

ORIGINAL ARTICLE

Normal Tissue Complication Probability (NTCP) modeling of late rectal bleeding following external beam radiotherapy for prostate cancer: A Test of the QUANTEC-recommended NTCP model

MITCHELL LIU¹, VITALI MOISEENKO², ALEXANDER AGRANOVICH¹,
ANAND KARVAT¹, WINKLE KWAN¹, ZIAD H. SALEH³, ADITYA A. APTE³
& JOSEPH O. DEASY³

¹Fraser Valley Centre, British Columbia Cancer Centre, Surrey, BC, Canada, ²Vancouver Centre, British Columbia Cancer Agency, Vancouver, BC, Canada and ³Department of Radiation Oncology and the Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO

Abstract

Purpose/background. Validating a predictive model for late rectal bleeding following external beam treatment for prostate cancer would enable safer treatments or dose escalation. We tested the normal tissue complication probability (NTCP) model recommended in the recent QUANTEC review (quantitative analysis of normal tissue effects in the clinic). *Material and methods.* One hundred and sixty one prostate cancer patients were treated with 3D conformal radiotherapy for prostate cancer at the British Columbia Cancer Agency in a prospective protocol. The total prescription dose for all patients was 74 Gy, delivered in 2 Gy/fraction. 159 3D treatment planning datasets were available for analysis. Rectal dose volume histograms were extracted and fitted to a Lyman-Kutcher-Burman NTCP model. *Results.* Late rectal bleeding (\geq grade 2) was observed in 12/159 patients (7.5%). Multivariate logistic regression with dose-volume parameters (V50, V60, V70, etc.) was non-significant. Among clinical variables, only age was significant on a Kaplan-Meier log-rank test ($p=0.007$, with an optimal cut point of 77 years). Best-fit Lyman-Kutcher-Burman model parameters (with 95% confidence intervals) were: $n = 0.068$ (0.01, +infinity); $m = 0.14$ (0.0, 0.86); and $TD50 = 81$ (27, 136) Gy. The peak values fall within the 95% QUANTEC confidence intervals. On this dataset, both models had only modest ability to predict complications: the best-fit model had a Spearman's rank correlation coefficient of $r_s = 0.099$ ($p = 0.11$) and area under the receiver operating characteristic curve (AUC) of 0.62; the QUANTEC model had $r_s=0.096$ ($p = 0.11$) and a corresponding AUC of 0.61. Although the QUANTEC model consistently predicted higher NTCP values, it could not be rejected according to the χ^2 test ($p = 0.44$). *Conclusions.* Observed complications, and best-fit parameter estimates, were consistent with the QUANTEC-preferred NTCP model. However, predictive power was low, at least partly because the rectal dose distribution characteristics do not vary greatly within this patient cohort.

Rectal injury with a late onset is a common side-effect of external beam radiotherapy for prostate cancer, and can lead to permanent changes in bowel function or chronic late rectal bleeding [1]. Identifying a reliably predictive model for late rectal bleeding would improve guidance for prostate cancer treatment planning, potentially reducing this clinically important side effect in conventional treatments, or enabling safer dose escalation. Published clinical and dose-volume determinants of radiation-induced late rectal injury have recently been reviewed by the QUANTEC group (quantitative analysis of normal tissue effects in the clinic) [1]. Using a combination

of meta-analysis and literature review, they identified NTCP model parameters judged to best describe the existing published data. However, the development of predictive risk models that account for the necessary details of delivered dose distributions is made difficult by variations in patient cohort characteristics, treatment technique details, and varying dose level and dose coverage practices within the radiation oncology community [2]. Independent validation tests of NTCP models, as for any clinical predictive model, are a key, yet often neglected, step towards potentially establishing models which are clinically useful and reliable [3]. The purpose of this report is

to test the QUANTEC-recommended model for late rectal bleeding against an independent, prospectively-collected, patient dataset.

Material and methods

Patients

With approval by the local human ethics board, consented patients, who had histologically confirmed, M0, prostate cancer, and were treated definitively with external beam radiation at the Fraser Valley Centre, British Columbia Cancer Agency, had relevant data prospectively stored in a database. In total, 161 patients, treated between February 2002 and October 2006, were included, though only 159 had evaluable treatment plans. Past medical history was elicited from patients, including diabetes, hypercholesterolemia, hypertension and current presence of hemorrhoids. Characteristics of the studied cohort are given in Table I.

