
Description of patient population	✔	
Description of complex clinical problems	✔	
Discussion of frequency of health problem	✔	
Discussion of current treatment patterns	✔	
Discussion of financial cost of the disease and its treatment	✔	
Description of professional groups to which the guideline is meant to apply	✔	
Rationale for excluding patient populations		Although the CPG-CM includes a description of the patient population and any assumptions about the target population, no explicit information is given about the rationale for excluding patient populations.
Clinical flexibility		
Discussion of tailoring guidelines	✔	
Discussion of patient preferences in health-care decisions	✔	
Discussion of methods of obtaining patient preferences	✔	
Description of ethical issues likely to arise in using the guideline		Although patient preferences are included in the CPG-CM, it does not require discussion of the ethical issues that may arise when the guideline is used. A decision analyst, however, could explore such topics in the background section of the guideline.
Reliability and reproducibility		
Independent review by experts or outside panels	✔	ALCHEMIST's CPGs do not have a set protocol for independent review. Clearly, such peer review would be needed before the adoption of the CPGs into clinical practice.

Table 8.1. Data requirements found in CPG-critiquing instruments. This table lists data elements that current CPG developers believe indicate a high-quality CPG. A checkmark in the CPG-CM column means that the corresponding data element is available in the CPG conceptual model (from the DM or the DM annotation editor) (Cluzeau et al. 1997, Institute of Medicine 1992, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993).

CPG data element	CPG CM	Explanation for those data elements not in design specification
Explanation of lack of independent review	✔	
Guidelines pretested in some manner		Similar to the peer-review requirement in that, before ALCHEMIST's guidelines are disseminated, the guideline developer must establish a method of piloting the guidelines.
Explanation of lack of pretesting	✔	
Validity		
Method of collecting scientific evidence is specifically described	✔	
Adequate references to sources of scientific evidence	✔	
General discussion of strength of scientific evidence	✔	
Explicit rating of the strength of the scientific evidence	✔	
If a formal method of synthesis is used, explicit description of the method	✔	
If applicable, the results of a formal synthesis of scientific evidence are explicitly reported	✔	
If applicable, the expert or group judgment techniques used for reaching professional consensus are explicitly described	✔	
If applicable, the strength of professional consensus resulting from use of group judgment techniques is reported	✔	
Qualitative description of health benefits		ALCHEMIST's CPG conceptual model provides only quantitative results.

Table 8.1. Data requirements found in CPG-critiquing instruments. This table lists data elements that current CPG developers believe indicate a high-quality CPG. A checkmark in the CPG-CM column means that the corresponding data element is available in the CPG conceptual model (from the DM or the DM annotation editor) (Cluzeau et al. 1997, Institute of Medicine 1992, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993).

CPG data element	CPG CM	Explanation for those data elements not in design specification
Qualitative description of potential harms or risks		ALCHEMIST's CPG conceptual model provides only quantitative results.
Quantitative information or estimates of health benefits	✔	
Health benefits projected in terms of life expectancy or similar measures	✔	
Quantitative information or estimates of potential harms or risks	✔	
Qualitative description of health costs or expenditures		ALCHEMIST's CPG conceptual model provides only quantitative results.
Quantitative information of health costs or expenditures	✔	
If health benefits projected in terms of life expectancy or similar measures, costs per unit of each identified benefit also estimated	✔	
Estimates of benefits, harms, and costs are consistent with the strength of provided evidence	✔	
Major recommendations made in the guideline	✔	
Discussion of strength of the scientific evidence for each major recommendation	✔	
Major recommendations consistent with strength of evidence	✔	
Other sets of guidelines identified		Although the decision analysts could provide links and references to existing guidelines in the references section of the CPG-CM, the model does not provide a slot for this needed information.

Table 8.1. Data requirements found in CPG-critiquing instruments. This table lists data elements that current CPG developers believe indicate a high-quality CPG. A checkmark in the CPG-CM column means that the corresponding data element is available in the CPG conceptual model (from the DM or the DM annotation editor) (Cluzeau et al. 1997, Institute of Medicine 1992, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993).

CPG data element	CPG CM	Explanation for those data elements not in design specification
Possible conflicts among existing guidelines discussed		Although the decision analyst could mention possible conflicts in the objective of the guideline, the CPG-CM does not identify this information explicitly.
Clarity		
Language describing the health condition is unambiguous	✔	
Language describing the options for management is unambiguous	✔	
Language describing each major recommendation is unambiguous	✔	
Recommendations are comprehensive and present when expected	✔	
Recommendations are consistent	✔	
Guidelines document uses clear headings, indexes, etc.	✔	
Guideline document has accurate summary or abstract		The CPG-CM does not provide an explicit abstract. The conceptual model, however, does include a menu of the main elements to permit easy navigation among the CPG.
Users can find recommendations easily	✔	
Schedule review		
Scheduled date for review or a procedure for arriving at such a date is provided		The CPG-CM indicates when the guideline was developed and the publication dates of the evidence used in formulating the recommendations. The conceptual model, however, does not indicate an explicit date for a scheduled review.

Table 8.1. Data requirements found in CPG-critiquing instruments. This table lists data elements that current CPG developers believe indicate a high-quality CPG. A checkmark in the CPG-CM column means that the corresponding data element is available in the CPG conceptual model (from the DM or the DM annotation editor) (Cluzeau et al. 1997, Institute of Medicine 1992, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993).

CPG data element	CPG CM	Explanation for those data elements not in design specification
Identification of the body responsible for reviewing and updating the guideline	✔	
Description of the parameters for which better information is needed	✔	
Description of any expected research or technical developments that could modify the policy		Although the evidence tables may highlight lack of evidence in a given area, the conceptual model does not indicate explicitly any expected research or technical developments that could modify the policy.
Multidisciplinary process		
Participation of persons in appropriate clinical and methodologic disciplines	✔	
Guideline document notes potential biases or conflicts of interest or indicates that they were detainment account	✔	
Balance of potential biases or conflicts of interest	✔	
Description of the methods used to solicit views of those not on the guidelines developments panel and to present those views to the panel	✔	
Identification of the agency responsible for the development of the guidelines	✔	
Implementation and dissemination		
Listing of possible methods for dissemination and implementation		The implementation and dissemination of the CPGs are not addressed in the CPG conceptual model; however, the use of the web interface enables almost universal access to developed CPGs.

Table 8.1. Data requirements found in CPG-critiquing instruments. This table lists data elements that current CPG developers believe indicate a high-quality CPG. A checkmark in the CPG-CM column means that the corresponding data element is available in the CPG conceptual model (from the DM or the DM annotation editor) (Cluzeau et al. 1997, Institute of Medicine 1992, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993).

CPG data element	CPG CM	Explanation for those data elements not in design specification
Criteria for monitoring compliance		The conceptual models do not capture the compliance of guideline users with the CPG.
Definition of measurable outcomes to be monitored	✔	
Identification of key elements which need to be considered by local guideline groups	✔	
Identification of key areas on which information for patients should be provided		Because the proposed users of ALCHEMIST's CPG are guideline developers at local institutions (rather than individual physicians or patients), the CPG-CM does not include additional information targeted to patient education. Such information clearly would be required before ALCHEMIST could be extended for use by individual physicians.

As a result of this evaluation, a future extension of my thesis work is to modify the conceptual framework. I will add a data element that describes the patients that should be excluded from following the recommendations of the CPG. Although my CPG conceptual model describes the targeted patient population it does not make explicit those populations that should be excluded. Also, my conceptual model indicates when the guideline was developed and the publication dates of the evidence used in formulating the recommendations. A data element that I will add however, is the expected lifetime of the recommendations. This element will help a local guideline implementer to determine if the CPG or its underlying DM needs updating.

8.1.2 Evaluation of the expressivity of the DM and CPG conceptual model

The expressivity of a conceptual model defines what subset of all objects the conceptual model can represent. In Chapter 7, I step through the DM-to-CPG process for my three example DMs. This process establishes that my DM conceptual model can express DMs that have sequential decision, dual utilities, and Markov processes. In this section, I concentrate on the expressivity of the CPG conceptual model.

8.1.2.1 Objective

My objective was to determine the expressivity of the CPG conceptual model.

8.1.2.2 Methods

I retrieved eight published CPGs from CPG-development organizations. The CPGs chosen (Table 8.2) are to make clinical decisions, are endorsed by a major clinical or governmental organization, and conform to the subset of CPGs described in Chapter 2. For each guideline listed, I attempted to express the information contained in these guidelines using my CPG conceptual model. I considered my CPG conceptual model sufficiently expressive if the published CPGs contain a subset of the CPG elements found in the CPG conceptual model. I kept track of any CPG knowledge that my CPG conceptual model could not represent or for which a modification of the CPG conceptual model would be required.

Table 8.2. Sample guidelines chosen to test expressivity of the CPG conceptual model.

Endorsing organization	CPG title
Agency for Health Care Policy and Research (AHCPR)	Management of acute pain
	Evaluation and management of early HIV infection

Table 8.2. Sample guidelines chosen to test expressivity of the CPG conceptual model.

Endorsing organization	CPG title
United States Preventive Services Task Force (USPSTF)	Screening for breast cancer
	Counseling to prevent gynecologic cancers
National Institutes of Health (NIH)	Consensus-development conference statement on cervical cancer
US Department of Health and Human Services: National Heart Attack Alert Program Coordinating Committee	Emergency department: Rapid identification and treatment of patients who have acute myocardial infarction
American Diabetes Association	Foot care in patients who have diabetes
Advisory Committee for Elimination of Tuberculosis	Preventive therapy for tuberculous infection in the United States

8.1.2.3 Results and discussion

Table 8.3 summarizes those elements of the eight guidelines that my CPG conceptual model could not express. Overall, the CPG conceptual model was able to capture the major elements of all the guidelines. In addition, I could add all the elements in Table 8.3 to the CPG conceptual model if future reviewers of my work decided that they are essential.

Table 8.3. Elements and explanation of existing guidelines that the existing CPG conceptual model could not be represent.

Guideline element missing from CPG conceptual model	Explanation
Executive summary	The CPG-CM does not provide an executive summary of the CPG. However, the structured menu of the CPG provides easy navigation among the numerous elements of the CPG.
Interactive links to the references cited	Although the CPG-CM provides the references used in the guideline development, it does not provide active links to the corresponding full text or Medline abstracts.
Topics to discuss with patients/ patient education	The CPG-CM does not include information targeted toward explaining the CPG to the patient.

Table 8.3. Elements and explanation of existing guidelines that the existing CPG conceptual model could not be represent.

Guideline element missing from CPG conceptual model	Explanation
Inclusion of non-traditional medical options	Although nontraditional medical therapies do not constitute an explicit element of the CPG-CM, if the developers consider such therapies to be one of the competing alternatives, the decision analyst may make them an explicit part of the underlying DM (and therefore of the CPG-CM).
Patients who have concurrent medical conditions	The CPG-CM is designed to represent a clinical decision for a described patient population. If a sub-population with such a set of concurrent conditions have a high prevalence among a given population, then a decision analytic team could develop an adapted DM and corresponding CPG could be produced for it.
Glossary of terms	The CPG-CM provides definitions of the numerous variables used in the underlying DM, although it currently does not provide a separate slot for additional terms.
Definition of high risk groups or special needs (e.g., children, elderly patients, pregnant women)	Although the CPG-CM describes the patient population, specific groups (such as high-risk patients, elderly patients, and pregnant woman) are not identified.

How does this expressivity translate into the proportion of published CPGs that the ALCHEMIST can express? Currently, the existence of a clinically valid DM limits the subset of guidelines that ALCHEMIST can represent. Sonnenberg and colleagues examined a set of CPGs published by the American College of Physicians (ACP) (American College of Physicians 1995). Their study showed that for 53% of these guidelines, published decision analyses address either one of the questions addressed by the guideline, or addressed the problem domain sufficiently that modification of the DM could be applied to answer guideline questions (Sonnenberg 1997). Therefore, the success of the ALCHEMIST system depends in part on the existence of a valid DM and a DM annotation editor to allow the transformation of existing DMs into the corresponding CPG.

8.2 Evaluation of the decision-model annotation editor

As I described in Chapters 3 and 4, the adoption of the ALCHEMIST system requires that prospective CPG developers — and specifically their decision-analytic teams — submit for review a structured annotation of their DM. If these prospective decision-analytic teams are to participate, the time and work required by the team to enter a DM directly into ALCHEMIST must be reasonable, and the analysts must be assured that the resulting CPGs will be fair and accurate. This section presents preliminary information on how time consuming DM annotating may be, based on my experience in entering the requested additional information for the SCD-MM directly into the DM annotation editor. In Section 8.3, I evaluate the quality of the generated CPG.

