INCIDENCE OF LATE RECTAL BLEEDING IN HIGH-DOSE CONFORMAL RADIOTHERAPY OF PROSTATE CANCER USING EQUIVALENT UNIFORM DOSE–BASED AND DOSE–VOLUME–BASED NORMAL TISSUE COMPLICATION PROBABILITY MODELS

MATTHIAS SÖHN, DIPL. PHYS.,* DI YAN, D.SC.,† JIAN LIANG, PH.D.,† ELISA MELDOLESI, M.D.,† CARLOS VARGAS, M.D.,‡ AND MARKUS ALBER, PH.D. *

*Section for Biomedical Physics, University Hospital for Radiation Oncology, Tübingen, Germany; †Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, MI; ‡Radiation Oncology, University of Florida, Gainesville, FL

Purpose: Accurate modeling of rectal complications based on dose–volume histogram (DVH) data are necessary to allow safe dose escalation in radiotherapy of prostate cancer. We applied different equivalent uniform dose (EUD)–based and dose–volume–based normal tissue complication probability (NTCP) models to rectal wall DVHs and follow-up data for 319 prostate cancer patients to identify the dosimetric factors most predictive for Grade ≥ 2 rectal bleeding.

Methods and Materials: Data for 319 patients treated at the William Beaumont Hospital with three-dimensional conformal radiotherapy (3D-CRT) under an adaptive radiotherapy protocol were used for this study. The following models were considered: (1) Lyman model and (2) logit-formula with DVH reduced to generalized EUD, (3) serial reconstruction unit (RU) model, (4) Poisson-EUD model, and (5) mean dose– and (6) cutoff dose–logistic regression model. The parameters and their confidence intervals were determined using maximum likelihood estimation.

Results: Of the patients, 51 (16.0%) showed Grade 2 or higher bleeding. As assessed qualitatively and quantitatively, the Lyman- and Logit-EUD, serial RU, and Poisson-EUD model fitted the data very well. Rectal wall mean dose did not correlate to Grade 2 or higher bleeding. For the cutoff dose model, the volume receiving > 73.7 Gy showed most significant correlation to bleeding. However, this model fitted the data more poorly than the EUD-based models.

Conclusions: Our study clearly confirms a volume effect for late rectal bleeding. This can be described very well by the EUD-like models, of which the serial RU- and Poisson-EUD model can describe the data with only two parameters. Dose–volume–based cutoff-dose models performed worse. © 2007 Elsevier Inc.

INTRODUCTION

The essential dose-limiting organs in prostate radiotherapy are the bladder and rectum. One of the most relevant side effects that can significantly compromise a patient’s quality of life is chronic rectal bleeding.

Conventional external beam radiotherapy (RT) treatment typically does not allow prostate doses beyond 65 to 70 Gy without an unacceptably high risk of rectal toxicity, although higher tumor doses are favorable for improved tumor control. The possibility of dose escalation beyond 70 Gy to the prostate is based on the volume–effect of rectum, i.e., the observation of increased tolerance to high doses if the high dose region is confined to a small volume. Technically, this becomes feasible because of conformal techniques such as three-dimensional conformal radiotherapy (3D-CRT) or intensity-modulated radiotherapy (IMRT), especially when aided by image-guided adaptive approaches.

Safe dose escalation necessitates accurate quantitative modeling of the volume effect based on the detailed dose–volume information provided by modern treatment planning systems. Numerous studies have established evidence of a significant correlation between parameters derived from rectal dose–volume histograms (DVHs) and toxicity (see

Reprint requests to: Matthias Söhn, Dipl. Phys., Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany; Tel: (+49) 7071-2986061; Fax: (+49) 7071-295920; E-mail: Matthias.Soehn@med.uni-tuebingen.de

Presented in part at the 48th Annual American Society for Therapeutic Radiology and Oncology Meeting (ASTRO), Philadelphia, PA, Nov. 5–9, 2006.

Supported in part by Deutsche Krebshilfe e.V. Grant No. 106280 and National Institutes of Health Grant No. RO1 CA091020.

