Int. J. Radiation Oncology Biol. Phys., Vol. 59, No. 2, pp. 353–365, 2004

0360-3016/04/$–see front matter

doi:10.1016/j.ijrobp.2003.12.033

CLINICAL INVESTIGATION **Prostate**

DOSE–VOLUME RESPONSE ANALYSES OF LATE RECTAL BLEEDING AFTER RADIOTHERAPY FOR PROSTATE CANCER

Susan L. Tucker, Ph.D.,* Rex Cheung, M.D., Ph.D.,† Lei Dong, Ph.D.,‡ H. Helen Liu, Ph.D.,‡ Howard D. Thames, Ph.D.,* Eugene H. Huang, M.D.,† Deborah Kuban, M.D.,† and Radhe Mohan, Ph.D.‡

Departments of *Biostatistics and Applied Mathematics, †Radiation Oncology and ‡Radiation Physics, The University of Texas M. D. Anderson Cancer Center, Houston, TX

Purpose: To compare the fits of various normal tissue complication probability (NTCP) models to a common set of late rectal toxicity data, with the aim of identifying the best model for predicting late rectal injury after irradiation.

Methods and Materials: Late toxicity data from 128 prostate cancer patients treated on protocol with three-dimensional conformal radiotherapy at The University of Texas M.D. Anderson Cancer Center (UTMDACC) were analyzed. The dose–volume histogram for total rectal volume, including contents, was obtained for each patient, and the presence or absence of Grade 2 or worse rectal bleeding within 2 years of treatment was scored. Five different NTCP models were fitted to the data using maximum likelihood analysis: the Lyman model, the mean dose model, a parallel architecture model, and models based on either a cutoff dose or a cutoff volume.

Results: All five of the NTCP models considered provided very similar fits to the UTMDACC rectal bleeding data. In particular, none of the more highly parameterized models (the four-parameter parallel model, three-parameter Lyman model, or three-parameter cutoff dose and volume models) provided a better fit than the simplest of the models, the two-parameter NTCP model describing rectal bleeding as a probit function of mean dose to rectum.

Conclusion: No dose–volume response model has yet been identified that provides a better description of the UTMDACC rectal toxicity data than the mean dose model. Because this model has relatively low predictive accuracy, the need to identify a better model remains.

Prostate cancer, Rectal toxicity, Normal tissue complication probability, Lyman model, Parallel model.

INTRODUCTION

Radiotherapy (RT) is one of the primary treatment modalities for localized prostate cancer, with doses of 65–70 Gy commonly prescribed when conventional treatment plans are used (1). Conformal techniques including three-dimensional-conformal RT (3D-CRT) and intensity-modulated RT (IMRT) have allowed dose escalation to the prostate beyond 70 Gy with improved prostate-specific antigen (PSA) control and acceptable toxicity (2–5). However, the need for additional improvement remains, both to increase tumor control and to improve the quality of life of prostate cancer survivors by reducing the risk of toxicity to adjacent normal tissues, including the rectum.

Many of the possible 3D-CRT and IMRT dose distributions to tumor and normal tissue are highly complex and cannot be adequately compared simply by visual inspection of the plans. Quantitative methods are vitally needed to assess and compare the vast number of possible dose distributions to select the safest and most effective plan for each patient. To this end, numerous studies have sought to quantify the dose–volume response relationship for late rectal injury after RT (6–21).

To date, no consensus has been reached on the optimal method for estimating the risk of late rectal injury as a function of dose and volume. In part, this is because of the difficulty involved in comparing the results of multiple analyses. Published studies have used various criteria to contour the rectum, a variety of different scoring systems to grade late rectal toxicity and have selected varying levels of injury to categorize patients into those with or without complications. The extent of patient follow-up has also varied among the studies, with a minimum follow-up time required for some studies but not for others. Additionally, in some studies, the time-to-complication data have been analyzed using the Cox proportional hazards model, and in

Reprint requests to: Susan L. Tucker, Ph.D., Department of Biostatistics and Applied Mathematics, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 237, Houston, TX 77030. Tel: (713)792-2613; Fax: (713)792-4262; E-mail: sltucker@mdanderson.org

Received Sep 12, 2003, and in revised form Dec 16, 2003. Accepted for publication Dec 19, 2003.

others, the complication data were dichotomized as a binary endpoint (rectal toxicity: yes vs. no). When the data were dichotomized, a cutoff time was imposed in some studies (e.g., rectal toxicity within 2 years) but not in others (any rectal toxicity during patient follow-up). These differences in data specification make it difficult to assess and compare the accuracy of the mathematical models themselves in predicting complication risk.

In a recent study (22), we used the Lyman normal tissue complication probability (NTCP) model (23) to perform a dose–volume response analysis of rectal toxicity in a cohort of prostate cancer patients treated with 3D-CRT at The University of Texas M.D. Anderson Cancer Center (UTMDACC). In the present study, we continued the analysis of these same data by investigating the fits of other complication probability models. By keeping the clinical data fixed while investigating models of increasing complexity, we hoped to gain a better understanding of the behavior and predictive accuracy of the various models. Ultimately, the goal is to identify the optimal quantitative method for comparing treatment plans with regard to the potential for late rectal toxicity.

METHODS AND MATERIALS

Patient population

The Institutional Review Board of the UTMDACC approved this retrospective analysis. In connection with an earlier study (11), the medical records were reviewed for all patients with localized, biopsy-confirmed prostate cancer treated with definitive 3D-CRT at the UTMDACC between 1992 and 1999, received no hormonal therapy, had a minimum follow-up of 24 months, and had 3D treatment plans that could be recovered from the institutional archives. Of the 163 patients meeting these criteria, 128 had been included in the 78-Gy arm of the UTMDACC dose-escalation trial (4); they were treated with 46 Gy at 2 Gy/fraction to the isocenter using a conventional four-field box technique, followed by a six-field 3D-CRT boost to 78 Gy using two lateral and four oblique fields. Because these 128 protocol patients were scored prospectively for rectal toxicity, they were selected as the population for the present data analysis, as well as for our previous modeling study (22).

Dose–volume histograms

The treatment plans were recovered from the institutional archives and analyzed to obtain a rectal dose–volume histogram (DVH) for each patient. The rectum was outlined 11 cm in length starting 2 cm below the inferiormost aspect of the ischial tuberosities. The entire rectal volume was contoured, including the rectal wall and contents. The dose bins for the DVH were 0.1 Gy in size for all patients, and the doses represent total doses, not corrected for fractionation. Unless otherwise specified, the term DVH always refers to the relative DVH, normalized to 100% of the outlined rectal volume for each patient.

Table 1. UTMDACC modified toxicity score for late rectal complications

Grade	Description
1	Excess bowel movements twice baseline; slight rectal discharge or blood
2	Two or more antidiarrheal agents weekly; two or fewer coagulations for bleeding; occasional steroids for ulceration; occasion dilation; intermittent use of incontinence pads; regular nonnarcotic or occasional narcotic for pain
3	Two or more antidiarrheal agents daily; three or more coagulations or any transfusion; prolonged steroids per enema; hyperbaric oxygen for bleeding/ulceration; regular dilation; persistent use of incontinence pads; regular narcotics for pain
4	Dysfunction requiring surgery; perforation; life-threatening bleeding

Abbreviation: UTMDACC = The University of Texas M.D. Anderson Cancer Center.

Assessment of rectal toxicity

Follow-up examinations for these 128 protocol patients were performed every 3–6 months during the first 2 years after RT, and annually thereafter. Late rectal complications, defined as those occurring ≥6 months after RT completion, were graded using a modified toxicity scale based on criteria from the Radiation Therapy Oncology Group (24), the Late Effects Normal Tissue Task Force (25), and Fox Chase Cancer Center (26) (Table 1). The endpoint selected for the present data analysis was Grade 2 or worse late rectal bleeding within 2 years of treatment. This was the same endpoint used for our previous modeling study (22). Because all patients in this cohort had at least 2 years of follow-up, the status of the endpoint (yes/no) was known for all patients.