I used the ALCHEMIST web interface to enter the SCD-MM decision model — a DM for which I have been the primary decision analyst, and, therefore, with whose structure, assumptions, and evidence I am familiar. This DM is complex (30 Markov states and over 100 variables), and therefore demonstrates the upper end of the time-requirement for the annotating task. Using ALCHEMIST's web interface, I took approximately 5 hours to enter the SCD-MM into ALCHEMIST. Compared to the time required for preparation of a traditional CPG or DM manuscript, this time requirement is minimal.

As is described in Chapter 3, the set of DMs that can be entered into the ALCHEMIST system is a restricted subset of all possible DMs. The main limitation is the requirement that the DM be structured with binary branches after each chance node. This restriction limits the number of existing DMs that a user could enter retrospectively into the ALCHEMIST system. However, a recent tutorial on medical decision analysis by Detsky and colleagues documents the following recommendations for a decision tree (Detsky et al. 1997b):

1. The tree must have balance[1].
2. The tree must have only two branches after each chance node.
3. The tree must have no embedded decision nodes.
4. The branches must be linked[2].
5. The tree must have symmetry.

These recommendations are designed to help people to produce DMs that function appropriately when sensitivity analyses are performed — yet trees that violate these recommendations can still be considered valid (e.g., the LC-EM violates recommendation #3). However, the ALCHEMIST system can represent trees that adhere to these same principles. ALCHEMIST is restricted to DMs that have binary branches after chance nodes purely so that it may satisfy the SMDM standard for flowchart algorithms. If a new flowchart standard were used this restriction could be relaxed easily.

8.3 Evaluation of the generated CPG

The evaluation of a CPG is an inherently subjective process. Each user of a guideline has a different background and biases and, therefore, will be looking for different information when using a CPG. Each implementer of a CPG will also have a different measure of a CPG's success — whether it be to increase quality of patient care, to decrease practice variation across institutions, to increase patient satisfaction, or to contain costs. The achievement of one of these goals is not necessarily correlated with the achievement of the others. Therefore, any assessment of the CPGs depends on the people who conduct the

1. The structure of outcomes in a decision analysis must reflect the tradeoff between risks and benefits

2. Linking refers to the explicit relationship among probabilities or utilities in the branches that ought to be related. Linkages are achieved by designing for the two branches probability or utility expressions that share common variables, thereby allowing both expressions to vary simultaneously when performing a sensitivity analysis on the common variable (Detsky et al. 1997b)

evaluation and on the factors that they consider (Eddy 1992). A CPG should be reviewed by people who were not involved in the design of the original guideline (independent); who have no stake in the guideline (impartial); and who are knowledgeable of both the medical and analytical aspects of the assessment problem (informed). (Eddy 1992)

I evaluated and critiqued ALCHEMIST's CPG in two ways: (1) I evaluated the quality of the CPG by comparing the CPG to established criteria, and (2) I evaluated the use of the CPG browser and custom-tailoring editor through a structured questionnaire and interview with 15 potential CPG users.

The experiments described in this section required the use of experienced guideline users as subjects. These guideline users were local physicians and health-services researchers who were familiar with CPG use. The guideline users were not asked to use actual patient data or to incorporate the recommended clinical strategies into their patient care; therefore, I did not envision any potential harm to the guideline user or to her future patients. However, because I used human subjects, my evaluation protocol was reviewed and approved by the Human Subjects Committee and informed consent was obtained from each user (Appendix E)

8.3.1 Critique with reference to established guideline criteria

My evaluation does not evaluate directly the efficacy of the generated CPG in achieving its intended goals — such as improving the health status of a given population, or reducing mortality. I also did not evaluate directly whether the generated CPGs increased physician compliance with CPGs. Both of these evaluations would require a clinical trial, which is beyond the scope of this work. Such a clinical trial, however, would be useful before CPGs such as those developed by ALCHEMIST are implemented in a clinical setting.

Instead, I evaluated the quality of the CPG using established criteria. Although a user's attitude toward a guideline may not necessarily reflect her adoption of the recommendation into clinical practice (Weingarten 1995, Tunis 1994, Woo 1985, Romm 1981, Litzelman 1993), I claim that the quality of a CPG serves as an indicator of the future effectiveness of that CPG (Eddy 1982, Eddy 1990a, Eddy 1990b, Eddy 1990c, Eddy 1990d, Eddy 1990e, Eddy 1990f, Basinski 1995).

8.3.1.1 Guideline criteria used

As I described in Section 8.1.1, the Institute of Medicine (IOM) published, in 1992, a set of criteria and an assessment instrument for evaluating CPGs (Institute of Medicine 1992). The IOM assessment instrument rates a guideline with reference to seven desirable attributes. Four of these attributes relate to substantive content: validity, reliability, clinical applicability, and clinical flexibility. The remaining three attributes relate to the process of guideline development or to the actual presentation of the guideline: clarity, multidisciplinary process, and scheduled review. In a separate study, Eddy listed the following five objectives of CPG development (Eddy 1990c).

1. **Accuracy:** The CPG should produce the intended outcomes, as seen by the people who designed the policies.
2. **Accountability:** The documentation provided for the CPG should enable users to review and understand the reasoning behind the CPG.
3. **Predictability:** The health and economic consequences of implementing the CPG should be able to be anticipated on both a societal and individual level.
4. **Defensibility:** The CPG should facilitate resolution of conflicts across policies.
5. **Usability:** The CPG should be able to be used in practice.

Several investigators have developed guideline evaluations based on the IOM criteria (Cluzeau and Littlejohns 1996, Cluzeau et al. 1997, Cluzeau et al. 1995, Scottish Intercollegiate Guidelines Network (SIGN) 1995, Sonnad et al. 1993). To develop an operationalized rating procedure, Sonnad and colleagues combined the IOM criteria with Eddy's objectives (Sonnad et al. 1993). In contrast to the other available evaluations, this rating procedure allows the guideline user to calculate a numeric score for a CPG, thereby allowing quantitative comparison of multiple CPGs. I therefore use the criteria developed by Sonnad and colleagues to assess the quality of the CPGs in my evaluation. The guideline-rating key is provided in Appendix F.

8.3.1.2 Objective

My objective was to evaluate the quality of an ALCHEMIST CPG, and to assess the usability of the ALCHEMIST browser and custom-tailoring editor.

8.3.1.3 Methods

After I determined ALCHEMIST's implementation to be sufficiently complete, I created — using the DM annotation editor — the global CPG for the LC-EM (Appendix D). I then solicited to serve as subjects 15 experienced guideline users affiliated with Stanford's Department of Health Research and Policy or Stanford Medical Informatics. None of the subjects had worked on constructing or testing ALCHEMIST's representation of DMs or CPGs, although one subject had worked extensively on the underlying lung-cancer DM. Ten of the 15 users were physicians, and 12 reported that they had had exposure to current clinical guidelines.

Subjects were asked to compare the quality of two guidelines: the LC-EM CPG produced by ALCHEMIST and a current CPG available over the web for the same clinical domain. The subjects chose the comparison CPG from a set of four current CPGs for the treatment of non–small-cell lung cancer.

I determined the set of comparison guidelines by performing a 1998 Medline search on publication type "Practice Guideline" and using the keyword term "non small cell." This search returned three guidelines (American Society of Clinical Oncology 1997, American

Thoracic Society and European Respiratory Society 1997, Ettinger et al. 1996), two of which were available in full text over the web (American Society of Clinical Oncology 1997, American Thoracic Society and European Respiratory Society 1997). Using the Medical Matrix index of medical resources (http://www.medmatrix.org/SPages/Practice_Guidelines.asp), I then performed a search of the World Wide Web. This search yielded two additional CPGs. Table 8.4 lists the endorsing society and URL for the four comparison guidelines. Each subject was presented with the list of these endorsing societies and was asked to choose which society's guideline would be the comparison. I allowed subjects this choice of current CPGs so as to mimic the choice that practitioners would have available and to allow the users to choose the CPG in which they had the greatest faith.

Table 8.4. Endorsing societies and locations of comparison CPGs.

Endorsing Society	Title of CPG	URL
American Society of Clinical Oncology	Clinical Practice Guidelines for the Treatment of Unresectable Non-Small-Cell Lung Cancer	https://asco.infostreet.com/prof/pp/html/m_lung.htm
National Cancer Institute	Non-Small Cell Lung Cancer	http://imsdd.meb.uni-bonn.de/cancernet/100039.html
Canadian Medical Association	Cancer Management: Lung Tumor Group	http://www.bccancer.bc.ca/cmm/08-1.html
American Thoracic Society	Pretreatment Evaluation of Non-Small-cell Lung Cancer	http://www.ajrccm.org/cgi/content/full/156/1/320#T6

Subjects were asked to complete the 15-item guideline-rating questionnaire for both the LC-EM CPG and the comparison CPG. The subjects were provided with a guideline-rating key that listed additional information on each question in the questionnaire. I randomized the order in which the subjects rated the two CPGs.

I calculated the quality score for a CPG in the following way:

1. For each question in the guideline-rating questionnaire, the user assigned a score of ND, 0, 1, or 2 to the guideline.
2. A score of ND meant that the user could not determine whether the criterion had been met from the guideline and accompanying materials, 0 meant that the user believed that the criterion was not met, 1 meant that the user believed that the criterion was partially or inadequately met, and 2 meant that the criterion was clearly fulfilled.
3. A weighted average of the individual sections and then the overall guideline was then calculated. The ND scores were counted as 0 in the overall score.

After the subjects completed the guideline-rating questionnaire for both guidelines, they used the CPG browser and custom-tailoring editor to perform one to three custom-tailoring tasks (Section 8.3.2). The subjects then completed an end-user computing-satisfaction questionnaire (adapted from Doll, 1988) (Doll and Torkzadeh 1988). In addition, I solicited from each subject answers to five open-ended questions. I include the complete questionnaire in Appendix F. The duration of the evaluation was unconstrained, and the total participation time for each subject was approximately 60 minutes.

The three outcomes of the user-evaluation study were (1) the difference between the quality score of the ALCHEMIST CPG and the comparison CPG, (2) the degree of user satisfaction with the CPG browser and custom-tailoring editor, and (3) the responses to the structured interview.

8.3.1.4 Results

In this section, I report the observed results from the quality scores of the CPG, the user satisfaction questionnaire, and the structured interview.

8.3.1.4.1 Quality score of the CPGs. Table 8.5 summarizes the main results of the evaluation. The mean score for the LC-EM CPG was 1.502 (on a scale from 0 to 2), compared to the comparison-CPG score of 0.987. The difference of 0.515 between these two scores was statistically significant ($p = 0.002$).

Table 8.5. Main results of guideline-rating questionnaire. Comparison 1 = American Society of Clinical Oncology CPG; Comparison 2 = National Cancer Institute CPG; Comparison 3 = Canadian Medical Association CPG; Comparison 4 = American Thoracic Society CPG.

Guideline	Mean	Median	Standard deviation	Difference between means	*p* value
LC-EM CPG	1.502	1.519	0.274		
Comparison CPG	0.987	0.986	0.461	0.515	0.002
Comparison 1 (*n*=5)	1.404	1.384	0.277	0.098	0.26
Comparison 2 (*n*=2)	1.074	1.074	0.209	0.428	0.12
Comparison 3 (*n*=5)	0.646	0.667	0.260	0.856	0.0002
Comparison 4 (*n*=3)	0.803	0.778	0.625	0.699	0.099

The set of comparison CPGs however, was diverse. Although the individual subsets are small, it is interesting to compare the results for the individual comparative CPGs. Of the 15 subjects, four subjects chose the American Society of Clinical Oncology CPG, two the National Cancer Institute CPG, six the Canadian Medical Association CPG, and three the American Thoracic Society CPG. In Table 8.5 I compare these individual CPGs with the LC-EM.

When we separate the group of guideline users who rated the ALCHEMIST LC-EM CPG first from the total group, the mean score for the LC-EM guideline is 1.550. Those subjects who rated the LC-EM guideline second gave it a mean score of 1.460. Thus, the order that the users rated the guidelines did not change significantly the scores on either of the CPGs.

The mean score for the LC-EM CPG was greater than that of the comparison CPG for all questions in the guideline-criteria instrument except for the question asking whether the guideline has been through some form of formal peer review (Table 8.6). Considering only the subscores of the questionnaire, the LC-EM CPG scores for usability, accountability, and accuracy were 1.683, 1.393, and 1.430, respectively; the comparison CPG scores were 1.192, 0.941, and 0.830. The differences between these means were statistically significant ($p < 0.05$) in all cases.

