Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2007 July 15; 68(4): 1212–1221. doi:10.1016/j.ijrobp.2007.03.064.

PREDICTING LUNG RADIOTHERAPY-INDUCED PNEUMONITIS USING A MODEL COMBINING PARAMETRIC LYMAN PROBIT WITHNONPARAMETRIC DECISION TREES

Shiva K. Das, Ph.D., Sumin Zhou, Ph.D., Junan Zhang, Ph.D., Fang-Fang Yin, Ph.D., Mark W. Dewhirst, D.V.M., Ph.D., and Lawrence B. Marks, M.D.

Department of Radiation Oncology, Duke University Medical Center, Durham, NC

Abstract

Purpose: To develop and test a model to predict for lung radiation-induced Grade 2+ pneumonitis.

Methods and Materials: The model was built from a database of 234 lung cancer patients treated with radiotherapy (RT), of whom 43 were diagnosed with pneumonitis. The model augmented the predictive capability of the parametric dose-based Lyman normal tissue complication probability (LNTCP) metric by combining it with weighted nonparametric decision trees that use dose and nondose inputs. The decision trees were sequentially added to the model using a “boosting” process that enhances the accuracy of prediction. The model's predictive capability was estimated by 10-fold cross-validation. To facilitate dissemination, the cross-validation result was used to extract a simplified approximation to the complicated model architecture created by boosting. Application of the simplified model is demonstrated in two example cases.

Results: The area under the model receiver operating characteristics curve for cross-validation was 0.72, a significant improvement over the LNTCP area of 0.63 ($p = 0.005$). The simplified model used the following variables to output a measure of injury: LNTCP, gender, histologic type, chemotherapy schedule, and treatment schedule. For a given patient RT plan, injury prediction was highest for the combination of pre-RT chemotherapy, once-daily treatment, female gender and lowest for the combination of no pre-RT chemotherapy and nonsquamous cell histologic type. Application of the simplified model to the example cases revealed that injury prediction for a given treatment plan can range from very low to very high, depending on the settings of the nondose variables.

Conclusions: Radiation pneumonitis prediction was significantly enhanced by decision trees that added the influence of nondose factors to the LNTCP formulation.

Keywords

Lung; Radiotherapy; Pneumonitis; Prediction; Models

INTRODUCTION

Radiation-induced pneumonitis is a serious problem in patients undergoing radiotherapy (RT) for tumors located in and around the thorax (1-8) because of the incidental radiation to the adjacent/surrounding healthy lung. Retrospective studies have noted an incidence of high-grade radiation pneumonitis of 15–36%, for various dose regimens (1-8). Understanding the

Reprint requests to: Shiva K. Das, Ph.D., Department of Radiation Oncology, Duke University Medical Center, Box 3295, Durham, NC 27710. Tel: (919) 681-5424; Fax: (919) 681-7183; E-mail: shiva.das@duke.edu.

Conflict of interest: none.

factors involved in causing radiation pneumonitis is important to be able to reduce its incidence, possibly by altering those factors that can be controlled.

Various studies have sought to determine predictors for radiation pneumonitis, dosimetric and otherwise. These studies have typically extracted one or more factors that correlated statistically with the incidence of radiation pneumonitis. Among such factors, the ones that appear with most frequency in published studies are the lung volume receiving >20 Gy (1, 9-15), lung volume receiving >30 Gy (9,14,16,17), and the mean lung dose (1,11,15,17-20). Other factors that have been identified are the lung volume receiving >5-13 Gy (21), >15 Gy (9), >40 Gy (14), and >45 Gy (14) and the tumor lobe location (22-24). In two recent studies, Kong *et al.* (1) identified the lung volume receiving >20 Gy and the mean lung dose, and Hope *et al.* (24) identified lower lobe tumor location, maximal lung dose, and minimal dose to the 35% volume receiving the highest doses as predictive of lung radiation pneumonitis.

The goal of the present work is in the same general vein as these earlier studies (*i.e.*, to predict for lung radiation-induced pneumonitis). The model we have created seeks to augment the predictive capability of the commonly used dose-based parametric Lyman normal tissue complication probability (LNTCP) metric (25) using nonparametric decision trees that account for the effect of other important dose and non-dose variables. Nonparametric predictors, which can flexibly tailor dose-response behavior to the input data, generally have an advantage over parametric predictors, which assume a fixed mathematical functional form for the dose-response behavior (26). Decision trees have the potential to divide the patient population and extract predictive trends that are only valid within each subpopulation. The flexible nature of nonparametric predictors, however, also allows them to detrimentally overfit to the data used to create them. This can result in poor predictive accuracy using other data. In the present study, the predictive accuracy of the model was realistically estimated using 10-fold cross-validation (27). The data were divided into 10 approximately equal groups, and each group was tested using the model built with the remaining 9 groups.

The process of building the model uses a process known as boosting (AdaBoost algorithm [28]) to enhance the predictive capability. Boosting combined together several weighted individual predictors into one model, with each individual predictor composed of the LNTCP metric and a decision tree. The complicated architecture that results from boosting, although capable of improving the model predictive capability, can also reduce interpretability of the influence of the variables on model outcome. To improve interpretability, and aid in dissemination, a simpler approximate model was created from the cross-validated results of the original model. This simpler, approximate model attempted to capture the response of the original model to important variables, allowing us to readily interpret the influence of combinations of variables on the response. The application of the simpler, approximate model is illustrated in 2 patient cases.

