Ranking Radiotherapy Treatment Plans Using Decision-Analytic and Heuristic Techniques

Nilesh L. Jain and Michael G. Kahn
Medical Informatics and Computer Science, Institute for Biomedical Computing,
Campus Box 8036, Washington University School of Medicine, St. Louis MO 63110

Abstract

Radiotherapy treatment optimization is done by generating a set of tentative treatment plans, evaluating them and selecting the plan closest to achieving a set of conflicting treatment objectives. The evaluation of potential plans involves making tradeoffs among competing possible outcomes. Multiattribute decision theory provides a framework for specifying such tradeoffs and using them to select optimal actions. Using these concepts, we have developed a plan-ranking model which ranks a set of tentative treatment plans from best to worst. Heuristics are used to refine this model so that it reflects the clinical condition of the patient being treated and the practice preferences of the physician prescribing the treatment. A figure of merit *is computed for each tentative plan, and is used to rank the plans. The approach described is very general and can be used for other medical domains having similar characteristics. The figure of merit can also be used as an objective function by computer programs that attempt to automatically generate an optimal treatment plan.*

Introduction

Cancer patients are treated using surgery, chemotherapy and radiotherapy, either alone or in combination. We are interested in those patients for whom radiotherapy is one of the selected treatment modalities. The goal of the radiation treatment is to uniformly irradiate all target volumes to the prescribed doses, and at the same time, to minimize radiation induced damage to any nearby normal tissue [11].

Radiotherapy treatment optimization is done by generating a set of tentative treatment plans, evaluating them and selecting the plan closest to achieving the treatment goal. The evaluation of tentative treatment plans is a difficult task because it involves comparing doses delivered to delineated target volumes and critical normal organs by each plan, and making tradeoffs among the competing possible clinical outcomes.

Building software tools to assist in this evaluation is one of the tasks of a multi-institutional research effort sponsored by the National Cancer Institute*. As part of that task, we are constructing a tool which ranks a set of tentative treatment plans from best to worst. Since the ranking of plans involves making tradeoffs among competing possible outcomes, the problem lends itself to the use of decision-theoretic concepts. This paper describes our plan-ranking model and how it differs from previous work done in this field.

Background

Expected utility decision making is a methodology for selecting a sequence of possible actions from many competing sequences of actions having different outcomes [5,12]. The desirability (utility) of each outcome is assessed and so are the probabilities of events that may occur due to a possible action. The sequence of actions which maximizes the expected utility is chosen. *Clinical decision analysis* has been well-studied and is known for its ability to handle the tradeoffs faced by the decision maker [10,16,17,18]. In particular, decision analysis has been used as a methodology for evaluating therapy plans [2] and for selecting among therapeutic options [4].

For some difficult decision problems, it is not possible to assign a single utility to the outcomes. In those cases, a multiattribute utility model is used; the outcomes are divided into component attributes and the utility of each attribute is assessed individually [6]. To use the axiom of maximizing expected utility, various combining functions have been used to derive a single utility from the individual attribute utilities. These functions depend on both the decision problem, and the range and units of the individual utility values.

In radiotherapy, decision analysis has been used for the optimization of treatment plans by Schultheiss, et. al. [13,14,15]. He constructed a multiattribute model where the attributes were the possible clinical complications of the treatment. For each complication, he calculated its utility by combining the probability that the complication occurs with the morbidity of that complication. The attributes were combined to obtain the overall *figure of merit (FOM)* of the plan.The *FOM* was computed using the formula:

$$FOM = \prod_i (1 - probability_i * weight_i) \qquad (1)$$

and was used as the objective function for an automatic optimization algorithm that attempted to obtain a statistically optimal treatment plan.

---

*The participating institutions are Washington University at St. Louis, University of Washington at Seattle and University of North Carolina at Chapel Hill

0195-4210/91/$5.00

The *probability* of complication was obtained from dose response models developed by Schultheiss. The *weight* was the physicians' subjective judgment of the morbidity of the complication. It was a combination of the severity of the complication, the clinical condition of the patient, and the practice preferences of the institution and the treating physician. Our model differs in that it provides a framework to independently assess all these factors.

All published accounts of Schultheiss' work have $weight_i = 1$, the highest possible morbidity. Thus, the *FOM* of a plan is simply the probability that no complication occurs. By doing so, he did not use the full power offered to him by his decision-analytic model.

Methodology

Given a set of tentative treatment plans, we seek to rank them from best to worst. In order to achieve this, we have constructed an adaptable plan-ranking model. Some of the key issues addressed by our model are:

- incorporating clinical distinctions among patients
- incorporating differences in practice preferences among physicians

The plan-ranking model is constructed by the integrating decision-analytic and heuristics techniques. Decision-analytic techniques are used to construct a *generic plan-ranking model.* This generic model is then refined using heuristics to address differences in the clinical features of the patients and the practice preferences of the physicians.