NTCP models

Several different dose–volume response models were fitted to the binary rectal bleeding data (Grade 2 or worse rectal bleeding within 2 years of treatment: yes/no). Each of the models considered has the same general form. First, a summary measure μ is extracted from the DVH, and second, the complication probability is modeled as a sigmoid (S-shaped) function of the summary measure μ.

Numerous mathematical expressions could be selected to model the sigmoid curve linking the summary measure μ to the complication probability. These include the probit model, the logistic model, the complementary log-log model, or any of these applied to a transformation of μ such as $\ln(\mu)$. For consistency, we used the probit model throughout this study, except as specified otherwise. The probit model has the form

$$\mathrm{NTCP}(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s(\mu-\mu_{50})} \exp(-u^2/2)du \qquad (1)$$

and includes two unknown parameters: a quantity s deter-

mining the slope of the sigmoid curve, and a parameter μ_{50} determining its position. μ_{50} corresponds to the value of μ for which NTCP(μ) = 50%. In the literature, the parameter s is sometimes replaced by the parameter $m = (s \cdot \mu_{50})^{-1}$ as a measure of slope.

Each NTCP model considered in the present study was fitted to the binary response data using maximum likelihood analysis (27), in which the model parameters are chosen to maximize the probability of occurrence of the observed data by maximizing the log-likelihood (LL) of the model fit. Confidence intervals for the estimated parameter values were derived using the profile likelihood method (27). All computations were performed using Stata (StataCorp, 2003, Stata Statistical Software, release 8.0. College Station, TX).

Lyman model. The well-known Lyman model (23) fits into the schema of the NTCP models described above. Lyman proposed the use of Eq. (1), with the upper limit of integration set equal to $[D - TD_{50}(V)]/[m \cdot TD_{50}(V)]$, to model the complication probability after irradiation of an organ fraction V to dose D. $TD_{50}(V)$, the dose corresponding to a 50% complication risk after irradiation of subvolume V, is assumed to be related to $TD_{50}(1)$ by a power law: $TD_{50}(V) = TD_{50}(1)/V^n$ for some parameter n (23). Various DVH reduction methods have been proposed to obtain a one-step DVH for use in the Lyman model when the dose to the organ is heterogeneous (28–31); these reduction methods correspond to various choices for the summary measure μ of the inhomogeneous DVH.

In the present study, we used the Lyman model applied to the effective dose D_{eff} defined by Mohan *et al.* (31):

$$D_{eff} = \left(\int \nu(D) \cdot D^{1/n} dD \right)^n \tag{2}$$

where *v(D)* is the function representing the differential DVH. In discretized form,

$$D_{eff} = \left(\sum_i \nu_i \cdot D_i^{1/n} \right)^n \tag{3}$$

where v_i is the volume of the dose bin corresponding to dose D_i. The Lyman model with D_{eff} as the summary measure of the DVH is mathematically equivalent to the Lyman model combined with the Kutcher-Burman DVH reduction scheme, in which the DVH is summarized as an effective volume (28). The Lyman model using either the effective dose or the effective volume as the summary measure of the DVH has three unknown parameters, n, m, and $TD_{50}(1)$; the latter is denoted here simply by TD_{50}.

Equivalent uniform dose. The equivalent uniform dose (EUD) is defined as

$$EUD = \left(\sum_i \nu_i \cdot D_i^a \right)^{1/a} \tag{4}$$

for some choice of parameter a (32, 33). If the EUD is fitted to the complication data using the probit link, the EUD model is mathematically equivalent, with the parameter $a = 1/n$, to the Lyman model based on either the effective dose or the effective volume. Because we restricted attention in the present study to the probit link, the EUD model was not considered further.

Mean dose. The mean dose (MD) to an irradiated structure is

$$MD = \sum_i \nu_i \cdot D_i \tag{5}$$

which corresponds to the special case of Eq. (3) in which $n = 1$. Fitting the MD to the complication data using the probit link function leads to a model with two unknown parameters: μ_{50}, denoted here by MD_{50}, and either s or m. Although the resulting model never fits the data better than the Lyman model, the improvement in fit resulting from inclusion of the additional parameter n in the Lyman model may or may not be statistically significant for a given data set.

Cutoff dose model. Dose–volume effects are often presented in terms of the proportion VD_c of an organ receiving doses greater than or equal to some cutoff dose, D_c (e.g., V60 [corresponding to a cutoff dose of 60 Gy]). Mathematically,

$$V(D) = \int_D^{D_{max}} \nu(u) du \tag{6}$$

where *v(D)* again represents the differential DVH and D_{max} is the maximal dose. The discretized form of Eq. (6) is

$$VD_c = \sum_{i \ni (D_i \geq D_c)} \nu_i \tag{7}$$

where the sum is over i such that $D_i \geq D_c$. In the present study, we fitted this model to the rectal bleeding data using the probit link. The resulting model has three unknown parameters: the optimal cutoff dose, D_c, the value of VD_c corresponding to a 50% complication risk, denoted here by $VD_c(50)$, and either s or m. We investigated the fit of this model based on both the relative volume (normalized to total rectal volume for each patient) and absolute rectal volume.

Cutoff volume model. By switching the roles of dose and volume in the cutoff dose model, a similar model is obtained with a cutoff volume V_c instead of a cutoff dose D_c. We then considered the minimal dose, DV_c, to the hottest volume of rectum of size V_c (e.g., D20 [corresponding to 20% of the organ]). Fitting the values of DV_c to the rectal bleeding data using the probit link, a model with three parameters is again obtained: the optimal cutoff volume, V_c, the value of DV_c corresponding to 50% complication probability, denoted

$DV_c(50)$, and either *s* or *m*. Again, both the relative volume and the absolute rectal volume were considered.

Parallel architecture model. A number of authors have described complication probability models for normal tissues having a parallel architecture, in which each small subvolume of the organ contributes independently to tissue function, and in which a complication occurs if the proportion of the organ damaged by radiation exceeds the functional reserve (34–36).

In its simplest form, the parallel model contains four parameters (37), which we denote here by s_D, D_{50}, s_f, and f_{50}. First, the model includes a local damage function *P(D)* representing the probability of destroying organ function in a small subvolume of tissue exposed to dose *D*. *P(D)* increases with increasing dose and takes on values between 0 and 1. Several authors have used a logistic function to model *P(D)* (37–39), but we have used the probit model here for consistency throughout in the choice of sigmoid curves. The parameters of *P(D)* are slope s_D and position D_{50} (see Eq. 1).

The fraction of the organ damaged by RT, f_{dam}, is calculated by integrating the differential DVH against the local damage function *P(D)*:

$$f_{dam} = \int_0^{D_{max}} v(D) \cdot P(D) dD \qquad (8)$$

or, in discretized form:

$$f_{dam} = \sum_i v_i \cdot P(D_i) \qquad (9)$$

Finally, the complication probability is expressed as a function of total organ damage, f_{dam}, using a second sigmoid curve to model the variation in functional reserve among patients. We again used a probit model, with parameters s_f and f_{50}. Note that as the slope (s_D) of the local damage curve *P(D)* becomes infinite, the parallel model converges to the cutoff dose model with $D_c = D_{50}$.

Akaike information criterion

For pairs of mathematical models in which one model is nested inside the other (e.g., as the mean dose model is nested inside the Lyman model, corresponding to the special case $n = 1$), the models can be compared using the likelihood ratio test. This test determines whether the larger model provides a significantly better fit to the data than the smaller model. For nonnested models, no such statistical comparison exists, but the models can be compared informally using the Akaike information criterion (AIC). This is a method proposed by Akaike (40, 41) for adjusting the LL by the number of parameters to assess whether the more complex model is worth the associated computational cost of estimating the additional parameters. The AIC is defined as AIC $= -2(\text{LL}) + 2\text{k}$, where k is the number of model parameters. Models with smaller values of AIC are considered to provide a better fit to the data than models with larger values of AIC.

RESULTS

Incidence of rectal bleeding

Twenty-nine (23%) of the 128 patients experienced Grade 2 or worse late rectal bleeding within 2 years of treatment. This included 1 patient who developed rectal bleeding after 5 months (instead of 6 months) and was also scored as having late complications. Patients with rectal bleeding represented most patients with late rectal complications during the first 2 years, because only 2 other patients were scored as having Grade 2 or worse late rectal complications during this period.