Conflict of interest: none.

Received Sept 6, 2006, and in revised form Oct 2, 2006. Accepted for publication Oct 9, 2006.
Table 1. Toxicity score for chronic rectal bleeding based on Common Terminology Criteria for Adverse Events (v. 3.0)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mild hemorrhage/bleeding; intervention (other than iron supplements) not indicated</td>
</tr>
<tr>
<td>2</td>
<td>Symptomatic and medical intervention or minor cauterization indicated</td>
</tr>
<tr>
<td>3</td>
<td>Transfusion, interventional radiology, endoscopic or operative intervention indicated</td>
</tr>
<tr>
<td>4</td>
<td>Life-threatening consequences; perforation/dysfunction requiring urgent intervention</td>
</tr>
</tbody>
</table>

Patient data

Data for 319 prostate cancer patients treated between 1999 and 2002 at the William Beaumont Hospital were used for this study. The characteristics of this patient population have been described in previous studies (3, 13). The patients were part of a Phase II dose-escalation study and underwent 3D-CRT with image-guided off-line correction under an ART protocol.

All patients had one pretreatment planning CT scan, daily portal images to determine and correct for setup errors, four additional CT scans during the first week of the treatment used for individual adaptation of the treatment plan, and weekly CT scans in the following to preclude undetected drifts. The (solid) rectum was contoured on the initial CT scan from the anal verge or ischial tuberosities (whichever was higher) to the sacroiliac joints or rectosigmoid junction (whichever was lower). Rectal wall was defined based on the solid rectum contours with 3- to 4-mm wall thickness.

The ART scheme used has been described elsewhere (14, 15). In short, a four-field box technique with 18 MV photons was used both for the initial treatment plan of the first week and the following adapted plan. In the first week, the patients were treated for a dose of 9 Gy to the target, where the planning target volume (PTV), was generated based on the clinical target volume (CTV), of the initial CT (prostate, or prostate + seminal vesicles) with a population-based margin of 1 cm. For the adapted plan, information from daily portal imaging and the five CT scans available after the first week of treatment were used to estimate setup error and individual prostate mobility, which allowed to define a (generally smaller) patient-specific PTV.

The final dose to the PTV was limited by dose-volume constraints of rectal wall and bladder based on the geometry of the initial planning CT image. For rectal wall these were: (1) $D_{50\%} = 75.6$ Gy, and (2) $D_{95\%} = 82$ Gy. The possible dose levels (minimal prostate dose) were chosen under the requirement to meet rectum (and bladder) constraints, and were as follows: 70.2, 72, 73.8, 75.6, 77.4, and 79.2 Gy.

For each patient the dose distributions of the initial and adapted plan were calculated using Pinnacle 6.2b (ADAC Laboratories, Milpitas, CA). An in-house developed software was used to calculate DVHs of the rectal wall. This software used the contours from the initial (planning) CT and calculated the overall dose as sum of initial and adapted (physical) dose distributions. The DVH-dose bin size was 0.1 Gy, with volume defined as relative (percentage) volume irradiated.

The rectal toxicity variable regarded in this analysis is chronic rectal bleeding. The follow-up scheme defined examinations at 3-month intervals during the first 2 years, and every 6 months from the second to the fifth year. As mentioned above, this study is based on the patient population analyzed in Vargas et al. (3, 13). However, for the current study, all patient files were re-examined to improve follow-up time. Complications were graded based on the Common Terminology Criteria for Adverse Events (CTCAE) v 3.0 (Table 1). Of the 331 patient datasets, 12 used by Vargas et al. could not be used because of technical problems in restoring dose distributions or lost, incomplete, or inconsistent follow-up information. The median clinical follow-up for the remaining 319 patients was 2.8 years (range, 0.1–6.4), with an interquartile range of 1.5 to 4.0 years (25th–75th percentile).

The NTCP models

An NTCP model assigns a complication probability for an organ at risk to a generally inhomogeneous dose distribution. The func-
tional form of such a model can be based on a mechanistic description of biologic processes, or might be designed to result in a phenomenologic fit of the data.

