Report Number: WUCS-93-12

1993-01-01

Research Proposal: Preference Acquisition through Reconciliation of Inconsistencies

Nilesh L. Jain

The quality of performance of a decision-support system (or an expert system) is determined to a large extent by its underlying preference model (or knowledge base). The difficulties in preference and knowledge acquisition make them a major focus of current research in decision-support and expert systems. Researchers have used various concepts to develop promising acquisition techniques. One of the concepts used is knowledge maintenence where the knowledge base is changed in response to incorrect or inadequate performance by the expert system. This dissertation investigates a preference acquisition technique based on the reconciliation of inconsistencies between the preference model and... **Read complete abstract on page 2.**

Research Proposal: Preference Acquisition through Reconciliation of Inconsistencies

Nilesh L. Jain

Complete Abstract:

The quality of performance of a decision-support system (or an expert system) is determined to a large extent by its underlying preference model (or knowledge base). The difficulties in preference and knowledge acquisition make them a major focus of current research in decision-support and expert systems. Researchers have used various concepts to develop promising acquisition techniques. One of the concepts used is knowledge maintenence where the knowledge base is changed in response to incorrect or inadequate performance by the expert system. This dissertation investigates a preference acquisition technique based on the reconciliation of inconsistencies between the preference model and the decision maker by allowing the decision maker to modify the preference model interactively. The technique can be used in the class of decision-support systems which objectively evaluate competing plans and select the best plan. The technique will be implemented in the domain of evaluating three-dimensional (3-D) radiation treatment plans. Another major aim of the dissertation is to develop a clinically-relevant objective plan-evaluation model for 3-D radiation treatment plans, and to build a clinical decision-support system to assist in that task using the new preference acquisition method.

Research Proposal:
Preference Acquisition through Reconciliation of Inconsistencies

Nilesh L. Jain

WUCS-93-12

April 8, 1993

Division of Medical Informatics
Campus Box 8005
Washington University School of Medicine
660 S. Euclid Ave
St. Louis MO 63110

Abstract

The quality of performance of a decision-support system (or an expert system) is determined to a large extent by its underlying preference model (or knowledge base). The difficulties in preference and knowledge acquisition make them a major focus of current research in decision-support and expert systems. Researchers have used various concepts to develop promising acquisition techniques. One of the concepts used is knowledge maintenance where the knowledge base is changed in response to incorrect or inadequate performance by the expert system. This dissertation investigates a preference acquisition technique based on the reconciliation of inconsistencies between the preference model and the decision maker by allowing the decision maker to modify the preference model interactively. The technique can be used in the class of decision-support systems which objectively evaluate competing plans and select the best plan. The technique will be implemented in the domain of evaluating three-dimensional (3-D) radiation treatment plans. Another major aim of the dissertation is to develop a clinically-relevant objective plan-evaluation model for 3-D radiation treatment plans, and to build a clinical decision-support system to assist in that task using the new preference acquisition method.

*The author was partially supported by National Library of Medicine Training Grant 5-T-15-LM07049. This work was partially supported by National Cancer Institute Contract N01-CM-97564.

Research Proposal: Preference Acquisition through Reconciliation of Inconsistencies

Nilesh L. Jain

1 Introduction

Intelligent computer-based systems used by people to perform various tasks fall into two broad categories — *expert systems* and *decision-support systems.* Expert systems, as the name suggests, attempt to capture the expertise of a person and make it available to non-experts. The expertise is represented in the form of a knowledge base, and various reasoning techniques have been developed to use this knowledge. Decision-support systems, on the other hand, assist people in making decisions. They contain the preferences of the decision-maker about the decisions to be made, and use the concepts of decision theory in order to make optimal decisions. It is obvious that the performance of these systems depends critically on the accuracy and completeness of the knowledge base or how accurately the preference model represents the preferences of the decision maker.

Over the past three decades, many expert systems and decision-support systems have been developed in research as well as in commercial settings. One of the main problems facing the developers of these systems is the acquisition of information to be contained in the systems — knowledge acquisition in the case of expert systems, and preference acquisition in the case of decision-support systems. Knowledge acquisition occurs through the interaction between the knowledge engineer developing the system and the domain expert whose expertise will be represented in the system; preference acquisition occurs through the interaction between the decision analyst developing the decision-support system and the decision maker whose preferences will be represented in the system. Knowledge acquisition is a major focus of current research on expert systems [6,47], and preference acquisition is a major focus of current research on decision-support systems [12, 44, 79]. To highlight the importance of the problem, Feigenbaum coined the phrase *knowledge acquisition bottleneck*, an expression that has now become cliché [19].

Traditionally, knowledge acquisition is performed by direct questioning of the expert, by interpretation of verbal protocols created by asking an expert to think aloud while working on real or simulated problems, by ethnographic methods which involves observing the expert directly in the workplace, and other methods [22]. In a recent review article, Musen has identified some of the causes of the knowledge acquisition bottleneck [47]. The first reason is that the knowledge that experts use has become tacit. Due to repeated practice, they can perform actions without being aware of the knowledge they are using. This poses a problem for the knowledge acquisition task because the experts can no longer provide the kind of knowledge needed for the knowledge base. The second reason is the problem of miscommunication as the knowledge engineer and the domain expert rarely share the same vocabulary. The third reason is the insufficiency of the expressive power of most knowledge representation schemes. The fourth reason is that the process of knowledge acquisition is actually the process of creating a model of the domain, and it includes simplifying assumptions because most models are approximations of the reality they are trying to represent. This leads to the brittleness in the knowledge base which is normally blamed on the inability of the expert to provide the required knowledge, but is actually caused by the nature of the simplifying assumptions made during the construction of the knowledge base.

In the last decade, researchers have developed innovative knowledge acquisition techniques for different classes of expert systems [4,7,15,33,37,42,48,78]. These techniques usually exploit features of the application domain, or attempt to solve one of the problems mentioned above. One common feature among most of these techniques is an attempt to eliminate the knowledge engineer from the knowledge acquisition process. The knowledge engineer instead becomes responsible for developing the knowledge acquisition technique and implementing the tool based on it. The expert then directly interacts with the knowledge acquisition tool to create the knowledge base. Research is also underway to develop metalevel knowledge acquisition tools, tools that will generate knowledge acquisition tools for a particular application domain [46, 56].

An interesting recent knowledge acquisition technique exploits the use of knowledge maintenance [7]. Defined broadly, knowledge maintenance is the addition, deletion, or modification of knowledge in the knowledge base. Reports of two of the best maintained expert systems suggest that knowledge maintenance is needed to update the knowledge base, improve the performance of the expert system, or add new functionality to the expert system [2, 8].

Traditionally, preference acquisition is performed by identifying the objectives of the decision, structuring the decision problem, and eliciting the desirability of the various outcomes [16]. The main problem with preference acquisition is that it asks the decision maker to consider hypothetical situations which may never occur in practice, making it very hard for the decision maker to express a true preference. Thus the preference model elicited from the decision maker is only approximately correct, and the decisions made using it need not necessarily represent the decisions that the decision maker would like to make. Numerous studies have been performed validating the fact that the acquired preference model does not necessarily represent the true preferences of the decision maker [23, 24, 32, 40, 55, 71, 72]. Recently, researchers have developed novel preference acquisition techniques for some classes of decision-support systems which elicit more accurate preference models [18].

Compared to knowledge acquisition, fewer researchers are working on preference acquisition. Thus there are many classes of decision-support systems for which efficient and accurate preference acquisition techniques have not been developed. One of the aims of this dissertation is to propose and implement a new preference acquisition technique for a particular class of decision-support systems — systems which objectively evaluate competing plans and select the best plan. The technique uses ideas from knowledge maintenance by changing the preference model whenever the recommendation of the system is not consistent with that of the decision maker [27]. The technique will be implemented in the domain of evaluating three-dimensional (3-D) radiation treatment plans. Thus, another aim of this dissertation is to construct a decision-support system for the objective evaluation of 3-D radiation treatment plans.

Section 2 provides basic background material on decision theory, some recent knowledge and preference acquisition methods, and radiation treatment planning. Section 3 briefly reviews a simple decision theoretic objective evaluation model for 3D radiation treatment plans and highlights its shortcomings which motivated this dissertation. Section 4 proposes a new decision theoretic model that eliminates some of the shortcomings of the previous model. Section 5 describes a new preference acquisition technique. Section 6 outlines the evaluation methodology for the preference acquisition technique and the decision-support system. Section 7 contains some concluding remarks.

2 Background

2.1 Decision Theory

Decision theory is a discipline of study combining ideas from operations research, statistics, and computer science [25,59,63]. It provides an explicit methodology for selecting an optimal action (or set of actions) from many competing actions which have different outcomes. There are four types of decision problems depending on the presence or absence of uncertainty, and on whether the decision problem has a single objective or multiple objectives.