DVH comparison for patients with and without bleeding

Figure 1 illustrates the rectal DVHs for patients in this study. Figure 1a shows the cumulative DVH (cDVH) for 20 patients selected at random. Figure 1a illustrates the amount of patient-to-patient variability in the cDVH and demonstrates that the cDVHs for most patients in the study have qualitatively similar shapes; this is a consequence of the consistency in treatment technique (see "Methods and Materials," "Patient population"). Figure 1b shows the average cDVH for those with and without bleeding; patients who experienced Grade 2 or worse late rectal bleeding within 2 years had, on average, a greater percentage of rectum exposed to each dose level for all except the very lowest dose levels. Figures 1c, and d show the mean differential DVH (averaged at each dose point in increments of 0.5 Gy) for the 29 patients with rectal bleeding and the 99 patients without rectal bleeding, respectively.

Lyman and mean dose models

Table 2 lists the parameters of the Lyman model derived in our earlier study (22). Although it was possible to derive a lower limit for the confidence interval on *n*, the upper confidence limit of ∞ reflects the fact that values of D_{eff}, and consequently the corresponding fit of the Lyman model, changed very little for values of *n* greater than about 4 or 5.

Table 2 also lists the parameters of the mean dose model fitted to the rectal bleeding data. For comparison with the Lyman model, the parameter $m = (s \cdot \mu_{50})^{-1}$ is listed in addition to the parameter *s*. As noted in "Methods and Materials", the Lyman model always provides a better fit to the data than the mean dose model, but for the present data, the improvement did not reach statistical significance ($p = 0.569$, likelihood ratio test). This lack of significant improvement was reflected in the fact that the confidence interval for the Lyman volume parameter *n* contained the value $n = 1$, corresponding to the mean dose model.

Cutoff dose model

The cutoff dose model was fitted to the rectal bleeding data by an exhaustive method in which we considered each

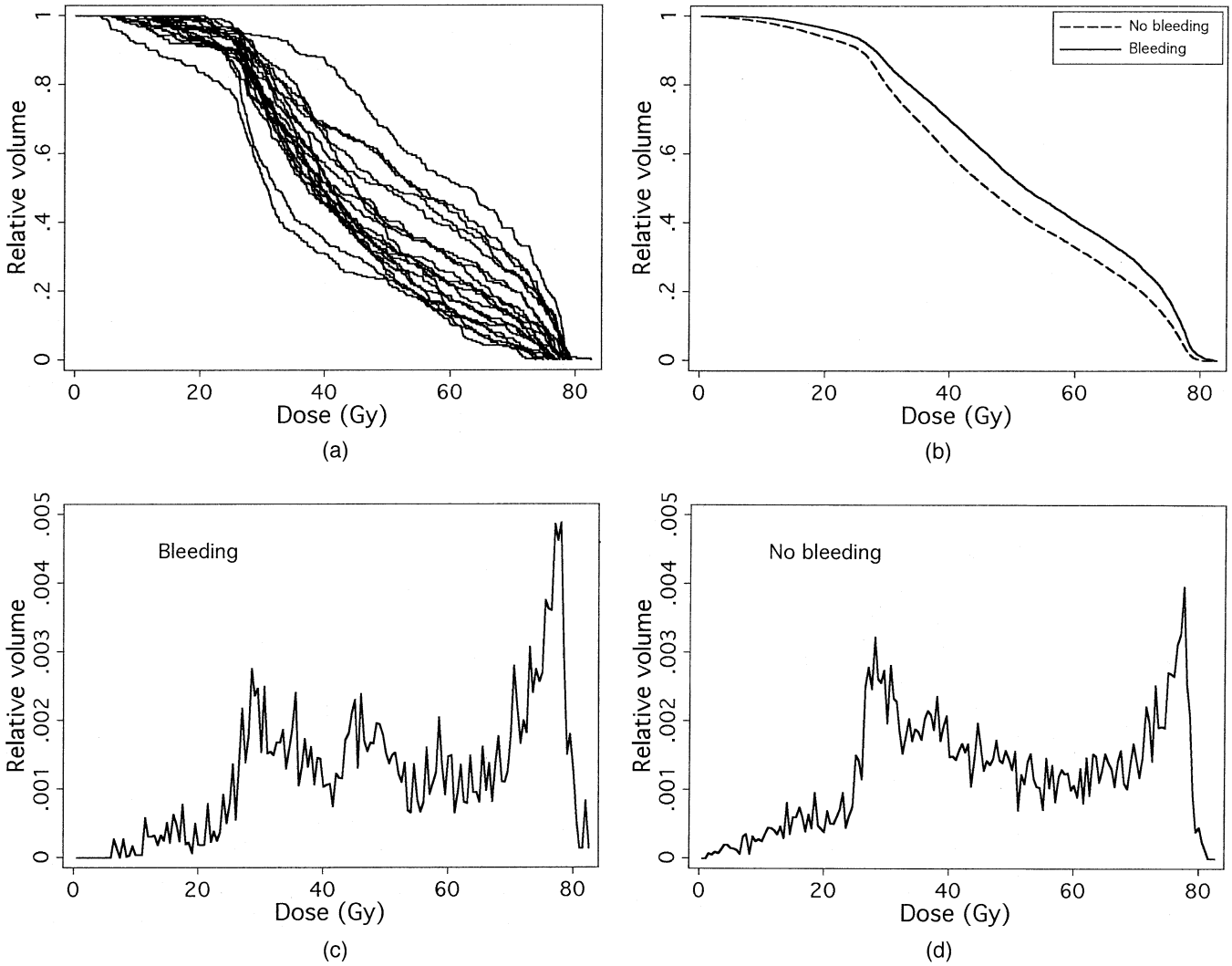


Fig. 1. (a) Rectal cumulative dose volume histogram (cDVH) curves for 20 randomly selected patients. (b) Average cDVH curves for patients with (solid curve) and without (dashed curve) rectal bleeding. (c) Average differential DVH for 29 patients with Grade 2 or worse late rectal bleeding within 2 years. (d) Average differential DVH for 99 patients without rectal bleeding during the first 2 years.

possible cutoff dose, D_c, from 5 to 80 Gy in increments of 0.1 Gy. This approach was used in lieu of a numeric optimization procedure to find the maximum likelihood since the cutoff dose model is not continuous in the parameter D_c because of the discretization of the measured DVH. Specifically, the function VD_c of Eq. (7) may have abrupt jumps at multiples of 0.1 Gy in the present study. We could have used an interpolation method to estimate $V(D_c)$ as a continuous function of dose and then used numeric optimization. However, we were interested in estimating the optimal D_c only to the closest 0.1 Gy, and the exhaustive method was feasible in that it did not require excessive computation time.

The parameters of the cutoff dose model are listed in Table 2; the parameter m is again given in addition to s, for comparison with the earlier models. The maximum likelihood estimate of D_c is 40.9 Gy, and irradiation of 76.1% of the rectum with doses of ≥40.9 Gy is estimated to correspond to a 50% risk of Grade 2 or worse late rectal bleeding by 2 years.

Figure 2 shows the LL of the cutoff dose model as a function of cutoff dose, D_c and illustrates the maximum likelihood occurring for D_c = 40.9 Gy. The profile likelihood confidence region for D_c corresponds to the values of D_c for which the likelihood curve lies above the dashed line in Fig. 2. The confidence region for D_c = 40.9 Gy is disjoint, consisting of several separate intervals along the dose axis where the curve in Fig. 2 dips below the horizontal line (Table 2).

