Two non-parametric methods for derivation of constraints from radiotherapy dose-histogram data

MA Ebert^{1,2}

SL Gulliford³

F Buettner⁴

Kerwyn Foo^{5,6}

A Haworth^{7,8}

A Kennedy¹

DJ Joseph^{1,9}

JW Denham¹⁰

¹ Department of Radiation Oncology, Sir Charles Gairdner Hospital, Western Australia, Australia

² School of Physics, University of Western Australia, Western Australia, Australia

³ Joint Department of Physics, Institute of Cancer Research and Royal Marsden National Health Service Foundation Trust, Sutton, United Kingdom

⁴ Institute of Computational Biology, Helmholtz-Zentrum Munich, Neuherberg, Germany

⁵ The Chris O'Brien Lifehouse, Camperdown, New South Wales, Australia

⁶ School of Medicine, University of Sydney, New South Wales, Australia

⁷ Department of Physical Sciences, Peter MacCallum Cancer Centre, Victoria, Australia

⁸ Sir Peter MacCallum Department of Oncology, University of Melbourne, Victoria

⁹ School of Surgery, University of Western Australia, Western Australia, Australia

¹⁰ School of Medicine and Public Health, University of Newcastle, New South Wales,
Australia

Correspondence address:

Dr Martin A Ebert

Department of Radiation Oncology

Sir Charles Gairdner Hospital

Hospital Ave

Nedlands, Western Australia 6009

Tel: 08 9346 4931

Fax: 08 9346 3402

Email: Martin.Ebert@health.wa.gov.au

Running Title: Derivation of radiotherapy constraints

Abstract

Dose constraints based on histograms provide a convenient and widely-used method for informing and guiding radiotherapy treatment planning. Methods of derivation of such constraints are often poorly described. Two non-parametric methods for derivation of constraints are described and investigated in the context of determination of dose-specific cut-points – values of the free parameter (e.g., percentage volume of the irradiated organ) which best reflect resulting changes in complication incidence. A method based on receiver operating characteristic (ROC) analysis and one based on a maximally-selected standardised rank sum are described and compared using rectal toxicity data from a prostate radiotherapy trial. Multiple test corrections are applied using a free step-down resampling algorithm, which accounts for the large number of tests undertaken to search for optimal cut-points and the inherent correlation between dose-histogram points. Both methods provide consistent significant cut-point values, with the rank sum method displaying some sensitivity to the underlying data. The ROC method is simple to implement and can utilise a complication atlas, though an advantage of the rank sum method is the ability to incorporate all complication grades without the need for grade dichotomisation.

Keywords: radiotherapy, dose volume histograms, treatment complications, toxicity, constraints

Introduction

Association of toxicity incidence with dose-volume or dose-surface area parameters has become part of the suite of methods for the presentation of normal tissue complications subsequent to radiotherapy clinical trials (e.g., [1-3]). By investigating the values (‘cut-points’) of the dependent variable (as a function of dose) that best discriminate responding from non-responding patients it is possible to generate clinically-relevant constraints. These constraints are used to optimise radiotherapy treatment planning of future patients based on a reduced representation of the full three-dimensional planned dose distribution (see for example, [4]).

Constraint derivation can be undertaken by collating dose-volume histogram (DVH) data* for the population and, at a specific dose, observing the volume at which the response event rate exceeds a specific percentage [1, 5-7]. The efficiency of a specified volumetric cut-point at a specific dose can be assessed via the contingency table corresponding to observed complication rates above and below that volume. Such a method has included assessment of the cut-point via χ^2 , Fisher’s exact test [8], log-rank [9, 10] or the Mantel-Haenszel test [11, 12]. Grouping of the population below or above the cut-point has been used to define categorical variables for logistic regression (e.g., [13]). In the scenario when such parametric test statistics are calculated for a multitude of cutpoints care has to be taken when determining the significance levels as the parametric test statistics rarely follow their nominal distribution [14, 15]. Simultaneously, corrections for multiple testing are rare, leading to exaggerated

* ‘volume’, and consequently DVH’s, are referred to here by default, though any quantity binned in the histogram (e.g., percentage volume, surface-area, image intensity, function) can be used instead.

significance levels and Type II error [16]. Additionally, as histogram points are usually
45 highly correlated, standard Bonferroni corrections for multiple testing will over-correct
significance values [10, 17]. Determination of significance is important when cut-points
are being used to generate hypotheses regarding dose-volume associations and the
pathological origins of treatment toxicity.

50 Methods to search for a cut-point have been implemented based on regression trees
[18], or an exhaustive search across discrete values of volume at each dose, selecting the
value that maximises the significance and/or odds-ratio of the resulting split [7]. Several
authors [19-21] have used receiver operating characteristic (ROC) curves as a method of
searching for the optimal volumetric cut-point at any specific dose and according to a
55 dichotomised outcome measure. Additionally, the process of searching for an optimal
cut-point involves testing multiple discrete volumetric splits – the associated multiple
testing needs to be accounted for when evaluating the probability of rejecting the
relevant null hypothesis.