The models considered in this work are of the general forms described below.

First, a summary measure $\mu$, such as the mean dose, an equivalent uniform dose (EUD), or similar, is calculated from the dose distribution. The quantity $\mu$ serves as a ranking function by imposing an order among individual plans according to their complication risk.

Next, a function, NTCP($\mu$), which assigns complication probabilities to the values of the summary measure is defined. Such a function is required (1) for continuously mapping $\mu$ to the interval [0, 1], while (2) preserving the ranking imposed by the numerical values of the summary measure. This leads to the class of sigmoid-type (S-shaped) functions, wherein the following probit, logit, Poissonian, and logistic formulae are used.

We applied six different models to our data, where the endpoint was chosen to be chronic rectal bleeding of Grade $\geq 2$. These models differ in the summary measure used and/or the functional form of the NTCP function as described in the following.

**Lyman-EUD model**

The most widely used phenomenological approach is the family of Lyman models (16–20), which uses the probit function

$$\text{NTCP}_{\text{probit}}(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(\mu-x)^2/2} dx$$

(1)

to map the summary measure $\mu$ to the interval [0, 1] of complication probabilities. This integral essentially yields the error function, where the parameter $s$ is the slope of the sigmoid response curve at the steepest point $\mu = \mu_{50}$, for which the NTCP function predicts 50% complication probability. Usually the slope parameter $s$ is replaced by its inverse $m$ according to $s = 1/(m \cdot \mu_{50})$.

Different DVH reduction schemes have been used for defining the summary measure $\mu$, such as an effective volume as in the LKB model (18) or an effective dose (21). In the following, the generalized equivalent uniform dose (22) is used, which defines EUD as Lebesgue $\alpha$-norm of the dose, i.e., in terms of the following power-law relationship:

$$\mu := \text{EUD}_{\alpha} = \left( \sum_i v_i D_i^\alpha \right)^{1/\alpha}$$

(2)

The sum is calculated over all bins $(v_i, D_i)$ of the differential DVH, and $\alpha$ is a parameter associated with the strength of the volume effect for the organ under consideration (range, $\alpha \in [1...\infty]$): For $\alpha \rightarrow \infty$ the EUD is the maximum dose (i.e., no volume effect), whereas for $\alpha = 1$ Eq. (2) gives the mean dose (large volume effect).

Summarizing, the Lyman-EUD model as used in our study is described by 3 parameters: $a, m,$ and EUD$_{50}$ (usually termed D$_{50}$).

**Logit-EUD model**

This model also uses the generalized EUD Eq. (2) as summary measure $\mu$, although it differs from the Lyman-EUD model in the choice of the NTCP function. Here the logit function

$$\text{NTCP}_{\text{logit}}(\mu) = \frac{1}{1 + (\mu_{50}/\mu)^a}$$

(3)

is chosen as sigmoid shape function (6, 10). Its two parameters $\mu_{50}$ (i.e., $D_{50}$) and $k$ are determined by the EUD, which causes a complication rate of 50%, and the slope of the NTCP curve here. Thus, together with the parameter $a$ of the EUD, this model has three parameters.

**Serial reconstruction unit model**

In contrast to the two previous phenomenologic NTCP models, the serial reconstruction unit model, which has been proposed recently by Alber and Belka (23), arises from certain general assumptions about the biological processes causing normal tissue complications.

The model regards radiation induced complications as the consequence of local failure of dynamic repair processes. As an assumption, the latter is attributed to the finite range of the repair mechanisms, which finds its correlate in the model by the description of finite-sized reconstruction units and their microscopic dose–response. Borrowing analogies from thermodynamics and statistical physics, the authors derive the following expression to describe the macroscopic dose–response in terms of the NTCP for homogeneous irradiation of the partial volume $V$ of an organ with the dose $D$:

$$\text{NTCP}_{\text{SRU}}(V, D) = 1 - \exp \left( -V \exp \left( \sigma(D - D_0) \right) \right)$$

(4)

where $\sigma$ is an organ specific sensitivity parameter and $D_0$ is a reference dose.