Cutoff volume model

The cutoff volume model was also fitted to the rectal bleeding data by an exhaustive method in which we considered each possible cutoff volume, V_c, from 1% to 99%, in increments of 1%. Table 2 lists the maximum

Table 2. Parameter estimates, profile-likelihood confidence intervals and log-likelihood values derived from fitting various NTCP models to rectal bleeding data

Model	Parameter estimates (95% CI)	LL
Lyman model*	n = 3.91 (0.031–∞) $TD(50)$ = 53.6 Gy (50.0–75.1) m = 0.156 (0.036–0.271)	−58.07
Mean dose model	$MD(50)$ = 56.3 Gy (53.4–61.6) s = 0.126 Gy^{-1} (0.056–0.187) m = 0.141 (0.098–0.238)	−58.24
Cutoff dose model	D_c = 40.9 Gy[†] VD_c (50) = 76.1% (65.1%–91.3%) s = 5.68 (3.14–12.8) m = 0.231 (0.148–0.728)	−58.06
Cutoff volume model	V_c = 16%[‡] DV_c (50) = 77.6 Gy (67.7–80.1) s = 0.175 Gy^{-1} (0.057–0.455) m = 0.074 (0.028–0.232)	−58.85
Parallel model	s_D = 0.011 Gy^{-1} (0–∞) $D(50)$ = −241 Gy (−∞–∞) S_f = 3574 (3.3–∞) f_{50} = 0.999 (0.681–1)	−57.87

Abbreviations: NTCP = normal tissue complication probability; CI = confidence interval; LL = log-likelihood.
* Results of Cheung *et al.* (22).
† 95% confidence interval for D_c consists of union of several disjoint intervals ranging from 29.5 to 75.9 Gy (Fig. 2).
‡ 95% CI for V_c consists of union of several disjoint intervals ranging from 6% to 72% (Fig. 3).

likelihood parameter estimates of the cutoff volume model. The optimal cutoff volume, V_c is estimated to be 16%, and irradiation of ≥16% of the rectum with doses ≥77.6 Gy is estimated to correspond to a 50% incidence of rectal bleeding. As indicated by the maximum LL value, this model does not fit the data as well as the cutoff dose model, which has the same number of parameters.

In Fig. 3, the LL is plotted as a function of cutoff volume, V_c, with the maximal likelihood occurring for V_c = 16%. The profile likelihood confidence interval for V_c is very wide and also consists of a union of disjoint intervals (Table 2).

Parallel architecture model

Our experience has shown that numeric convergence is often difficult to achieve when fitting the parallel model to normal tissue toxicity data using an automated optimization procedure. Therefore, in the present study, we first used a grid search to identify reasonable starting values for the optimization. For D_{50} values ranging from 5 to 80 Gy in increments of 5 Gy, and for values of $\log_{10}(s)$ ranging from −3 to 2 in increments of 0.1, the corresponding values of f_{dam} were computed and fitted to the rectal bleeding data using probit analysis.

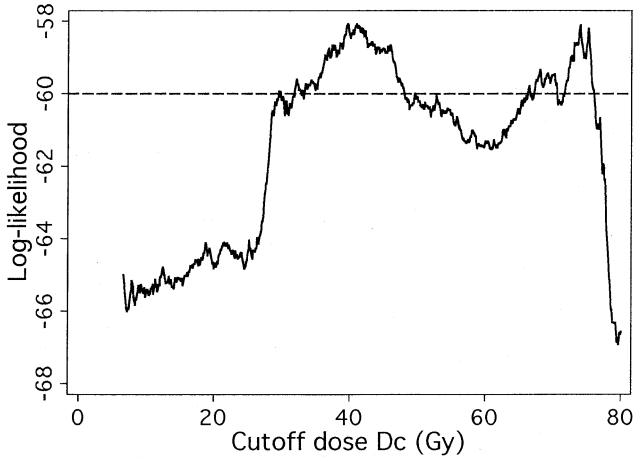


Fig. 2. Log-likelihood of fitted cutoff dose model plotted against value of cutoff dose, D_c. Doses with LL values above dashed horizontal line lie within the 95% profile-likelihood confidence region for optimal value (40.9 Gy) of D_c.

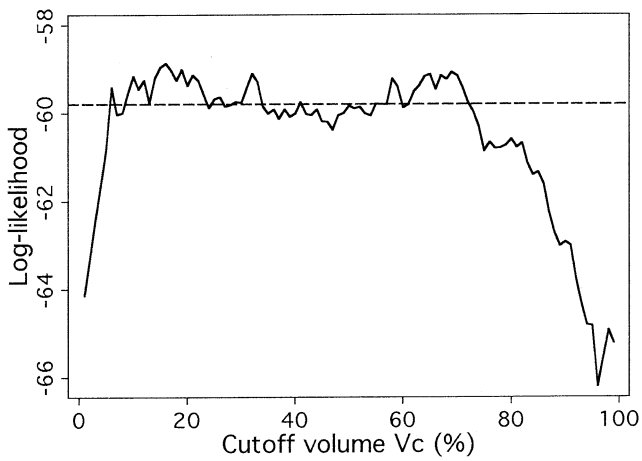
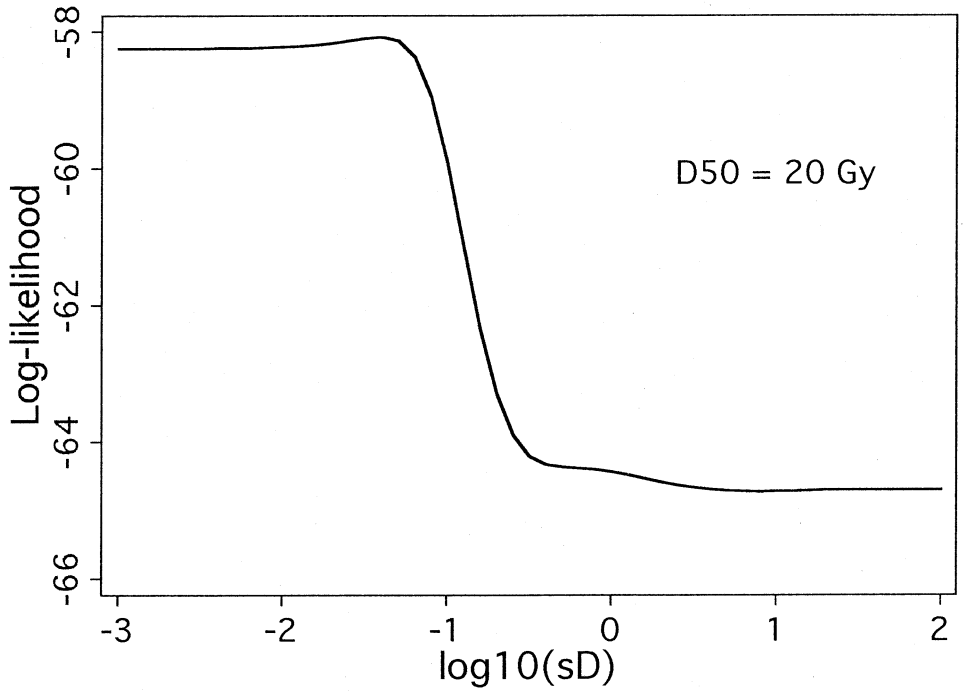
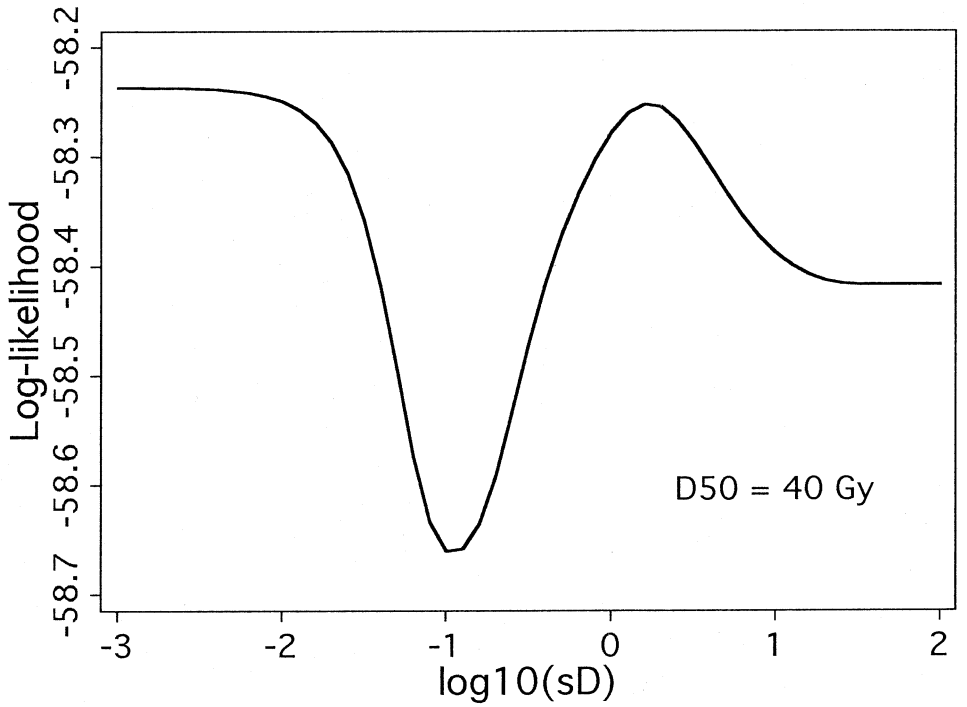