METHODS AND MATERIALS

Patient characteristics

The patient database consisted of 234 patients with lung cancer treated with RT at our institution, of whom 43 were diagnosed with Grade 2+ radiation-induced pneumonitis during follow-up (every 3-4 months after treatment). Radiation-induced pneumonitis was graded as Grade 0, no increase in pulmonary symptoms due to RT; Grade 1, radiation-induced symptoms not requiring initiation or an increase in steroids and/or oxygen; Grade 2, radiation-induced pulmonary symptoms requiring initiation or an increase in steroids; Grade 3, radiation-induced pulmonary symptoms requiring oxygen; and Grade 4, radiation-induced pulmonary symptoms requiring assisted ventilation or resulting in death. The patient characteristics and treatment details are described in Table 1. The RT plans consisted of anteroposterior primary beams

delivering 40–45 Gy, followed by off-cord parallel opposed boost beams or multiple non-coplanar, nonaxial beams (29). The treatment sessions were once daily (1.8–2.0 Gy/fraction) or twice daily (1.25 Gy/fraction to the clinical target volume and 1.6 Gy/fraction to the gross volume with the fractions separated by a minimum of 6 h).

The total available set of variables for these patients consisted of dose and nondose variables. The dose variables consisted of the lung dose–volume histogram (in 2-Gy bins), lung generalized equivalent uniform dose (EUD) (30), computed from the lung dose–volume histogram [$EUD = (\sum_i V_i D_i^\alpha / \sum_i V_i)^{1/\alpha}$, where V_i is the volume at dose D_i for various values of the exponent α , ranging from 0.4 to 4 (in increments of 0.1)], mean heart dose, and treatment fraction size. The nondose variables were age, gender, race, histologic type (squamous cell cancer, adenocarcinoma, non-small-cell cancer, small cell cancer, large cell cancer, other), tumor stage (I through IV), central/peripheral tumor location in lung, upper/lower/middle tumor location in the lung, right/left lobe tumor location in the lung, chemotherapy schedule with respect to RT (before, concurrent, before and concurrent, after, concurrent and after, none), surgery (yes vs. no), once or twice-daily RT, pre-RT forced expiratory volume in 1 s (expressed as the percentage of the predicted normal), pre-RT diffusion capacity of carbon monoxide in the lung (DLCO), and pre-RT DLCO expressed as a percentage of the predicted normal (DLCO%). Note that only some subset of the total available set of variables will be selected by the model.

Model building and testing

The model was composed of a sequence of weighted individual decision prediction units (Fig. 1). Each prediction unit in the sequence was built using patient weights that were computed according to the predictive capability of the previous unit in the sequence. The first prediction unit in the sequence was built with equal patient weights. The process of changing the patient weights for each prediction unit in the sequence, and computing the weight associated with each prediction unit, used the AdaBoost algorithm (described in detail later). The AdaBoost algorithm attempts to sequentially increase the model predictive capability (“boosting”) by adding on each successive predictive unit. In this study, each unit (with the exception of the first unit) was the sum of the LNTCP metric and a decision tree; the first unit was composed of only the LNTCP metric. The LNTCP metric is described in the Appendix. This choice of composition of the decision prediction unit was motivated by the rationale that the nonparametric decision tree component would augment the predictive capability of the more rigidly parametric, but widely used, LNTCP metric. Thus, the decision tree model, which is capable of using dose and nondose variables as inputs, would compensate for the shortcomings in the LNTCP metric, which uses only dose in a parametric formulation.

A predictive model is typically built from a subset of patients termed the training set and then evaluated by testing it on the remaining patients (test set). To realistically evaluate the model predictive capability on all patients, we used a 10-fold cross-validation procedure (27). In this procedure, the patients were randomly divided into 10 groups of approximately equal size (each group contained approximately one-tenth of the injured and uninjured patients). The model was built using data from nine groups of patients and then tested on the remaining group. Each group, in turn, served as the test group. Figure 2 illustrates this concept. The combined result of testing on all groups, with each group tested one at a time, was then used to estimate the model predictive capability. The metrics used for model evaluation were the area under the receiver operating characteristic (ROC) curve (31), sensitivity (true-positive fraction), and specificity (true-negative fraction).

The process of model building using the AdaBoost boosting algorithm is shown in Fig. 3. The algorithm built the units (U) in sequence, denoted by the loop. Within the loop, the prediction error, ε , from a prediction unit was computed as the sum of the individual patient errors

multiplied by the corresponding patient weights. Individual patient error was computed as the deviation of the prediction from 1 for patients with injury, or 0 for patients without injury. The prediction error ε was then used to compute the weight of the prediction unit and compute the patient weights required to build the next prediction unit in the sequence. In essence, the AdaBoost algorithm lowered the weights of the patients for the next prediction unit in accordance with the error incurred by the current prediction unit; thus, patients with higher error were reduced less, and *vice versa*. Once built, the model estimate of injury for any patient was obtained as the weighted sum of the individual prediction unit outputs.