Table 8.6. Summary descriptive statistics for the subjects' responses to individual questions in the guideline-rating questionnaire. An asterisk (*) indicates that the difference between means was statistically significant ($p < 0.05$).

Question	Mean	Median	Standard deviation	Difference between means
Usability				
Is the intervention clearly defined?				
• LC-EM CPG	1.867	2	0.3519	
• Comparison CPG	1.467	1	0.516	0.400*
Is the desired goal of the intervention clearly stated?				
• LC-EM CPG	1.867	2	0.3519	
• Comparison CPG	1.267	1	0.704	0.600*
Does the guideline state explicitly the population to which the statements apply?				
• LC-EM CPG	2.000	2	0	
• Comparison CPG	1.333	1	0.724	0.667*

Table 8.6. Summary descriptive statistics for the subjects' responses to individual questions in the guideline-rating questionnaire. An asterisk (*) indicates that the difference between means was statistically significant ($p < 0.05$).

Question	Mean	Median	Standard deviation	Difference between means
Does the guideline identify the specifically known or generally expected exceptions to this recommendation?				
• LC-EM CPG	1.400	2	0.828	
• Comparison CPG	1.200	1	0.775	0.200
Is guidance offered about how to modify the guideline for differing clinical circumstances?				
• LC-EM CPG	1.467	2	0.743	
• Comparison CPG	0.867	1	0.743	0.600*
Accountability				
Was the guideline developed in a multidisciplinary process?				
• LC-EM CPG	1.600	2	0.632	
• Comparison CPG	1.400	2	0.737	0.200
Is the evidence used in drawing conclusions included?				
• LC-EM CPG	1.267	1	0.704	
• Comparison CPG	1.133	1	0.915	0.133
Can you determine the process used to synthesize evidence and develop the guideline?				
• LC-EM CPG	1.133	1	0.743	
• Comparison CPG	0.733	1	0.799	0.400
Are the assumptions used in guideline development clearly stated?				
• LC-EM CPG	1.933	2	0.258	
• Comparison CPG	0.600	0	0.737	1.333*
Is there a procedure for scheduled reviews included in this guideline?				
• LC-EM CPG	1.133	1	0.516	
• Comparison CPG	0.600	0	0.828	0.533*

Table 8.6. Summary descriptive statistics for the subjects' responses to individual questions in the guideline-rating questionnaire. An asterisk (*) indicates that the difference between means was statistically significant ($p < 0.05$).

Question	Mean	Median	Standard deviation	Difference between means
Accuracy				
Are intermediate events and their relationship to final outcomes clearly stated including intermediate events?				
• LC-EM CPG	1.333	2	0.816	
• Comparison CPG	1.000	1	0.655	0.333
Are the methods of measurement for the intervention and the outcomes clearly stated?				
• LC-EM CPG	1.600	2	0.632	
• Comparison CPG	1.133	1	0.743	0.467
Is the method used in linking the intervention and guideline clearly stated?				
• LC-EM CPG	1.733	2	0.704	
• Comparison CPG	0.733	1	0.799	1.00*
Has the guideline been through some form of formal peer review?				
• LC-EM CPG	0.933	1	0.884	
• Comparison CPG	1.133	1	0.915	-0.200
Are patient preferences explicitly considered in development of the guideline?				
• LC-EM CPG	1.800	2	0.561	
• Comparison CPG	0.400	0	0.632	1.400*

The subject's responses for almost all questions — for both CPGs — varied from 0 to 2. This variability among subjects indicates that the **reliability** of the CPG-critiquing tool was poor. I use the coefficient of variation to express the reliability of the responses. This value is simply the standard deviation of the set of repeated measurements divided by the set of measurement's mean, ordinarily expressed as a percentage. Although the standard deviation itself is a measure of variability, we put it in perspective by relating it to the mean, because the larger the measured values, the more absolute variability we expect (Friedman 1994). The reliability of the LC-EM scores was 18.2%.

8.3.1.4.2 User satisfaction with the ALCHEMIST CPG. Using an ordinal scale of 1 to 5 where 5 is ideal; the subjects rated ALCHEMIST's ease of use at 4.76, the usefulness of the content at 3.98, and the format of the presentation at 4.64. Table 8.7 summarizes these results.

Table 8.7. Responses to the user-satisfaction section of the guideline-rating questionnaire.

Statement about ALCHEMIST	Subject agreement with statement	Agreement (%)[a]
The information is clear	4.333	93
ALCHEMIST provides the precise information you need	3.733	73
ALCHEMIST is user friendly	4.200	87
The information content meets your needs	3.933	73
The output is presented in a useful format	4.600	100
ALCHEMIST is easy to use	4.800	100
ALCHEMIST provides sufficient information	3.467	67
ALCHEMIST provides reports that seem to be just about exactly what you need	3.733	73

a. *Agreement* is defined as the percentage of subjects who assigned either a 4 or 5 to the question.

8.3.1.4.3 Responses to the structured interview. Overall, the subject's experience with the ALCHEMIST system was extremely positive. All subjects who held medical degrees said they would consider using a guideline system such as ALCHEMIST to help them in their clinical practice. The free-form comments from the subjects revealed four broad themes. I list and explore these themes here.

1. *Need for clearer information about exceptions to the CPG:* Several subjects expressed a desire for more information about when the CPG would not apply to their patient population or when there would be other exceptions to treatment recommendations. As I noted in Section 8.1.1.3, these exclusion criteria should be added to my CPG conceptual framework. One subject noted that, because an expert panel does not review each custom-tailored algorithm, she would question the algorithm's accuracy. Clearly, ALCHEMIST requires methods for ensuring the validity of the tailored guidelines; I discussed ALCHEMIST's approach to custom tailoring in Section 6.5.
2. *Need for additional evidence for best-estimate values:* Subject expressed a desire to view the evidence tables on which the CPG was based. Although the evidence tables are part of the ALCHEMIST CPG-CM (Section 4.3), this capability was not implemented when the user evaluation took place. The inclusion of these evidence tables would have strengthened the accountability scores of the ALCHEMIST guideline. One subject said that approval of the CPG by a recognized clinical organization would have increased her faith in the CPG. Although I did not solicit such approval for my example DMs, I envision that it would be part of the peer-review process and would be required before a CPG was implemented into a clinical setting.
3. *Need for help functionality:* Numerous subjects said that they would have liked a help function to describe nonclinical terms such as QALY, societal perspective, tornado diagram, and time-tradeoff. One subject also wanted a more detailed introduction to the site and its functionality. Because my users were for the most part familiar with decision analysis, their concern with this terminology makes the addition of a help functionality to the CPG all the more important.

4. *Need for additional features:* Other suggestions by the subjects pointed to additional features that could be implemented in future work. Such features include a method for indicating visually in the flowchart algorithm the strength of the recommendations (i.e., by the thickness of the line), providing links to the relevant Medline citations, and incorporation of an explanation facility. I discuss the possibility of an explanation module into the ALCHEMIST system in Section 9.2.3.

One further suggestion by one subject was that ALCHEMIST enable a guideline user to suggest an algorithm for the system to critique through evaluation of the underlying DM. This feature would change the overall purpose of the ALCHEMIST CPG from producing the optimal CPG for a given clinical population to critiquing potential CPGs for the population. Although this critiquing could be beneficial to guideline user, it is not the main objective of the ALCHEMIST CPG system. It is however, a way that future researchers could adapt my work.

8.3.1.5 Discussion

There are several factors that confound the findings of this evaluation study. Chief among them is that the users were not blinded to the origin or development method of the CPGs. To thus blind the users, I would have had to eliminate the custom-tailoring abilities of the ALCHEMIST CPG — and thus to eliminate a core element of the system's design. Therefore, I biased the evaluation toward the comparison set of CPGs by allowing the subject to choose the professional society who endorsed the CPG, and by highlighting the greater level and detail of evidence that was presented in the comparison CPGs.

I discussed in Section 8.3.1.4.1 the variability of the subject's responses to the guideline-rating questionnaire in rating the LC-EM CPG. There was a poor interrater reliability of the Sonnad CPG-assessment tool. Although the users had access to the guideline-rating key, they were not trained to use the instrument, and, therefore, there were still several instances when the subjects were confused about what the meaning of the questions was or how to adapt the questionnaire's language to the specific guideline that they were assessing.

My evaluation of the quality of the LC-EM CPG and the user satisfaction with the ALCHEMIST system demonstrated that a guideline created automatically from a DM produced a high-quality and usable CPG as compared to current CPGs. The subjects were generally pleased with the system's presentation of information, with the usefulness of that information, and with the system's ease of use. Because only one CPG-rating instrument was used by only 15 subjects on only the two CPGs, the conclusions and the generalizability of this study are limited. Nevertheless, this study complements the evaluation of the properties of my DM and CPG conceptual models (Section 8.1), and lays the groundwork for future evaluations of automatically generated CPGs.

8.3.2 Evaluation of the CPG custom-tailoring and updating editor

In the final part of my evaluation, I looked at ALCHEMIST's custom-tailoring and updating abilities. I restricted all changes to the CPG to variable changes (e.g., updating the prior probability of mediastinal metastases in the LC-EM), as opposed to structural changes (e.g., adding positron emission tomography (PET) examinations as a possible diagnostic alternative). I evaluated the feasibility of custom tailoring and updating a CPG using the ALCHEMIST system, and the accuracy of the changed CPG (under certain restrictions).

8.3.2.1 Objective

My objective was to evaluate the feasibility and accuracy of ALCHEMIST's custom-tailoring abilities.

8.3.2.2 Methods

To evaluate the feasibility of updating the CPG, I asked the 15 subjects to change input variables of the LC-EM CPG to reflect different clinical scenarios. To help the subjects determine which variables to change, I permitted them to look at the listed sensitive variables and the tornado diagram. For this evaluation, the subjects were able to change any combination of the values listed in the input-variables table. Each subject then recorded whether the algorithm was sensitive to these changes and what effect the changes had on the (quality-adjusted) expected utility. The subjects completed one to three such scenarios.

8.3.2.3 Results and discussion

I considered the CPG accurate if the new CPG produced the same expected outcomes and flowchart algorithm as those obtained directly by the DM and through manual computation. The clinical validity of the new CPG was not assessed. The 15 subjects completed 38 such scenarios, all of which corresponded to the expected utility and algorithm calculated through manual derivation.

All the subjects said that they were pleased with the custom-tailoring abilities of the ALCHEMIST system. Several subjects attempted to predict how changes in the input variables would affect the guideline, and made subsequent changes to test their intuitions. Of note, several subjects — through the use of the custom-tailoring abilities — realized that a change in the CPG did not necessarily correspond to a change in the expected utility (or vice versa).

An additional evaluation metric that was not part of my thesis work uses real clinical scenarios (and population distributions) to determine the expected change in overall health benefit if a guideline user were able to custom tailor a CPG. From such an analysis, I could determine the expected value of custom tailoring for a given CPG.

8.4 Assertions not evaluated

For clarity, I state here explicitly the evaluation tasks that I did not undertake as part of my thesis work. To evaluate the implementation assertions 1 through 3, I would have had to design and perform a randomized clinical trial comparing the use of ALCHEMIST to current standard of care, or to other existing CPGs. Such work is beyond the scope of my thesis.

1. I did not evaluate whether patients' health outcomes were improved because of my CPGs.
2. I did not evaluate whether physician behavior would change because of the implementation of the ALCHEMIST CPGs.
3. I did not evaluate what the cost effectiveness of the ALCHEMIST system is.

All three evaluations are important aspects of the measurement of the quality and effectiveness of a CPG in improving quality of care. Even if guideline raters judge a CPG to be high quality, there are numerous criteria that must be met before the CPG can be linked to improvement of patient outcomes. Guidelines must be disseminated and understood by physicians (Pierre et al. 1991). These physicians must agree with the guideline recommendations (Burack and Liang 1987), and must translate the guideline recommendations into their clinical practice (Grilli et al. 1991, Kosecoff et al. 1987, Lomas et al. 1989, Pierre et al. 1991).

I assumed that the initial DM used in creating the CPG was valid and was created by an experienced decision analyst. Therefore,

4. I did not evaluate to what extent the initial DM was valid.

I envision the assessment of the validity of the DMs to be part of the peer-review process and therefore not a task that ALCHEMIST must address.