Fig. 3. Log-likelihood of cutoff volume model plotted against cutoff volume, V_c. Volumes with LL values above dashed horizontal line lie within 95% profile-likelihood confidence region for optimal value (16%) of V_c.



(a)



(b)

Fig. 4. Partial results of grid search for parameters of parallel model. Log-likelihood plotted against s_D for (a) D_{50} = 20 Gy or (b) 40 Gy.

Using the results of the grid search, Fig. 4 shows plots of LL versus s_D for two different values of D_{50} (D_{50} = 20 or 40 Gy). Figure 4 illustrates that the LL is constant or nearly constant over large ranges of parameter values, which explains why numeric convergence is sometimes so difficult to achieve with the parallel model. In particular, it can easily be shown that for any value of D_{50}, the likelihood surface becomes flat as s_D approaches infinity (Fig. 4). This means that for many starting values, the optimization routine will converge to a false maximum likelihood corresponding to

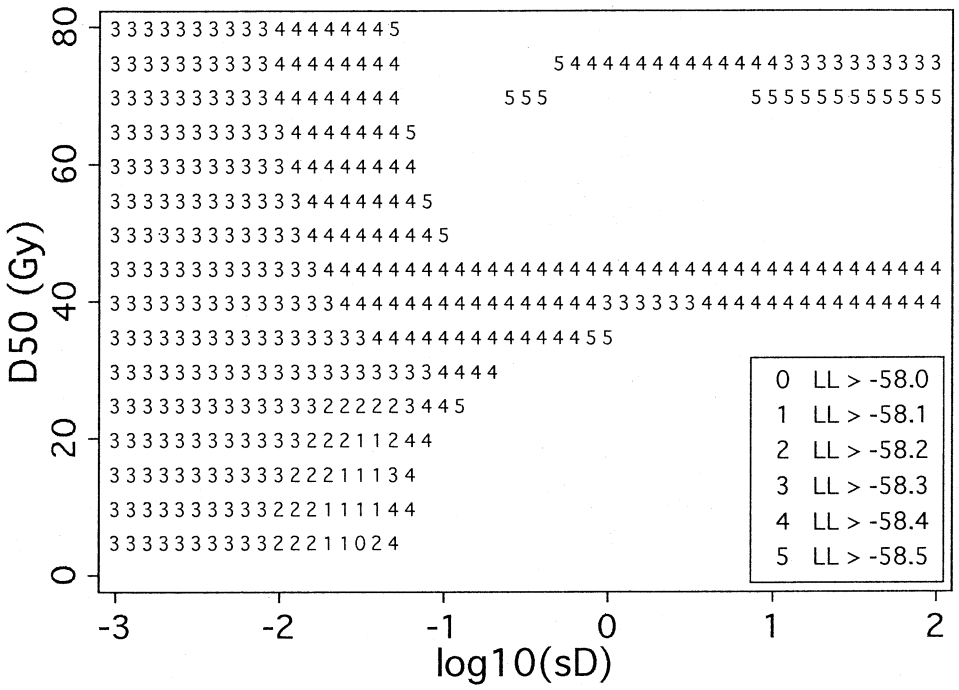


Fig. 5. Contour plot illustrating the results of the grid search for parameters of the parallel model. Symbols 0 through 5 indicate log-likelihoods of decreasing value. 0: LL >-58.0; 1: LL $\in (-58.1, -58.0]$; 2: LL $\in (-58.2, -58.1]$; 3: LL $\in (-58.3, -58.2]$; 4: LL $\in (-58.4, -58.3]$; 5: LL $\in (-58.5, -58.4]$.

$s_D = \infty$. As noted previously, $s_D = \infty$ corresponds to the cutoff dose model.

Figure 5 is a contour plot showing the regions in which the largest values of LL were obtained during the grid search for parameters of the parallel model. The largest value of LL was obtained for $D_{50} = 5$ Gy and $\log_{10}(s_D) = -1.5$, and therefore, these were used as starting values for the optimization procedure.

The resulting best-fitting parameter estimates for the parallel model (Table 2) are biologically unrealistic in that the D_{50} estimate is negative; this implies that the probability of local damage is >50% even for a dose of 0 Gy. The estimated local damage function, *P(D)*, is >99.5% for the entire dose range relevant to this study (0–80 Gy). To force the local damage function to be zero at zero dose, we refitted the parallel model using a version of *P(D)* expressed as a probit function of ln(dose) instead of dose. The resulting estimate of *P(D)* was again essentially flat, taking on values in the range of 49.9–50.1% between 1 and 80 Gy. Although this model has the correct behavior at zero dose, it rises much too steeply <1 Gy to be biologically reasonable. Thus, although the parallel model can be fitted to the rectal bleeding data, the resulting fit is not biologically meaningful.

As indicated in Table 2, the confidence intervals for the parameters s_D and D_{50} of the parallel model, describing the shape of the local damage function, appear to encompass all possible parameter values. In other words, we were unable to find parameter values falling outside the region of 95% certainty. Hence, the parameter values corresponding to the optimal fit of the parallel model are numerically meaningless in that they cannot be statistically resolved from any other possible values. Apparently, the present data set was simply too small to achieve a meaningful fit with the parallel model.

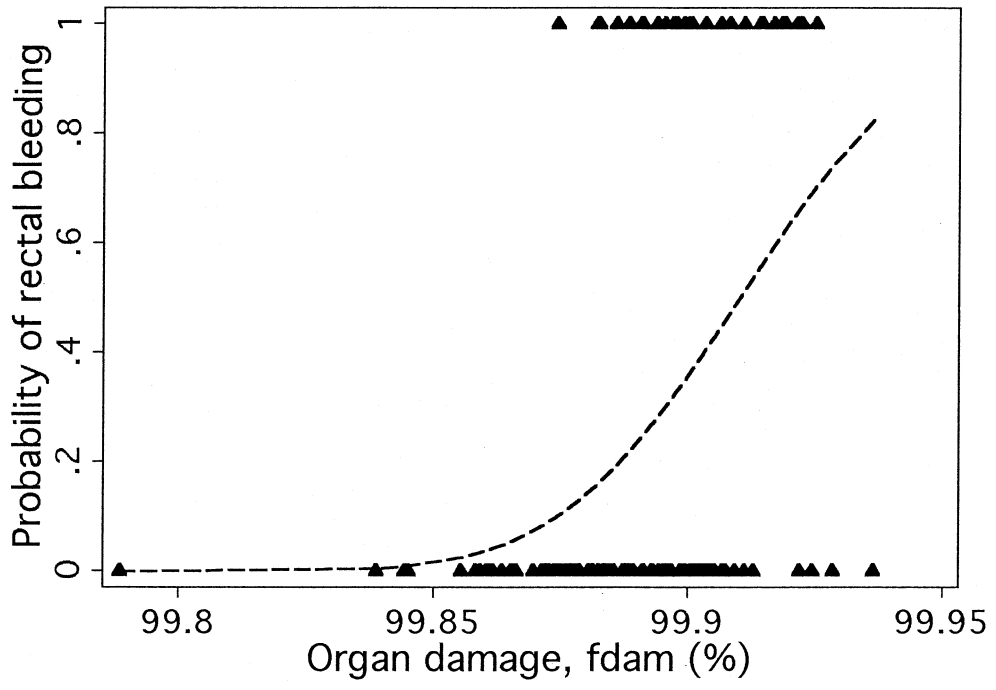
Comparisons among the models

Table 3 lists the LL and the value of the AIC for each of the five dose–volume response models considered in the present study. The AIC adjusts the LL by the number of model parameters, as described in "Methods and Materials." Although the four-parameter parallel model has the largest likelihood overall, the slight improvement in fit compared with the simpler models is not judged to be worth the additional computational effort, as assessed by the AIC. Among the three-parameter models, the Lyman and cutoff dose models provided somewhat better fits than the cutoff volume model. However,