The trees constituting the prediction units in the AdaBoost algorithm (Figs. 1 and 3) were built according to the flowchart in Fig. 4. The tree originated at a root node, which split the patient population into two branches, with the aim of maximally separating the two classes according to the gini index split threshold criterion (27). The method of splitting the patient population at a node is shown in Fig. 5 and explained in the next paragraph. The split was considered successful if any reduction occurred in the gini index after the split, with respect to the gini index before splitting. This is equivalent to a positive value for the following expression:

$$N_S \left(1 - p_{inj,S}^2 - p_{uninj,S}^2 \right) - N_L \left(1 - p_{inj,L}^2 - p_{uninj,L}^2 \right) - N_R \left(1 - p_{inj,R}^2 - p_{uninj,R}^2 \right)$$

where subscripts *inj* and *uninj* refer to the injured and uninjured patients, respectively; subscripts *S*, *L*, and *R* refer to the node being split and the left and right branches, respectively; and *p* and *N* is the proportion and number of patients, respectively. The three bracketed terms in this expression are the gini indexes for the node being split and the left node and right node, respectively. The left and right branches were designated as predicting for injury and no injury, respectively. The nodes at the ends of the branches were, in turn, passed through the node-splitting algorithm (loop in Fig. 4). A node that could not be split was designated as terminal. When all nodes were terminal, tree building was deemed complete. A patient who travels down the tree is predicted to incur a probability of injury corresponding to the terminal node at which the patient stops. This probability of injury is proportional to the number of training set injured patients who have stopped at the same terminal node. Because of the left and right branch designations, left branch terminal nodes had a greater probability of injury than their corresponding right branch terminal nodes.

The process of node splitting is shown in Fig. 5. In typical tree implementations (32), a single variable and its threshold are used to split the patient population at a node (*e.g.*, patients at a node aged >55 years branching to the left and those aged ≤55 years branching to the right). A drawback of this typical implementation is the high variance of the resulting decision tree (27), which is manifested as poor predictive capability on an independent test set of patients. The high variance can be attributed to three aspects of this typical node-splitting strategy: (1) although only one variable is selected to split a node, several other variables with split criteria that are almost as good are not selected; (2) a small number of “non-representative” patients can influence the value of the split threshold for a selected variable; and (3) the binary allocation of outcome to the left/right branches rigidly categorizes a patient, although the left branch might only be suggestive of a greater analog injury probability than the right branch. The node-splitting procedure adopted in the present study (Fig. 5) attempted to reduce the effect of these issues, so as to create a more robust, lower variance, tree. Each of the variables available to the model was evaluated on the set of patients at that node. The candidate variables for splitting a node were ranked from best to worst, with the best variable achieving the greatest reduction in the gini index. For each continuous variable, the split threshold was computed by fitting normal distributions to the injured- and uninjured-node patients and choosing the value

that maximally separated the distributions. This method of selecting the split threshold attempted to reduce its dependence on a small “nonrepresentative” subgroup of patients. Furthermore, an internal validation process was used to select split variables that were robust, by selecting their split thresholds from one portion of the training set and testing the thresholds on the remaining portion of the training set (Fig. 5, dashed box). Once the candidate variables were ranked, those that achieved a gini index reduction of $\geq 80\%$ of the maximal gini reduction were selected as the node split variables (80% was selected, during the course of model building, as the level at which trees built from the training set had the least variance). As a consequence of this node split variable selection method, only some subset of the available variables were chosen by the model.

The selection of input split variables for a node took into account “missing” data (*i.e.*, patient variable data that were not documented/collected). For our patient data set, the numbers of patients whose data were missing for a particular variable were as follows: mean heart dose, 7; histologic type, 1; stage, 6; upper/lower/middle tumor location, 1; pre-RT forced expiratory volume in 1 s, 29; pre-RT DLCO, 44; and pre-RT DLCO%, 44. The split threshold for a variable with missing data was computed from the available data. If a variable with missing data was chosen as an input split variable at a node, the patients with missing data corresponding to this variable traveled down both branches originating at the node.

To reduce the possibility of overfitting, not more than three node splits were allowed for each component decision tree. Also, certain rules were implemented to prevent the model from splitting variables in an “unrealistic” manner (*e.g.*, a higher lung volume above a given dose predicting for a reduced incidence of injury). These rules were set as a directional parameter during the tree building process: a direction of +1 for a variable indicated that the variable had to be above a split threshold to predict for injury, a direction of -1 indicated that the variable had to be below a split threshold to predict for injury, and a direction of 0 indicated that values either above or below a split threshold could be used to predict for injury. All dose variables and stage were assigned a direction of +1. The pre-RT forced expiratory volume in 1 s, pre-RT DLCO, and pre-RT DLCO% were assigned a direction of -1. All other variables were assigned a direction of 0.

Model simplification

An advantage of a single decision tree, such as CART (Classification and Regression Trees) (32), is its easy interpretability. The AdaBoost algorithm, however, although capable of increasing model predictive capability by stringing together a sequence of decision trees, also reduces this interpretability. In this study, the results of cross-validated testing on the patients were used to create an approximate, simplified, but nevertheless easily interpretable, model. The rationale for this simplified, approximate, model was the interpretability of the influence of variables used in the actual model and the ease of dissemination. The simplified model was created by varying the input variables over a range of values and fitting the corresponding cross-validated patient outputs from the actual model to a function of the inputs. Application of the simplified model is illustrated in 2 patients with lung cancer. In both patients, multiple retrospective RT plans with increasing normal lung radiation involvement were created. The influences of dose and nondose variables on injury prediction are given.