For inhomogeneous dose distributions an equivalent uniform dose, which would give the same macroscopic dose–response when applied homogeneously to the whole organ ($V = 1$), can be defined as

$$\text{EUD}_{\text{SRU}} = \frac{1}{\sigma} \log \left( \sum_i v_i \exp \left( \sigma D_i \right) \right)$$

(5)

Consequently, the NTCP function then reads:

$$\text{NTCP}_{\text{SRU}}(\text{EUD}_{\text{SRU}}) = 1 - \exp \left( -\exp \left( \sigma(\text{EUD}_{\text{SRU}} - D_0) \right) \right)$$

(6)

Summarizing, the serial reconstruction unit model has the two parameters $\sigma$ and $D_0$ to be fitted. Note that in contrast to the previously described models, which have the volume effect parameter $a$ of the power-law EUD as a third parameter, here the sensitivity parameter $\sigma$ is inherently coupled to the same value in the EUD and NTCP function.

**Poisson-EUD model**

Similarly to the serial reconstruction unit model, this model uses mechanistic concepts to describe predominantly serial tissue dose–response. Assuming that complication is a consequence of local
dose–response of noninteracting subunits, the following NTCP-function can be derived based on Poissonian statistics (24):

\[
\text{NTCP}_{\text{poisson}}(\text{EUD}) = 1 - \exp \left[ -\left( \frac{\text{EUD}}{D_0} \right)^a \right] = 1 - \exp \left[ -\ln 2 \cdot \left( \frac{\text{EUD}}{D_{50}} \right)^a \right]
\]

(7)

with a reference dose \(D_0\) (or a dose \(D_{50}\) causing 50% complication probability) and a volume–effect (steepness) parameter \(a\). The EUD is given by Eq. (2), where, according to this model, the exponent of EUD and the steepness parameter of the NTCP function have the same value. Thus, unlike the Lyman- and Logit-EUD models, the Poisson-EUD model has only two parameters.

**Mean dose logistic regression model**

An association of the rectal mean dose with chronic rectal bleeding has been reported by some authors (7, 8). To test for such an association based on our data, we used logistic regression as a standard method from statistics for this purpose. In terms of the general NTCP model scheme presented above, here the NTCP function is given by the two-parametric logistic function

\[
\text{NTCP}_{\text{logistic}}(\mu) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \mu)}
\]

(8)

with the mean dose \(D_{\text{mean}} = \sum_i V_i D_i\) as summary measure \(\mu\). Note that \(D_{\text{mean}}\) is a special case of the EUD, Eq. (2), for fixed parameter \(a = 1\). Thus, this model has only the two parameters \(\beta_0\) and \(\beta_1\).

**Cutoff dose logistic regression model**

In classical, i.e., nonbiological treatment planning approaches, dose limitation to organs at risk (OAR) is usually implemented in terms of dose–volume constraints. For a given treatment-technique, the most relevant dose level(s) predictive for toxicity can be determined by retrospectively fitting a sigmoid type NTCP function to outcome data of a patient population.

In this phenomenological approach, the summary measure \(\mu\) is given by the proportion \(V_{D_{c}}\) of the OAR receiving doses equal to or above a (cutoff) dose level \(D_{c}\). In the present study, \(V_{D_{c}}\) is regarded as relative volume, formally: \(V_{D_{c}} = \sum_{i | D_i \geq D_{c}} V_i \), where \(V_i\) is the discretized form of the differential DVH, and only dose bins \(i\) with \(D_i \geq D_{c}\) are used in the sum.

For given value \(D_{c}\) we used logistic regression Eq. (8) to test for correlation of \(V_{D_{c}}\) and chronic rectal bleeding, resulting in two fit parameters \(\beta_0\) and \(\beta_1\). This was systematically repeated for all possible \(D_{c}\) up to 85 Gy in increments of 0.1 Gy to assess the significance of such a logistic regression model for different cutoff doses. Thus, altogether the model has the three parameters \(D_{c}\), \(\beta_0\), and \(\beta_1\).