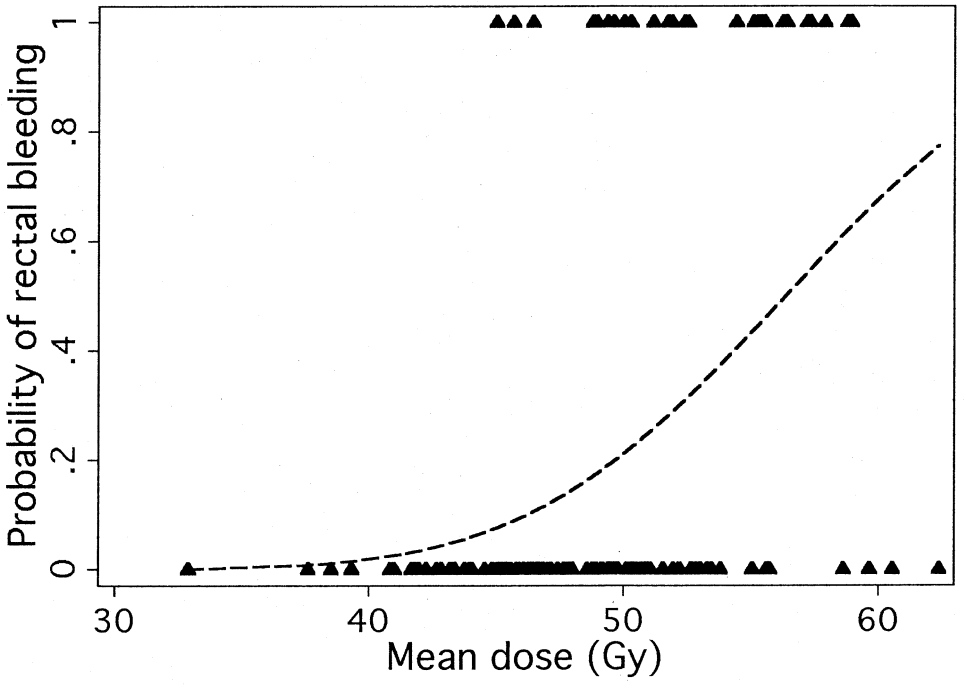
Table 3. Comparison of NTCP models using Akaike information criterion

Model	Parameters (n)	LL	AIC
Lyman model	3	−58.07	122.1
Mean dose model	2	−58.24	120.5
Cutoff dose model	3	−58.06	122.1
Cutoff volume model	3	−58.85	123.7
Parallel model	4	−57.87	123.7

Abbreviation: AIC = Akaike information criterion; other abbreviations as in Table 2.



(a)



(b)

Fig. 6. (a) Predicted probability of late rectal bleeding as a function of organ damage, f_{dam}, estimated using parallel model. (b) Predicted probability of late rectal bleeding as a function of mean dose to rectum. Symbols represent patient outcome (NTCP = 1 or NTCP = 0 for patients with and without bleeding, respectively) plotted as a function of f_{dam} or mean dose. Dashed curves represent model fits.

neither was markedly better than the simplest of the five models, the two-parameter mean dose model.

The values of f_{dam} computed from the parameters of the parallel model are shown in Fig. 6a, together with the sigmoid function linking f_{dam} to the probability of rectal bleeding, according to the fit of the parallel model presented in Table 2. The observed outcome for each patient is plotted as 1 if rectal bleeding occurred and 0 otherwise. Figure 6b shows the same type of plot for the mean dose and its corresponding sigmoid link. The two panels are very similar

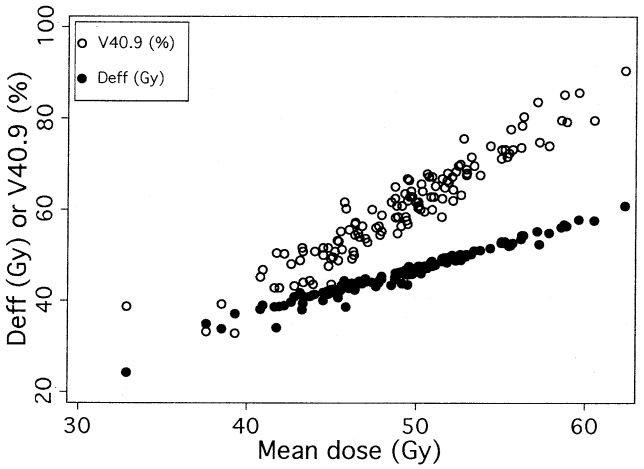


Fig. 7. Correlations between mean dose (MD) and D_{eff} or V40.9.

in appearance, reflecting the fact that the values of MD and f_{dam} for each patient were highly correlated (Pearson correlation coefficient $r = 0.983$, $p < 0.001$), although quite different in scale.

The MD was also highly correlated with each of the other summary measures of the DVH considered in the present study. The Pearson correlation coefficient for mean dose versus D_{eff}, V40.9, and D16 was $r = 0.983$, $r = 0.952$, and $r = 0.831$, respectively ($p < 0.001$ in each case). Figure 7 illustrates the first two of these correlations. The high degree of correlation among all the various summary measures of the DVH considered here explains the similarity in the values of maximum LL obtained from the various model fits (Table 2) and explains why none of the other models is judged to fit the data any better than the mean dose model.

Absolute rectal volumes

All the analyses described above were based on the relative DVH, normalized to the total rectal volume. Each of the models was also fitted to the data using the same numeric expression, but with the absolute rectal DVH substituted in place of the relative DVH. As shown in Table 4, the models based on absolute rectal DVH did not fit the data as well as those based on relative DVH.

Table 4. Maximum log-likelihood values obtained by fitting each NTCP model using absolute rectal DVH instead of relative DVH

Model	LL
Lyman model	−64.27
Mean dose	−67.41
Cutoff dose model	−65.22
Cutoff volume model	−63.72
Parallel architecture model	−59.39

Abbreviations: NTCP = normal tissue complication probability; DVH = dose–volume histogram; LL = log-likelihood.

DISCUSSION

We fitted five different NTCP models to the late rectal toxicity data from 128 patients treated for prostate cancer at the UTMDACC with 3D-CRT without hormonal therapy. The goal of the study was to help identify the most accurate model for predicting the risk of late rectal toxicity as a function of the dose and volume of rectum exposed. The availability of such a model would be of enormous benefit for future treatment planning, particularly as IMRT becomes more and more widely used.

The endpoint used in this study was Grade 2 or worse late rectal bleeding within 2 years of treatment. The 2-year time point appears to be a good one for analyses of rectal toxicity, because the vast majority of Grade 2 or worse late rectal toxicities occur by 2 years (11, 20), and 2 years of follow-up is sufficiently short to obtain mature clinical data within a reasonable time frame. We elected to restrict our analyses to Grade 2 or worse late rectal bleeding instead of all Grade 2 or worse late rectal toxicities for the sake of consistency. We believed that different dose–volume response relationships might describe different late events. However, nearly all (94%) of the Grade 2 or worse late rectal events observed in this patient population were rectal bleeding. Thus, exclusion of the other toxicities most likely made little practical difference in terms of model fitting.

