Int. J. Radiation Oncology Biol. Phys., Vol. 64, No. 4, pp. 1275–1286, 2006

0360-3016/06/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2005.11.022

PHYSICS CONTRIBUTION

MULTIVARIABLE MODELING OF RADIOTHERAPY OUTCOMES, INCLUDING DOSE–VOLUME AND CLINICAL FACTORS

Issam El Naqa, Ph.D., Jeffrey Bradley, M.D., Angel I. Blanco, M.D., Patricia E. Lindsay, Ph.D., Milos Vicic, Ph.D., Andrew Hope, M.D., and Joseph O. Deasy, Ph.D.

Department of Radiation Oncology, Washington University, St. Louis, MO

Purpose: The probability of a specific radiotherapy outcome is typically a complex, unknown function of dosimetric and clinical factors. Current models are usually oversimplified. We describe alternative methods for building multivariable dose–response models.

Methods: Representative data sets of esophagitis and xerostomia are used. We use a logistic regression framework to approximate the treatment–response function. Bootstrap replications are performed to explore variable selection stability. To guard against under/overfitting, we compare several analytical and data-driven methods for model-order estimation. Spearman's coefficient is used to evaluate performance robustness. Novel graphical displays of variable cross correlations and bootstrap selection are demonstrated.

Results: Bootstrap variable selection techniques improve model building by reducing sample size effects and unveiling variable cross correlations. Inference by resampling and Bayesian approaches produced generally consistent guidance for model order estimation. The optimal esophagitis model consisted of 5 dosimetric/clinical variables. Although the xerostomia model could be improved by combining clinical and dose–volume factors, the improvement would be small.

Conclusions: Prediction of treatment response can be improved by mixing clinical and dose–volume factors. Graphical tools can mitigate the inherent complexity of multivariable modeling. Bootstrap-based variable selection analysis increases the reliability of reported models. Statistical inference methods combined with Spearman's coefficient provide an efficient approach to estimating optimal model order. © 2006 Elsevier Inc.

Treatment response modeling, Normal tissue complication probability, Logistic regression, Model selection, Information theory, Radiotherapy.

INTRODUCTION

The era of three-dimensional treatment planning in radiotherapy has opened the way for patient-specific, individualized treatment planning decisions based on estimates of the risks of complications vs. increases in local control (1–3). This type of individualization of treatment planning can be accomplished, however, only with predictive models validated against relevant outcome data sets. This has been partially enabled by the recent widespread adoption of three-dimensional data-based treatment planning. Thus, the store of data that relates outcomes to dose distributions is rapidly increasing (4–9). However, there are many open issues in treatment response model building based on complex factors, such as the following: How complicated should the models be? How should models be built? How can the predictive power of models best be tested before completely new data sets are gathered?

Radiotherapy outcomes are usually characterized by tumor local control probability (TCP), and the side effects to the surrounding tissues are quantified by normal tissue complication probabilities (NTCP), as illustrated in Fig. 1. The *x* axis will be referred to here as the "damage metric," whereas the *y* axis is the response. The basic idea is that, as increasing damage is done to the tissue, there is an increasing probability of a clinical complication. Currently, data relevant to various NTCP end points are expanding at a more rapid pace than data and modeling for TCP (4). Variations in normal tissue dose–volume histograms (DVHs) from patient to patient are significantly greater than tumor DVH variations, thus providing more valid input to define models meant to study the effect of dose nonuniformity. Moreover, proliferative tumors often regress rapidly during therapy, are highly functionally heterogeneous, and are more difficult than normal tissues to define radiobiologically. This work therefore focuses on NTCP modeling, though the

Reprint requests to: Joe Deasy, Ph.D., Department of Radiation Oncology, Division of Bioinformatics and Outcomes Research, Washington University School of Medicine, 4921 Parkview Place, Box 8224, St. Louis, MO 63110. Tel: (314) 362-1420; Fax: (314) 362-8521; E-mail: deasy@wustl.edu

This work was supported in part by NIH grant R01 CA 85181.

Received Aug 2, 2005, and in revised form Nov 7, 2005. Accepted for publication Nov 17, 2005.