**Fitting procedure**

The method of choice for fitting such models to sparse, dichotomous response data (0/1, if patient shows bleeding of Grade 2 or ≥ 2, respectively) is maximum likelihood estimation (Jackson et al. (25) and references therein; see also Ref. 6). In this method the optimal model parameters are determined such as to maximize the probability of occurrence of the observed data, which is given by the so-called likelihood function \(L\). Because of its smallness, numerically this is usually implemented as maximization of the natural logarithm \(\ln(L)\), the LogLikelihood (LL). In our implementation, the software package Mathematica version 5.0 (Wolfram Research Inc., Champaign, IL) was used. To reduce calculation time, the DVH discretization was changed to 0.5 Gy when fitting the Lyman-, Logit-, Poisson-EUD, and serial reconstruction unit models.

Uncertainties of the model parameters were assessed using the variance–covariance matrix of the parameters calculated around the maximum of LL as described in (25). The 68% confidence intervals for the parameters can be estimated by the square root of the diagonal elements.

The goodness of fit of each individual model has been quantified in two ways. The first method follows Jackson et al. (25). If the actual (maximized) value of the LL for the NTCP model fitted to the observed patient population is denoted by LL_{obs}, the probability \(P\) of obtaining a value smaller than LL_{obs} (i.e., a worse fitting model) purely by chance can be assessed based on the statistical distribution of LL. This can be obtained from analytical formulae for the mean (LL) and variance \(\Delta_{L}^2\) (see Eq. A3 and A4 in Jackson et al.) under the assumption of a normally distributed LL. If this probability turns out to be too large, the model “overfits” the data; if it is too small, the model does not fit the data well. According to Rancati et al. (6) values between 30% and 70% indicate a satisfactory fit.

Additionally, a chi-square goodness-of-fit test was performed for each model: A histogram of the observed patient data was calculated, and the resulting group complication rates \(f_{i, \text{obs}}\) were compared with the corresponding rates \(f_{i}^{(\text{exp})}\) predicted by the respective NTCP model by determining \(\chi^2 = \sum_{i} (f_{i, \text{obs}} - f_{i}^{(\text{exp})})^2/f_{i, \text{obs}}^2\), which approximately follows a chi-square distribution with \(N - 1 - n\) degrees of freedom (\(N = \text{number of histogram bins}; n = \text{number of model parameters}\)).

**Intermodel comparison: Akaike information criterion**

A widely used measure allowing to compare competing models is the AIC (26, 27); see also (7). Generally, for fits of different models to a given dataset a larger likelihood value indicates a better fit to the data. The AIC quantifies the tradeoff between the model’s quality of fit (associated with the likelihood value) and its complexity (expressed by the number of model parameters \(n\)), and is defined as AIC = −2LL + 2n. Models with smaller AIC values are considered to provide a better (in the sense of more efficient) fit to the data than models with larger AIC.

**RESULTS**

Of the 319 patients, 45.5, and 1 (14.1%, 1.6%, and 0.3%) showed chronic rectal bleeding of Grade 2, 3, and 4, respectively. Thus, altogether 51 patients (16.0%) showed chronic rectal bleeding of Grade ≥ 2, which is the endpoint for the following model fits.

**EUD-based models, mean dose model**

Parameter estimates and their 68% CIs for the Lyman-, Logit-, Poisson-EUD, and serial reconstruction unit model are given in Table 2. The plots of the corresponding NTCP-curves are shown in Fig. 1a to 1d. As obvious in these plots by comparison with the observed complication rates, all four models fit the data very well. Quantitatively, this is
manifest by the small values of $\chi^2$: For example, the fit of
the serial reconstruction unit model (Fig. 1c) resulted in a
$\chi^2$ of 0.03, where the upper limit of $\chi^2$ for an acceptable fit is
5.99 according to Chi-square statistics (2 degrees of freedom; $\alpha = 0.05$). As described later here, the goodness of fit
was also determined based on the LL, (LL) and $S_{LL}$. These
values together with the resulting probability P of obtaining
a smaller LL are given in Table 2. Again, for all four models
P indicates acceptable fits.