RESULTS AND DISCUSSION

Table 2 compares the lung dose–volume histogram metrics and the LNTCP metric for all patients in the training set (overestimate of predictive capability) and for 10-fold cross-validation (realistic estimate of predictive capability). All metrics appeared to be robust, reflected in the minimal drop of ROC area between the overestimate and realistic estimate. Also, all metrics had approximately equal predictive capability, with the cross-validated ROC

areas ranging from 0.61 to 0.64. The cross-validated LNTCP achieved an ROC area of 0.63, with a sensitivity and specificity of 60% and 59%, respectively. The maximum likelihood estimated parameters for the LNTCP metric were $\alpha = 1.94$, uniform dose for 50% probability of injury = 29.6 Gy, and $\sigma = 18.8$ Gy (see Appendix). An exponent $\alpha > 1$ implies that the LNTCP metric selects an equivalent uniform dose greater than the mean lung dose ($\alpha = 1$). The mean lung dose metric yielded the greatest cross-validated ROC area of 0.64. In previous work from our institution (17) on a subset of the patients in this study, the variables lung volume receiving >20 Gy, >30 Gy, and >40 Gy, mean lung dose, and LNTCP were all significantly associated with radiation pneumonitis on univariate analysis ($p < 0.05$). Table 2 supports that result in the sense that these variables yielded approximately equal areas under the cross-validated ROC curves.

Figure 6 shows the ROC area of the model as a function of the number of predictive units constituting the model for all patients in the training set (overestimate of predictive capability) and 10-fold cross-validation (realistic estimate of predictive capability). The ROC areas increase rapidly beyond the first predictive unit (first predictive unit = LNTCP metric; Fig. 1). After approximately 11 predictive units, very little change occurred between the overestimate (ROC area, 0.77) and realistic estimate (ROC area, 0.72) of model predictive capability. Thus, the first 11 predictive units were used to constitute the model. The sensitivity and specificity of the cross-validated model was 70% and 68%, respectively. Improvement in the model compared with the LNTCP metric (cross-validated ROC area, 0.72 vs. 0.63) was significant ($p = 0.0054$; significance testing used ROCKIT [33]). The variables used as model inputs were LNTCP; lung volume receiving >50 Gy and >60 Gy; EUDs for exponent $\alpha = 2, 3$, and 4; gender; histologic type (squamous or otherwise); chemotherapy (before RT or otherwise); and once/twice-daily treatment.

The simplified, approximate model is shown in Fig. 7. The right and left branches of the decision tree correspond to the true and false states, respectively, of the node variables. The injury output for a patient is the sum of one-half the LNTCP metric and the value of the tree node at which the patient terminated. Within the range of input parameters used to create this simplified model, the influence of certain inputs (lung volume receiving >50 Gy and >60 Gy and EUDs for exponent $\alpha = 2, 3$, and 4) was small enough to be discarded. The simplified model separated the lung dose variables (LNTCP) and nondose variables. The ROC area of the simplified model was 0.75, greater than the corresponding cross-validated area of the actual model (0.72). This implies that the simplified model, in the process of approximation, created a more optimistic overfit to the data. The variables used by the model were LNTCP, chemotherapy before RT (true/false), once-daily treatment (true/false), female gender (true/false), and squamous cell histologic type (true/false). In a previous analysis on a subset of the patients in this study (17), none of these variables, with the exception of LNTCP, were significant on univariate or multivariate analysis. The reason for this discrepancy was likely the use of decision trees in the model developed in the present study—decision trees are able to incorporate nodal split variables that only discriminate on the subset of data at the node (compared with univariate/multivariate variables that are required to discriminate on all data).

As seen in Fig. 7, the highest to lowest injury output occurred for the following scenarios: (1) chemotherapy before RT, once-daily treatment, and female gender; (2) chemotherapy before RT, once-daily treatment, and male gender; (3) chemotherapy before RT, twice-daily treatment, and female gender; (4) no chemotherapy before RT, squamous cell histologic type, and male gender; (5) no chemotherapy before RT, squamous cell histologic type, and female gender; (6) chemotherapy before RT, twice-daily treatment, and male gender; and (7) no chemotherapy before RT and histologic type other than squamous cell. Although injury output can be used in a relative sense to compare different plans for the same patient, or to compare different patients, a more easily interpretable measure is the number of patients in the Duke

data set with less or more injury than the patient being evaluated. Figure 8 can be used for this purpose. Figure 8 shows normal fits to the simplified model output for injured and uninjured patients in the Duke database. The fits were significant for normality at a level of 0.05, using the Lilliefors test in MATLAB (Mathworks, Natick, MA). The fit parameters for the model output were mean of 0.521, standard deviation of 0.076 for injured patients and mean of 0.437 and standard deviation of 0.10 for uninjured patients. Application of this simplified, approximate model to 2 example patients (Male A, Female B) is illustrated. The locations of the primary and boost target volumes are shown in Fig. 9 on transverse and coronal planes through the patients. Three-dimensional conformal treatment plans (15-MV photons, 0.8-cm conformal margin for all beams) were generated on the ECLIPSE RT planning system (Varian Medical Systems, Palo Alto, CA).