Radiotherapy planning and treatment

All patients had computed tomography-simulations in a supine position, intended to be with a full bladder. A Clinical Target Volume (CTV) was defined as the prostate; typical margins from the CTV to the Planning target volume (PTV) were 10 mm in all directions. Organs of interest considered during planning included the bladder and rectum. The rectum was

contoured as a “solid” organ (i.e., including the rectal content); the superior limit was the rectosigmoid junction; the inferior limit was the anal verge. Patients were treated to an isocentric dose of 74 Gy in 37 fractions. Planning was performed using the CadPlan or Eclipse (Varian, Palo Alto, CA, USA) treatment planning systems. Radiation therapy was delivered with 18 MV photon beam treatments from Elekta SL20 or Varian EX linear accelerators. Of 159 evaluable patients, 72 received two-course treatment starting with a pelvic field arrangement using a four-field box technique. The pelvic field dose was 44 or 46 Gy in 2 Gy fractions; the balance was given using a 3D planned, conformal four-field technique with a beam arrangement of two lateral beams plus right and left anterior oblique beams.

Treatment planning data extraction

Treatment plan information for each patient, including scans, doses, and contoured structures, was exported using DICOM and then converted into the CERR (Computational Environment for Radiotherapy Research) [4] format for further analysis. Once each plan was converted, a second in-house developed software tool was used to identify doses and structures for extraction on a case-by-case basis as well as to add together dose distributions from the two treatment courses as required. Two plans could not be correctly extracted, as identified by cross checking rectal mean doses with treatment plan mean doses, leaving 159 for analysis. All plans were individually reviewed visually after conversion, within CERR. Physical doses, rather than biologically-equivalent doses, were used, consistent with the QUANTEC model.

Statistical analysis

The endpoint of interest was defined as late rectal bleeding of grade ≥ 2 . Patients were scored according to late rectal bleeding alone, such that one or more treatments for rectal bleeding (usually coagulation treatments) resulted in being scored as grade two or greater. All patients were followed every six months for the first three years, and thereafter annually. Median follow-up time was 32 months. Of the 159 evaluable patients, 12 (7.5%) had late rectal bleeding of grade ≥ 2 .

Kaplan-Meier analysis was used to test the significance of clinical variables (age, hemorrhoids, diabetes, hypercholesterolemia, hypertension), as well as use of a pelvic field course. For age, the cut-point of maximum significance according to the log-rank test was determined. Spearman's rank correlation coefficient was used to test the significance of plan ranking. The consistency of the QUANTEC-recommended

Table I. Patient characteristics.

Patients	159
Mean age at RT (range), years	70.4 (34.4, 82.5)
Pelvic field (four-field box)	
No	87 (55%)
Yes	72 (45%)
Hormone therapy	108 (68%)
Gleason score	
6	51 (32%)
7	80 (50%)
8	14 (9%)
9	14 (9%)
Presenting PSA	
≤ 10 ng/ml	85 (53%)
$10 < \text{and } \leq 20$ ng/ml	44 (28%)
> 20 ng/ml	30 (19%)
Diabetes	
Yes	29 (18%)
No	130 (82%)
Hypercholesterolemia	
Yes	31 (19%)
No	128 (81%)
Hypertension	
Yes	63 (40%)
No	96 (60%)
Haemorrhoids	
Yes	12 (8%)
No	147 (92%)

model with observations was tested using the χ^2 test. The statistical significance of dose volume thresholds (V_x , meaning the relative volume receiving more than x Gy's) was tested using logistic regression. Dose-volume cut-points were tested every 5 Gy (i.e., 5, 10, 15, ... 50, 55, 60, ... Gy). A separate best-fit Lyman-Kutcher-Burman (LKB) model was determined using the Matlab statistics toolbox (ver. 7.0; The Mathworks, Inc., Natick, MA, USA) routine NLIN-FIT. This method minimizes the least-squares fit of the prediction curve to the observed binary response rates. The area under the receiver operating characteristic curve (AUC) was used to characterize model predictive power. Calibration was assessed graphically and using the χ^2 test.