Factors other than the radiation dose and volume may also influence the risk of late rectal toxicity after 3D-CRT. For example, the presence of pre-existing hemorrhoids has been implicated in the risk of late rectal bleeding in a larger cohort of patients from which the present subset was drawn (11). We were unable to identify any patient or clinical factors that were significantly associated with the risk of late rectal bleeding in the present cohort. For this reason we considered models based only on dose and volume. However, other factors could, in principle, be included in the modeling by assuming the dosimetric and volumetric parameters to be dependent on patient characteristics.

Each of the five models considered in this study had the same general form: a summary measure μ was extracted from the cumulative DVH, and the probit model was used to link μ to the probability of late rectal bleeding. The summary measures considered here were the D_{eff}, the MD, the relative volume of rectum receiving more than a specified cutoff dose, the minimal dose to the hottest volume of rectum of a specified (cutoff) size, and total organ damage, f_{dam} (from the parallel architecture model). The results showed that according to the AIC, none of the models provided a fit markedly better than the simplest of the models considered, the mean dose model. Although three of the other models led to somewhat larger values for the maximum LL, the LL values were not sufficiently larger, as judged by the AIC, to offset the added computational effort involved in fitting those models to data. As illustrated in Fig. 7, the summary measures of the DVH for all models were highly correlated with one another, meaning that there was actually very little difference among the model fits to

the present data. It should be noted, however, that the predictions of the various models outside the present range of data might be very different. For example, the presence of small, but very highly dosed, "hot spots" would increase the MD and lead to a predicted increased risk of rectal bleeding according to the MD model. However, the presence of such a hot spot would not significantly change the estimate of f_{dam} in the parallel model, and thus would not lead to a prediction of higher risk according to the parallel model. Therefore, the potential superiority of one of the models over the others would be much more apparent if data having a much wider variation in dose–volume distributions than in the present study had been analyzed (Fig. 1a).

The present study revealed some interesting features of the parallel architecture model, which we have found in past experience to be quite difficult to fit to complication data. In particular, it can be shown analytically that the derivative of the parallel NTCP model with respect to the parameter s_D approaches zero as s_D approaches infinity. This means that the optimization procedure will often converge to $s_D = \infty$ (the threshold dose model) even though this may not correspond to the overall maximum likelihood. In addition, there may be other broad ranges of parameter values for which the likelihood surface is relatively flat, which will also make achievement of convergence difficult. In addition, we found that the parallel architecture model with four parameters could be fitted to our data (with 128 patients and 29 events) almost equally well for nearly every possible set of parameter values. This implied that very large data sets might be required to fit even the simplest version of the parallel model with any certainty.

Although none of the other models considered here was markedly better than the MD model, the MD does not in fact provide a very precise prediction of rectal toxicity in the present study. The MDs ranged from 32.9 to 62.3 Gy (median 49.5). The value of MD corresponding to a 50% risk of Grade 2 or worse rectal bleeding was 56.3 Gy (95% confidence interval 53.4 to 61.6). If we had used this value of MD as a cutoff value for predicting late rectal bleeding in the present data set i.e., by predicting rectal bleeding for patients with MD $\geq$56.3 Gy and no bleeding for patients with MD $<$ 56.3 Gy, we would have been correct in 101 of 128 cases (79% accuracy), with a sensitivity of 21% (6 of 29) and a specificity of 96% (95 of 99). In clinical practice, it may be most important to have an NTCP model with high sensitivity, so that most patients likely to have adverse reactions can be identified prospectively and have treatment plans modified accordingly. If we increased sensitivity in the present case, for example by predicting bleeding among those with a $\geq$15% risk (corresponding to MD = 48.0 Gy), we would have had a sensitivity of 90% (26 of 29), with an accuracy of 59% (75 of 128) and specificity of 49% (49 of 99). Hence, the sensitivity would be increased at the expense of reduced overall accuracy and greatly reduced specificity.

We compared the predictive accuracy of the MD model to that of other models applied to the present data. For example, if V40.9 $\geq$76%, corresponding to an estimated bleeding risk of $\geq$50%, the resulting sensitivity, specificity, and accuracy was 17%, 95%, and 77%, respectively. Using the cutoff V40.9 $>$58%, corresponding to a $\geq$15% estimated risk of rectal bleeding, the sensitivity, specificity, and accuracy was 90%, 49%, and 59%, respectively. These values are very close to those obtained using the MD, as described above. If we instead considered the commonly used clinical guideline of keeping V70 $<$25%, and used V70 $\geq$25% as the criterion for predicting rectal bleeding, the sensitivity, specificity, and accuracy were all approximately equal for these data: sensitivity 69% (20 of 29), specificity 70% (69 of 129), and accuracy 69% (89 of 128).

An overall picture of the tradeoff between sensitivity and specificity for a predictive model is provided by the receiver operating characteristic curve, in which sensitivity is plotted against 1 − specificity as the estimated risk level from the model is varied from 0 to 100% (42). The area under the receiver operating characteristic curve (AUC) is often used to quantify the accuracy of the predictive model, with area 0.5 corresponding to a random prediction (e.g., a coin toss) and 1.00 corresponding to a perfect prediction. Using a nonparametric method to compute the AUC, the AUC for all five models considered in this study were identical to two digits: AUC = 0.77, with a 95% confidence interval (43) ranging from 0.67 or 0.68 to 0.86 or 0.87 in all cases. This similarity in AUC values further illustrates that no substantial differences exist among the fits of the five models considered with the UTMDACC rectal bleeding data. The AUC corresponding to V70 (with cutpoints other than 25%) was nearly the same: AUC = 0.76 (95% confidence interval 0.65–0.86).

Although we expect always to have some errors in prediction because of unknown factors and stochastic effects, we would hope for more precise predictors of rectal bleeding than provided by the NTCP models considered in the present study. Intuitively, one might expect measures of organ damage based on absolute DVHs to be more strongly associated with a risk of toxicity than those based on relative DVHs. That is, the patient with a larger organ volume might be expected to have a greater functional reserve than the patient with a smaller organ volume. However, in the present analysis, we found that each of the models considered fitted the data better when used with the relative rectal DVH than with the absolute DVH (Table 4). This was probably because of much of the variation in absolute rectal volumes from patient to patient was a result of variations in rectal filling, and was therefore unlikely to affect the risk of rectal toxicity.

CONCLUSION

Although it is possible that some other NTCP model based on the total rectal DVH may fit the data better than the models we considered, a more likely scenario is that much better predictive ability will be achieved if we exclude the portion of rectal volume attributed to rectal filling and consider models based instead on the dose–rectal wall histogram or dose–

surface histogram (44–47). Correction of the total dose to account for differences in fractionation in various segments of the treated organ (48, 49) and models that take organ motion into account are also expected to improve our ability to predict the likelihood of late rectal toxicity substantially. Our plan is to incorporate each of these into the modeling in turn to identify which have the greatest influence on the predictive accuracy. The ultimate goal is to identify an NTCP model of sufficient accuracy for clinical use in the design of optimal individualized treatment plans.