For Patient A (male), two RT plans were generated: A1 and A2. Plan A2 involved a greater radiation dose to the normal lung to satisfy the target prescription goal of 40 Gy to the primary target and 66 Gy to the boost target. The beam directions used in the plans were anteroposterior–posteroanterior for the primary target and 20° off vertical parallel-opposed for the boost target (Plan A1) and right–left parallel opposed for both primary and boost targets (Plan A2). Considering the location of the targets, Plan A2 would likely not be considered clinically acceptable owing to the large volume of normal lung involvement. Table 3 ranks the plans in order from lowest to highest injury output for permutations of the nondose parameters used in the simplified model. The last two columns of Table 3 list the percentages of injured/uninjured patients in the Duke database with lower/higher injury output than the patient being evaluated. The percentages were computed from the normal fits described in the previous paragraph. The LNTCP metric for Plan A1 was lower than that for Plan A2 (0.50 vs. 0.73). However, for certain combinations of nondose variables (*e.g.*, nonsquamous cell histo-logic type, no chemotherapy before RT), Plan A2 had the second lowest prediction of injury (36% of injured Duke database patients below and 29% of uninjured Duke database patients above). In general, Plan A2 had a greater prediction of injury. Various combinations of Plans A1 and A2 and the nondose parameters predicted for a wide range of injury, ranging from 3% to 97% of injured Duke database patients below Patient A (3–64% for Plan A1 and 36–97% for Plan A2). The lowest prediction of injury was for Plan A1 with nonsquamous cell histologic type and no chemotherapy before RT. The greatest prediction of injury was for Plan A2 with chemotherapy before RT and once-daily treatment.

For Patient B (female), two sets of plans were created: one set with the primary and boost targets treated to 40 Gy and 66 Gy, respectively, and another set with corresponding doses of 40 Gy and 74 Gy. In each set, four plans with increasing lung involvement were created (two beams, four beams, six beams, and eight beams). The beam directions used by the plans were as follows: two-beam plan, anteroposterior–post-eroanterior; four-beam plan, two-beam plan directions plus two fields angled at 45° on either side of the anteroposterior direction; six-beam plan, four-beam plan directions plus two fields angled at 45° on either side of the posteroanterior direction; and eight-beam plan, six-beam plan directions plus two fields in the right–left direction. Table 4 shows the various dose and nondose parameters, arranged in row order from lowest to highest injury prediction. The smallest model injury prediction (27% of Duke database injured patients below patient B) corresponded to the smallest LNTCP for the four-field, 66-Gy plan, with nonsquamous cell histologic type and no chemotherapy before RT. In general, however, a smaller LNTCP does not necessarily correspond to lower model injury prediction. For example, the eight-field, 74-Gy plan with the highest LNTCP of 0.82 (nonsquamous cell histologic type and no chemotherapy before RT) had a lower injury prediction than the four-field, 66-Gy plan with the lowest LNTCP of 0.70 (chemotherapy before RT and once-daily treatment). The model injury predictions for these two plans were 57% vs. 95% of the Duke database injured patients below Patient B. The greatest injury

prediction corresponded to the eight-field, 74-Gy plan, with chemotherapy before RT and once daily treatment (99% of Duke database injured patients below Patient B).

CONCLUSION

The lung radiation pneumonitis prediction model we have developed is composed of a weighted sequence of prediction units, with each prediction unit composed of the sum of the LNTCP metric and a decision tree. To improve the interpretability of this complicated model architecture and aid in dissemination a simplified approximation to this model was developed. The simplified, approximate model is composed of the LNTCP metric and a decision tree that uses the following nondose variables: chemotherapy before RT, histologic type (squamous cell or otherwise), once/twice-daily treatment, and gender. When prospectively applied, the injury output from the simplified model can be used to compare the patient being evaluated against the normal fits to the Duke database (percentage of Duke database injured/uninjured patients with lower/higher injury than the patient being evaluated). Application of the simplified injury prediction model to 2 patients showed that various settings of nondose variables, in combination with the extent of normal lung involvement in the RT plan, can predict for a wide range of radiation pneumonitis injury.

Acknowledgements

Supported by National Institutes of Health Grants R01 CA115748 and R01 CA69579.

APPENDIX

Lyman normal tissue complication probability

The Lyman normal tissue complication probability (LNTCP) is a commonly used parametric metric that quantifies the probability of injury for a given uniform dose to an organ by the integral of a gaussian probability function from $-\infty$ to the uniform dose (25). The parameters defining the gaussian function are the mean (the uniform dose for 50% probability of injury) and standard deviation (σ). Because the lung dose distribution is nonuniform, the equivalent uniform dose was approximated using the generalized equivalent uniform dose formulation (30) as uniform dose = $(\sum_i V_i D_i^\alpha / \sum_i V_i)^{1/\alpha}$, where V_i is the volume at dose D_i . Thus, the three unknown parameters are the uniform dose for 50% probability of injury, σ , and the generalized equivalent uniform dose exponent α . These three parameters were fitted to the data using the maximum likelihood estimation technique (34).