Finally, although I included the content of the DM annotation editor in my evaluation of the conceptual framework (Section 8.1), and detailed a portion of the DM annotation requirements (Section 8.2), my research does not concentrate on interface design, and, therefore, I did not evaluate the ease of use of the DM annotation editor by decision analysts, or otherwise assess the interface design:

5. I did not evaluate the ease-of-use or efficiency of the DM annotation-editor interface

8.5 Summary

In this chapter, I described my evaluation of the ALCHEMIST framework and system. To evaluate the contribution of my work, I studied the conceptual models and mapping algorithm, the generated CPG, and ALCHEMIST's custom-tailoring and updating abilities. In my final chapter, I conclude by describing the principal contributions of my work, limitations of my research, and areas for future work.

Chapter 9

Summary, Contributions, and Future Work

In this chapter, I summarize ALCHEMIST (Section 9.1), discuss the current limitations and future extensions of my research (Section 9.2), and describe the contributions of my work to medical informatics, to decision theory, and to health research and policy (Section 9.3).

9.1 A decision-analytic approach to CPG development

Current difficulties faced by guideline users who want to create and maintain clinical-practice guidelines (CPGs) can be attributed to the large amount of required resources (e.g., experts in clinical medicine and in evidence synthesis) that developers need to create high-quality CPGs, and to the static nature of most produced CPGs. In my thesis work, I have explored the reasons for these two problems that impede CPG success. I developed a new approach that allows people to create, disseminate, and tailor CPGs using normative decision models (DMs). As part of my work, I developed the ALCHEMIST system, which automates the DM-to-CPG creation process and allows for tailoring of the generated CPG.

How will ALCHEMIST help guideline developers to create high-quality CPGs that can be custom tailored to specific patient populations? A local guideline developer, using my ALCHEMIST system, can review a globally generated CPG, including its recommendations, underlying evidence, and assumptions. Using the underlying DM, ALCHEMIST can adapt the CPG to reflect the guideline user's specific site or patient population. A guideline developer can also use ALCHEMIST to update automatically the CPG over time as the DM or evidence evolves. With ALCHEMIST, CPG creation is evidence based, adaptable to local specifications, and able to be maintained.

To build such an automated CPG-creation system, I specified DM and CPG conceptual models. I also designed a DM annotation editor to retrieve missing information from the decision analyst to allow the creation and automatic updating of a CPG. This conceptual framework allows ALCHEMIST to produce CPGs based on DMs from various clinical domains, using different DM representations that have a variety of attributes, and representing different types and purposes of CPGs. My approach is designed to improve CPG applicability, relevance, and acceptance by local clinicians and guideline developers, and thus to promote high-quality and cost-effective health care. In my evaluation of the ALCHEMIST system, health-services researchers rated the quality of ALCHEMIST's CPG to be significantly higher than that of currently available guidelines. The subjects also gave high marks to the functionality of the CPG browser and to the system's custom-tailoring abilities.

9.2 Future work

I intend the work presented in this dissertation to lay a foundation for the use of DMs by guideline developers in the automated creation of CPGs. There are therefore many areas for future work. I discussed in Section 8.1.1 extensions of the conceptual framework that will be needed before a system such as ALCHEMIST could be incorporated into clinical practice. In this Section 9.2.1 through 9.2.7, I discuss additional areas of future work that would be required.

9.2.1 Expansion to other types of decision models

ALCHEMIST requires the DM to be in a Decision Maker format. A decision analyst can easily port SMLtree files over to the Decision Maker software file format; therefore, this restriction does not impose a strong limitation if the decision analyst uses conventional decision-tree representations. Although my DM conceptual model can depict influence-diagram representations, the web-based interface requires a Decision Maker file; therefore, use of influence diagrams to create the CPG automatically would require additional software development, but would increase greatly the number of potential users, and, therefore, would increase ALCHEMIST's usefulness.

9.2.2 Incorporation of a controlled clinical vocabulary

ALCHEMIST ignores vocabulary issues. ALCHEMIST does not place any restrictions on the naming conventions that the decision analyst uses when he builds the underlying DM. Also, ALCHEMIST's DM annotation editor queries the decision analyst for free-text definitions for variables and for certain CPG attributes (e.g., the guideline objective). Before a guideline developer could integrate ALCHEMIST with a computer-based patient record and be able to extract automatically the needed patient-specific input data, ALCHEMIST would require the addition of a structured vocabulary to both the DM modeling process and the DM annotation editor. Although such a vocabulary would restrict in some ways the expressiveness of the decision analyst, it would allow easier incorporation of the generated CPG into a hospital information system, and would increase the amount of knowledge that could be extracted automatically from the DM.

9.2.3 Incorporation of an explanation module

The explanation abilities of ALCHEMIST are limited. The presentation of the evidence on which the recommendations are based, and the facilities for the guideline user to make changes to the data and to view resulting CPGs, may give ALCHEMIST a way to explain the resulting recommendations. However, ALCHEMIST does not provide explicit textual output that steps through the reasoning behind the creation of the CPG. A potential extension of my work would combine my generated CPGs with an explanation module (Jimison 1988, Langlotz et al. 1988, Suermondt and Cooper 1992).

9.2.4 Incorporation of a utility-assessment module

Currently, ALCHEMIST treats utility variables the same as any other input variable. An extension of my work would incorporate research on computer-based utility assessment to help the guideline user to determine her patient's utilities (Lenert et al. 1995, Nease et al. 1996, Sanders et al. 1994, Sumner et al. 1991). As an example, George Scott and colleagues have developed SecondOpinion, a web-based program for eliciting patient preferences for relevant health states (Scott et al. 1997). SecondOpinion explains the desired health states to the user, assesses the user's preferences, detects and corrects logic errors in the elicited preferences, and provides feedback to the user on the implications of his preference-assessment results. The incorporation of the SecondOpinion system or of a similar program into ALCHEMIST could help a guideline user to elicit preferences from her patient. Such a preference-assessment method would need to be incorporated before ALCHEMIST could be used to recommend or assist in patient-specific (in contrast to institution- or population-specific) decisions.

9.2.5 Modification of the DM structure

ALCHEMIST restricts tailoring and updating to changes of the input-variable value (e.g., changing the value for the prior probability of mediastinal metastases). Changes to the structure of the model (e.g., adding a positron-emission-tomography (PET) scan as a diagnostic-test choice) are not possible. If it is to be clinically useful, ALCHEMIST may need to allow such structural changes. If guideline users wanted to add a new diagnostic test, they would need to know how the results of that test affected the results of any subsequent tests, and whether there are additional outcomes that the guideline would now need to consider (e.g., what is the risk of death from a PET scan? what additional costs would the patient incur?). Ensuring that the underlying DM is still valid after a guideline user makes a structural change is an area for extensive research.

9.2.6 Integration with web-based host site for sharable guidelines

As part of the InterMed Collaboratory, Stanford, Harvard, and Columbia are collaborating on the creation of methods for sharing clinical guidelines, and development tools for authoring guidelines. An area of possible future research would to work with members of the InterMed team to develop a mapping between ALCHEMIST's generated CPGs and the GLIF language. This mapping would allow CPGs to be maintained at a central resource, custom tailoring to be done on the central CPG resource, with updated CPGs — expressed in GLIF — then downloaded for local application.

9.2.7 Incorporation of a randomized clinical trial-bank

There are efforts underway to establish a worldwide network of standardized, structured knowledge bases for randomized clinical trials (RCTs) (Sim 1997). These trial banks will contain all the information necessary for reasoning about RCTs, and will be accessible to systems such as ALCHEMIST through an open applications-programming interface (API). Although users of my ALCHEMIST system can now review the evidence tables for variables used in the CPG creation, they cannot examine the details of the studies on which the recommendations were based. If ALCHEMIST were linked directly to trial-bank entries of the supporting RCTs, ALCHEMIST users could examine in detail the design, execution, and results of the supporting RCTs, and could examine and use subgroup results to custom tailor the CPG. As new relevant RCTs are published, the trial-bank API would allow ALCHEMIST to identify these trials and to incorporate their results into the CPGs automatically.

9.3 Contributions

My work combines ideas from medical informatics, decision analysis, and health policy to produce a methodology for the automated creation of evidence-based CPGs. The design of the CPG and DM conceptual models combines extensive domain knowledge about the proper structure of CPGs and about the information within DMs. Using ALCHEMIST, I demonstrated that it is possible to transform DMs into CPGs, and to perform automatic updating and tailoring of CPGs. Using my evaluation of the generated CPG browser and custom-tailoring editor, I generated pilot data that will be helpful to guideline developers who design quantitative studies that compare the use of automated CPGs with that of existing guidelines. Descriptions follow of the specific contributions to the domains of medical informatics, decision theory, and health policy.

9.3.1 To medicine and medical informatics

In this dissertation, I have detailed the application of informatics methods to the creation and use of CPGs. Current CPGs, although promoted by policy makers and health-care institutions, have been hindered in effectiveness by their static nature, variety of creation methods, and requirement for intensive resources (for evidence-based CPGs). I have provided a method for combining decision-analytic techniques with a computer-based representation and interface that supports doing this work in medical informatics.

I based my research on my long-term goal of using decision analysis to create CPGs that guideline users are able to use, to custom tailor, and to maintain. My mapping between the DM and CPG conceptual models provides an automated, quantitative link between the clinical data from the literature and CPGs, allowing users to view how changes to the clinical data affect the resulting CPGs and their patient-management strategies. In addition, ALCHEMIST allows the automated construction of CPGs that explicitly represent the uncertainties and evidence inherent in clinical decision-making problems, allowing users to determine the degree to which the resulting CPG is pertinent to their patient population, and thus whether it requires tailoring.

9.3.2 To decision theory

In reviewing current DM taxonomies, I identified three main shortcomings: (1) their inability to represent dual-utility DMs, (2) their inability to model Markov processes, and (3) their inability to model the DM's underlying assumptions. My DM conceptual model combines previous DM research with these additional modeling capabilities to create an extended taxonomy of the knowledge within DMs.

Although DMs provide a normative approach to clinical decision making, guideline developers often do not use DMs to create CPGs. My translation of a DM into a CPG demonstrates an automated method for providing evidence-based CPGs to users unfamiliar with the technical and mathematical details of a DM; it thereby allows guideline developers to obtain the benefits of evidence-based CPG creation.

Although, in my research, I concentrated on clinical DMs, the DM conceptual model and the transformation of the DM into algorithmic form are domain independent, and can be applied to fields other than medicine.

9.3.3 To health research and policy

In reviewing current approaches to CPG creation, I explored and categorized CPG shortcomings and noted that a computer-based method for creating CPGs from DMs can alleviate many of these limitations. I combined ideas from the literature on CPG creation, dissemination, implementation, and maintenance to justify what elements CPGs should have, and I designed a detailed, complete, and extensible CPG conceptual model that includes this knowledge. This conceptual model will provide insight to other health researchers who are interested in the properties of successful CPGs.

In addition, my ALCHEMIST system resolves the tension between performing comprehensive high-quality analyses centrally and accommodating legitimate variation in circumstances, practice settings, and patient preferences.

9.4 Concluding remarks

As health-care costs have continued to rise, policy makers have recognized the need for cost-effective health care. These policy makers consider CPGs to be a potential method for improving quality of patient care, because CPGs provide a means to review patient management and formal descriptions of appropriate levels of care. Existing CPG-development projects have had limited success. In this dissertation, I have argued that a computer-based system that creates CPGs automatically from evidence-based DMs could enhance greatly the efficacy of the generated CPG. I have further argued that conceptual models of DMs and CPGs are necessary for such a successful CPG-creation system. I believe that my work will increase the usefulness of both DMs and CPGs, and will enable a broader audience to incorporate systematic analyses into both policy and clinical decisions. In an era when great importance is placed on defending clinical practice with rigorous supporting evidence, my work offers a powerful approach for mapping from such evidence to familiar guidelines suitable for clinical adoption.