Concerning the mean dose logistic regression model, it
turns out that $D_{mean}$ does not significantly correlate to
chronic rectal bleeding of Grade $\geq 2$ ($p = 0.11$). The worse
fit quality of this model in comparison with the four above-
mentioned EUD-based models is also expressed by its sig-
ificantly lower LL value (Table 2). Thus, this model is not
considered further in this study.

Cutoff dose model

Figure 2a shows the LL values of different logistic re-
gression model fits when varying the cutoff dose $D_c$. For all
dose levels in the range $D_c \geq 50$ to 80 Gy, a significant
correlation ($\alpha < 0.05$) of the relative volume irradiated with
doses $\geq D_c$ and chronic rectal bleeding of Grade $\geq 2$ was
found for our patient population. The curve has maxima at
$D_c = 73.7$ Gy and 79.6 Gy. Figure 2b depicts the model fit
for $D_c = 73.7$ Gy: As obvious, the fit quality is slightly
worse than for the EUD-models. However, both the value of
P in Table 2 and a $\chi^2$ of 1.06 (upper limit according to
Chi-square statistics [1 degree of freedom, $\alpha = 0.05$]: 3.84)
show that the fit is acceptable. Formally, this is also the case
for the model with $D_c = 79.6$ Gy (plot not shown; $\chi^2 = 2.24$
$< 3.84$). However, as only a part of our patient population
receives doses above 79.6 Gy to nonvanishing volumes of
rectal wall, the distribution of the summary measure $V_{79.6}$

\[
\text{Table 2. Chronic rectal bleeding grade } \geq 2: \text{ Estimated parameter values for the six normal tissue}
\text{complication probability models, observed LogLikelihood (LL) values, estimated LL distribution,}
\text{and resulting probability of obtaining a smaller LL value than observed}
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter estimates (68% CI)</th>
<th>LL</th>
<th>(LL) ( \pm S_{LL} )</th>
<th>P (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyman-EUD model</td>
<td>$a = 11.9 \pm 3.8, m = 0.108 \pm 0.027$</td>
<td>-134.5</td>
<td>-134.6 ( \pm 10.5 )</td>
<td>50.4</td>
</tr>
<tr>
<td>Logit-EUD model</td>
<td>$a = 12.1 \pm 3.8, k = 15.4 \pm 4.5$</td>
<td>-134.5</td>
<td>-134.5 ( \pm 10.5 )</td>
<td>50.0</td>
</tr>
<tr>
<td>Serial RU model</td>
<td>$a = 0.179 \pm 0.047, D_0 = 80.6 \pm 0.9$</td>
<td>-134.5</td>
<td>-135.6 ( \pm 10.6 )</td>
<td>54.0</td>
</tr>
<tr>
<td>Poisson-EUD model</td>
<td>$a = 13.5 \pm 3.8, D_0 = 78.5 \pm 0.6$</td>
<td>-134.5</td>
<td>-135.6 ( \pm 10.6 )</td>
<td>54.1</td>
</tr>
<tr>
<td>Mean dose model</td>
<td>$D_c = 73.7$ Gy, $\beta_0 = -2.88 \pm 0.34, \beta_1 = 0.050 \pm 0.013$</td>
<td>-136.1</td>
<td>-136.1 ( \pm 10.7 )</td>
<td>50.0</td>
</tr>
<tr>
<td>Cutoff dose model</td>
<td>$D_c = 79.6$ Gy, $\beta_0 = -2.10 \pm 0.16, \beta_1 = 0.068 \pm 0.015$</td>
<td>-135.3</td>
<td>-135.3 ( \pm 10.7 )</td>
<td>50.0</td>
</tr>
</tbody>
</table>

**Abbreviations:** EUD = equivalent uniform dose; RU = reconstruction unit. Dash indicates that model is not significant.

**Comparison of the models**

Both visually from Fig. 1 and according to the LL values in
Table 3, the four EUD-based NTCP models fit the data of
chronic rectal bleeding of Grade $\geq 2$ equally well. However,
as the Akaike information criterion shows, both two-
parametric NTCP models (serial reconstruction unit model
and Poisson-EUD model) have the lowest AIC-values (Table
3) and thus, can be considered to fit the data most
efficiently.