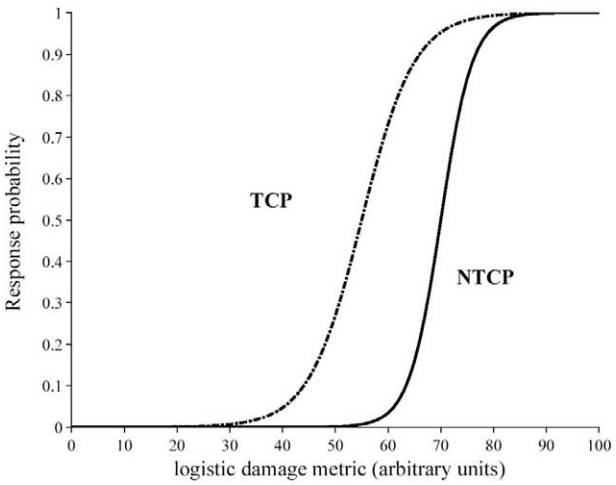

Fig. 1. Sigmoidally shaped response curves (for tumor control probability of normal tissue complication probability) are constructed as a function of a linear weighting of various factors, for a given dose distribution, which may include multiple dose–volume metrics as well as clinical factors. The units of the *x* axis may be thought of as "equivalent dose" units.

presented techniques could be easily integrated into TCP modeling.

Many dose–volume outcome models have been proposed in the literature. The Lyman–Kutcher–Burman model was one of the earliest proposed (6, 7, 10) and uses a simple power law, with an exponent that controls the volume effect. The often-used generalized equivalent uniform dose equation (also called the generalized mean dose) is the same as the dose–volume reduction model used by Lyman–Kutcher–Burman, with a slight change of parameter definition (11). Other models have been proposed that attempt to explicitly model tissue architecture (12–15). All of these models use information only about the dose distribution and fractionation. However, it is well known that the probability of a complication may be affected also by multiple clinical prognostic factors, such as surgery, diabetes, smoking, age, anemia, gender, etc. (16).

Highly simplified models applied to the data, although useful, have the fundamental limitation that they cannot follow the presumably complicated shape of the outcome probability surface as a function of dose and other correlated factors. It seems unlikely that any single mathematical function (mean dose, maximum dose, generalized mean dose, etc.) will alone provide the best approximation for the true NTCP or TCP response curve. Buckland *et al.* (17) described their view of biostatistical modeling, which we believe applies also to radiotherapy: *"Our philosophy is that truth is high (effectively infinite) dimensional. The more information that is gathered, the greater is the model complexity that the data can support. If data are sparse, they can support only a simple model with few parameters. In our view, model selection is the process of identifying the best approximating model, accepting that the data can never support, and we can never identify, the true model."*

It seems likely that in all end points of interest for NTCP and TCP modeling, there is no "correct model." There is only an outcome probability surface that, as a result of highly complicated radiobiologic and physiologic effects, is a very complicated, nonlinear function of treatment-, patient-, and disease-related factors.

For this reason, the present paper concentrates, not on proposing new equations as a basis for modeling (although there are surely new models to be developed), or on fitting simple model equations with few parameters to data, but on methods to find the best (or nearly best) approximation of the data in hand with respect to the true response surface. This is referred to in the "data mining" community as the principle of parsimony (18).

We propose doing this by approximating the complication probability surface using variables selected from a pool of terms that *a priori* might have a causal relationship with the complication surface. Models that mix dose-based metrics with other patient- or disease-based prognostic factors are relatively new, but have been discussed by others (19–21). Our proposal is a "data mining" and model building approach.

A key issue is how to estimate the predictive power of alternative candidate models, that is, to answer the question: Which model would work best on unseen samples generated by the source complication probability density function learned from the patients' population? On the one hand, underfitting (too few fitted variables) should be avoided, because it will lead to weaker predictive power than is possible. On the other hand, overfitting (fitting the "noise" as well as the underlying biologic effect) will also lead to suboptimal predictive power. We study several methods of judging prediction power using analytical models based on information theory and data resampling approaches such as bootstrap and cross-validation methods.

In addition, the question of variable selection and whether some variables are included in the model by accident is always an issue. We address this issue using graphical and quantitative analyses of the variable selection process applied to bootstrap data replicates, as explained below.

The techniques are applied to recently accrued clinical data sets: esophagitis after lung radiotherapy and reduced salivary function (xerostomia) after head-and-neck radiotherapy. These analyses are primarily to demonstrate the modeling techniques and can be considered adjuncts to analyses presented elsewhere (22, 23). The esophagitis analysis primarily mixes variables of different dosimetric types, whereas the xerostomia analysis mixes a well-tested dose–volume model with candidate clinical factors.

MODELING METHODS

Data collection using CERR

Data collection and analysis were aided greatly by CERR (A Computational Environment for Radiotherapy Research) (24). CERR (pronounced "sir") is an open source, freely available general treatment plan analysis package based on Matlab, a high-