REFERENCES

1. Kong F-M, Hayman JA, Griffith KA, et al. Final toxicity results of a radiation-dose escalation study in patients with non-small-cell lung cancer (NSCLC): Predictors for radiation pneumonitis and fibrosis. *Int J Radiat Oncol Biol Phys* 2006;65:1075–1086. [PubMed: 16647222]
2. Maguire PD, Marks LB, Sibley GS, et al. 73.6 Gy and beyond: Hyperfractionated, accelerated radiotherapy for non-small-cell lung cancer. *J Clin Oncol* 2001;19:705–711. [PubMed: 11157021]
3. Rosenzweig KE, Mychalczak B, Fuks Z, et al. Final report of the 70.2-Gy and 75.6-Gy dose levels of a phase I dose escalation study using three-dimensional conformal radiotherapy in the treatment of inoperable non-small cell lung cancer. *Cancer J* 2000;6:82–87. [PubMed: 11069224]
4. Bedini AV, Tavecchio L, Gramaglia A, et al. Radiotherapy and concurrent continuous infusion of cisplatin with adjuvant surgery in nonresectable Stage III lung carcinoma: Short- and long-term results of a phase II study. *Int J Radiat Oncol Biol Phys* 1999;45:613–621. [PubMed: 10524413]
5. Segawa Y, Takigawa N, Kataoka M, et al. Risk factors for development of radiation pneumonitis following radiation therapy with or without chemotherapy for lung cancer. *Int J Radiat Oncol Biol Phys* 1997;39:91–98. [PubMed: 9300744]

6. Oral EN, Bavbek S, Kizir A, et al. Preliminary analysis of a phase II study of paclitaxel and CHART in locally advanced non-small cell lung cancer. *Lung Cancer* 1999;25:191–198. [PubMed: 10512130]
7. Fu XL, Jiang GL, Wang LJ, et al. Hyperfractionated accelerated radiation therapy for nonsmall cell lung cancer: Clinical phase I/II trial. *Int J Radiat Oncol Biol Phys* 1997;39:545–552. [PubMed: 9336130]
8. Byhardt RW, Scott C, Sause WT, et al. Response, toxicity, failure patterns, and survival in five Radiation Therapy Oncology Group (RTOG) trials of sequential and/or concurrent chemotherapy and radiotherapy for locally advanced non-small-cell carcinoma of the lung. *Int J Radiat Oncol Biol Phys* 1998;42:469–478. [PubMed: 9806503]
9. Tsujino K, Hirota S, Kotani Y, et al. Radiation pneumonitis following concurrent accelerated hyperfractionated radiotherapy and chemotherapy for limited-stage small-cell lung cancer: Dose-volume histogram analysis and comparison with conventional chemoradiation. *Int J Radiat Oncol Biol Phys* 2006;64:1100–1105. [PubMed: 16373082]
10. Lind PA, Wennberg B, Gagliardi G, et al. ROC curves and evaluation of radiation-induced pulmonary toxicity in breast cancer. *Int J Radiat Oncol Biol Phys* 2006;64:765–770. [PubMed: 16257129]
11. Chang DT, Olivier KR, Morris CG, et al. The impact of heterogeneity correction on dosimetric parameters that predict for radiation pneumonitis. *Int J Radiat Oncol Biol Phys* 2006;65:125–131. [PubMed: 16427214]
12. Graham MV, Purdy JA, Emami B, et al. Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small-cell lung cancer (NSCLC). *Int J Radiat Oncol Biol Phys* 1999;45:323–329. [PubMed: 10487552]
13. Tsujino K, Hirota S, Endo M, et al. Predictive value of dose-volume histogram parameters for predicting radiation pneumonitis after concurrent chemoradiation for lung cancer. *Int J Radiat Oncol Biol Phys* 2003;55:110–115. [PubMed: 12504042]
14. Rancati T, Ceresoli GL, Gagliardi G, et al. Factors predicting radiation pneumonitis in lung cancer patients: A retrospective study. *Radiother Oncol* 2003;67:275–283. [PubMed: 12865175]
15. Jenkins P, D'Amico K, Benstead K, et al. Radiation pneumonitis following treatment of non-small-cell lung cancer with continuous hyperfractionated accelerated radiotherapy (CHART). *Int J Radiat Oncol Biol Phys* 2003;56:360–366. [PubMed: 12738310]
16. Nielsen M, Hansen O, Vach W. Attempts to predict the long-term decrease in lung function due to radiotherapy of non-small cell lung cancer. *Radiother Oncol* 2006;78:165–168. [PubMed: 16414136]
17. Hernando ML, Marks LB, Bentel GC, et al. Radiation-induced pulmonary toxicity: A dose-volume histogram analysis in 201 patients with lung cancer. *Int J Radiat Oncol Biol Phys* 2001;51:650–659. [PubMed: 11597805]
18. Martel MK, Tenhaken RK, Hazuka MB, et al. Dose-volume histogram and 3-D treatment planning evaluation of patients with pneumonitis. *Int J Radiat Oncol Biol Phys* 1994;28:575–581. [PubMed: 8113100]
19. Kwa SLS, Lebesque JV, Theuws JCM, et al. Radiation pneumonitis as a function of mean lung dose: An analysis of pooled data of 540 patients. *Int J Radiat Oncol Biol Phys* 1998;42:1–9. [PubMed: 9747813]
20. Moiseenko V, Craig T, Bezjak A, et al. Dose-volume analysis of lung complications in the radiation treatment of malignant thymoma: A retrospective review. *Radiother Oncol* 2003;67:265–274. [PubMed: 12865174]
21. Yorke ED, Jackson A, Rosenzweig KE, et al. Correlation of dosimetric factors and radiation pneumonitis for non-small-cell lung cancer patients in a recently completed dose escalation study. *Int J Radiat Oncol Biol Phys* 2005;63:672–682. [PubMed: 15939548]
22. Seppenwoolde Y, De Jaeger K, Boersma LJ, et al. Regional differences in lung radiosensitivity after radiotherapy for non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2004;60:748–758. [PubMed: 15465191]
23. Graham MV, Purdy JA, Emami B, et al. Preliminary results of a prospective trial using three dimensional radiotherapy for lung cancer. *Int J Radiat Oncol Biol Phys* 1995;33:993–1000. [PubMed: 7493861]
24. Hope AJ, Lindsay PE, El Naqa I, et al. Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters. *Int J Radiat Oncol Biol Phys* 2006;65:112–124. [PubMed: 16618575]