Appendix A

Clinical-Practice Guideline Conceptual Model

Guideline Tasks: Task	Guideline Tasks: Sub-task	CPG Conceptual Model Component
1. Are the recommendations valid?	1. Were all important options and outcomes clearly stated?	a. Options considered
		b. Health outcomes considered
		c. Economic outcomes considered
		d. Definition of terms
	2. Was an explicit and sensible process used to identify, select, and combine evidence?	a. Method used to identify evidence
		b. Method used to combine evidence
		c. Evidence tables
		d. References used in CPG development
		e. Use of expert opinion
		f. Description of event pathway
		g. Modeling assumptions
	3. Was an explicit and sensible process used to consider the relative value of different outcomes?	a. Utility of outcomes

Guideline Tasks		
Task	**Sub-task**	**CPG Conceptual Model Component**
	4. Is the guideline likely to account for important recent developments?	a. Publication date of guideline
		b. Publication date of evidence used
		c. Methods used for updating and maintaining CPG
		d. Sources used
		e. Sensitivity analysis ranges tested
	5. Has the guideline been subject to peer review and testing?	a. Peer review process
		b. Pretesting of CPG
		c. Comparison to existing CPGs
2. What are the recommendations	1. Are practical, clinical important, recommendations made?	a. Clinical flowchart algorithm
		b. Expected outcomes based on CPG
		c. Comparison of CPG to existing recommendations
		d. Burden of disease
		e. Methods for implementing CPG
	2. How strong are the recommendations?	a. Sensitivity analyses
		b. Level of evidence used
		c. Grading of recommendations
	3. What is the quality of the investigators that provide the evidence for the recommendations?	a. Funding sources
		b. Endorsement by professional societies
		c. Authors of CPG
		d. Clinical specialties represented in CPG development team
		e. Potential biases
	4. What is the magnitude and consistency of positive outcomes relative to negative outcomes?	a. Balance sheet of benefits, harms, and costs
		b. Expected outcomes
		c. Flowchart algorithm
		d. Marginal cost-effectiveness analysis
	5. What is the relative value placed on different outcomes?	a. Marginal cost-effectiveness of strategies
		b. Patient preferences for different outcomes

Guideline Tasks		
Task	**Sub-task**	**CPG Conceptual Model Component**
	6. What is the effect of uncertainty associated with the evidence and values used in the guidelines?	a. Best-estimate variable values
		b. Sensitivity analysis ranges
		c. Variables to which the CPG is sensitive
		d. Tornado diagram
3. Will the recommendations help you in caring for your patients?	1. Is the primary objective of your guideline consistent with your objective?	a. Objective of guideline
		b. Targeted health problem
		c. Targeted CPG user
	2. Are the recommendations applicable to your patients?	a. Targeted patient population
		b. Instructions for tailoring guideline to specific populations
		c. Patient preferences
		d. Method for obtaining patient preferences
		e. Sensitive variables
		f. Ranges used in sensitivity analyses
		g. Perspective of CPG

Appendix B

Decision Model Conceptual Model

Decision Modeling Task	DM Conceptual Model Component
1. Define the decision problem	a. Modeling assumptions
	b. Structure of decision model
	c. Patient population
	d. Perspective of the analysis
2. Identify the decision alternatives	a. Available alternatives
	b. Test alternatives
	c. Treatment alternatives
3. List the possible clinical outcomes of each decision alternative	a. Health benefits for each alternative
	b. Health disutilities for each alternative
	c. Economic benefits for each alternative
	d. Economic costs for each alternative
4. Represent the sequence of events leading to the clinical outcomes by a series of chance and decision nodes	a. Structure of the decision model
	b. Children of decision nodes
	c. Children of chance nodes

Decision Modeling Task	DM Conceptual Model Component
5. Choose a time horizon and discount rate for the problem	a. Cycle length for Markov models
	b. Time horizon for acute models
	c. Discount rate for health benefits
	d. Discount rate for costs
6. Determine the probability of each chance outcome	a. Best-estimate variable values
7. Assign a value to each clinical outcome	a. Analytical results
	b. Sensitivity analyses
	c. Threshold analyses
	d. Tornado diagram
8. Additional information	a. Decision analyst
	b. Date decision model was produced

Appendix C

Naming Conventions for Model Variables

(Based on (Nease and Owens 1996)) By convention, all probability variables should begin with a lowercase ‘p’ (e.g., pDie); all utility variables should begin with a lowercase ‘u’ (e.g., uDie); and all switches should begin with a lowercase ‘s’. (A switch is a variable that may take one of a finite number of values. For example, sIndCst might include indirect costs when set at one, and exclude them when set at zero). In Markov analyses, the names of Markov states should be fully capitalized (e.g., WELL). For clarity, other variables should avoid the use of these conventions (e.g., all variables beginning with a lowercase ‘c’ should be cost variables).

The following naming conventions should also be used:

Variable	Meaning
InitAge	Age of the patient at the start of the analysis (an input), in years
Age	Current age of the patient, in years
dCost	Annual discount rate for costs (e.g., 0.05 = 5% annual discount rate)
dHealth	Annual discount rate for health benefits (e.g., 0.05 = 5% annual discount rate)
pMale	The proportion of the cohort that is male (may be excluded in analyses involving a single sex)

Variable	Meaning
CycLen	Length of Markov cycle, in years
tHoriz	Time horizon for the analysis, in years

Appendix D

Global CPG: Lung-Cancer Effectiveness Model

The following pages are the global LC-EM CPG that the subjects used in evaluation of the ALCHEMIST system.

Appendix E

Human Subjects: Informed Consent Form

Consent Form

FOR QUESTIONS ABOUT THIS STUDY, CONTACT:

Gillian Sanders
Medical School Office Building X215
Stanford University
Stanford CA 94305
email: sanders@smi.stanford.edu
phone: (415) 725-3395

DESCRIPTION:
You are invited to participate in a research study on the automated creation of clinical practice guidelines (CPGs) from decision models. You have been selected as a possible participant in this study because you have been identified as a potential guideline user.

If you decide to participate, you will be asked to use a guideline-rating key to score two guidelines. You will be asked to perform some guideline-related tasks using a web-based CPG interface and to evaluate the ease of this system.

RISKS AND BENEFITS:
There is no risk to you or to your patients. The benefit of this study is that it will help us determine the quality, benefit, and use of an automated CPG system based on decision models.

TIME INVOLVEMENT:
Your participation in this experiment will take approximately 45 minutes to one hour.

PAYMENTS:
You will not receive any payment for your participation

SUBJECT'S RIGHTS:
If you have read this form and have decided to participate in this project, please understand your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time without penalty. You have the right to refuse to answer particular questions. Your individual privacy will be maintained in all published and written data resulting from this study.

We will not track any information regarding your identity and will, therefore, assure your confidentiality in our study and in any data published as a result of this study.

If you have any questions about your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact – anonymously, if you wish – the Administrative Panels Office, Stanford University, Stanford, CA (USA) 94305-5532 (or by phone (415) 723-4697 – you may call collect)

The extra copy of this consent form is for you to keep.

SIGNATURE_____________________________________ **DATE**______________________

Appendix F

Clinical-Practice Guideline Evaluation

Clinical-Practice Guideline Evaluation

Please describe yourself, by checking all that apply:

1.	Sex	❑ Female	❑ Male			
2.	Training Level	❑ Pre-Med	❑ MD	❑ MS	❑ PhD	❑ Other
3.	Do you use or have exposure to current clinical practice guidelines?	❑ Yes	❑ No			

For the given clinical-practice guideline (Guideline No.____), please use the following scale to answer questions 4-18:

ND = Cannot determine whether or not the criterion has been met from the guideline and accompanying materials
0 = Criterion is not met
1 = Criterion is partially or inadequately met
2 = Criterion is clearly fulfilled

(You may refer to the attached *Guideline-Rating Key* for a more detailed description of the requirements)

Usability

4	Is the intervention clearly defined?	❑ ND	❑ 0	❑ 1	❑ 2
5	Is the desired goal of the intervention clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
6	Does the guideline state explicitly the population to which the statements apply?	❑ ND	❑ 0	❑ 1	❑ 2
7	Does the guideline identify the specifically known or generally expected exceptions to this recommendation?	❑ ND	❑ 0	❑ 1	❑ 2
8	Is guidance offered about how to modify the guideline for differing clinical circumstances?	❑ ND	❑ 0	❑ 1	❑ 2

Accountability

9	Was the guideline developed in a multidisciplinary process?	❑ ND	❑ 0	❑ 1	❑ 2
10	Is the evidence used in drawing conclusions included?	❑ ND	❑ 0	❑ 1	❑ 2
11	Can you determine the process used to synthesize evidence and develop the guideline?	❑ ND	❑ 0	❑ 1	❑ 2
12	Are the assumptions used in guideline development clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
13	Is there a procedure for scheduled reviews included in this guideline?	❑ ND	❑ 0	❑ 1	❑ 2

Accuracy

14	Are intermediate events and their relationship to final outcomes clearly stated including intermediate events?	❑ ND	❑ 0	❑ 1	❑ 2
15	Are the methods of measurement for the intervention and the outcomes clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
16	Is the method used in linking the intervention and guideline clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
17	Has the guideline been through some form of formal or peer review?	❑ ND	❑ 0	❑ 1	❑ 2
18	Are patient preferences explicitly considered in development of the guideline?	❑ ND	❑ 0	❑ 1	❑ 2

For the given clinical-practice guideline (Guideline No.____), please use the following scale to answer questions 4-18:

ND = Cannot determine whether or not the criterion has been met from the guideline and accompanying materials
0 = Criterion is not met
1 = Criterion is partially or inadequately met
2 = Criterion is clearly fulfilled

(You may refer to the attached *Guideline-Rating Key* for a more detailed description of the requirements)

Usability

4	Is the intervention clearly defined?	❑ ND	❑ 0	❑ 1	❑ 2
5	Is the desired goal of the intervention clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
6	Does the guideline state explicitly the population to which the statements apply?	❑ ND	❑ 0	❑ 1	❑ 2
7	Does the guideline identify the specifically known or generally expected exceptions to this recommendation?	❑ ND	❑ 0	❑ 1	❑ 2
8	Is guidance offered about how to modify the guideline for differing clinical circumstances?	❑ ND	❑ 0	❑ 1	❑ 2

Accountability

9	Was the guideline developed in a multidisciplinary process?	❑ ND	❑ 0	❑ 1	❑ 2
10	Is the evidence used in drawing conclusions included?	❑ ND	❑ 0	❑ 1	❑ 2
11	Can you determine the process used to synthesize evidence and develop the guideline?	❑ ND	❑ 0	❑ 1	❑ 2
12	Are the assumptions used in guideline development clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
13	Is there a procedure for scheduled reviews included in this guideline?	❑ ND	❑ 0	❑ 1	❑ 2

Accuracy

14	Are intermediate events and their relationship to final outcomes clearly stated including intermediate events?	❑ ND	❑ 0	❑ 1	❑ 2
15	Are the methods of measurement for the intervention and the outcomes clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
16	Is the method used in linking the intervention and guideline clearly stated?	❑ ND	❑ 0	❑ 1	❑ 2
17	Has the guideline been through some form of formal or peer review?	❑ ND	❑ 0	❑ 1	❑ 2
18	Are patient preferences explicitly considered in development of the guideline?	❑ ND	❑ 0	❑ 1	❑ 2

ALCHEMIST
CPG Browser and Customization Editor Evaluation

Customization Tasks: Please perform the following tasks for up to three scenarios.

SCENARIO 1:

1. Change any of the given input variables to values within the high and low range. List the chosen variables and their new values here:
2. Click on the “Update Guideline” button at the bottom of the guideline
3. Is the clinical-practice guideline sensitive to these changes?
   - ❑ Yes
   - ❑ No
4. What is the expected utility for the new clinical-practice guideline? ☐

SCENARIO 2:

5. Change any of the given input variables to values within the high and low range. List the chosen variables and their new values here:
6. Click on the “Update Guideline” button at the bottom of the guideline
7. Is the clinical-practice guideline sensitive to these changes?
   - ❑ Yes
   - ❑ No
8. What is the expected utility for the new clinical-practice guideline? ☐

SCENARIO 3:

9. Change any of the given input variables to values within the high and low range. List the chosen variables and their new values here:
10. Click on the “Update Guideline” button at the bottom of the guideline
11. Is the clinical-practice guideline sensitive to these changes?
    - ❑ Yes
    - ❑ No
12. What is the expected utility for the new clinical-practice guideline? ☐

Overall Evaluation: Please answer questions 19-26 using the following scale:

1 = Almost never
2 = Some of the time
3 = Almost half of the time
4 = Most of the time
5 = Almost always

19	Is the information clear?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
20	Does ALCHEMIST provide the precise information you need?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
21	Is ALCHEMIST user-friendly?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
22	Does the information content meet your needs?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
23	Do you think the output is presented in a useful format?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
24	Is ALCHEMIST easy to use?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
25	Does ALCHEMIST provide sufficient information?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
26	Does ALCHEMIST provide reports that seem to be just about exactly what you need?	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5

Please answer questions the questions 27-31 in the space provided.

23. What information, if any, did you feel was missing from the guideline layout?

24. Would you consider using such a guideline system to help you in your clinical care?

25. What are your concerns, if any, for such a system as this?

26. Any additional features which you would like for such a system?

27. Any additional comments or concerns?

THANK YOU VERY MUCH!!!