Normal Tissue Complication Probability Modeling

We test the NTCP model suggested by the QUANTEC review of late rectal bleeding dose-volume factors: the commonly-used LKB model, given by the following equations:

$$\text{NTCP} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx, \quad (1)$$

$$t = \frac{\text{gEUD} - \text{TD50}}{m \times \text{TD50}}, \quad (2)$$

$$\text{gEUD} = \left(\sum_i v_i d_i^{1/n} \right)^n, \quad (3)$$

where the i th bin of the DVH has dose (d_i) and relative volume (v_i), and gEUD stands for 'generalized equivalent uniform dose,' and the sum is over all the bins of the DVH. The LKB model provides a straightforward method for identifying the most statistically important part of the DVH, i.e., the high dose region (resulting in $n \ll 1$), the mean dose ($n \sim 1$), or something between (intermediate n). The parameter m is inversely proportional to the steepness of the dose response; TD50 fixes the 50% point of the dose-response curve. The QUANTEC-recommended parameters [1] to predict \geq grade 2 toxicity are (95% confidence intervals): $n = 0.09$ (0.04–0.14); $m = 0.13$ (0.1 – 0.17); and $\text{TD50} = 76.9$ (73.7–80.1) Gy.

Results

The actuarial incidence of late-rectal bleeding is shown in Figure 1. The only variable significant on Kaplan-Meier analysis was age (see Figure 2). An optimal cut point of 77 years yielded a significant log-rank test ($p = 0.007$). Using gEUD with the QUANTEC n value of 0.09, no cut-points reach

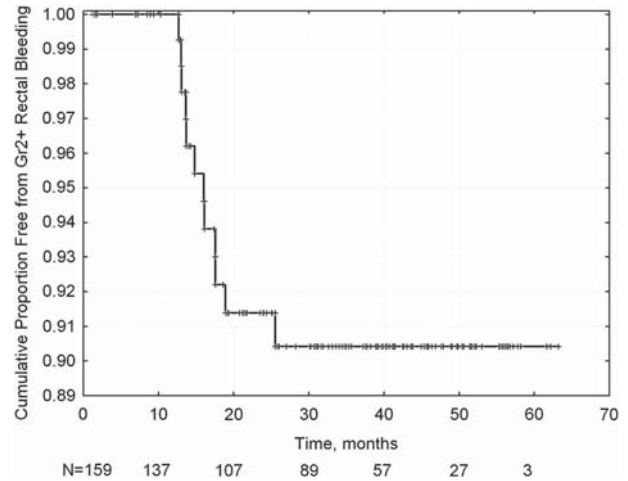


Figure 1. Kaplan-Meier actuarial estimate of late rectal bleeding \geq grade 2. Ninety percent of cases are observed within two years of the end of radiotherapy.

significance, although a gEUD value of 62.6 Gy approaches significance on the log-rank test ($p = 0.053$). Multivariate logistic regression yielded non-significant correlation between dose-volume parameters (V_{60} , V_{70} , etc.) and late rectal bleeding.

Best-fit Lyman-Kutcher-Burman model parameters (with 95% confidence intervals) were: $n = 0.068$ (0.01, +infinity); $m = 0.14$ (0.0, 0.86); and $\text{TD50} = 81$ (27, 136) Gy; yielding a non-significant Spearman correlation coefficient of $r_s = 0.098$ ($p = 0.11$). The AUC was 0.62. The QUANTEC-recommended model had $r_s = 0.096$ ($p = 0.11$), and AUC of 0.61.

Figure 3 gives the calibration of the models: the observed rate of bleeding, binned by gEUD values (using the best fit value of $n = 0.07$). The risk of bleeding among most of the patients, according to both observations and the best-fit model, is modest ($< 10\%$). Figure 3 also plots the QUANTEC-recommended model (ignoring the small error due to the change in optimal n values used to bin the x-axis values). Observations are always below the QUANTEC-recommended model predictions. However, the corresponding χ^2 statistic, with four degrees of freedom for error, is 3.81, which does not reject the model ($p = 0.44$).

Discussion

We note that the present study is limited to late rectal bleeding, which can be scored in a relatively objective fashion. Other endpoints related to late rectal injury (such as fecal incontinence) may have different dose-volume-tolerance relationships (see [5]).