25. Lyman JT. Complication probability as assessed from dose volume histograms. *Radiat Res* 1985;104:S13–S19.
26. Das SK, Baydush AH, Zhou S, et al. Predicting radiotherapy-induced cardiac perfusion defects. *Med Phys* 2005;32:19–27. [PubMed: 15719950]
27. Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: Data mining, inference, and prediction. Springer-Verlag; New York: 2002.
28. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput System Sci* 1997;55:119–139.
29. Kocak Z, Yu X, Zhou SM, et al. The impact of pre-radiotherapy surgery on radiation-induced lung injury. *Clin Oncol* 2005;17:210–216.
30. Niemierko A. A generalized concept of equivalent uniform dose (EUD) [Abstract]. *Med Phys* 1999;26:1100.
31. Swets, J.; Pickett, R. Evaluation of diagnostic systems: Methods from signal detection theory. Academic Press; New York: 1982.
32. Breiman, L.; Friedman, J.; Olshen, R., et al. Classification and regression trees. Pacific Grove; Wadsworth: 1984.
33. Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. *Med Decis Making* 1998;18:110–121. [PubMed: 9456215]
34. DeGroot, MH.; Schervish, MJ. Probability and statistics. Vol. 3rd ed.. Addison-Wesley; Boston: 2002.

Fig. 1.

Schematic of boosted model. Prediction units (U) were weighted (w) and summed to create model. Each prediction unit, with the exception of the first, combined the Lyman normal tissue complication probability (LNTCP) and a decision tree; the first prediction unit was composed of only the Lyman normal tissue complication probability. Model output is value reflecting extent of injury predicted—greater value implies more severe injury.

Fig. 2.

Schematic of 10-fold cross validated testing. Each numbered box (group) contains approximately one-tenth of injured and uninjured patients. Each group was tested using model built with the remaining nine groups. Net result from testing all groups (enclosed in dashed box on right) represents realistic estimate of model predictive capability.

Fig. 3.
AdaBoost algorithm for synthesizing model (schematically illustrated in Fig. 1). LNTCP = Lyman normal tissue complication probability.

Fig. 4.
Flowchart for building trees constituting prediction units.

Fig. 5.
Flowchart for variable selection and variable threshold determination for node splitting (see Fig. 4).

Fig. 6. Receiver operating characteristics (ROC) area of model as function of number of predictive units constituting model, for all patients in training set (overestimate of predictive capability) and 10-fold cross-validation (realistic estimate of predictive capability).

Fig. 7. Simplified, approximate model. Injury output for patient is sum of one-half Lyman normal tissue complication probability (LNTCP) metric and value of tree terminal node at which patient terminated. Right/left branches of decision tree correspond to true/false states of node variables. Chemo pre-RT = chemotherapy before radiotherapy; tx = treatment.

Fig. 8.
Normal fits to output of simplified model for (a) injured and (b) uninjured patients in Duke database.

Fig. 9. Transverse and coronal planes through Patients A and B showing primary and boost target volumes. PTV = planning target volume.

Table 1

Patient characteristics and treatment details

Characteristic	Radiation pneumonitis	
	Yes	No
Patients (<i>n</i>)	43	191
Age (y)		
Mean	62	64
Range	43–83	27–87
Gender (%)		
Male	47	57
Female	53	43
Race (%)		
White	93	82
Black or Asian	3	18
Pre-RT FEV ₁ (%)		
Mean	62	63
Range	15–99	16–127
Pre-RT DLCO (%)		
Mean	69	60
Range	32–111	11–129
Fractionation (%)		
Once daily	77	69
Twice daily	23	31
Radiation dose (Gy)		
Once daily		
Mean	65.9	62.5
Range	45.0–86.4	30.0–85.2
Twice daily		
Mean	75.6	71.8
Range	56.4–86.2	45.0–86.4
Histologic type (%)		
SCLC	9	10
NSCLC	91	89
Squamous cell	35	26
Adenocarcinoma	19	19
Other	0	1
Tumor stage (%)		
I-II	14	20
III-IV	86	77
Recurrent	0	3
Tumor location (%)		
Central/peripheral	65/35	60/40
Upper/lower/middle	58/26/16	61/21/17

Characteristic	Radiation pneumonitis	
	Yes	No
Right/left	56/44	59/41
Chemotherapy schedule (%)		
Concurrent with RT	7	14
Before RT	72	42
Before + concurrent	5	6
Concurrent + after	0	2
After	0	4
No chemotherapy	16	32

Abbreviations: RT = radiotherapy; SCLC = small cell lung cancer; NSCLC = non-small-cell lung cancer; FEV₁ = forced expiratory volume in 1 s; DLCO = diffusion capacity of carbon monoxide.