In the simplest case, the decision problem has a single objective and there is no uncertainty. The desirability (*utility* in decision theoretic jargon) of each possible outcome is assessed, and the action with the highest utility is considered to be the optimal action. Slightly more complicated are the decision problems where there is uncertainty about the effects of some of the actions. All the possible events due to each such action are elicited along with the probability of the occurrence of these events. Again, the utility of each possible outcome is assessed, and the expected utility of each action is computed. Normative decision theory states that the action which maximizes the expected utility is the optimal action. These decision problems are called single-attribute decision problems and the utilities represent the preferences of the decision maker.

Most real-life decisions involve choosing among available alternative plans in order to fulfill conflicting multiple objectives. For such non-trivial decision problems, it is difficult to assign a single utility to each outcome; so an outcome is divided into a number of meaningful component attributes which contribute to the overall utility of the outcome. The resulting model is called a multiattribute decision model [36, 74]. The utility of each component attribute is assessed as in the single-attribute case. The decision maker's preferences are obtained in terms of attributes weights which signify the trade-offs that he is willing to make among these attributes. The overall utility for each outcome is obtained by combining the utilities and weights of all the attributes using additive, multiplicative, or other suitable combining functions. If there is uncertainty about the effects of some actions, probabilities of all possible events due to the action are elicited, and the expected overall utility for each action is computed. Normative decision theory states that the sequence of actions which maximizes the overall utility is the one that should be chosen. These techniques have been used in many real-life planning situations including managing quality of computing services [60], siting waste facilities [69], energy system development [73], and many others [1, 10].

Decision theory provides many techniques for the elicitation of single-attribute utilities and attribute weights for multiattribute decision models. Psychophysical and psychological scaling techniques such as direct rating, category estimation, ratio estimation, and curve drawing which require numerical estimation of the strength of preference are used to elicit the utility of outcomes in decision problems with no uncertainty [3, 9, 70]. Difference measurement techniques such as difference standard sequences and bisection which require indifference between pairs of outcomes are also used to elicit the utility of outcomes in these problems [38, 52, 70]. Standard gamble based techniques, such as variable probability method and variable certainty equivalent which require indifference between a gamble and a sure outcome, are used to elicit the utility of outcomes in decision problems involving uncertainty [36, 59]. Multiattribute weight elicitation techniques include ranking, direct rating, ratio estimation, and swing weights which require numerical estimation [14], cross-attribute indifference and cross-attribute strength of preference which require indifference or strength of preference judgments comparing at least two attributes [13, 38], and analytic hierarchy process which requires pairwise strength of preference judgments [62].

Clinical decision analysis has been well-studied and has been shown to handle effectively the problems faced in routine medical decision making [35, 50, 68, 75]. A *clinical decision-support system* is any computer program designed to help health professionals in making clinical decisions [67]. One class of medical decision making problems addressed by clinical decision analysis is the selection of the optimal therapy plan from competing therapy plans. Recently, clinical decision analysis has been used to choose among competing therapy plans in infectious diseases [41], oncology [28, 30, 77], and in the intensive care unit [17].

The main aim of this dissertation is to investigate a new preference acquisition technique. The medium for the investigation will be clinical decision-support systems which use decision theoretic techniques to evaluate competing treatment plans and to select the best plan based on the treatment preferences of the physician. The major bottleneck in such systems is the acquisition of the physician preferences (utilities and weights). The decision theoretic preference acquisition techniques mentioned above, although grounded in a well-developed theory, are difficult to apply in practice as they ask the physician to consider hypothetical situations which may not occur in practice. Because of the artificial conditions used in standard preference acquisition techniques, the acquired preferences will only be approximately correct. This dissertation proposes a new preference acquisition technique based on the clinical use of the decision-support system, and the reconciliation of the inconsistencies between the recommendations of the system and the physician's judgment.

2.2 Knowledge and Preference Acquisition

Section 1 mentions some of the problems in knowledge and preference acquisition. Researchers have been working on innovative methods to solve these problems. In this section, I will describe two recent methods, one in knowledge acquisition and one in preference acquisition, and mention the key idea provided by the technique toward my proposed preference acquisition technique.

2.2.1 Ripple Down Rules

Ripple down rules is a knowledge acquisition technique based on knowledge maintenance which adds or modifies a rule in the knowledge base whenever the expert system makes an inadequate or incorrect recommendation [7]. A key philosophical idea underlying the system is that experts never provide knowledge about how they reach a particular judgment, instead they provide a justification of why the judgment is correct, and this justification usually depends on the context of the case. Thus, the rules in a knowledge base must also contain a description of the context in which they can be applied.

One way of doing this is to organize the knowledge base as a binary tree with a rule at each node, and the branches indicate whether the rule was satisfied or not. Thus subsequent rules along the *satisfied* branch indicate that the expert system made an incorrect recommendation due to a wrong condition or wrong context information. Subsequent rule along the *unsatisfied* branch indicate that the expert system was unable to provide any recommendation.

The key idea provided by this methodology is that the preference model should be updated whenever the recommendation of the decision-support system does not agree with the recommendation of the decision maker.

2.2.2 Simulated Decision Scenarios

Simulated decision scenarios is a preference acquisition technique which infers the preference model of the decision maker based on the decisions made by him/her on simulated decisions [18]. This technique applies to the class of decision-support systems which have a preference model and an environmental or process model which predicts the outcome of a decision alternative.

The method comprises of three phases. In the first phase, the decision analyst must identify the objectives of the decision problems, the attributes of the outcomes and their respective ranges. Based on these, a parameterized preference model is constructed which includes all the basic elements of the finished model, captures the qualitative characteristics of the decision problem, and has parameters which will determine the exact preference model for the decision maker. The functional form of the individual single attribute utility function and the overall combining function are determined using traditional preference acquisition methods.

In the second phase, the analyst creates a set of simulated decision problems in the application domain, and determines the responses of the decision maker for each of them. The cases must reflect the range of possibilities that the system is intended to handle, should cover the range of outcomes as well as the range of available options. A computer-based acquisition tool is created to present these cases to the decision maker. The process model is embedded in the tool and provides the outcomes of the decision alternatives. The decision maker is allowed to explore the decision alternatives and their corresponding outcomes, and selects the optimal decision.

In the final phase, the analyst refines the preference model by adjusting the parameters so as to minimize the difference between the recommendations made by the mode and the decision maker. Mathematical optimization techniques are used for this step. The result is a preference model which represents the preferences of the decision maker.

The key idea provided by this methodology is that one can perform preference acquisition during the actual use of the decision-support system.

2.3 Radiation Treatment Planning

The clinical domain in which the preference acquisition technique will be implemented is the evaluation of three-dimensional (3-D) radiation treatment plans. Cancer patients are treated using surgery, chemotherapy, and radiotherapy, either alone or in combination. The application domain for this dissertation will focus on those patients for whom radiotherapy is one of the selected treatment modalities. The intent of the treatment can be curative or palliative. The task of radiation treatment planning (RTP) is assisted by computer-based treatment planning systems. These systems are able to calculate and display the dose distributions through the radiation treatment field.

Once radiotherapy is selected, the patient undergoes radiologic scanning using computed tomography (CT) or magnetic resonance imaging (MRI) to determine the internal anatomical structure. The radiation oncologist then has to contour the tumor and the possible areas of microextension on the images. Each of these areas is called a target volume. All organs present near the target volumes are also contoured, and are called normal tissues. The radiation oncologist prescribes the radiation type (photon, electron, x-ray, etc.), and doses for each target volume based on the cell type and severity of the tumor and the radiation tolerance of the normal tissues. The goal of radiation treatment is to irradiate uniformly all target volumes to their prescribed doses, and at the same time, to minimize radiation induced damage to the nearby healthy normal tissues [51]. Treatment planning then involves designing a set of beams which satisfy the treatment goal as closely as possible. The ideal treatment plan would deliver all of the prescribed dose to the target volumes while simultaneously delivering little or no radiation to the adjacent normal tissues. Since the ideal treatment plan generally is not attainable, several potential plans are generated. Each potential plan must make trade-offs in the doses delivered to the target volumes and the normal tissues.

Evaluation of radiation treatment plans forms an integral part of the clinical responsibilities of a radiation oncologist. Current evaluation techniques are largely subjective in nature, based primarily on the dose distribution in some of the critical tissues that may be exposed to radiation. Current standard-of-practice radiation treatment planning (RTP) systems generate two-dimensional (2-D) dose distributions, and plan evaluation is relatively easy because it involves examining one dose distribution in a single, or at most, only a few transverse planes. The rapid improvement in computer technology, both in terms of computation speed and high-resolution graphics, has allowed for better dose-computation algorithms and dose-display techniques. State-of-the-art treatment planning systems are capable of calculating 3-D dose distributions. 3-D radiation treatment planning provides a better and more realistic view of the anatomical relationships and dose distributions [53, 54].