The Generic Plan-Ranking Model

The plan-ranking problem is formulated as a multiattribute decision problem. Each attribute represents a specific clinical issue in a plan. Typical issues are non-eradication of tumor, and damage to nearby normal tissues. For each issue, we compute its *utility* as a number from 0 to 1. A utility of 0 for an issue means the plan addresses that issue in an undesirable manner, and 1 means the plan addresses that issue in a desirable manner.

In order to compare and rank different plans, we need to combine the utilities of all the issues to get the overall utility for the plan. Different issues represent different target volumes or normal tissues. Damage to any one of them is not influenced by damage to any of the others. Thus, all the issues are independent of each other. Since the issues are independent of each other and have utilities from 0 to 1, the multiplicative multiattribute model is appropriate. Thus, the overall utility of a plan, also known as its *figure of merit (FOM)* is obtained by taking the product of the utilities of all the issues:

$$FOM = \prod_i utility_i \qquad (2)$$

For issues such as the non-eradication of the tumor, or the damage to nearby normal tissues, we use numeric models for calculating the probability of their occurrence. These models have been developed by radiation physicists to characterize the dose response of tumors and normal tissues. These models use *Dose Volume Histograms (DVH)* [1], a plot of the dose-volume frequency distribution in an organ. *DVH*s provide a graphical summary of the dose distribution in a volume of interest. For the tumor, the numeric model computes the probability that the tumor is eradicated. This is known as the *Tumor Control Probability (TCP).* We currently are using the *TCP* model by Goitein [3]. For each normal tissue, the numeric model computes the probability that some clinical complication occurs due to radiation to that tissue. It is known as the *Normal Tissue Complication Probability (NTCP).* We currently are using the NTCP models by Kutcher [7] and Lyman [8,9].

Not all issues have the same clinical relevance in the evaluation of the treatment plans. Thus, to obtain the utility of an issue, we combine the probability of the occurrence of an untoward event with the clinical relevance of the issue in the plan. When the probability of an untoward event is high and the issue is important, we want the utility to be low. When the probability of an untoward event is low or the issue is irrelevant, we want the utility to be high. One function which demonstrates this behavior is

$$utility_i = 1 - probability_i * weight_i \qquad (3)$$

The *probability* is the likelihood of occurrence of an untoward event for an issue. A probability of 0 indicates an untoward event will not occur and 1 indicates an untoward event will occur. The *weight* indicates the clinical relevance of an issue. A weight of 0 means the issue is irrelevant and 1 means it is important.

The model described so far is similar to the model proposed by Schultheiss. The next section describes our modifications and enhancements to the model.

The Complete Plan-Ranking Model

The previous section describes the *generic plan-ranking model.* However, the same generic model cannot be applied to all patients and may not be acceptable to all physicians. The model needs to incorporate clinical distinctions among patients as well as the differences in practice preferences of physicians. We model these differences by suitably modifying the *weight.* For each issue, we define its weight to be the function of two quantities — the *prototypical weight* and the *modifier.* Thus, we have

$$weight_i = f(prototypical\ weight_i,\ modifier_i) \qquad (4)$$

The *prototypical weight* for an issue represents its relevance for an *average* patient. A prototypical weight of 0 means the issue is irrelevant and 1 means it is important. Intuitively, given an average patient, the prototypical weight reflects the relevance of that issue for the physicians at a specific institution. It is the consensus weight for that issue among the physicians at the institution.

However, not all patients are the same. The clinical condition of a patient may cause the relevance of an issue to be different from that for an average patient. Also, not all the physicians at an institution have the same practice preferences. Some of them may have different opinions for the relevance of an issue, different from the one expressed by the prototypical weight. These adjustments are modeled by the *modifiers*.

The modifier adjusts the prototypical weight to obtain the weight for an issue. The modifier ranges from -1 to 1. A modifier less than 0 means the relevance of the issue must be decreased; a modifier equal to 0 means the relevance must remain the same; and a modifier greater than 0 means the relevance must be increased. One possible combining function *f()* for weight having the above behavior is:

$$\begin{aligned} wt_i &= p\_wt_i + (1 - p\_wt_i) * mod_i & mod_i > 0 \\ &= p\_wt_i & mod_i = 0 \\ &= p\_wt_i + p\_wt_i * mod_i & mod_i < 0 \end{aligned} \quad (5)$$

where wt is weight, p_wt is prototypical weight and mod is modifier.

The modifiers represent the application of clinical heuristics in the plan-ranking model. The modifiers are derived from a set of heuristics obtained from physicians, and will be represented as IF-THEN rules.