Guideline Rating Key

This guideline-rating procedure developed by Sonnad and colleagues (Sonnad S, McDonald TW, Nease RF, Oleske J, Owens DK. An evaluation of the methodology of guidelines for zidovudine therapy in HIV disease. *Medical Decision Making*. 1993;13:398.) is based on a combination of the criteria proposed by David M. Eddy and by the Institutes of Medicine (1990). Each numbered component is rated ND, 0, 1, or 2; where:

ND = Cannot determine whether or not the criterion has been met from the guideline and accompanying materials
0 = Criterion is not met
1 = Criterion is partially or inadequately met
2 = Criterion is clearly fulfilled.

The numbered items are then averaged to obtain a score for the overall rating of the guideline.

Usability

Clarity

QUESTION 4: Is the intervention clearly defined?
A clear definition of the intervention should include a clear statement of the intervention procedure (e.g., a list of criteria for the user to identify before implementing the procedure, steps of the procedure itself are clearly understandable) and a clear statement of interventions goal (e.g. reduction of hospitalized days for HIV infected persons). Measurement methods for the intervention should also be included.

2 = If the definition, procedure, and measurement method are all included
1 = Any of the above are missing
0 = Any 2 or more of the above are missing

QUESTION 5: Is the desired goal of the intervention stated clearly in terms of health outcomes (harms or benefits) and economic outcomes?
E.g., broad statement of an increase in the overall QALY/cost ratio for patients under the protocol in the guideline versus the current intervention procedures.

2 = A clear and specific statement of the outcomes (in terms of health status change with or without economic outcomes)
1 = Nonspecific or non health status outcomes that are clearly states
0 = No statement of desired outcomes rates a 0.

Applicability

QUESTION 6: Does the guideline state explicitly the populations to which statements apply?

2 = Explicit statement
1 = Unclear definition of population
0 = No statement

Flexibility

QUESTION 7: Does the guideline identify the specifically known or generally expected exceptions to its recommendations?

Ideally sub populations for which the guideline must be altered will be identified. These subpopulations may be defined demographically or by health status measures. There should be an explicit statement about exceptions. General statements about using clinical judgment are not acceptable for fulfilling this criterion.

0 = Sub populations are explicitly designated
0 = Mention of possible exceptions or a statement that there shouldn't be exceptions,
0 = No mention

QUESTION 8: Is guidance offered about how to modify the guideline for differing clinical circumstances?

2 = Specific modifications to make in various clinical circumstances
1 = General or possible modifications without linkage to specific clinical situations
0 = No modification guidance

Accountability/ Defensibility

Multidisciplinary Process

QUESTION 9: Was the guideline developed in a multidisciplinary process?

Inclusion in the development process only of physicians from varied disciplines is preferable to single discipline guidelines, but ideally the development process should include representatives of all affected groups (e.g. nurses, hospital administrators, pharmacists, computer/data support staff).

2 = Teams that include representatives from other disciplines
1 = Inclusion of a variety of physicians only
0 = None of the above

Documentation

QUESTION 10: Is the evidence used in drawing conclusions included?

This should include references to studies, names of panel members (in the case of consensus opinions being used as evidence), indication of whether evidence is direct or indirect, and reasons for not incorporating evidence contrary to the final guideline recommendations. Evidence used in developing the guideline should be clearly described along with justification for their use in the particular setting of the guideline.

2 = Methods, justification, and contrary evidence all are presented
1 = Any one of the above are missing
0 = Two or more of the above are missing.

QUESTION 11: Can you determine the process used to synthesize evidence and develop recommendations for the guideline?
It should be clear whether global-subjective judgment, evidence-based judgment, outcomes-based judgment, or preference-based judgment was used in the development of the guideline (see Eddy [Eddy, 1990 #206] for full definition of these approaches). Rationale for the method used should also be present

2 = Level is clearly stated and rationale is present
0 = It is unclear, but possible to determine what level of importance incorporation was used
0 = Not indicated.

QUESTION 12: Are the assumptions used in guideline development clearly stated?
These might include assumptions about treatment mortality/morbidity for new interventions, assumptions about degree of patient compliance with guideline.

1 = Assumptions are explicitly stated
1 = Assumptions are implicit, but discernible
0 = Nothing is said

QUESTION 13: Is there a procedure for scheduled reviews included in the guideline?

2 = Explicit schedule and procedure
1 = Mention of review, but no procedure
0 = No mention

Accuracy/ Predictability

Intervention/Outcome

QUESTION 14: Are the intermediate events and their relationship to final outcomes clearly stated including intermediate events?

2 = Quantitative methodologies for establishing these linkages
1 = Logical chains or qualitative reasoning
0 = No established linkages

QUESTION 15: Are the methods of measurement for the outcome clearly stated?
E.g. what scale of quality adjustment will be used, will cost be included in the outcome, is there a method of keeping track of how often the intervention was not correctly administered.

2 = Possible to reproduce the same results about outcomes (e.g. same ratios or utilities from description rate
1 = Described, but not reproducible
0 = Rate not mentioned

QUESTION 16: Is the method used in linking the intervention and guideline clearly stated?
This should include whether quantitative or qualitative methods were employed.

2 = Quantitative methodologies
1 = Qualitative methodologies
0 = No statement of methodology

Review method

QUESTION 17: Has the guideline been through some form of external or peer review?
This would include review of the guideline by experts or peers outside of the development process of the guideline.

2 = It has been through review with suggestions incorporated
1 = Informal review process
0 = No external review

Patient preferences

QUESTION 18: Are patient preferences explicitly considered in development of the guideline?
This should include information on how patient preferences were obtained and how they were incorporated in guideline development.

2 = Quantitative assessments of any type
0 = The guideline explicitly incorporates preferences regardless of method
0 = No mention of patient preferences.

Bibliography

Abendroth, T. W., and R. A. Greenes. 1989. Computer presentation of clinical algorithms. *MD Computing* Sept-Oct: 295-299.

Abendroth, T. W., R. A. Greenes, and E. A. Joyce. 1988. Investigations in the use of clinical algorithms to organize medical knowledge. *SCAMC* : 90.

Agency for Health Care Policy and Research. 1998. Invitation to submit guidelines to the National Guideline Clearinghouse (TM). [Web document] http://www.ahcpr.gov/fund/ngcguidl.htm

American College of Physicians. 1992. Guidelines for counseling postmenopausal women about preventive hormone therapy. *Ann Intern Med* 117: 1038-41.

American College of Physicians. 1994. Screening for ovarian cancer: Recommendations and rationale. *Ann Intern Med* 121: 141-42.

American College of Physicians. 1995. *Clinical practice guidelines (1995 edition)*. American College of Physicians, Philadelphia.

American Medical Association. 1996. *Directory of practice parameters: titles, sources, and updates*. American Medical Association, Chicago.

American Society of Clinical Oncology. 1997. Clinical practice guideline for the treatment of resectable non-small-cell lung caner. *Journal of Clinical Oncology* 15: 2996-3018.

American Thoracic Society, and European Respiratory Society. 1997. Pretreatment evaluation of non-small-cell lung cancer. *American Journal of Respiratory and Critical Care Medicine* 156: 320-32.

Audet, A. M., S. Greenfield, and M. Field. 1990. Medical practice guidelines: current activities and future directions. *Ann Int Med* 113: 703-714.

Barnes M., Barnett G.O. 1995. An architecture for a distributed guideline server. *Proceedings of the nineteenth annual Symposim on Computer Applications in Medical Care.* 233-7.

Basinski, A. S. H. 1995. Evaluation of clinical practice guidelines. *Canadian Medical Association Journal* 153: 1575-1581.

Beck, J. R., and S. G. Pauker. 1983. The Markov process in medical prognosis. *Med Decis Making* 3: 419-458.

Brook, R. 1989. Practice guidelines and practicing medicine. *JAMA* 262: 3027-30.

Burack, R. C., and J. Liang. 1987. The early detection of cancer in the primary care setting: factors associated with the acceptance and completion of recommended procedures. *Prev Med* 16: 739-51.

Carlson, K. J., S. J. Skates, and D. E. Singer. 1994. Screening for ovarian cancer. *Ann Intern Med* 121: 124-32.

Carter, A. O., R. N. Battista, M. J. Hodge, and et al. 1995a. Report on activities and attitudes of organizations active in the clinical practice guidelines field. *Can Med Assoc J* 153: 901-907.

Carter, A. O., R. N. Battista, M. J. Hodge, and et al. 1995b. Proceedings of the 1994 Canadian clinical practice guidelines network workshop. *Canadian Medical Association Journal* 153.

Chassin, M. R., R. H. Brook, J. Keesey, A. Fink, J. Kosecoff, and et al. 1986. Variations in the use of medical and surgical services by the Medicare population. *N Engl J Med* 314: 285-90.

Cimino J. J., Socratous S.A., Clayton P.D. 1995. Automated guidelines implemented via the World Wide Web. *Proceedings of the Nineteenth Annual Symposim on Computer Applications in Medical Care.* 941.

Clemen, R. T. 1996. *Making hard decisions: an introduction to decision analysis (2nd edition)*. Duxbury Press, Belmont.

Cluzeau, F., and P. Littlejohns. 1996. Towards valid clinical guidelines: development of a critical appraisal instrument. *Health Care Risk Report* 2: 16-18.

Cluzeau, F., P. Littlejohns, J. Grimshaw, and G. Feder. 1997. Appraisal instrument for clinical guidelines. St. George's Hopsital Medical School, London.

Cluzeau, F., P. Littlejohns, J. Grimshaw, and A. Hopkins. 1995. Appraising clinical guidelines and the development of criteria - a pilot study. *Journal of Interprofessional Care* 9: 227-235.

Cohen, S. J., M. Weinberger, S. L. Hui, W. M. Tierney, and C. J. McDonald. 1985. The impact of reading on physicians' nonadherance to recommended standards of medical care. *Soc Sci Med* 21: 909.

Conway, A. C., R. B. Keller, and D. E. Wennberg. 1995. Partnering with physicians to achieve quality improvement. *Joint Commission Journal on Quality Improvement* 21: 619-26.

Detsky, A. S., G. Naglie, M. D. Krahn, D. Naimark, and D. A. Redelneier. 1997a. Primer on medical decision analysis: Part 1 - Getting started. *Med Decis Making* 17: 123-125.

Detsky, A. S., G. Naglie, M. D. Krahn, and D. A. Redelmeier. 1997b. Primer on medical decision analysis: Part 2 - Building a tree. *Med Decis Making* 17: 126-135.

Deibel, S. R. A. 1996. [Web document] Introduction to the InterMed common Guideline Model and Guideline Interchange Format. http://dsg.harvard.edu/public/intermed/glif_overview.htm

Doll, W. J., and G. Torkzadeh. 1988. The assessment of end-user computing satisfaction. *MIS Quarterly* 12: 259-274.

Eagle K.A., Mulley A.G., Skates S.J. et al. 1990. Length of stay in the intensive care unit: effects of practice guidelines and feedback. *JAMA*. 264: 922-7.

ECRI. 1997. Health care standards, 1997. *Plymouth Meeting, PA*. ECRI, Plymouth, PA.

Eddy D. M.. 1982. Clinical policies and the quality of clinical practice. *N Engl J Med* 307: 343-347.

Eddy, D. M. 1990a. Comparing benefits and harms: the balance sheet. *JAMA* 263: 2493-2505.

Eddy, D. M. 1990b. Designing a practice policy, Standards, guidelines, and options. *JAMA* 263: 3077-3084.

Eddy, D. M. 1990c. Guideline for policy statements: the explicit approach. *JAMA* 263: 2239-43.

Eddy, D. M. 1990d. Practice policies: guidelines for methods. *JAMA* 263: 1839-1841.

Eddy, D. M. 1990e. Practice policies: what are they? *JAMA* 263: 877-80.

Eddy D. M., 1990f. Anatomy of a decision. *JAMA* 263: 441-3.

Eddy, D. M. 1991a. Screening for breast cancer. Pages 229-254 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Eddy, D. M. 1991b. Screening for cervical cancer. Pages 255-285 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Eddy, D. M. 1991c. Screening for colorectal cancer. Pages 286-311 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Eddy, D. M. 1991d. Screening for lung cancer. Pages 312-325 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Eddy, D. M. 1992. *A manual for assessment of health practices and designing practice policies: the explicit approach.* The American College of Physicians, Philadelphia, PA.

El-Sadr, W., J. M. Oleske, B. D. Agins, K. A. Bauman, C. L. Brosgart, G. M. Brown, and et al. January 1994. *Evaluation and management of early HIV infection. Clinical Practice Guideline No. 7.* Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, AHCPR Publication No. 94-0572. Rockville, MD.