A major difficulty in 3-D RTP is that the radiation oncologist must decipher a huge amount of planning data. Making an unambiguous conclusion about the merits of one plan over another is a difficult task, and thus far objective plan-evaluation methodologies that reflect actual clinical practice have been non-existent. Three-dimensional (3-D) displays of dose distributions are difficult to interpret and 3-D dose distributions have led to data overload [45]. To help combat this data overload, radiation physicists have developed ways of summarizing the dose display in the form of dose-volume histograms (DVH) [11], and radiation biologists have developed quantitative models which use the DVH to compute the Tumor Control Probability (TCP) for target volumes [21] and the Normal Tissue Complication Probability (NTCP) for normal tissues [39]. The National Cancer Institute is currently sponsoring a multi-institutional research effort† to implement these models as software tools to be used in radiation treatment plan evaluation [57, 58]. I believe that objective plan-evaluation models which use the results of these quantitative models can lead to one of their possible appropriate uses in plan evaluation.

In addition, 3-D planning has spurred the use of unconventional beam arrangements with plans sometimes using non-coplanar beams. In 2-D planning, in which the beam axes are either in or parallel to the plane of the paper, the treatment planner or radiation oncologists could make rational assumptions regarding the dose distribution in other regions. However, when beam directions are oblique to the viewed plane and non-coplanar beams are used, the treatment planner's intuition is much less reliable and the dose distribution must be checked throughout the irradiated volume. Finally, because 3-D planning is not constrained to using radiotherapy beam arrangements that adhere to a co-planar geometry about the transverse patient axis, there is a much larger number of possible solutions to the planning problem. Consequently, the issue of plan optimization has gained renewed interest [20, 43, 49]. Ideally the 3-D planning system should be able to optimize the treatment plan automatically, or at least semi-automatically, with user guidance. To do this, the computer must calculate a large number of different plans, compare them, and determine which is the best. Thus, 3-D planning systems must provide a methodology of ranking plans, so that one can be said to be better than another.

Looking into the future of designing radiation treatment plans, the number of options to be considered will be greatly increased with the unrestricted use of non-coplanar beams and the likely need to significantly increase the number of treatment fields to obtain 3-D conformal plans that offer a significantly improved probability of local-regional control and non-significant increase in complications. The determination of an optimal combination of a large number of beams via interactive modification is virtually impossible without computer-aided plan-evaluation tools such as my proposed objective plan-evaluation tool. Therefore I maintain that the full clinical potential of 3-D RTP cannot be realized without the implementation of software tools needed for plan evaluation.

3 Simple Decision Theoretic Model

I have used decision theoretic techniques to develop an objective plan-evaluation model for ranking competing radiation treatment plans [26, 28–31]. This model was based on work done by Schultheiss [64–66].

†Radiotherapy Treatment Planning Tools Contract. The collaborating institutions are University of North Carolina at Chapel Hill, University of Washington at Seattle, and Washington University at St. Louis.

3.1 The Model

The plan-ranking problem was formulated as a multiattribute decision problem. Each attribute represented a specific clinical issue that may appear in a treatment plan. Typical attributes (clinical issues) were non-eradication of tumor and radiation induced damage to healthy normal tissues appearing in the treatment field. For each issue, its *utility* was computed as a number from 0 to 1. A utility of 0 for an issue meant the plan addressed that issue in an undesirable manner, and 1 meant the plan addressed that issue in a desirable manner.

In order to compare and rank different plans, the utilities of all issues in a specific plan needed to be combined to obtain an overall utility for that plan. If any one issue is addressed by the plan in an undesirable manner, the overall utility of the plan had to be reduced significantly. Since the utilities of the issues ranged from 0 to 1, a multiplicative multiattribute model was appropriate to achieve this. Thus, the overall utility of a plan, also known as its *figure of merit (FOM)*, was obtained by taking the product of the utilities of all issues:

$$FOM = \prod_{i}^{issues} utility_i \tag{1}$$

Not all issues had the same clinical relevance in the evaluation of the treatment plans. For example, in radiation treatment of lung cancer, the clinical relevance of radiation-induced myelitis markedly exceeds the clinical relevance of skin telangiectasia. Thus, to obtain the utility of an issue, the probability of the occurrence of that issue was combined with the clinical relevance of the issue in the plan. When the probability of the issue was high and the issue was important, the utility had to be low (undesirable). When the probability of the issue was low or the issue was irrelevant, the utility had to be high (desirable). One function which demonstrated this behavior was:

$$utility_i = 1 - probability_i * weight_i \tag{2}$$

In Equation 2, the *probability* was the likelihood of occurrence of the issue. A probability of 0 indicated the issue will not occur, and 1 indicated the issue will occur. The issue associated with each target volume was treatment failure. In order to obtain its probability, I used the TCP model by Goitein [21], which computes the probability that the tumor is eradicated. The issue associated with each of the normal tissues appearing in the treatment field was the occurrence of a radiation-induced clinical complication. In order to obtain its probability, I used the NTCP model by Kutcher [39], which computes the probability that some clinical complication occurs due to radiation. The *weight* indicated the clinical relevance of an issue. A weight of 0 meant the issue was irrelevant, and 1 meant it was important.

Thus, FOM was computed as:

$$FOM = \prod_{i}^{issues} (1 - probability_i * weight_i) \tag{3}$$

The plans were ranked based on their FOM values. Additionally, the utilities of the issues were used to determine which issue should be improved to increase the FOM of the plan.

3.2 Preference Acquisition

My initial model development focused on the following three tumor sites — prostate (stage C), lung and head-and-neck. Using local radiation oncologists as my domain experts, I obtained a set of clinically relevant issues for each of the three tumor sites. For the target volumes, the issue was treatment failure. For the normal tissues, the issue was a set of clinical endpoints observed if the treatment dose exceeded the threshold dose for that tissue.

The preference (weight) acquisition methodology was a variant of the direct rating method. Three worksheets were designed, one for each tumor site, and they contained the clinically relevant issues for that site (Figure 1). Each issue on the worksheet represented a treatment plan having the following characteristics: for all the other issues on the worksheet, the probability of their occurrence was the normal probability seen in routine clinical setting, and for the specific issue that this plan represented, the probability of its occurrence was *double* the normal probability. Normal probabilities were elicited by consensus from local radiation oncologists. Thus, for each issue on the worksheet, there was a hypothetical treatment plan which performed worse on that issue and performed normally on all other issues, making the evaluation of the plan sensitive to only that issue.

Tumor Site: *PROSTATE (stage C)*

LEVEL OF ENTHUSIASM

Treatment Volumes with issue being a doubling of risk of treatment failure

1	Target volume 1 - microscopic diseases	
2	Target volume 2 - gross enlarged nodes with appropriate margin	
3	Target volume 3 - gross tumor with appropriate margin	

Critical Structures with issue being a doubling of risk of the listed clinical endpoints

4	Bladder - clinically significant cystitis with volume loss	
5	Connective tissue - severe fibrosis	
6	Femoral head - necrosis	
7	Intestine - obstruction, perforation, fistula	
8	Rectum - fistula, severe proctitis	
9	Testicles - infertility	

Guidelines

Level of Enthusiasm is a number from 0 to 100

While filling the Level of Enthusiasm for an issue, assume that you are considering a plan that has twice the normal chance of complication for that issue, and the normal chance of complication for all the other issues and this is the only plan available to you

0 means you will not use that plan

100 means you will use that plan without any hesitation

Figure 1: An adapted version of the worksheet used for Prostate (stage C) tumors

---

Having defined such a set of hypothetical treatment plans, the radiation oncologists were asked to express their enthusiasm for such a plan by assigning it a *Level of Enthusiasm.* Level of Enthusiasm was a number from 0 to 100. A Level of Enthusiasm of 0 meant that the plan was unacceptable (the clinician will not administer it to the patient). A Level of Enthusiasm of 100 meant that the plan was acceptable (the clinician will administer it with the same enthusiasm as a plan that had the normal probability of occurrence of the issue). Intermediate values were used to indicate levels of enthusiasm lying between the two extremes. While assigning level of enthusiasm to a plan, the clinicians were told to assume that this was the only plan available to them to treat the patient. This was to eliminate the bias that would occur if they assumed the presence of the other plans, some of which were better than the plan they were considering.

A value for the Level of Enthusiasm was converted to a weight using the formula:

$$weight = \frac{100 - Level\ of\ Enthusiasm}{100} \tag{4}$$

From Equation 4, a Level of Enthusiasm of 100 (no hesitation) for an issue resulted in a weight of 0 (irrelevant); whereas a Level of Enthusiasm of 0 (would not administer plan) resulted in a weight of 1 (important).