Thus, the *complete plan-ranking model* is conceptualized as:

$$FOM = \prod_{i}^{issues} (1 - probability_i * f(p\_wt_i,\ mod_i)) \quad (6)$$

It is comprised of the following building blocks:

probabilities: These values are obtained from numeric models such as *TCP* and *NTCP*. Their values are limited by the underlying assumptions of the model. Although we have currently incorporated only one *TCP* model and two *NTCP* models, the plan-ranking model can incorporate other *TCP* or *NTCP* models, or other numeric models that may be developed in the future.

prototypical weights: These values come from a pre-established database, indexed by tumor site and the clinical features of the patient. They consist of consensual knowledge reflecting institutional practice preferences. Their values are limited by the degree to which patients with similar clinical features can be treated in a similar manner. Although our current database will reflect the practice preferences of the radiation oncologists at the Mallinckrodt Institute of Radiology at the Washington University School of Medicine, similar databases can be constructed at other institutions and used when the plan-ranking model is being used at that institution.

modifiers: These values adjust the prototypical weights to incorporate clinical distinctions between a patient and the average patient, as well as the practice preferences of the physician which differ from the institutional practice preferences. Their values are limited by the degree to which physicians can identify patient-specific and treatment-specific features that modify their personal approach to selecting a treatment plan. Although we are acquiring the practice preferences of local radiation oncologists, other institutions can acquire the practice preferences for their physicians and use them when the plan-ranking model is being used by that physician at that institution.

Plan-Ranking Tool Architecture

We are implementing a plan-ranking tool based on the model described above. The general architecture of the tool is given in Figure 1.

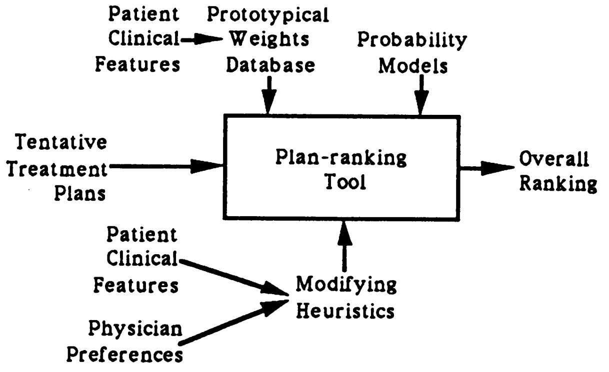

Figure 1: Plan-Ranking Tool Architecture

The tool uses the set of tentative treatment plans along with the clinical features of the patient and the practice preferences of the physician to give the overall plan ranking. This ranking is obtained by sorting the plans based on their *figure of merit*, calculated using Equation 6. With this design architecture and the *FOM* model, we have a methodology for incorporating all of the concerns that go into the evaluation of treatment plans.

Current Status

Our initial model development is focused on three tumor sites — prostate, lung and head-and-neck. From the radiation oncologists serving as our domain experts, we have obtained a set of clinically relevant issues for each of these tumor sites. Since the clinical relevance of an issue is used to make tradeoffs when comparing different plans having different sets of possible complications, we need a preference ordering on the issues. In order to obtain these, we have developed a set of worksheets in which we ask a group of radiation oncologists to rank order the issues. There is one worksheet for each tumor site. Each worksheet contains the list of clinically relevant issues specific to that tumor site. The intent of the ranking is to obtain a preference ordering on the issues; higher ranking means higher morbidity, or a less preferred outcome.

However, this approach poses some problems. There is no easy way to convert the ranking into prototypical weights because differences in the ranks of two issues do not quantify the difference in the clinical relevance of those issues. Also, the physicians are comfortable in comparing pairs of issues, but they find it difficult to give an overall ranking for all issues.

To rectify this, we have changed our approach. Instead of ranking issues, we asked the physicians to give us a *level of concern* for each issue. The level of concern ranges from 0 to 100 and is intended to be a measure of the clinical relevance of an issue. In the case of normal tissues, the issue is the worst possible clinical endpoint resulting from the irradiation of the organ to a dose above its threshold dose. For the tumor volumes, the issue is the non-eradication of the tumor. Figure 2 shows some of the issues and the guidelines that appear on the worksheet for head-and-neck tumors handed out to the physicians.

The physicians feel much more comfortable with this worksheet. However, we have found that their concern for a normal tissue sometimes depends on the fraction of its volume that is being irradiated above its threshold dose (for example, brain in Figure 2). This is because the chance of recovery from the associated clinical endpoint as well as the pain and discomfort felt by the patient depend on the volume of the normal tissue suffering from the endpoint. In the case of certain paired organs, the concern of the physicians depends on whether both organs receive doses above their threshold, or only one of them does (for example, eye lens in Figure 2). This is because losing functionality in only one of the organs leaves the other organ functional, and does not incapacitate the patient. However, losing functionality in both the organs of the pair deprives the patient of some of the capabilities he had, and thus making it a different outcome than losing functionality in only one of the organs.