Table 2

Comparison of ROC area for various dose metrics

Metric	ROC area for all patients in training set (overestimate of model predictive capability)	10-fold Cross-validation (realistic estimate of model predictive capability)		
		ROC area	Sensitivity (%)	Specificity (%)
V ₁₀	0.625	0.625	53	55
V ₂₀	0.631	0.627	56	55
V ₃₀	0.622	0.621	58	56
V ₄₀	0.609	0.606	60	56
V ₅₀	0.622	0.621	58	58
V ₆₀	0.615	0.611	60	60
Mean lung dose	0.642	0.642	60	58
LNTCP	0.650	0.630	60	59

Abbreviations: ROC = receiver operating characteristics; V₁₀ through V₆₀ = lung volume receiving >20 Gy through >60 Gy; LNTCP = Lyman normal tissue complication probability.

Table 3
Comparison of simplified model injury prediction for combinations of radiotherapy plan and various settings of nondose parameters, for male patient A

Plan name	Histologic type	Chemotherapy before RT	Once/twice-daily treatment	LNTCP	Injury output* (simplified model)	% Injured patients below	% Uninjured patients above
A1	Nonsquamous	No	Either	0.50	0.38	3	72
A2	Nonsquamous	No	Either	0.73	0.49	36	29
A1	Squamous	No	Either	0.50	0.50	37	28
A1	Any	Yes	Twice	0.50	0.51	43	24
A1	Any	Yes	Once	0.50	0.55	64	13
A2	Squamous	No	Either	0.73	0.61	88	4
A2	Any	Yes	Twice	0.73	0.62	91	3
A2	Any	Yes	Once	0.73	0.66	97	1

Abbreviations: RT = radiotherapy; LNTCP = Lyman normal tissue complication probability.

* Model injury output value; based on injury output value, last two columns are normal fit predictions for Patient A, relative to Duke database population (percentage of injured/uninjured patients in Duke database with lower/higher injury output than Patient A).

Table 4

Comparison of simplified model injury prediction for combinations of radiotherapy plan and various settings of nondose parameters for female patient B

Beams (n)	Total dose (Gy)	Histologic type	Chemotherapy before RT	Once/ twice daily RT	LKB NTCP	Injury _{5%} output	% Injured below	% Uninjured above
4	66	Nonsquamous	No	Either	0.70	0.47	27	36
2	66	Nonsquamous	No	Either	0.71	0.48	30	33
6	66	Nonsquamous	No	Either	0.72	0.49	33	31
8	66	Nonsquamous	No	Either	0.74	0.50	37	28
4	74	Nonsquamous	No	Either	0.78	0.52	47	22
2	74	Nonsquamous	No	Either	0.79	0.52	50	20
6	74	Nonsquamous	No	Either	0.80	0.53	53	18
8	74	Nonsquamous	No	Either	0.82	0.53	57	16
4	66	Squamous	No	Either	0.70	0.59	83	6
2	66	Squamous	No	Either	0.71	0.60	85	5
4	66	All	Yes	Twice	0.70	0.60	86	5
6	66	Squamous	No	Either	0.72	0.61	87	5
2	66	All	Yes	Twice	0.71	0.61	88	4
8	66	Squamous	No	Either	0.74	0.61	89	4
6	66	All	Yes	Twice	0.72	0.62	90	4
8	66	All	Yes	Twice	0.74	0.63	92	3
4	74	Squamous	No	Either	0.78	0.63	93	2
2	74	Squamous	No	Either	0.79	0.64	94	2
4	66	All	Yes	Once	0.70	0.64	95	2
4	74	All	Yes	Twice	0.78	0.65	95	2
6	74	Squamous	No	Either	0.80	0.65	95	2
2	66	All	Yes	Once	0.71	0.65	96	2
2	74	All	Yes	Twice	0.79	0.65	96	2
8	74	Squamous	No	Either	0.82	0.65	96	2
6	66	All	Yes	Once	0.72	0.66	96	1
6	74	All	Yes	Twice	0.80	0.66	96	1
8	74	All	Yes	Twice	0.82	0.66	97	1
8	66	All	Yes	Once	0.74	0.67	97	1

Beams (n)	Total dose (Gy)	Histologic type	Chemotherapy before RT	Once/ twice daily RT	LKB NTCP	Injury * output	% Injured below	% Uninjured above
4	74	All	Yes	Once	0.78	0.68	98	1
2	74	All	Yes	Once	0.79	0.69	99	1
6	74	All	Yes	Once	0.80	0.70	99	0
8	74	All	Yes	Once	0.82	0.70	99	0

Abbreviations as in Table 3.

* Model injury output value; based on injury output value, last two columns are normal fit predictions of Patient B relative to Duke database population (percentage of injured/uninjured patients in Duke database with lower/higher injury output than patient B).