3.3 Shortcomings

Schultheiss proposed a model similar to the one above. However, all published accounts of Schultheiss' work have $weight_i = 1$, the highest possible morbidity for all possible clinical complications. Thus, the FOM he computed was:

$$FOM = \prod_{i}^{issues} (1 - probability_i) \tag{5}$$

which is simply the probability that no complication occurs. By fixing the weights of all the attributes to 1, Schultheiss did not exploit the ability of his decision-analytic formula to model the effects of trade-offs among complications of differing morbidity. The initial focus of our work was to extend Schultheiss' model to overcome some of the limitations we had identified in his work. A significant shortcoming in his model was that the utility of an attribute depended on its trade-off weight, whereas multiattribute utility theory recommends that the utility of an attribute should be independent of its weight. Since our initial aim was to remove the other limitations of his work, we used his model as our starting point, inheriting a utility function which depended on the attribute weight. Section 4 contains a new model where the utility of an attribute is independent of its weight.

During the preliminary evaluation, I observed that one of the rankings was incorrect due to a high weight given by the radiation oncologist for one of the issues. My current model does not have the capability of recovering from the inadequacies of the acquisition methodology, or the inability of the radiation oncologist to provide me with proper preference information. The incorrect weight may be due to the infamiliarity of the radiation oncologists with the preference acquisition technique. Section 5 contains a new preference acquisition which gets away from this shortcoming.

The NTCP values for the issues rarely exceeded 0.03. This meant that the utility of the issues was at least 0.97 and all the trade-off decisions were being made in the narrow interval from 0.97 to 1.00. This was also reflected in the preliminary results as some of the plans differed by as little as 0.001 in their figure of merit values. Similarly, for some of the patients, the best achievable TCP was less than 0.75. This meant that the utility for that target volume could not exceed 0.75, whereas most of the other issues would have a higher utility. This affected the model's ability to select issues for improvement as the target volume had already been optimized to the best achievable TCP. The new model in Section 4 eliminates this shortcoming.

4 New Decision Theoretic Model

Any objective plan-evaluation model should incorporate the preferences of the decision maker, and also fine tune these preferences based on the conditions or abilities of the other people involved in the process. Hence, I seek to model the treatment preferences of the radiation oncologist. But these preferences can be affected by the clinical condition of the patient. Schultheiss proposed incorporating the treatment preferences of the radiation oncologist into his model. However, no objective plan-evaluation models in the literature incorporate the clinical condition of the patient. Thus, the evaluation of radiation treatment plans should include the preferences of the radiation oncologist who is prescribing the treatment, and clinical condition of the patient who is going to receive the treatment. I will now describe a new decision theoretic model which I propose to investigate through this research.

The plan-ranking problem is again formulated as a multiattribute decision problem. Each attribute represents a specific clinical issue that may appear in a treatment plan. Typical attributes (clinical issues) are non-eradication of tumor and radiation-induced damage to nearby healthy normal tissues appearing in the treatment field. For each issue, its *utility* is computed as a number from 0 to 1. The *utility* will encode the treatment preference of the radiation oncologist for that issue. A *utility* of 0 means for an issue means that the plan addresses the issue in an undesirable manner, and 1 means the plan addresses that issue in a desirable manner.

For each issue, its utility will be a function of the dose distribution in the tissue represented by that issue. The objective of the issue depends on the type of tissue represented by it. In the case of a target volume, the objective is to irradiate it uniformly to the prescribed dose. For a normal tissue appearing in the treatment field, the objective is to minimize the dose delivered to it. However, it is impractical if not impossible to enumerate all the possible dose distributions for a tissue. This makes it impossible to elicit utility functions for the issue based on the dose distribution in the associated tissue. I will use *proxy attributes* in order to elicit the utility functions. A proxy attribute is one that reflects the degree to which an associated objective is met but does not directly measure the objective [36]. The proxy attribute will be the probability of a bad outcome for the issue. This probability will come from TCP and NTCP models which compute the tumor control probability or the normal tissue complication probability based on the dose distribution in and other radiobiological characteristics of the tissue. The objective of each issue will be to minimize the chance of a bad outcome for that issue, and *utility* will be a function of the probability of a bad outcome for that issue. I acknowledge that it is unusual to have a utility function based on the probability of the attribute. However it is appropriate in this case as it matches the objective for the issue which is to minimize the chance of a bad outcome. Also instances of such attributes have been used by Keeney and Raiffa [36] where they use the probability of death as a possible attribute in a hypothetical medical decision problem (pp. 40, 53–55). The utility function will be described in more detail in the Section 4.1.

Different issues have different levels of morbidity. The morbidity of the issue will be represented by its *weight* which will be from 0 to 1. The *weight* will be used to make trade-offs among the different issues having different levels of morbidities. Furthermore, since the clinical condition of the patient affects the trade-offs made among the various issues, the *weight* will encode the clinical condition of the patients. A *weight* of 0 means that the issue is irrelevant, and 1 means that the issue is important. My proposed methodology for eliciting the weights will be described in more detail in Section 4.2 and Section 5.

In order to compare and rank competing treatment plans, the *utility* and *weight* of the issues need to be combined to obtain an overall aggregate utility for the plan. Let the contribution of each issue to the aggregate utility be called its *score*. The *score* is a number from 0 to 1. When the utility of an issue is low and the issue is important, *score* should be low. Whereas, if the utility of the issue is high or the issue is irrelevant, *score* should be high. One function which has this behavior is:

$$score_i = 1 - (1 - utility_i) * weight_i \tag{6}$$

Whenever the score for any one issue is low, aggregate utility for the plan should be low. Since the scores are between 0 and 1 for all the issues, a suitable aggregation model which has this behavior is the multiplicative multiattribute model. Thus, the aggregate utility of the plan, known as its *figure of merit (FOM)*, is obtained by taking the product of the scores of all the issues:

$$FOM = \prod_{i}^{issues} score_i \tag{7}$$

Thus, the new objective plan-evaluation model is:

$$FOM = \prod_{i}^{issues} (1 - (1 - utility_i) * weight_i) \tag{8}$$

Notice how Equation 8 is similar to my previous model (Equation 3), but the utility (goodness) and weight (importance) are more clearly separated in my new model.

I have presented an intuitive development of the multiplicative multiattribute model. I will now relate this model to the theoretical multiplicative multiattribute model as described in the standard textbooks [36,74]. The standard multiplicative model is expressed as:

$$1 + wF = \prod_{i=1}^{n}(1 + ww_iu_i) \tag{9}$$

where the model has n attributes, F is the aggregate utility, u_i is the utility of attribute i, w_i is its weight $(0 \leq w_i \leq 1)$ and w is the interaction parameter. There are three cases for w:

1. If $\sum_{i=1}^{n} w_i < 1$, then $0 < w < \infty$.

2. If $\sum_{i=1}^{n} w_i = 1$, then $w = 0$ and we obtain the additive multiattribute model.

3. If $\sum_{i=1}^{n} w_i > 1$, then $-1 \leq w < 0$.

Furthermore, the value of w can be determined from the values of the w_i's by solving the following equation:

$$1 + w = \prod_{i=1}^{n}(1 + ww_i) \tag{10}$$

This equation is obtained by considering the perfect plan where the utilities of all the attributes are 1 and the aggregate utility is also 1.

In Section 4.2, it will be seen that in my new objective plan-evaluation model $\sum_{i=1}^{n} w_i > 1$, and that the weight elicitation methodology assigns a weight of 1 to the most important attribute. Thus, one of the terms in the product term in Equation 10 becomes $(1 + w)$ making $w = -1$ the solution to the equation for my model. Substituting this value in Equation 9, we obtain the following form for the multiplicative multiattribute model:

$$F = 1 - \prod_{i=1}^{n}(1 - w_iu_i) \tag{11}$$

This form of the model is a variant of the form I am using. The difference can be accounted for by the different behavior I want from the interaction of the utility and weight for an attribute. Also, in this case, I assume the plan to begin with a figure of merit of 1 and then penalize it for bad performance on the attributes. The standard multiplicative model instead starts with a value of 0 for the aggregate utility and adds to it for good performance on the attributes.

4.1 Physician Factors

In my new model, the treatment preferences of the radiation oncologist are encoded in the utilities of the issues. The utility of an issue measures how closely the objective for that issue is being met. In the case of radiation treatment planning, the objective for each issue is to minimize the probability of any untoward clinical event associated with that issue. Thus, the probability of occurrence of the issue is an important factor in determining its utility. In our earlier model, we used this probability directly and we observed that due to the low NTCP values (0-3%), we were never using the entire range of our utility function (0-1). My new utility function seeks to map the narrow range of observed probabilities onto the entire range of the utility. This method of scaling the acceptable levels of an issue onto the entire utility range is one of the common ways of constructing utility functions [74].