60 As an alternative, Buettner *et al* [17, 22] chose the Wilcoxon rank sum as a test statistic
for determining the optimal cut-point. This optimal cut-point is determined as the split
which maximises the test statistic. In this approach, the distribution of the test statistic is
determined within a general framework of a permutation test, allowing for reliable
estimates of statistical significance by taking the dependency structure between the cut-
65 points into account. An advantage of the rank sum approach is the lack of necessity to
dichotomise outcomes, instead, using all available outcome grades.

We have investigated two above-mentioned methods for derivation of histogram-based constraints – the use of ROC curves and the use of maximally-selected Wilcoxon rank sum – in combination with more rigorous derivation of significance levels. We present these methods together with a brief investigation of their use with sample trial data.

METHODS AND MATERIALS

Figure 1 illustrates the problem of identifying the optimal volume cut-point as a function of dose from collated histogram data. Sample data, compiled from DVHs for the anorectum for 702 participants of the TROG 03.04 RADAR (Randomised Androgen Deprivation and Radiotherapy) trial [23, 24] is shown. Histograms are colour-scaled according to the peak LENT SOMA grade (0 to 4) of rectal bleeding at any point during late follow-up [25]. A more complete summary of the rectal bleeding outcomes relative to anorectum DVH is provided in the supplementary appendix in the form of complication atlases [26]. A volumetric cut-point must, at a specific dose (60 Gy equivalent dose when delivered in 2 Gy fractions (EQD2) in Figure 1), provide the optimal prediction of increase in events across the range of potential volumetric split values.

Cutpoint derivation using receiver operating characteristic (ROC) analysis

ROC analysis involves determining the change in sensitivity and specificity resulting from the dichotomisation of a continuous variable (being in this case, percentage volume[†]). It is possible to undertake ROC analysis using the differential form of a

[†] For convenience, we refer to relative ‘volume’ based histograms, though the description applies generally to all histograms such as surface area histograms, absolute and relative.

complication atlas as described previously [8, 26, 27], and which requires binning DVH
90 data into dose-volume ‘pixels’. Here we treat volume, v , at each dose bin of the DVH,
as the collection of discrete values across all patient datasets.

At each split value of volume, c , at dose d , the intent is to establish the efficiency with
which a value of $v \geq c$ predicts a complication Y . Sensitivity and specificity correspond
95 respectively to the resulting true positive rate (the number of true positive predictions as
a fraction of all patients with the complication), and true negative rate (the number of
true negatives as a fraction of all patients without the complication). The ROC curve at
 d is the resulting plot of sensitivity as a function of $(1 - \text{specificity})$. The area under the
curve (AUC) provides a non-parametric estimate of the increased probability of a
100 patient experiencing a complication relative to one not experiencing the complication
will have a larger value of the predictive variable, representing a non-parametric two-
sample statistic, or rank sum based on a binary indicator [28-31]. For example, an AUC
of 0.6 indicates that a patient with the complication incidence of interest will have a
higher value of the predictive variable (volume) than will one without the complication,
105 60% of the time.

A completely predictive dose-volume distribution at dose d would have a perfectly
defined cut-point giving sensitivity = 1, $(1 - \text{specificity}) = 0$, and an $\text{AUC}(d)$ of 1.0. The
value of c that provides the closest approach to ideal is the value that provides the
110 closest point to $(0,1)$ on the ROC curve, and this is frequently identified as the optimal
cut-point [19]. An alternative is to identify the cut-point as the value of c , corresponding
to the point with the maximum vertical distance to the line representing a complete

inability of the variable to differentiate patients experiencing the complication. This maximises the difference between sensitivity and (1-specificity), $\Delta S_{max}^Y(d)$, at the optimal cut-point $c_{optimal}$, and is known as the ‘Youden Index’ [32]. These two criteria for the optimal cut-point are only equivalent under specific conditions [33]. Figure 2a shows distributions of sensitivity, specificity and $\Delta S^Y(d, c)$ for the sample data of Figure 1, together with the resulting optimal cut-point.

Cut-point derivation using maximally-selected standardised rank sum

The method of identifying variable values which maximise a standardised test statistic has also been utilised for non-parametric derivation of cut-point values [17, 34]. In this case, the Wilcoxon rank sum, T , provides a suitable test statistic for a given volume value, c , used to split the cohort into two groups:

$$T^Y(d, c) = \sum_{i=1}^N g_i(d, c) R_i^Y \quad (2)$$

Here, for patient i , $g_i(d, c) = 0$ when the patient’s DVH at d has $v \geq c$, with $g_i(d, c) = 1$ otherwise. R_i^Y is the rank of patient i according to the complication Y , typically indicated (as in the examples presented here) by ordinal values. Note that when reduced to a dichotomous complication, the rank sum statistic is equivalent to the ROC-derived AUC [28].