The observed rectal bleeding dose-volume response is statistically consistent with the QUANTEC recommendations. However, the predictive power in this dataset is limited, partly because rectal

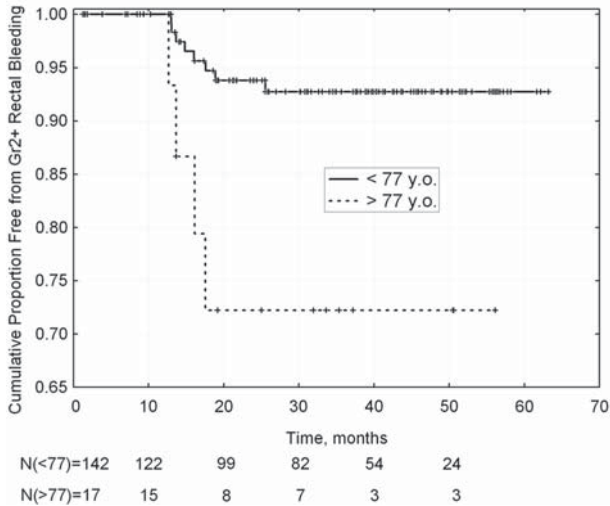


Figure 2. Freedom from \geq grade 2 rectal bleeding in patients aged over (or under) 77 years when commencing RT ($p=0.007$; log-rank test).

dose distribution characteristics were similar between most patients. The best-fit LKB parameters are similar to the QUANTEC parameters: both models, with small n values, emphasize the importance of the hottest 10% or so of the dose-volume histogram (see also [6]), although the confidence interval on the value of n for the current dataset is very wide. In this sense, we have confirmed the QUANTEC recommendations.

However, both the QUANTEC-preferred LKB model parameters and the best-fit LKB model

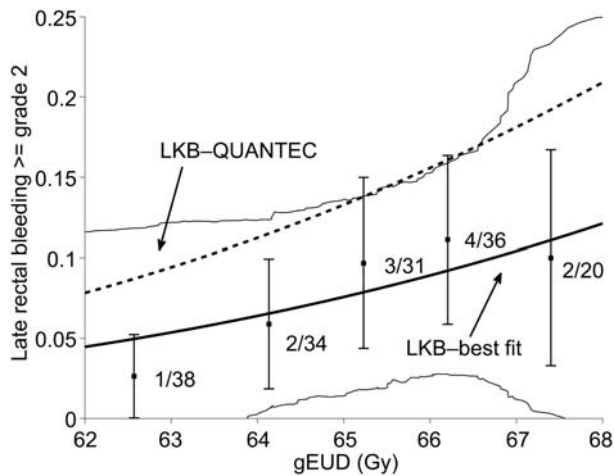


Figure 3. Incidence of late rectal bleeding as a function of gEUD. Data was binned in 1 Gy intervals according to computed gEUD values with the best fit n parameter value of 0.068. Error bars are binomial 68% confidence intervals. Only three plans fell outside the gEUD range 62 to 68, and were not included. The estimated 95% confidence interval for the best-fit model is shown by thin solid lines. The QUANTEC-recommended LKB model is shown as the dashed line (ignoring a small error due to the slightly different n values). Observed rates are smaller than QUANTEC predicted rates at all gEUD levels.

parameters yield poor predictive power in this dataset (AUC values of 0.61 and 0.62 for the best-fit and QUANTEC models, respectively). These values are less than those typically considered clinically useful (often taken as $AUC > 0.8$). This is perhaps not surprising, given the somewhat small spread in gEUD values in this dataset. In contrast, much higher variations in the risk of late rectal bleeding could be possible in other clinical protocols, so this should not be taken to mean that the model has no meaningful clinical role.

Although this dataset is clearly too small to accurately establish the effect of age, the fact that a cut point could be found that is significant on a log-rank test (uncorrected for multiple comparisons) supports further exploration of this variable as a part of a general NTCP model, especially with larger datasets. Considering recent interest in hypofractionation for prostate cancer [7], future efforts should be made to build datasets with varying prescription fraction sizes.

A generally unresolved issue is whether the LKB NTCP model really captures all the relevant toxicity information, compared to, say, dose-volume constraint guidelines. The recent clinical trial data analysis by Gulliford et al. [8] showed (as seen in the QUANTEC review) that the mid-dose region (40–60 Gy) also correlates to the risk of late rectal bleeding, though not as strongly as the region above 60 Gy. It is currently unresolved whether this is primarily because mid-dose volumes naturally correlate with high-dose volumes, or if there is a truly independent radiobiological contribution to increased risk due to changes in mid-dose volumes. In large enough datasets, this could be approached by testing the statistical significance of, say, mean dose as an additional term to gEUD with a typically low n -value.