For any issue i in a treatment plan, I believe that the radiation oncologist considers two key probabilities – the lower threshold probability p_i^l, and the upper threshold probability p_i^u. For any issue i, p_i^l represents the highest probability of occurrence that the radiation oncologist is willing to ignore. For some critical clinical issues, p^l can be 0. For any issue i, p_i^u represents the lowest probability of occurrence above which the radiation oncologist will reject the treatment plan categorically. Thus, for any issue i, the range $[0, p_i^u]$ represents the range of complication probabilities that the radiation oncologist is willing to consider while selecting a treatment plan for a patient. Let U be the utility function. Because the objective for any issue is to minimize the probability of its occurrence, part of the utility function over the probability p looks as follows:

$$\begin{aligned} U(p) &= 1 \qquad && 0 \leq p \leq p^l \qquad && \text{(ignore the issue)} \\ &= 0 && p^u \leq p \leq 1 && \text{(reject the plan)} \end{aligned} \tag{12}$$

The region of interest is $p^l \leq p \leq p^u$ in which $U(p)$ goes from 1 to 0. There are three possible ways this can happen – at a linear rate, at an exponentially increasing rate, or at an exponentially decreasing rate. The three scenarios are shown in Figure 2.

Note that the utility function for the same issue may be different for two different tumor sites. This is due to the fact that the organ being considered may be at a higher risk for damage in one of the tumor sites due to its proximity to the tumor than in the other tumor site. For a given tumor site, I will obtain the list of all target volumes and normal tissues that appear in the treatment field for the tumor site. Worksheets will be designed for each tumor site to contain the list of issues that have been elicited from the radiation oncologists. For each issue on the worksheet, two probabilities p^l and p^u will be elicited. I believe that the radiation oncologists will be able to give me these probabilities quite easily as they must regularly be considering subjective values for the probability of complication in an organ while selecting treatment plans. Eliciting these two probabilities will give me the range of probabilities over which the utility goes from 1 to 0. Over this range, I will elicit the utility of the issue in two parts – I will first obtain the shape of the curve, and then obtain its steepness.

In order to obtain the shape of the curve, I will provide the following verbal description of the three curves in Figure 2. For the utility decreasing at a linear rate (Figure 2(a)), the preference for the issue goes steadily from 1 to 0 as the probability goes from p^l to p^u. For the utility decreasing at an exponentially increasing rate (Figure 2(b)), the preference for the issue is quite high for probabilities slightly over p^l,

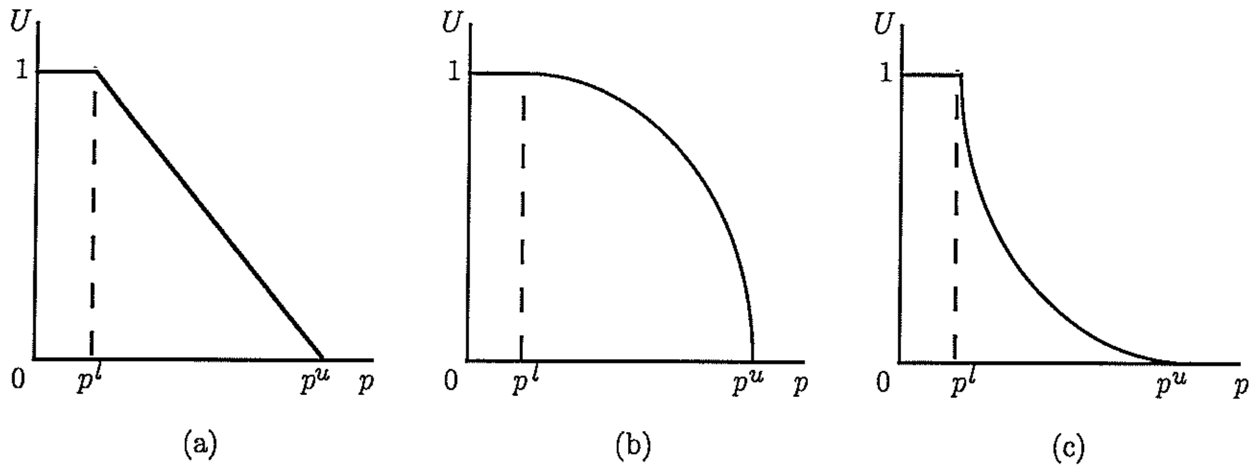

Figure 2: Three possible shapes for utility function U in the range $[p^l, p^u]$. The utility is decreasing at a (a) linear, (b) exponentially increasing, and (c) exponentially decreasing rate.

but the preference rapidly approaches 0 for probabilities approaching p^u. For the utility decreasing at an exponentially decreasing rate (Figure 2(c)), the preference for the issue starts becoming very low even for probabilities slightly over p^l. The radiation oncologist will be asked to pick the curve which best reflects how his/her preference for that issue changes with increasing probability of complication.

If the radiation oncologist picks either of the last two cases, then the steepness of the curve has to be obtained. This can be done by determining a point on the curve and calculating the rest of the curve. However, this will be a difficult task as there is no way of calibrating the intermediate points on the curve so that all the radiation oncologists use it in a consistent manner. Instead, I will approximate the process by presenting the radiation oncologist with three curves of varying steepness. The radiation oncologist will then be asked to select which of these curves best represents his/her preference for the issue. The first of these curves will be a slight deviation from the linear case. The second curve will be quarter of a circle with radius equal to the length of the x-axis from p^l to p^u. The third curve will be even steeper being almost flat (vertical or horizontal) at the two extremes. Thus, there are seven possible utility curves which can be seen in Figure 3.

The methodology described above is a variant of the category estimation technique for eliciting the utility function for an attribute [14]. I will elicit these utilities for the following three tumor sites – prostate, lung and head-and-neck – that have been chosen for evaluating the performance of our objective plan-evaluation model.

4.2 Patient Factors

The clinical condition of the patient will be encoded through the weights for the issues as it affects the trade-offs to be made among the various issues. For each tumor site, I will elicit a list of patient factors which can affect the trade-offs being made among the various treatment issues. These clinical conditions include the stage of the cancer, the age of the patient, the presence of some other concurrent illness such as diabetes, the functional capacity of an organ such as the kidney, etc. Patients will be classified into categories depending on the presence or absence of the relevant clinical conditions. I will use a simple rule-based expert system

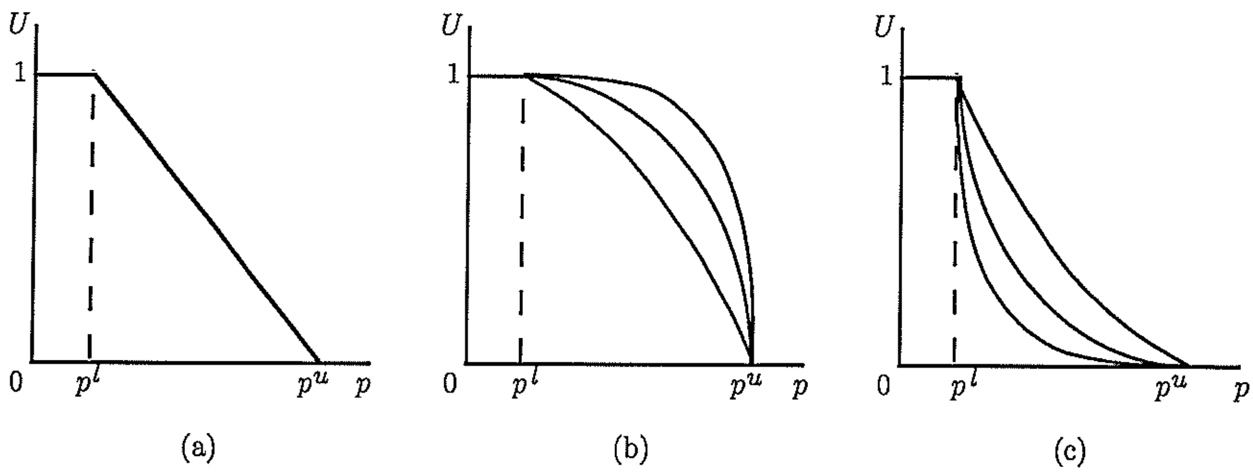

Figure 3: Seven possible utility functions in the range $[p^l, p^u]$. The utility is decreasing at a (a) linear, (b) exponentially increasing, and (c) exponentially decreasing rate.

to determine the patient category depending on his/her clinical conditions. A set of weights will be elicited for each patient category.