T can be standardised by its mean $\mu(d, c)$ and variance $V(d, c)$ (which can be derived as described by Buettner *et al* [17] or Laussen and Schumacher [34]):

$$T^{*Y}(d, c) = \frac{|T^Y(d, c) - \mu^Y(d, c)|}{\sqrt{V^Y(d, c)}} \quad (2)$$

The value of c giving the maximum value of this standardised rank sum,

$T_{max}^{*Y}(d) = \text{maximum}\{T^{*Y}(d, c) \forall c\}$, is the optimal cut-point occurring at $c_{optimal}$. A

135 calculation based on the sample data of Figure 1 is shown in Figure 2b.

Calculation of Test Significance

In an intensive computational environment, a reliable estimation of test significance can be made using a permutation or re-sampling procedure to determine the distribution of
 140 the derived test measure and identify a significance (p-value) as the proportion of permutations giving better fits than the identified cut-point. This can be achieved via a Monte Carlo process by randomly allocating the observed complication grades across all N patients, deriving both $\Delta S^Y(d, c)$ and $T^{*Y}(d, c)$ as functions of dose and every tested split value of volume, c , a large number of times and obtaining p-values for every
 145 dose/split combination [17]. In the examples presented here, where the splits are represented by discrete patient DVH values for up to 702 patients, and at 80×1 Gy dose intervals, both methods require up to 60,320 individual tests.

A multiple test correction for significance is required. As histogram data at different
 150 doses are extracted from the same dose distributions, histogram values at different doses will be correlated. The free step-down resampling method provides a suitable multiple test correction in this context [35], and this was implemented for the examples presented here. This requires recalculation of all individual test values for every sample

described above. For practical computation, the number of samples needs to be
 155 minimised, and we found p-values varied minimally above approximately 500 samples.

Implementation

The ROC, rank sum and significance calculation methods were implemented in Matlab
 (Mathworks, Natick MA). DVH data (using percentage volume) were analysed in 1 Gy
 160 EQD2 dose intervals, with splits tested at every unique volume across all patient DVHs
 at each dose. Selected values of $c_{optimal}$ corresponded to $\Delta S_{max}^Y(d)$ and $T_{max}^{*Y}(d)$. In order
 to examine the sensitivity of ROC derived cut-points to the underlying DVH data, at
 each dose-volume, the value of $\Delta S^Y(d, c)$ was displayed as a colour scale on the dose-
 volume atlas, with each interval representing a ‘pixel’. For investigating T^*
 165 distributions, $T^*(d, c)$ was also displayed as a colour scale on the dose-volume atlas.

Investigation with sample data

To compare the predictions of each method we focus on incidence of rectal bleeding,
 across the entire late follow-up period (> 3 months post radiotherapy) for RADAR trial
 170 patients (median 72 months). DVH data were generated, for the combined treatment
 phases of 754 archived prostate radiotherapy patients treated by 23 centres (for which
 702 patients had completed bleeding assessments and a baseline grade of 0), via the
 SWAN system [36]. Physical dose values comprising the dose grid for each phase were
 converted to EQD2, for $\alpha/\beta = 3.1$ Gy, using:

$$EQD2 = D \left(\frac{\alpha/\beta + D/n}{\alpha/\beta + 2} \right) \quad (3)$$

where D is the dose for the treatment phase delivered in n fractions. The converted dose grids were then summed and DVHs derived in 1 Gy EQD2 bins for the ‘anorectum’ structure (which extends cranially from where the rectum turns horizontally to the sigmoid colon, and caudally to the level of the ischial tuberosities). Cut-point values at each 1 Gy interval were calculated with complication defined as greater than base grade (i.e., reported grade of at least 1) using the ROC method, and with inclusion of all ordinal grades for the maximally selected rank sum method.

Investigation with random outcomes data

We wished to verify that any patterns and significance being observed were due to the influence of dose-volume on complication incidence and not merely a consequence of the general dose/volume distribution of DVH data. For this purpose, random permutations of the same patient outcomes data were used to generate cut-point distributions on the anorectum DVH data using identical methods to those described above.