Compared with these models, the AIC-values of the two
cutoff-dose models ($D_c = 73.7$ Gy and 79.6 Gy) are con-
siderably larger, which is caused by both the smaller LL
values (indicating worse fit quality itself) and the larger
number of model parameters. Thus the EUD-based models
provide a quantitatively better and more efficient descrip-
tion of our dataset.

**DISCUSSION AND CONCLUSIONS**

In this study, six dose–volume–based or EUD-like NTCP
models were fitted to late rectal bleeding data of a large,
consistently treated patient population (319 patients),
thereby aiming to identify the most accurate approach for
quantifying the risk of chronic rectal bleeding based on
planned dose distributions of rectum and allowing statisti-
cally robust estimation of the corresponding model param-
eters.

Our results clearly confirm the volume effect for chronic
rectal bleeding of Grade 2 or worse. Quantitatively, this can
be described very well with the four EUD-like models,
where the serial reconstruction unit and the Poisson-EUD model fitted the data most efficiently according to the smallest AIC values, as they need only two parameters to describe the dataset. Thus according to our data, these two models can be considered to provide the most concise approach to quantifying the risk of chronic rectal bleeding of Grade 2 or worse. The cutoff-dose model, a directly dose–volume based model, generally fitted the data worse, but still found significant correlation of rectal wall relative volume above single cutoff-dose levels \( D_c \) in the range \( \sim 50 \) to \( 80 \) Gy (most significant for \( D_c = 73.7 \) and \( 79.6 \) Gy) for our patient population. Mean dose did not correlate to late rectal bleeding of Grade \( \geq 2 \).

For the three models incorporating the power–law EUD Eq. (2), i.e., the Lyman-, Logit-, and Poisson-EUD model, the parameter describing the volume effect was found to be in the order of \( a \sim 12 \) (Table 2). Accounting for uncertainties in parameter estimation, this is consistent to Burman et al. (5), who found \( n \approx 0.12 \) (different definition of volume effect parameter: \( a = 1/n \), i.e., \( a = 8.3 \)). It is also in agreement with the data published by Skwarchuk et al. (28) which yield \( a = 10.3 \) (fit based on Fig. 3 of the cited paper).

---

**Fig. 1.** Predicted probability of chronic rectal bleeding of Grade \( \geq 2 \) according to the different equivalent uniform dose–based normal tissue complication probability (NTCP) models: (a) Lyman–EUD model Eq. (1); (b) Logit-EUD model Eq. (3); (c) serial reconstruction unit model Eq. (6); (d) Poisson-EUD model Eq. (7). NTCP is plotted as function of the corresponding equivalent uniform dose, EUD\(_{PL}\) Eq. (2) or EUD\(_{SRU}\) Eq. (5). The x-symbols represent reply toxicity (1/0 for patients with/without toxicity, respectively). For each NTCP-model, the observed toxicity rates are shown in centers of equally sized bins (except for two bins in the low incidence region, which were combined). Errors shown are binomial. \( \chi^2 \) of the fit and the upper threshold according to Chi-square statistics (\( \alpha = 5\% \)) are given for each model.
However, it is in contradiction to findings of Rancati et al. (6) \((a = 4.3, 68\% \text{ CI } 3.6\text{—}5.6)\) and Tucker et al. (7) \((a = 0.3, \text{ with large uncertainty } a \sim 0\text{—}32.3 \text{ [95\% CI]}\).

Because of the ART-protocol used, which defined different prescription dose levels depending on individual organ geometry and mobility, the patient population shows a wide range of DVH shapes and thus large variability of dose–volume combinations, which is advantageous for robust fitting of NTCP models. However, possible statistical biases are introduced by the treatment technique and specific characteristics of the patient population itself.