To obtain patient-specific weights, a hypothetical patient is described to the radiation oncologist based on the patient category for which the weights are being elicited. Then, the radiation oncologist is asked to select from the worksheet a single issue that he/she would consider to be the most critical issue for such a patient. This selection has to be made based on the morbidities of the complications related with each of the issues. Let this critical issue be i_c. It is assigned a weight of 1 (most important). Now, for every other issue i on the worksheet, two hypothetical plans p_1 and p_2 are described. Table 1 contains the probabilities of complication for the issues in these plans.

Table 1: Probabilities of complication for the issues in the hypothetical plans p_1 and p_2 being used to elicit the *weight* of issue i.

	Issue i_c	Issue i	Other issues
Plan p_1	$p^l_{i_c}$	p^u_i	0
Plan p_2	$p^u_{i_c}$	p^l_i	0

The radiation oncologist is asked which of these two plans does he/she prefer. Three cases are possible:

1. Plans p_1 and p_2 are equivalent. Since the issues have complementary utilities in the two plans, they must have the same weight in order to obtain the same figure of merit. Thus, the weight of issue i is 1 in this case.

2. Plan p_1 is preferred over plan p_2. In this case, the radiation oncologist will be asked to give a probability p of complication of the issue i_c in plan p_2 that will make the two plans equivalent. We are trying to improve plan p_2 till it becomes as good as plan p_1. The weight of issue i can then be calculated from the

following formula obtained by equating the figures of merit of the two plans:

$$weight_i = 1 - U_{i_c}(p) \tag{13}$$

where U_{i_c} is the utility function for issue i_c.

3. Plan p_2 is preferred over plan p_1. This is inconsistent because it implies that issue i is more important than issue i_c which is not true since issue i_c is the most critical issue.

The methodology described above is a variant of the trade-off technique or the cross-attribute indifference technique for constructing attribute weights [13, 38]. This methodology can be applied to different patient categories. As in the case of the utility functions, I will use the same three tumor sites, and will elicit the weights from the radiation oncologists from whom I will elicit utility functions. I will outline a complementary methodology for refining the weights for different patient categories in the section on updating of model parameters through interactive feedback in Section 5.

5 Proposed Preference Acquisition Technique

In Section 3.3, I described an incorrect ranking which was produced by the model due to a high weight for an issue. By suitably modifying the weight, the ranking produced by the model agreed with the ranking of the radiation oncologist. Such inconsistencies will be present initially in the preference model because the radiation oncologists are being asked to perform a task which is quite different from anything that they have done before. The non-clinical environment of the acquisition will provide only an approximation of the preference model of the radiation oncologists. Extended clinical use will provide the radiation oncologists with better insight into the model parameters, and enable them to express their preferences more accurately. In order to allow this, I am proposing a methodology which allows reconciliation of the inconsistencies between the recommendations of the preference model and the decision maker through interactive feedback so that the preference model eventually converges to its correct form. We will also see how this methodology can be applied to obtain the weights for various patient categories after having obtained the weights for an average patient.

The objective plan-evaluation model will be implemented as an interactive computer-based tool. The tool will be linked to Mallinckrodt Institute of Radiology's existing 3-D RTP system so that it can share patient data with the tool, and can offer a single platform to all the people involved in the design and evaluation of radiation treatment plans. This will enable the clinical use of the objective plan-evaluation model for retrospective evaluation of the actual radiation treatment plans designed during the treatment planning.

The tool will have an interactive graphical user interface which will provide the user (the radiation oncologist who prescribed the treatment) with the ability to examine all the parameters (probabilities, utilities, and weights) that were used in the evaluation of the plans, and also allow him/her to modify some of those parameters. The user will be provided with a description of the patient as well as details about the plans that are being evaluated simultaneously. The tool will use the user's utility functions and weights to rank these plans. If the user feels that the ranking does not match the ranking that he/she would have assigned to those plans, then the user can examine the probabilities, utilities, and weights that were used in that ranking. Either the user may be convinced that the ranking produced by the tool is actually correct, or the user may want to change some of the model parameters in order to get the correct ranking.

Assuming that the utility functions are correct, the incorrect rankings will be due to inappropriate weights for the issues. This assumption is made to make the acquisition practical. It is justified because the utility is based on the probability of complication, and the radiation oncologists already have acceptable and unacceptable ranges of these values that they use in subjective plan evaluation. The tool will assist the user in determining the issues that are having the most significant impact in causing the incorrect ranking and the user can change the weights of those issues appropriately until the ranking produced by the tool matches the ranking that he/she had in mind. The tool will record the changes in the weights, and make appropriate changes to the database of weights. Each time the radiation oncologist changes the weights, he/she is giving new insights into the way that he/she makes trade-offs among the various issues. A critical hypothesis is that, after using the tool to evaluate the plans for a large number of patients, the weights will converge to reflect the true trade-offs that the radiation oncologist makes while evaluating radiation treatment plans, and this will represent the true preference model of the radiation oncologist.

Typically, the weight of an issue can be expected to vary in the manner shown in Figure 4. The initial weight for the issue is w_i, and the converged value is w_f. Note that the weight has not changed for the last X patients. For the preference model to converge, none of the issue weights for this patient category should change for the last X patients.

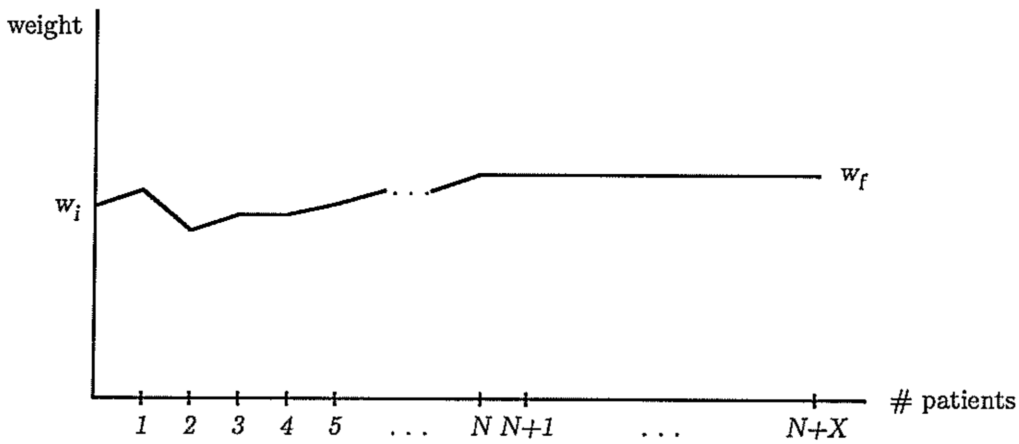

Figure 4: Typical variation in the weight of an issue before it converges to its true value.

Since there is no prior experience with this acquisition technique, I do not know what value of X will be appropriate. I hypothesize that lower values of X will lead to preference models which do not perform as well as preference models obtained through higher values of X. At the same time, the number of patients needed for convergence grows rapidly as X increases (trivially, the minimum number of patients needed is X itself), and the acquisition technique may become impractical beyond some value of X because of the inordinately high number of patients whose data is needed for convergence of the preference model. So, I will try to empirically determine a value or range of values of X which provide an acceptable performance by the preference model, and are practical for the acquisition. In order to do this, I will obtain a series of preference models P_X for $X \geq 1$. Let P_0 be the set of initial weights obtained using the traditional method described in Section 4.2. P_X is the set of issue weights obtained when no weights are changed for X consecutive patients for the first time. In Section 6, I will show how this can be used to empirically determine acceptable value(s) of X.

The ability to change the weights can also be used to elicit the weights for the various patient categories. This can be done by eliciting the weights for the average patient using the methodology described in the Section 4.2. This set of weights can then be duplicated for all the known patient categories. Through continued clinical use of the tool, the user will change the weights for the various patient categories until they converge to their actual values.

6 Evaluation

The performance of any clinical decision-support system can be evaluated by comparing it with the performance of the experts or by measuring the improvement in the health outcomes with use of the system. Since it currently is impractical to do a performance evaluation based on the health outcomes, the performance of my objective plan-evaluation model will be evaluated by comparing it with the performance of the experts. In this case, the experts are the radiation oncologists who currently evaluate radiation treatment plans using manual and subjective methods. I will evaluate my new objective plan-evaluation model by comparing rankings produced by radiation oncologists with rankings produced by my tool using the utility functions and weights of the same radiation oncologist. The evaluation will be done using treatment plans from the following three tumor sites – prostate, lung and head-and-neck. The treatment plans to be used will be actual plans designed during the treatment planning process for that patient and the evaluation will be done retrospectively so as not to interfere with the actual patient care. The radiation oncologists will be blinded to the identity of the patient.

For each P_X, the performance evaluation will be conducted in two phases. The first phase of the evaluation of the performance of my new objective plan-evaluation model will go hand in hand with the updating of the weights through interactive feedback. During this phase, in case the rankings do not match, the radiation oncologist will be asked to change the weights as was described in Section 5. Once the weights have converged to obtain a P_X, the evaluation will enter the second phase where the performance of the objective plan-evaluation model with P_X will be compared to the performance of the corresponding radiation oncologist. The performance of the objective plan-evaluation model will be quantified using two different criteria.