REFERENCES

1. D'Amico AV, Whittington R, Malkowicz SB, *et al.* Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998;280:969–974.
2. Hanks GE, Lee WR, Hanlon AL, *et al.* Conformal technique dose escalation in prostate cancer: Improved cancer control with higher doses in patients with pretreatment PSA >10 ng/mL. *Int J Radiat Oncol Biol Phys* 1996;35:523–868.
3. Kupelian PA, Mohan DS, Lyons J, *et al.* Higher than standard radiation doses (> or = 72 Gy) with or without androgen deprivation in the treatment of localized prostate cancer. *Int J Radiat Oncol Biol Phys* 2000;46:567–574.
4. Pollack A, Zagars GK, Starkschall G, *et al.* Prostate cancer radiation dose response: Results of the M.D. Anderson phase III randomized trial. *Int J Radiat Oncol Biol Phys* 2002;53: 1097–1105.
5. Zelefsky MJ, Fuks Z, Hunt M, *et al.* High dose radiation delivered by intensity modulated conformal radiotherapy improves the outcome of localized prostate cancer. *J Urol* 2001; 166:876–881.
6. Benk VA, Adams JA, Shipley WU, *et al.* Late rectal bleeding following combined X-ray and proton high dose irradiation for patients with stages T3-T4 prostate carcinoma. *Int J Radiat Oncol Biol Phys* 1993;26:551–557.
7. Boersma LJ, van den Brink M, Bruce AM, *et al.* Estimation of the incidence of late bladder and rectum complications after high-dose (70–78 Gy) conformal radiotherapy for prostate cancer, using dose–volume histograms. *Int J Radiat Oncol Biol Phys* 1998;41:83–92.
8. Bonta DV, Fontenia E, Lu Y, *et al.* A variable critical-volume model for normal-tissue complication probability. *Med Phys* 2001;28:1338–1343.
9. Emami B, Lyman J, Brown A, *et al.* Tolerance of normal tissues to therapeutic irradiation. *Int J Radiat Oncol Biol Phys* 1991;21:109–122.
10. Hartford AC, Niemierko A, Adams JA, *et al.* Conformal irradiation of the prostate: Estimating long-term rectal bleeding risk using dose-volume histograms. *Int J Radiat Oncol Biol Phys* 1996;36:721–730.
11. Huang EH, Pollack A, Levy L, *et al.* Late rectal toxicity: Dose–volume effects of conformal radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2002;54:1314–1321.
12. Jackson A, Skwarchuk MW, Zelefsky MJ, *et al.* Late rectal bleeding after conformal radiotherapy of prostate cancer (II): Volume effects and dose-volume histograms. *Int J Radiat Oncol Biol Phys* 2001;49:685–698.
13. Kutcher GJ, Leibel SA, Ling CC, *et al.* New wine in an old bottle? Dose escalation under dose-volume constraints: A model of conformal therapy of the prostate. *Int J Radiat Oncol Biol Phys* 1996;35:415–416.
14. Mameghan H, Fisher R, Mameghan J, *et al.* Bowel complications after radiotherapy for carcinoma of the prostate: The volume effect. *Int J Radiat Oncol Biol Phys* 1990;18:315–320.
15. Roach M, Pickett B, Weil M, *et al.* The "critical volume tolerance method" for estimating the limits of dose escalation during three-dimensional conformal radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 1996;35:1019–1025.
16. Schultheiss TE, Hanks GE, Hunt MA, *et al.* Incidence of and factors related to late complications in conformal and conventional radiation treatment of cancer of the prostate. *Int J Radiat Oncol Biol Phys* 1995;32:643–649.
17. Schultheiss TE, Lee WR, Hunt MA, *et al.* Late GI and GU complications in the treatment of prostate cancer. *Int J Radiat Oncol Biol Phys* 1997;37:3–11.
18. Skwarchuk MW, Jackson A, Zelefsky MJ, *et al.* Late rectal toxicity after conformal radiotherapy of prostate cancer (I): Multivariate analysis and dose-response. *Int J Radiat Oncol Biol Phys* 2000;47:103–113.
19. Smit WGJM, Helle PA, van Putten WLJ, *et al.* Late radiation damage in prostate cancer patients treated by high dose external radiotherapy in relation to rectal dose. *Int J Radiat Oncol Biol Phys* 1990;18:23–29.
20. Storey MR, Pollack A, Zagars G, *et al.* Complications from radiotherapy dose escalation in prostate cancer: Preliminary results of a randomized trial. *Int J Radiat Oncol Biol Phys* 2000;48:635–642.
21. Wachter S, Gerstner N, Goldner G, *et al.* Rectal sequelae after conformal radiotherapy of prostate cancer: Dose-volume histograms as predictive factors. *Radiother Oncol* 2001;59:65–70.
22. Cheung R, Tucker SL, Ye J, *et al.* Characterization of rectal normal tissue complication probability after high-dose external beam radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2004;58:1513–1519.
23. Lyman JT. Complication probability as assessed from dose-volume histograms. *Radiat Res* 1985;104:S13–S19.
24. Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). *Int J Radiat Oncol Biol Phys* 1995;31:1341–1346.
25. Denekamp J, Letschert J, *et al*. Pavey JJ. Late effects toxicity scoring: The SOMA scale. *Int J Radiat Oncol Biol Phys* 1995;31:1043–1047.
26. Hanlon AL, Schultheiss TE, Hunt MA, *et al.* Chronic rectal bleeding after high-dose conformal treatment of prostate cancer warrants modification of toxicity scales. *Int J Radiat Oncol Biol Phys* 1997;38:59–63.
27. Morgan BJT. Analysis of quantal response data. New York: Chapman & Hall, 1992.
28. Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method. *Int J Radiat Oncol Biol Phys* 1989; 16:1623–1630.
29. Lyman JT, Wolbarst AB. Optimization of radiation therapy, III: A method of assessing complication probabilities from dose-volume histograms. *Int J Radiat Oncol Biol Phys* 1987; 13:103–109.
30. Lyman JT, Wolbarst AB. Optimization of radiation therapy, IV: A dose–volume histogram reduction algorithm. *Int J Radiat Oncol Biol Phys* 1987;17:433–436.
31. Mohan R, Mageras GS, Baldwin B, *et al.* Clinically relevant optimization of 3-D conformal treatments. *Med Phys* 1992; 19:933–944.
32. Niemierko A. Reporting and analyzing dose distributions: A concept of equivalent uniform dose. *Med Phys* 1997;24:103–110.

33. Wu Q, Mohan R, Niemierko A, *et al.* Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *Int J Radiat Oncol Biol Phys* 2002;52:224–235.
34. Wolbarst AB, Chin LM, Svensson GK. Optimization of radiation therapy: Integral response of a model biological system. *Int J Radiat Oncol Biol Phys* 1982;8:1761–1769.
35. Niemierko A, Goitein M. Modeling of normal tissue response to radiation: The critical volume model. *Int J Radiat Oncol Biol Phys* 1992;25:135–145.
36. Jackson A, Kutcher GJ, Yorke ED. Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation. *Med Phys* 1993;20: 613–625.
37. Jackson A, Ten Haken RK, Robertson JM, *et al.* Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int J Radiat Oncol Biol Phys* 1995; 31:883–891.
38. Yorke ED, Jackson A, Rosenzweig KE, *et al.* Dose-volume factors contributing to the incidence of radiation pneumonitis in non–small-cell lung cancer patients treated with three-dimensional conformal radiation therapy. *Int J Radiat Oncol Biol Phys* 2002;54:329–339.
39. Seppenwoolde Y, Lebesque JV, de Jaeger K, *et al.* Comparing different models that predict the incidence of radiation pneumonitis. *Int J Radiat Oncol Biol Phys* 2003;55:724–735.
40. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Second international symposium on information theory. Budapest: Academiai Kiado, 1973: p. 267–281.
41. Akaike H. A new look at statistical model identification. *IEEE Trans Automa Contr* 1974;19:716–722.
42. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:26–36.
43. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating curves: A nonparametric approach. *Biometrics* 1988;44:837–845.
44. Lu Y, Song PY, Li SD, *et al.* A method of analyzing rectal surface area irradiated and rectal complications in prostate conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 1995; 33:1121–1125.
45. Li S, Boyer A, Lu Y, *et al.* Analysis of the dose-surface histogram and dose-wall histogram for the rectum and bladder. *Med Phys* 1997;24:1107–1116.
46. MacKay RI, Hendry JH, Moore CJ, *et al.* Predicting late rectal complications following prostate conformal radiotherapy using biologically effective doses and normalized dose-surface histograms. *Br J Radiol* 1997;70:517–526.
47. Meijer GJ, van den Brink M, Hoogeman MS, *et al.* Dose-wall histograms and normalized dose–surface histograms for the rectum: A new method to analyze the dose distribution over the rectum in conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 1999;45:1073–1080.
48. Lee SP, Leu MY, Smathers JB, *et al.* Biologically effective dose distribution based on the linear quadratic model and its clinical relevance. *Int J Radiat Oncol Biol Phys* 1995;33:375–389.
49. Wheldon TE, Deehan C, Wheldon EG, *et al.* The linear-quadratic transformation of dose-volume histograms in fractionated radiotherapy. *Radiother Oncol* 1998;46:285–295.