In order to accomodate these, we have augmented our model. Additional issues are defined to handle the case when both members of paired organs receive doses above their threshold. Also, for the organs in which the concern depends on the fraction of the total volume being irradiated, three numbers are acquired instead of one — for less than one-third of the volume, for less than two-thirds of the volume and for up to the total volume. Note that this augmentation does not violate our initial assumption that the issues are independent. This is because only one of the set of issues for any normal tissue will be present in the *FOM*.

Tumor site: *HEAD & NECK*	LEVEL OF CONCERN
Target volume 3 – gross tumor	

Critical Structures	VOLUME		
Volumes above threshold dose	1/3	2/3	3/3
Brain – necrosis infarction			
Eye lens (unilateral) – cataract			
Eye lens (bilateral) – cataract			

Guidelines:

Level of concern is any number from 0 to 100.
Two or more issues can have same *Level of concern*

Here is an approximate calibration:
100 - cannot ignore (critical)
75 - high concern
50 - moderate concern
25 - mild concern
0 - no concern

Figure 2: Part of a worksheet for head & neck tumors

Our current worksheets incorporate these changes. They are being used to gather data from physicians, and the data will be used to construct the database of prototypical weights.

Conclusion

Selection of the optimal treatment plan from a set of tentative treatment plans involves making tradeoffs on the possible outcomes due to those plans. Previous investigators have employed decision-theoretic techniques to develop a plan-ranking model for a decision problem having the above characteristics. Most models incorporate the morbidities of the various outcomes, the clinical conditions of the patient, and the practice preferences of the institution and the physician into a single factor so that these aspects of the decision problem cannot be examined separately.

Our model provides a framework by which each of the above factors can be specified individually and then combined together to form a plan-ranking model.

This ensures that the model constructed suits both the patient being treated and the physician treating the patient. Note that the model is independent of the domain for which it was developed. The concepts of probabilities, prototypical weights and modifiers can be used in any other medical domain where the plan-ranking problem has similar characteristics. Thus, this model can be used to solve an important class of problems in medical decision making — the ranking of a set of tentative therapeutic plans from best to worst. The figure of merit computed by the model also can be used as the objective function for optimization algorithms which try to find an optimal treatment plan automatically.

Acknowledgments

We thank Dr. James Purdy and Dr. Bahman Emami of the Mallinckrodt Institute of Radiology for their help with the design of the clinical issues worksheet. We also thank the residents, fellows, and clinical faculty members in the Division of Radiation Oncology for their comments and answers to our questionnaires. We acknowledge the support of the NCI sponsored Collaborative Working Group, part of the multi-institutional research effort organized under the Radiotherapy Treatment Planning Tools Contract, for their critique of an earlier version of our model. This project is supported in part by NLM Training Grant 5-T-15-LM07049 and NCI Contract N01-CM-97564.

References

[1] RE Drzymala, L Brewster, J Chu, M Goitein, W Harms, R Mohan, and M Urie. Dose-volume histograms. NCI Contract Report for N01-CM-47696: Evaluation of 3-D Treatment Planning for Photon Beams, 1988.

[2] BR Farr and LM Fagan. Decision-theoretic evaluation of therapy plans. In *Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care*, pages 188–192, 1989.

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

[4] GA Gorry, JP Kassirer, A Essig, and WB Schwartz. Decision analysis as the basis for the computer-aided management of acute renal failure. *American Journal of Medicine*, 55:473–484, 1973.

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

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

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

[8] JT Lyman. Complication probability as assessed from dose-volume histograms. *Radiation Research*, 104:S13–S19, 1985.

[9] JT Lyman and AB Wolbarst. Optimization of radiation therapy, III: A method of assessing complication probabilities from dose-volume histograms. *International Journal of Radiation Oncology, Biology and Physics*, 13:103–109, January 1987.

[10] DK Owens and HC Sox, Jr. Medical decision making: Probabilistic medical reasoning. In EH Shortliffe, LE Perreault, G Wiederhold, and LM Fagan, Editors, *Medical Informatics: Computer Applications in Health Care*, pages 70–116. Addison-Wesley Publishing Company, Reading MA, 1990.

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

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

[13] TE Schultheiss. Treatment planning by computer decision. In *Proceedings of the Eighth International Conference on the Use of Computers in Radiation Therapy*, pages 225–229, 1984.

[14] TE Schultheiss and AM El-Mahdi. Statistical decision theory applied to radiation therapy treatment decisions. In *Proceedings of the Sixth Symposium on Computer Applications in Medical Care*, pages 978–982, 1982.

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

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

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

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