Because radiation oncologists treat their patients with the best possible plan, it is important that my model's first-choice plan agree with the radiation oncologist's first-choice plan. Thus, the number of times the model's first-choice plan agrees with the radiation oncologist is a useful indicator of the potential clinical usefulness of my tool. I will record the number of times the model is used to evaluate plans, and the number of times its first-choice ranking agrees with the radiation oncologist. During the second phase of the evaluation, the fraction of correct rankings will be called C_X for the preference model P_X. This will be the first criterion for the evaluation.

The plan-ranking model can also be a useful pedagogical tool for residents who seek to learn the trade-offs among competing plans that would be made by more experienced radiation oncologists. I believe that the number of times the model agrees with the radiation oncologist on the ranking of *all* competing plans is one indicator of its ability to accurately represent the trade-offs made by that physician. Unlike the clinical metric where only the first-choice ranking is critical, for instructing less experienced physicians on how senior radiation oncologists make trade-offs among clinical issues, it is important that the entire ranking is correct. I will record the number of times the model is used to evaluate plans, and the number of times its *entire* ranking agrees with the radiation oncologist. For incorrectly ranked plans, I will calculate a penalty score, where the penalty will depend on the rank assigned to that plan by the radiation oncologist. A plan incorrectly ranked lower will receive a higher penalty. The penalty score for an incorrect ranking will indicate how far it is from the actual ranking. This will be the second criterion for the evaluation.

The usefulness of the preference acquisition technique can be proved by showing that $C_X > C_0$ for some $X \geq 1$. That is, some preference model obtained by changing the weights has better clinical performance than the initial preference model.

The second phase of the performance evaluation of the different P_Xs will be compared to determine which value of X produces a better preference model. As stated earlier, I expect C_X to be better for higher values of X. For all these P_Xs, I will also determine the number of patient cases needed for the convergence, N_X. Obviously N_X will increase for higher values of X. The effectiveness E_X of the preference model P_X can be measuring the increase in the fraction of clinically correct rankings per patient data needed. That is,

$$E_X = \frac{C_X - C_0}{N_X} \tag{14}$$

Equation 14 is similar to the cost-effectiveness measures used in health policy research [76]. The value of X having the highest E_X would represent the best trade-off between the clinical performance of the preference acquisition method and the number of patient cases needed to achieve it.

7 Conclusion

Through this research, I am proposing to implement and demonstrate a new technique for preference acquisition. This technique can be used for acquiring individual preference models to be used in a decision-support system. Also, after the elicitation of the initial approximate preference model, the decision analyst is not needed for the reconciliation of the preference model. One of the biggest criticism of decision-support systems has been their lack of transportability as they encode the preferences of a single or a group of individuals who may not be the intended users. My acquisition method will enable the easy acquisition of individual preference models, thus making this system available to people outside of the institution where it was developed. I have already installed my initial objective plan-evaluation tool at the University of North Carolina in Chapel Hill and the University of Washington in Seattle using the preferences of local radiation oncologists.

I envision many potential uses for an objective plan-evaluation model. When embedded into an interactive plan-evaluation environment, an objective model can be used by radiation oncologists for evaluating and selecting treatment plans. It can also be used by medical physicists, dosimetrists and others involved in the process of manually optimizing radiation treatment plans. As mentioned in Section 6, an objective plan-evaluation model also has tremendous pedagogical value. Residents can learn from the evaluation skills of the senior radiation oncologists who would otherwise be hard-pressed for time to instruct the residents on the difficult art of decision-making. The numerical results of an objective plan-evaluation model can be used as an objective function by computer programs that try to obtain an optimal treatment plan by using mathematical optimization techniques. Qualitative and symbolic results based on an objective plan-evaluation model can be used by computer programs that try to obtain an optimal treatment plan using artificial intelligence techniques [5,34]. In the future, with the promise of real-time 3D dose calculation [61], the treatment planners may use an objective plan-evaluation model to obtain an instantaneous evaluation as they move the beams during the design of a treatment plan.

References

[1] Aksoy Y. Interactive multiple objective decision making: A bibliography (1965–1988). *Management Research News*, 13:1–8, 1990.

[2] Bachant J, McDermott J. R1 revisited: Four years in the trenches. *AI Magazine*, 5(3):21–32, 1984.

[3] Baird JC, Noma E. *Fundamentals of Scaling and Psychophysics.* John Wiley and Sons, New York, NY, 1978.

[4] Bennett JS. ROGET: A knowledge-based system for acquiring the conceptual structure of a diagnostic expert system. *Journal of Automated Reasoning*, 1:49–74, 1985.

[5] Berger J. ROENTGEN: Case-based reasoning and radiation therapy planning. In Frisse ME, ed., *Proceedings of the Sixteenth Symposium on Computer Applications in Medical Care*, pages 210–214. McGraw-Hill, New York, NY, 1992.

[6] Boose JH. A survey of knowledge acquisition techniques and tools. *Knowledge Acquisition*, 1:3–37, 1989.

[7] Compton P, Edwards G, Kang B, Lazarus L, Malor R, Preston P, Srinivasan A. Ripple down rules: Turning knowledge acquisition into knowledge maintenance. *Artificial Intelligence in Medicine*, 4:463–475, 1992.

[8] Compton P, Horn K, Quinlan JR, Lazarus L, Ho K. Maintaining an expert system. In Quinlan JR, ed., *Applications of Expert Systems: Volume 2*, pages 366–384. Addison-Wesley, London, 1989.

[9] Coombs CH. *A Theory of Data.* John Wiley and Sons, New York, NY, 1964.

[10] Corner JL, Kirkwood CW. Decision analysis applications in the operations research literature, 1970–1989. *Operations Research*, 39:206–219, 1991.

[11] Drzymala RE, Mohan R, Brewster L, Chu J, Goitein M, Harms W, Urie M. Dose-volume histograms. *International Journal of Radiation Oncology, Biology, Physics*, 21:71–78, 1991.

[12] Dyer JS, Fishburn PC, Steuer RE, Wallenius J, Zionts S. Multiple criteria decision making, multiattribute utility theory: The next ten years. *Management Science*, 38:645–654, 1992.

[13] Dyer JS, Sarin RK. Measurable multiattribute value functions. *Operations Research*, 27:810–822, 1979.

[14] Edwards W. How to use multiattribute utility measurement for social decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 7:326–340, 1977.

[15] Eshelman L, Ehret D, McDermott J, Tan M. MOLE: A tenacious knowledge-acquisition tool. *International Journal of Man-Machine Studies*, 26:41–54, 1987.

[16] Farquhar PH. Utility assessment methods. *Management Science*, 30:1283–1300, 1984.

[17] Farr BR, Fagan LM. Decision-theoretic evaluation of therapy plans. In Kingsland LC, ed., *Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care*, pages 188–192. IEEE Computer Society Press, Washington, DC, 1989.

[18] Farr BR, Shachter RD. Representation of preferences in decision-support systems. *Computers and Biomedical Research*, 25:324–335, 1992.

[19] Feigenbaum EA. Knowledge engineering: The applied side of artificial intelligence. *Annals of the New York Academy of Sciences*, 426:91–107, 1984.

[20] Goitein M. The comparison of treatment plans. *Seminars in Radiation Oncology*, 2:246–256, 1992.

[21] Goitein M, Schultheiss TE. Strategies for treating possible tumor extension: Some theoretical considerations. *International Journal of Radiation Oncology, Biology, Physics*, 11:1519–1528, 1985.

[22] Hart A. *Knowledge Acquisition for Expert Systems.* Second edition, McGraw-Hill, New York, NY, 1992.

[23] Hershey JC, Kunreuther HC, Schoemaker PJH. Sources of bias in assessment procedures for utility functions. *Management Science*, 28:936–954, 1982.

[24] Hershey JC, Schoemaker PJH. Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science*, 31:1213–1231, 1985.

[25] Howard RA, Matheson JE, eds. *Readings on the Principles and Applications of Decision Analysis.* Strategic Decisions Group, Menlo Park, CA, 1983.

[26] Jain NL, Kahn MG. Ranking radiotherapy treatment plans using decision-analytic and heuristic techniques. In Clayton PD, ed., *Proceedings of the Fifteenth Symposium on Computer Applications in Medical Care*, pages 1000–1004. McGraw-Hill, New York, NY, 1991.

[27] Jain NL, Kahn MG. A methodology for reconciliation of inconsistencies in physicians' preferences through clinical use of decision-support systems (abstract). *Medical Decision Making*, 12:349, 1992.