Ettinger, D. S., J. D. Cox, R. J. Ginsberg, R. Komaki, M. G. Kris, R. B. Livingston, and D. J. Sugarbaker. 1996. NCCN Non-small-cell lung cancer practice guidelines. *The National Comprehensive Cancer Network Oncology* 10: 81-111.

Evidence-Based Medicine Working Group. 1992. Evidence-based medicine. A new approach to the practice of medicine. *JAMA* 268: 2420-5.

Fahs, M. C., J. Mandelblatt, C. Schechter, and C. Muller. 1992. Cost effectiveness of cervical cancer screening for the elderly. *Ann Intern Med* 117: 520-7.

Fisher, E. S., H. G. Welch, and J. E. Wennberg. 1992. Prioritizing Oregon's hospital resources. An example based on variations in discretionary medical utilization [see comments]. *JAMA* 267: 1925-31.

Fishman, N., and M. Bronstein. 1975. Is mediastinoscopy necessary in the evaluation of lung cancer? *Annals of Thoracic Surgery* 20: 678-686.

Fletcher, R. H., and S. W. Fletcher. 1990. Clinical practice guidelines. *Ann Intern Med* 113: 645-6.

Fridmsa, D. B., J. H. Gennari, and M. A. Musen. 1996. Making generic guidelines site-specific. Pages 597-601 in J. J. Cimino, ed. *AMIA Annual Fall Symposium*. Hanley & Belfus, Washington, DC.

Friedman, G. D. 1994. *Primer of epidemiology (4th edition)*. McGraw-Hill, Inc., New York, NY.

Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein, eds. 1996. *Cost-effectiveness in health and medicine*. Oxford University Press, New York, NY.

Goldman, L. 1996. The value of cardiology. *Med Decis Making* 335: 1918-9.

Gottlieb, K., H. N. Sokol, K. O. Murrey, and S. C. Schoenbaum. 1992. Algorithm-based clinical quality improvement. Clinical guidelines and continuous quality improvement. *HMO Practice* 6: 5-12.

Gould, M. K., R. F. Nease, H. C. Sox, and D. K. Owens. 1997. A decision model for mediastinal staging in non-small cell lung cancer. *Journal of General Internal Medicine* 12: 61.

Grady, D., S. M. Rubin, D. B. Petitti, C. S. Fox, D. Black, B. Ettinger, V. L. Ernster, and S. R. Cummings. 1992. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med* 117: 1016-37.

Granata, A. V., and A. L. Hillman. 1998. Competing practice guidelines: Using cost-effectiveness analysis to make optimal decisions. *Ann Intern Med* 128: 56-63.

Grilli, R., G. Apolone, S. Marsoni, A. Nicolucci, P. Zola, and A. Liberati. 1991. The impact of patient management guidelines on the care of breast, colecteral, and ovarian cancer patients in Italy. *Med Care* 29: 50-63.

Grimshaw, J. M., and I. T. Russell. 1993. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 342: 1317-1322.

Hayward, R. S., M. C. Wilson, S. R. Tunis, E. B. Bass, and G. Guyatt. 1995. User's guides to the medical literature. VIII. How to use the clinical practice guidelines. A. Are the recommendations valid? The Evidence-Base Medicine Working Group. *JAMA* 274: 570-4.

Hayward, R. S., M. C. Wilson, S. R. Tunis, G. H. Guyatt, K. A. Moore, and E. B. Bass. 1996. Practice guidelines. What are internists looking for? *Journal of General Internal Medicine* 11: 176-8.

Hayward, R. S. A. 1995. Proposal to develop and evaluate a prototype registry and database of clinical practice guidelines. Guideline Appraisal Project, McMaster University, Hamilton, Ontario.

Hayward, R. S. A., G. H. Guyatt, K. A. Moore, K. A. McKibbon, and A. O. Carter. 1997. Canadian physicians' attitudes about and preferences regarding clinical practice guidelines. *CMAJ* 156: 1715-23.

Hayward, R. S. A., and A. Laupacis. 1993. Initiating, conducting, and maintaining guidelines development programs. *Can Med Assoc J* 148: 507-512.

Hayward, R. S. A., M. D. Wilson, and S. R. Tunis. 1993. More informative abstracts of articles on clinical practice guidelines. *Ann Intern Med* 118: 731-737.

Health Services Research Group. 1992a. Small-area variations: What are they and what do they mean? *Canadian Medical Association Journal* 146: 467-70.

Health Services Research Group. 1992b. Standards, guidelines, and clinical policies. *Can Med Assoc J* 146: 833-837.

Holland, R. R. 1975. Decision tables: Their use for the presentation of clinical algorithms. *JAMA* 233: 455-457.

Hollenberg, J. P. 1984. The decision tree builder: an expert system to simulate medical prognosis and management. *Medical Decision Making* 4: 531.

Howard, R. 1988. Decision analysis. Practice and promise. *Management Science* 34: 679-695.

Institute of Medicine. 1992. *Guidelines for clinical practice: From development to use*. National Academy Press, Washington, DC.

Iscoe, N. A., V. Goel, K. Wu, G. Fehringer, E. J. Holowaty, and C. D. Naylor. 1994. Variation in breast cancer surgery in Ontario. *Canadian Medical Association Journal* 150: 345-52.

Jimison, H. B. 1988. A representation for gaining insight into clinical decision models. *Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care* : 110-113.

Kaegi, L. 1996. From paper to practice to point of care: Reports from a Zitter Group Conference on Implementing Guidelines. *Journal on Quality Improvement* : 551-589.

Kamae, I. R., and R. A. Greenes. 1991. A computational model of approximate bayesian inference for associating clinical algorithms with decision analyses. *SCAMC* : 691-5.

Kassirer, J. P. 1993. The quality of care and the quality of measuring it [editorial]. *N Engl J Med* 329: 1263-65.

Keller, R. B., D. N. Soule, J. E. Wennberg, and D. F. Hanley. 1990. Dealing with geographic variations in the use of hospitals. The experience of the Maine Medical Assessment Foundation Orthopaedic Study Group. *Journal of Bone and Joint Surgery* 72: 1286-93.

Kosecoff, J., D. E. Kanouse, W. H. Rogers, L. McCloskey, C. M. Winslow, and R. H. Brook. 1987. Effects of the National Institutes of Health Consensus Development Program on physician practice. *JAMA* 258: 2708-2713.

Krahn, M. D., G. Naglie, D. Naimark, D. A. Redelmeier, and A. S. Detsky. 1997. Primer on medical decision analysis: Part 4 - Analyzing the model and interpreting the results. *Med Decis Making* 17: 142-151.

Kulikowski, C. A. 1996. AIM: Quo Vadis? *JAMIA* 3.

Langlotz, C. 1989. A decision theoretic approach to heuristic planning. *Medical Information Sciences*. Stanford University, Stanford.

Langlotz, C. P., E. H. Shortliffe, and L. M. Fagan. 1988. A methodology for generating computer-based explanations of decision-theoretic advice. *Med Decis Making* 8: 290-303.

Larsson, S. 1976. Mediastinoscopy in bronchogenic cancer: A study of 486 cases with special reference to indications and limitations. *Scan J Cardiovasc Surg* 19 (Suppl): 1-23.

Leape, L. L. 1990. Practice guidelines and standards. An overview. *QRB Qual Rev Bull* 16: 42-49.

Lenert, L. A., D. Michelson, C. Flowers, and M. R. Bergen. 1995. IMPACT: an object-oriented graphical environment for construction of multimedia patient interviewing software. *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care* : 319-23.

Liem, E. B., J. S. Obeid, P. Shareck, L. Sato, and R. A. Greenes. 1995. Representation of clinical practice guidelines through an interactive World-Wide-Web interface. Pages 223-227. *SCAMC*, New Orleans, LA.

Littenberg, B., A. M. Garber, and H. C. Sox, Jr. 1991. Screening for hypertension. Pages 22-46 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Litzelman D.K., Dittus R.S., Miller M.E., Tierney W.M. 1993. Requiring physicians to respond to computerized reminders improves their compliance with preventive care protocols. *J Gen Intern Med* 8: 311-7.

Lomas, J., G. M. Anderson, K. Domnick-Pierre, and et. al. 1989. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med* 321: 1306-11.

Lomas, J., and R. B. Haynes. 1988. A taxonomy and critical review of tested strategies for the application of clinical practice recommendations: From 'official' to 'individual' clinical policy. *American Journal of Preventive Medicine* 4 (suppl): 77-94.

Maiman, L. A., P. Greenland, N. G. Hildreth, and C. Cox. 1991. Patterns of physicians' treatments for referral patients from public cholesterol screening. *Am J Prev* 7: 273-9.

Manton, K. G., and E. Stallard. 1988. *Chronic disease modeling: Measurement and evaluation of the risks of chronic disease processes*. Oxford University Press, New York, NY.

Margolis, C. Z. 1983. Uses of clinical algorithms. *JAMA* 249: 627.

McNeil, B. J., R. Weichselbaum, and S. G. Pauker. 1978. Fallacy of the five-year survival in lung cancer. *N Engl J Med* 299: 1397-1401.

Melton, J. L., III, D. M. Eddy, and C. C. Johnston. 1991. Screening for osteoporosis. Pages 202-28 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Musen M.A., Tu S.W., Das A.K., Shahar Y. 1996. EON: A component-based approach to automation of protocol-directed therapy. *JAMIA*, 3: 367-388.

Mutnick, A. H., and B. A. Szymusiak-Mutnick. 1996. Locally derived clinical practice guidelines using a decision analysis model. *AACN Clin Issues* Aug 7: 448-55.

Naglie, G., M. D. Krahn, D. Naimark, D. A. Redelmeier, and A. S. Detsky. 1997. Primer on medical decision analysis: Part 3 - Estimating probabilities and utilities. *Med Decis Making* 17: 136-141.

Naimark, D., M. D. Krahn, G. Naglie, D. A. Redelmeier, and A. S. Detsky. 1997. Primer on medical decision analysis: Part 5 - Working with Markov processes. *Med Decis Making* 17: 152-159.

National Centre for Reviews and Dissemination. 1994. Implementing clinical practice guidelines: can guidelines be used to improve clinical practice? NHS Centre for Reviews and Dissemination, York.

Nease, R. F., L. H. Hynes, and B. Littenberg. 1996. Automated utility assessment of global health. *Qual Life Res* 5: 175-82.

Nease, R. F., and D. K. Owens. 1991. Decision models as an aid to short-term evaluation of clinical guidelines [abstract]. *Med Decis Making* 11: 325.

Nease, R. F., and D. K. Owens. 1994. A method for estimating the cost-effectiveness of incorporating patient preferences into practice guidelines. *Medical Decision Making* 14: 382-392.

Nease, R. F., and D. K. Owens. 1996. Naming conventions for model variables.

Nease, R. F., and D. K. Owens. 1997. Use of influence diagrams to structure medical decisions. *Medical Decision Making* 17: 263-275.

Oddone, E. Z., G. Samsa, and D. B. Matchar. 1994. Global judgments versus decision-model-facilitated judgments: are experts internally consistent? *Medical Decision Making* 14: 19-26.

Ohno-Machado L., Gennari J.H., Murphy S., et al. 1998. The guideline interchange format: A model for representing guidelines. *JAMIA* (in press).

Owens, D. K. 1994. Commentary on "Clinical guidelines and medical practice: A review". *Ann Internal Med* 120: 85.

Owens, D. K. 1998. Spine update. Patient preferences and the devlopment of practice guidelines. *Spine* 23: 1073-9.

Owens, D. K., and R. F. Nease. 1991a. Evaluating the impact of variation in clinical circumstance on optimal guideline development: the expected value of customization [abstract]. *Med Decis Making* 11: 330.

Owens, D. K., and R. F. Nease. 1991b. A model-based approach for prioritizing information acquisition during guideline development [abstract]. *Med Decis Making* 11: 330.

Owens, D. K., and R. F. Nease. 1993. Development of outcome-based practice guidelines: a method for structuring problems and synthesizing evidence. *Joint Commission Journal on Quality Improvement* 19: 248-263.

Owens, D. K., and R. F. Nease. 1997. A normative analytic framework for development of practice guidelines for specific clinical populations. *Med Decis Making* 17: 409-426.

Owens, D. K., R. F. Nease, and R. A. Harris. 1993. Use of cost-effectiveness and value of information analyses to customize guidelines for specific clinical practice settings [abstract]. *Med Decis Making* 13: 395.

Owens, D. K., G. D. Sanders, R. A. Harris, K. M. McDonald, P. A. Heidenreich, A. D. Dembitzer, and M. A. Hlatky. 1997a. Cost effectiveness of implantable cardioverter defibrillators relative to amiodarone for prevention of sudden cardiac death. *Annals of Internal Medicine* 126: 1-12.