This is indeed the case as becomes most evident for the results of the cutoff-dose model, Fig. 2a: The (local) maxima of the LL at \(D_c = 73.7\) and 79.6 Gy are strongly influenced by the different prescription dose levels defined by the ART protocol \((70.2, 72, 73.8, 75.6, 77.4, \text{ and } 79.2 \text{ Gy})\), whereas the finding that all models with \(D_c \text{ in the range } 50\text{—}80 \text{ Gy} \) are formally significant can be traced back to the four-field box treatment technique used, which induces correlations of all DVH dose-bins in this range. Quantitatively, such correlations can be assessed by principal component analysis (29, 30); a detailed investigation of the use of DVH– principal component analysis in the context of NTCP modeling will be presented in a subsequent publication.

In this context, a comparison with results from other publications is elucidating. Both Jackson et al. (2) and Tucker et al. (7) found significant correlations of intermediate doses \(\sim 40\text{ to } 45 \text{ Gy} \) with toxicity in contradiction to our findings. This is likely because of differences in the treatment technique (6-field 3D-CRT vs. 4-field box in combination with 6-field 3D-CRT boost), which induces specific correlations between dose-bins. Thus an important conclusion is that results of DVH-based models like the cutoff-dose model are superimposed by characteristics of the treatment technique and patient population, which compromises inferences about radiobiological effects and extrapolability of results gained from a certain treatment technique to others.

With regard to the mean dose model, our results are in contrast to the studies of Zapatero et al. (8), Tucker et al. (7), who found a correlation of mean dose and chronic rectal bleeding of Grade \(\geq 2\). Again, this might be because of differences in the treatment technique or the grading scheme used; in this context, it should be mentioned that Grade \(\geq 3\) bleeding correlated to mean dose for our population \((p = 0.045; \text{ data not shown})\).

With respect to the magnitude of the volume effect for rectum, our results can be compared with values published for other organs (5): Lung \((a \approx 1.1)\) and liver \((a \approx 3.1)\) are typical organs with a large volume effect, while e.g., spinal cord shows only a small volume effect \((a \approx 20.0)\). Thus, for our

Table 3. Comparison of the normal tissue complication probability models using the Akaike information criterion (AIC value)

<table>
<thead>
<tr>
<th>Model</th>
<th>(n)</th>
<th>LL</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyman-EUD model</td>
<td>3</td>
<td>-134.54</td>
<td>275.1</td>
</tr>
<tr>
<td>Logit-EUD model</td>
<td>3</td>
<td>-134.51</td>
<td>275.0</td>
</tr>
<tr>
<td>Serial RU model</td>
<td>2</td>
<td>-134.53</td>
<td>273.1</td>
</tr>
<tr>
<td>Poisson-EUD model</td>
<td>2</td>
<td>-134.55</td>
<td>273.1</td>
</tr>
<tr>
<td>Mean dose model</td>
<td>2</td>
<td>-138.89*</td>
<td>—</td>
</tr>
<tr>
<td>Cutoff dose model (73.7 Gy)</td>
<td>3</td>
<td>-136.08</td>
<td>278.2</td>
</tr>
<tr>
<td>Cutoff dose model (79.6 Gy)</td>
<td>3</td>
<td>-135.34</td>
<td>276.7</td>
</tr>
</tbody>
</table>

* Model is not significant \((p > 0.05)\).
data, the EUD-like models suggest a rather small volume effect of rectal bleeding (a ≈ 13.5 for the Poisson-EUD model, Table 2). As illustration of this, a plan should not have more than 2.4% (1.7%) of rectal volume irradiated with 80 Gy (82 Gy) to have the same EUD (and thus same NTCP) as an otherwise identical plan with 72 Gy to 10% of the volume.

If the presented EUD-based models are used for dose optimization, e.g., for IMRT dose escalation, it is always preferable to err toward smaller volume effects. In consequence, a EUD-based optimization of prostate RT with the parameters derived here will severely penalize high dose regions in the rectum, perhaps for some patients to an extent that prevents satisfying PTV coverage. Hence, the data suggest that safe dose escalation to the prostate can be achieved only by image guided and adaptive strategies to reduce the extent of the PTV.

REFERENCES