[28] Jain NL, Kahn MG. Ranking radiotherapy treatment plans using decision-analytic and heuristic techniques. *Computers and Biomedical Research*, 25:374–383, 1992.

[29] Jain NL, Kahn MG, Drzymala RE, Emami B, Purdy JA. Ranking radiotherapy treatment plans using decision-analytic and heuristic techniques (abstract). *Medical Physics*, 18:641, 1991.

[30] Jain NL, Kahn MG, Drzymala RE, Emami B, Purdy JA. Objective evaluation of 3-D radiation treatment plans: A decision-analytic tool incorporating treatment preferences of radiation oncologists. *International Journal of Radiation Oncology, Biology, Physics*, In press, 1993.

[31] Jain NL, Kahn MG, Drzymala RE, Purdy JA. Evaluation of radiotherapy treatment plans using decision-analytic and heuristic techniques (abstract). *Medical and Biological Engineering and Computing*, 29 (Suppl.):437, 1991.

[32] Johnson EJ, Schkade DA. Bias in utility assessments: Further evidence and explanations. *Management Science*, 35:406–424, 1989.

[33] Kahn G, Nowlan S, McDermott J. Strategies for knowledge acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7:511–522, 1985.

[34] Kalet IJ, Paluszyński W. Knowledge-based computer systems for radiotherapy planning. *American Journal of Clinical Oncology*, 13:344–351, 1990.

[35] Kassirer JP, Moskowitz AJ, Lau J, Pauker SG. Decision analysis: A progress report. *Annals of Internal Medicine*, 106:275–291, 1987.

[36] Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* John Wiley and Sons, New York, NY, 1976.

[37] Klinker G, Bentolila J, Genetet S, Grimes M, McDermott J. KNACK—Report-driven knowledge acquisition. *International Journal of Man-Machine Studies*, 26:65–79, 1987.

[38] Krantz DH, Luce RD, Suppes P, Tversky A. *Foundations of Measurement (Volume 1).* Academic Press, New York, NY, 1971.

[39] Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method. *International Journal of Radiation Oncology, Biology, Physics*, 16:1623–1630, 1989.

[40] Laskey KB, Fischer GW. Estimating utility functions in the presence of error. *Management Science*, 33:965–980, 1987.

[41] Magid D, Schwartz B, Craft J, Schwartz JS. Prevention of Lyme disease after tick bites. A cost-effectiveness analysis. *New England Journal of Medicine*, 327:534–541, 1992.

[42] Marcus S, McDermott J. SALT: A knowledge acquisition language for propose-and-revise systems. *Artificial Intelligence*, 39:1–37, 1989.

[43] Mohan R, Mageras GS, Baldwin B, Brewster LJ, Kutcher GJ, Leibel S, Burman CM, Ling CC, Fuks Z. Clinically relevant optimization of 3-D conformal treatments. *Medical Physics*, 19:933–944, 1992.

[44] Moskovitz H, Bunn D. Decision and risk analysis. *European Journal of Operational Research*, 28:247–260, 1987.

[45] Munzenrider JE, Brown AP, Chu JCH, Coia LR, Doppke KP, Emami B, Kutcher GJ, Mohan R, Purdy JA, Shank B, Simpson JR, Solin LJ, Urie MM. Numerical scoring of treatment plans. *International Journal of Radiation Oncology, Biology, Physics*, 21:147–163, 1991.

[46] Musen MA. *Automated Generation of Model-based Knowledge-Acquisition Tools.* Pitman, London, 1989.

[47] Musen MA. An overview of knowledge acquisition. In David JM, Krivine JP, Simmons R, eds., *Second Generation Expert Systems*, pages 415–438. Springer-Verlag, Berlin, 1993.

[48] Musen MA, Fagan LM, Combs DM, Shortliffe EH. Use of a domain model to drive an interactive knowledge-editing tool. *International Journal of Man-Machine Studies*, 26:105–121, 1987.

[49] Niemerko A, Urie M, Goitein M. Optimization of 3D radiation therapy with both physical and biological end points and constraints. *International Journal of Radiation Oncology, Biology, Physics*, 23:99–108, 1992.

[50] Pauker SG, Kassirer JP. Decision analysis. *New England Journal of Medicine*, 316:250–258, 1987.

[51] Perez CA, Brady LW, eds. *Principles and Practice of Radiation Oncology.* J. B. Lippincott, Philadelphia, PA, 1987.

[52] Pfanzagl J. *Theory of Measurement.* John Wiley and Sons, New York, NY, 1968.

[53] Photon Treatment Planning Collaborative Working Group. Evaluation of high energy photon external beam treatment planning: Project summary. *International Journal of Radiation Oncology, Biology, Physics*, 21:3–8, 1991.

[54] Photon Treatment Planning Collaborative Working Group. Three-dimensional dose calculations for radiation treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, 21:25–36, 1991.

[55] Pitz GF, Sachs NJ. Judgment and decision: Theory and application. *Annual Review of Psychology*, 35:138–163, 1984.

[56] Puerta AR, Egar JW, Tu SW, Musen MA. A multiple-method knowledge-acquisition shell for the automatic generation of knowledge-acquisition tools. *Knowledge Acquisition*, 4:171–196, 1992.

[57] Purdy JA, Kalet IJ, Chaney EL, Zink S. Radiotherapy treatment planning tools: A collaborative project of the National Cancer Institute (abstract). *Medical and Biological Engineering and Computing*, 29 (Suppl.):319, 1991.

[58] Purdy JA, Kalet IJ, Chaney EL, Zink S. Radiotherapy treatment planning tools: A collaborative project of the National Cancer Institute (abstract). *Medical Physics*, 18:666–667, 1991.

[59] Raiffa H. *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Random House, New York, NY, 1968.

[60] Roeseler A, Von Mayrhauser A. Using service-level indices to manage the quality of computing services: A case study. *Journal of Systems and Software*, 18:139–145, 1992.

[61] Rosenberger FU, Matthews JW, Johns GC, Drzymala RE, Purdy JA. Use of transputers for real time dose calculation and presentation for 3-D radiation treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, In press, 1993.

[62] Saaty TL. *The Analytic Hierarchy Process.* McGraw-Hill, New York, NY, 1980.

[63] Savage LJ. *The Foundations of Statistics.* Dover, New York, NY, 1972.

[64] Schultheiss TE. Treatment planning by computer decision. In Cunningham JR, Ragan D, Van Dyk J, eds., *Proceedings of the Eighth International Conference on the Use of Computers in Radiation Therapy*, pages 225–229. IEEE Computer Society Press, Silver Spring, MD, 1984.

[65] Schultheiss TE, El-Mahdi AM. Statistical decision theory applied to radiation therapy treatment decisions. In Blum BI, ed., *Proceedings of the Sixth Symposium on Computer Applications in Medical Care*, pages 978–982. IEEE Computer Society Press, Los Angeles, CA, 1982.

[66] Schultheiss TE, Orton CG. Models in radiotherapy: Definition of decision criteria. *Medical Physics*, 12:183–187, 1985.

[67] Shortliffe EH. Clinical decision-support systems. In Shortliffe EH, Perreault LE, Wiederhold G, Fagan LM, eds., *Medical Informatics: Computer Applications in Health Care*, pages 466–502. Addison-Wesley, Reading, MA, 1990.

[68] Sox, Jr. HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making.* Butterworths, Boston, MA, 1988.

[69] Swallow SK, Opaluch JJ, Weaver TF. Siting noxious facilities: An approach that integrates technical, economic, and political considerations. *Land Economics*, 68:283–301, 1992.

[70] Torgerson WS. *Theory and Methods of Scaling.* John Wiley and Sons, New York, NY, 1958.

[71] Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*, 211:453–458, 1981.

[72] Tversky A, Sattath S, Slovic P. Contingent weighting in judgment and choice. *Psychological Review*, 95:371–384, 1988.

[73] Tzeng GH, Shiau T, Lin CY. Application of multicriteria decision making to the evaluation of new energy system development in taiwan. *Energy*, 17:983–992, 1992.

[74] von Winterfeldt D, Edwards W. *Decision Analysis and Behavioral Research.* Cambridge University Press, Cambridge, UK, 1986.

[75] Weinstein MC, Fineberg HV. *Clinical Decision Analysis.* W. B. Saunders, Philadelphia, PA, 1980.

[76] Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine*, 296:716–721, 1977.

[77] Whitmore, Jr. WF, Adolfsson J, Steineck G. Conservative management of localized prostatic cancer. *American Journal of Clinical Oncology*, 15:446–452, 1992.

[78] Wielinga B, Schreiber AT, Breuker J. KADS: A modeling approach to knowledge acquisition. *Knowledge Acquisition*, 4:5–53, 1992.

[79] Winkler RL. Research directions in decision making under uncertainty. *Decision Sciences*, 13:517–533, 1982.