Owens, D. K., R. D. Schacter, and R. F. Nease. 1997b. Representation and analysis of medical decision problems with influence diagrams. *Med Decis Making* 17: 241-62.

Owens, D. K., and H. C. J. Sox. 1990. Medical decision making: Probabilistic medical reasoning. Pages 70-116 in E. Shortliffe, L. Perreault, L. Fagan, and G. Widerhold, eds. *Medical Informatics: Computer Applications in Health Care*. Addison-Wesley, Reading, MA.

Owens, D. K., H. C. J. Sox, and S. K. Inouye. 1989. Strategies for test selection in the staging of lung neoplasms. *Seminars in Respiratory Medicine* 10: 195-202.

Oxman, A. D. 1993. Coordination of guidelines development. *CMAJ* 148.

Parmigiani, G., M. Ancukiewicz, and M. D. (submitted for publication). Decision models in clinical recommendations development: the Stroke Prevention Policy Model. .

Physician Payment Review Commission. 1995. Annual report to congress 1995. US Government Printing Office, Washington DC.

Pierre, K. D., E. Vayda, J. Lomas, M. W. Enkin, and G. M. Anderson. 1991. Obstetrical attitudes and practices before and after the Canadian Consensus Conference Statement on Cesarean Birth. *Soc Sci Med* 32: 1283-9.

Reinwald, L. T., and R. M. Soland. 1966. Conversion of limited entry decision tables to optimal computer programs I: Minimum average processing time. *Journal of the ACM* 13: 339-358.

Reinwald, L. T., and R. M. Soland. 1967. Conversion of limited entry decision tables to optimal computer programs II: Minimum storage requirement. *Journal of the ACM* 14: 742-755.

Romm, F. J., S. W. Fletcher, and B. S. Hulka. 1981. The periodic health examination: comparison of recommendations and internists' performance. *South Med J* 74: 265-71.

Sanders, G. D., A. D. Dembitzer, P. A. Heidenreich, K. M. McDonald, and D. K. Owens. 1996. Presentation and explanation of medical decision models using the world wide web. Pages 60-64. *Proceedings of the 1996 AMIA Annual Fall Symposium*, Washington, DC.

Sanders, G. D., C. G. Hagerty, F. A. Sonnenberg, M. A. Hlatky, and D. K. Owens. 1998. Distributed decision support for guideline development using a web-based interface: prevention of sudden cardiac death. Stanford University, Stanford.

Sanders, G. D., R. A. Harris, M. A. Hlatky, and D. K. Owens. 1995. Prevention of sudden cardiac death: a probabilistic model for decision support. Pages 258-262. *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care*, New Orleans, LA.

Sanders, G. D., D. K. Owens, N. Padian, A. B. Cardinalli, A. N. Sullivan, and R. F. Nease. 1994. A computer-based interview to identify HIV risk behaviors and to assess patient preferences for HIV-related health states. *Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care* : 20-24.

Schapira, M. M., D. B. Matchar, and M. J. Young. 1993. The effectiveness of ovarian cancer screening. A decision analysis model. *Ann Intern Med* 118: 838-43.

Schoenbaum, S. C., and L. K. Gottlieb. 1990. Algorithm based improvement of clinical quality. *BMJ* 301: 1374-1376.

Scott, G. C., D. J. Cher, and L. A. Lenert. 1997. SecondOpinion: Interactive web-based access to a decision model. *Proc AMIA Annual Fall Symposium* : 769-773.

Scottish Intercollegiate Guidelines Network (SIGN). 1995. Draft criteria for appraisal of clinical guidelines for national use. [Web document].

Shapiro, D. W., R. D. Lasker, A. B. Bindman, and P. R. Lee. 1993. Containing costs while improving quality of care: The role of profiling and practice guidelines. *Ann Rev Publ Health* 14: 219-41.

Shiffman, R. N. 1991. A critique of published medical logic representation schemata: how decision tables can help. [Abstract]. *Section on Computers, American Academy of Pediatrics* New Orleans.

Shiffman, R. N. 1995. Guideline maintenance and revision: 50 years of the Jones criteria for diagnosis of rheumatic fever. *Archives of Pediatrics and Adolescent Medicine* 149: 727-32.

Shiffman, R. N., and R. A. Greenes. 1991. Use of augmented decision tables to convert probabilistic data into clinical algorithms for the diagnosis of appendicitis. *SCAMC* : 686-90.

Shiffman, R. N., and R. A. Greenes. 1992. Rule set reduction using augmented decision table and semantic subsumption techniques: application to cholesterol guidelines. *SCAMC* : 339-43.

Shiffman, R. N., and R. A. Greenes. 1994. Improving clinical guidelines with logic and decision table techniques: application to hepatitis immunization recommendations. *MDM* 14: 245-54.

Shiffman, R. N., L. L. Leape, and R. A. Greenes. 1993. Translation of appropriateness criteria into practice guidelines: application of decision table techniques to the RAND criteria for coronary artery bypass graft. *SCAMC* : 248-52.

Shortliffe, E. H. 1997. Intermed Grant Proposal. Stanford University, Stanford.

Shortliffe, E.H. 1990 Medical informatics and clinical decision making: The science and the Pragmatics. *Med Dec Making* 11: S2-S14.

Sim, I. 1996. Sharable database of randomized clinical trials: In support of computer-assisted evidence-based medicine. *Section on Medical Informatics*. Stanford University, Stanford.

Sim, I. 1997. Trial Banks: An informatics foundation for evidence-based medicine. Pages 246. *Medical Information Sciences*. Stanford University, Stanford.

Singer, D. E., J. H. Samet, C. M. Coley, and D. M. Nathan. 1991. Screening for diabetes mellitus. Pages 154-178 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Society for Medical Decision Making (SMDM) Committee on Standardization of Clinical Algorithms (CSCA). 1992. Proposal for clinical algorithm standards. *Med Dec Making* 12: 149-54.

Sonnad, S., T. W. McDonald, R. F. Nease, J. Oleske, and D. K. Owens. 1993. An evaluation of the methodology of guidelines for zidovudine therapy in HIV disease. *Medical Decision Making* 13: 398.

Sonnenberg, F. A. 1997. Decision analytic support for clinical guidelines. UMDNJ/R.W.J. Medical School, New Brunswick, NJ.

Sonnenberg, F. A., and J. R. Beck. 1993. Markov models in medical decision making: a practical guide. *Medical Decision Making* 13: 322-38.

Sonnenberg, F. A., C. G. Hagerty, and C. A. Kulikowski. 1994. An architecture for knowledge based construction of decision models. *Med Decis Making* 14: 27-39.

Sonnenberg, F. A., and S. G. Pauker. 1987. Decision maker: an advanced personal computer tool for clinical decision analysis. *Proceedings of the Eleventh Annual Symposium Computer Applications in Medical Care*. IEEE Computer Society, Washington, DC.

Sonnenberg, F. A., and S. G. Pauker. 1997. Decision Maker 7.0: Operating Manual. Division of Clinical Decision Making, New England Medical Center, Boston.

Sox, H. C., M. A. Blatt, M. C. Higgins, and K. I. Marton. 1988. *Medical decision making*. Butterworth-Heinemann, Boston, MA.

Sox, H. C., Jr., A. M. Garber, and B. Littenberg. 1991a. The resting electrocardiogram as a screening test: A clinical analysis. Pages 47-80 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Sox, H. C., Jr., B. Littenberg, and A. M. Garber. 1991b. The role of exercise testing in screening for coronary artery disease. Pages 81-112 in D. M. Eddy, ed. *Common Screening Tests*. American College of Physicians, Philadelphia.

Sox, H. C., and S. H. Woolf. 1993. Evidence-based practice guidelines from the US Preventive Services Task Force. *JAMA* 269: 2678.

St. Louis, M., K. J. Rauch, L. R. Petersen, J. E. Anderson, C. A. Schable, and T. J. Dondero. 1990. Seroprevalence rates of human immunodeficiency virus infection at sentinel hospitals in the United States. The Sentinel Hospital Surveillance Group. *N Engl J Med* 323: 213-8.

Stoufflet P., Deibel S., Traum J., Greenes R.A. 1995. State transition method of modeling clinical encounters. *American Medical Informatics Association Spring Congress*.

Stoufflet P.E., Ohno-Machado L. Deibel S.R.A., Lee D., Greenes R.A. 1996. GEODE-CM: A state transition framework for clinical management. *JAMIA Symposium supplements.* 924.

Suermondt, H. J., and G. F. Cooper. 1992. An evalutation of explanations of probabilistic inference. *SCAMC* : 579-85.

Sumner, W., R. Nease, and B. Littenberg. 1991. U-titer: A utility assessment tool. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care* : 701-5.

Tunis, S. R., R. S. A. Hayward, M. C. Wilson, and et al. 1994. Internists' attitudes about clinical practice guidelines. *Ann Intern Med* 120: 956-63.

US Congress Office of Technology Assessment. 1994. Idenitifying health technologies that work: Searching for evidence. US Government Printing Office, Washington DC.

US Department of Health and Human Service. 1995. *Using clinical practice guidelines to evaluate quality of care. Volume 1: Issues*. U.S. Dept. of Health and Human Services, Rockville, MD.

US Preventive Services Task Force. 1996. *Guide to clinical preventive services*. Williams & Wilkins, Baltimore.

Wachtel T.J., O'Sullivan P. 1990. Practice guidelines to reduce testing in the hospital. *J Gen Intern Med.* 5: 335-41.

Walker, D. R., M. O. Howard, and M. D. Lambert. 1994. Medical practice guidelines. *West J Med* 161: 39-44.

Wears, R. L., P. H. Stenklyft, and R. C. Luten. 1994. Using decision tables to verify the logical consistency and completeness of clinical practice guidelines: Fever without source in children under three. *Academic Emergency Medicine* 1: A35.

Weingarten S., Agocs L., Tankel N., et al. 1993. Reducing lengths of stay for patients hospitalized with chest pain using medical practice guidelines and opinion leaders. *Am J Cardiol*. 71: 259-62.

Weingarten, S. R., M. S. Riedinger, L. Conner, and et al. 1994. Practice guidelines and reminders to reduce duration of hospital stay for patients with chest pain. An interventional trial. *Ann Intern Med* 120: 257-63.

Weingarten S., Sotone E., Hayward R. et al. 1995. The adoption of preventive care practice guidelines by primary care physicians: Do actions match intentions. *J Gen Intern Med* 10: 138-144.

Weinstein M.C., Stason W.B. 1977. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 296 (13): 716-21.

Welch, W. P., M. E. Miller, H. G. Welch, E. S. Fisher, and J. E. Wennberg. 1993. Geographic variation in expenditures for physicians' services in the United States [see comments]. *New Engl J Med* 328: 621-7.

Wellman, M. P., M. H. Eckman, C. Fleming, S. L. Marshall, F. A. Sonnenberg, and S. G. Pauker. 1989. Automated critiquing of medical decision trees. *Med Decis Making* 9: 272-284.

Wennberg, J., and A. Gittelsohn. 1973. Small area variations in health care delivery. *Science* 182: 1102-8.

Wilson, M. C., R. S. Hayward, S. R. Tunis, E. B. Bass, and G. Guyatt. 1995. User's guides to the medical literature. VIII. How to use clinical practice guidelines. B. what are the recommendations and will they help you in caring for your patients? The Evidence-Based Medicine Working Group. *JAMA* 275: 1630-2.

Woo B., Woo B., Cook E.F., Weisberg M., Goldman L 1985. Screening procedures in the asymptomatic adult. Comparison of physician’s recommendations, patients’ desires, published guidelines, and actual practice. *JAMA* 254: 1480-4.

Woolf, S. H. 1990. Practice guidelines: A new reality in medicine. I. Recent developments. *Arch Intern Med* 150: 1811-18.

Woolf, S. H. 1991. Manual for clinical practice guideline development: A protocol for expert panels convened by the office of quality and effectiveness in health care. AHCPR, Rockville, MD.

Woolf, S. H. 1992. Practice Guidelines: A new Reality in Medicine. II. Methods of Developing Guidelines. *Arch Intern Med* 152: 946-952.

Woolf, S. H. 1998. Do clinical practice guidelines define good medical care? The need for good science and the disclosure of uncertainty when defining "best practices". *Chest* 113: 166S-171S.

Woolf, S. H., R. N. Battista, G. M. Anderson, A. G. Logan, and E. Wang. 1990. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods of reviewing evidence and developing clinical practice recommendations. *J Clin Epidemiol* 43: 891-905.