It is well known that the true rectal dose-volume histograms vary significantly due to day-to-day anatomical changes and setup variations [9,10]. In addition, the overall accuracy may be off due to any systematic deviation away from the treatment planning anatomical representation. Thus, even if only small hot spots can cause a complication, such complications might be more likely if the nearby hot region is large, thereby making it more likely that at least some of the rectum receives an injurious dose due to anatomical shifts. Thus, the ‘true’ n -value could be significantly smaller.

The results shown in Figure 3 seem to support the idea that, although the QUANTEC-LKB model is statistically consistent with observations in this patient cohort, it may overestimate risk. In our view, a likely cause of this potential ‘miscalibration’ is the difference between the planned dose distributions and the true accumulated dose distribution [11]. An important clinical lesson is that NTCP derived values

should be considered to have significant uncertainties, unless a detailed calibration test has been carried out locally.

Interestingly, Muren et al. have shown that, at least for some patient cohorts, rectum DVH's derived with a margin for motion may correlate better to late rectal bleeding than conventional pre-treatment DVH's [12]. Thus, continued progress in defining more predictive models of late rectal bleeding risk may require the incorporation of geometrical uncertainties, or even methods to estimate the true accumulated dose distribution.

Acknowledgements

This research was partially supported by US NIH grant R01 CA85181.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- [1] Michalski JM, Gay H, Jackson A, Tucker SL, Deasy JO. Radiation dose-volume effects in radiation-induced rectal injury. *Int J Radiat Oncol Biol Phys* 2010;76(3 Suppl): S123–9.
- [2] Deasy JO, Bentzen SM, Jackson A, Ten Haken RK, Yorke ED, Constine LS, et al. Improving normal tissue complication probability models: The need to adopt a data-pooling culture. *Int J Radiat Oncol Biol Phys* 2010;76(3 Suppl):S151–4.
- [3] Bentzen SM. From cellular to high-throughput predictive assays in radiation oncology: Challenges and opportunities. *Semin Radiat Oncol* 2008;18:75–88.
- [4] Deasy JO, Blanco AI, Clark VH. CERR: A Computational Environment for Radiotherapy Research. *Med Phys* 2003; 30:979–85.
- [5] Peeters ST, Lebesque JV, Heemsbergen WD, van Putten WL, Slot A, Dielwart MF, Koper PC. Localized volume effects for late rectal and anal toxicity after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2006;64:1151–61.
- [6] Fiorino C, Valdagni R, Rancati T, Sanguineti G. Dose-volume effects for normal tissues in external radiotherapy: Pelvis. *Radiother Oncol* 2009;93:153–67.
- [7] Fowler JF. The radiobiology of prostate cancer including new aspects of fractionated radiotherapy. *Acta Oncol* 2005; 44:265–76.
- [8] Gulliford SL, Foo K, Morgan RC, Aird EG, Bidmead AM, Critchley H, et al. Dose-volume constraints to reduce rectal side effects from prostate radiotherapy: Evidence from MRC RT01 Trial ISRCTN 47772397. *Int J Radiat Oncol Biol Phys* 2010;76:747–54.
- [9] Lebesque JV, Bruce AM, Kroes AP, Touw A, Shouman RT, van Herk M. Variation in volumes, dose-volume histograms, and estimated normal tissue complication probabilities of rectum and bladder during conformal radiotherapy of T3 prostate cancer. *Int J Radiat Oncol Biol Phys* 1995;33:1109–19.
- [10] Muren LP, Ekerold R, Kvinnsland Y, Karlsdottir A, Dahl O. On the use of margins for geometrical uncertainties around the rectum in radiotherapy planning. *Radiother Oncol* 2004; 70:11–9.
- [11] Jaffray DA, Lindsay PE, Brock KK, Deasy JO, Tome WA. Accurate accumulation of dose for improved understanding of radiation effects in normal tissue. *Int J Radiat Oncol Biol Phys* 2010;76(3 Suppl):S135–9.
- [12] Muren LP, Karlsdottir A, Kvinnsland Y, Wentzel-Larsen T, Dahl O. Testing the new ICRU 62 'Planning Organ at Risk Volume' concept for the rectum. *Radiother Oncol* 2005;75: 293–302.