RESULTS

Investigation with sample data

The cut-point values derived from DVH data for the anorectum using the ROC method, and for incidence of rectal bleeding at any point during follow-up, are shown in Figure 3a, plotted over the underlying distribution of $\Delta S^Y(d, c)$. Optimal cut-points derived from $\Delta S^Y_{max}(d)$ are shown as black circles when the multiple test-corrected p-value is

greater than 0.05, and as white circles for $p \leq 0.05$. The corresponding values derived using maximally selected rank sum are shown in Figure 3b, plotted over the underlying distribution of $T^{*Y}(d, c)$. These examples have been selected out of a large number of combinations of structures and complications available in the RADAR data set and which will be presented in a separate publication due to their particular clinical significance. Particularly for the rank sum method, the underlying distribution of the test statistic can display some local maxima with change in split volume, c , at each dose, d . This can result in cut-point values displaying some discontinuities with variation in dose. Cut-point values can be related to observed complication rates by referring to the complication atlases provided in the supplementary appendix.

Investigation with random outcomes data

Figure 3c and Figure 3d were obtained based on a random permutation of the outcomes data used to generate Figure 3a and Figure 3b. Although results for just one such permutation are shown, they are representative of all results for such permutations – being inconsistency of the ROC and T_{max}^{*Y} derived cut-points (with a tendency towards the median DVH value) and absence of significance at any dose, as well as the absence of consistent areas of high $\Delta S^Y(d, c)$ or $T^{*Y}(d, c)$.

DISCUSSION

In deriving cut-points as constraints, we are trying to identify prognostic dosimetric indices that will distinguish patients likely to experience a complication from those likely not to and to associate significance with those indices. The two methods

examined here have allowed derivation of such indices with some regions of consistency. The method using maximally-selected rank sum (T_{max}^{*Y}) shows sensitivity to the structure of the underlying dataset, occasionally resulting in considerable discontinuity. Extreme discontinuities may need to be filtered by restricting the minimum number of patients falling either side of a potential split. A distinct advantage of this method however is inclusion of all details of toxicity grades rather than a simple dichotomisation. In the examples shown in Figure 3 and from a comparison across other toxicities and structures in the RADAR dataset this does not have any distinct impact on the resulting cut-points, levels or regions of significance. It should be noted that high-grade toxicities were relatively rare in the RADAR trial [23].

Although the principal aim is the derivation of dose-volume indices to guide treatment planning for future patients, identification of significant cut-points can also aid the association of dose-volume with specific complications, their severity grade and particular anatomical definitions. Such association has been undertaken on the complete RADAR trial data set and will be the subject of a future publication. It is important to keep in mind however that data from a single trial is usually representative of a relatively narrow range in dose fractionation and treatment techniques, and extrapolation of observations to alternative situations should be undertaken with caution. Comparing Figure 1a with Figure 3, and having examined distributions for other endpoints and structures, it is noted that the optimal cut-points tend to reflect the median DVH. This is also reflected in the values derived with randomly distributed outcomes in Figure 3c and Figure 3d. This is to be expected as, unless there is a strong increase in toxicity with increasing volume, the optimal discriminating split will tend to

occur when there are equal populations either side of the split. This suggests a strong influence of the actual population distribution of DVH data, and highlights the necessity for caution when applying derived constraints under alternative irradiation conditions.

Two non-parametric methods for deriving cut-points from DVH data have been presented. These allow the inclusion of either dichotomous or multiple complication grades and form the basis of robust estimates of efficient cut-points, for use as dose constraints, and appropriate levels of significance.

ACKNOWLEDGEMENTS

We acknowledge funding from Cancer Australia and the Diagnostics and Technology Branch of the Australian Government Department of Health and Ageing (grant 501106), the National Health and Medical Research Council (grant 1006447) and the University of Western Australia (Research Collaboration Award). We thank the Cancer Council Western Australia for a John Nott Fellowship for travel support. We are grateful to Rachel Kearvell, Michelle Krawiec, Sharon Richardson and the RADAR trial team for providing high quality trial data.

260 REFERENCES

1. Michalski, J.M., et al., *Radiation Dose-Volume Effects in Radiation-Induced Rectal Injury*. Int J Radiat Oncol Biol Phys, 2010. **76**(3, Supplement 1): p. S123-S129.
- 265 2. Jackson, A., 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**(3): p. 685-698.
3. Bentzen, S.M., et al., *Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues*. Int J Radiat Oncol Biol Phys, 2010. **76**(3 Suppl): p. S3-9.
- 270 4. Zarepisheh, M., et al., *A moment-based approach for DVH-guided radiotherapy treatment plan optimization*. Phys Med Biol, 2013. **58**(6): p. 1869-87.
5. Marks, L.B., et al., *Use of Normal Tissue Complication Probability Models in the Clinic*. Int J Radiat Oncol Biol Phys, 2010. **76**(3, Supplement): p. S10-S19.
- 275 6. Fiorino, C., et al., *Clinical and Dosimetric Predictors of Late Rectal Syndrome After 3D-CRT for Localized Prostate Cancer: Preliminary Results of a Multicenter Prospective Study*. Int J Radiat Oncol Biol Phys, 2008. **70**(4): p. 1130-1137.
7. Snyder, K.M., et al., *Defining the risk of developing grade 2 proctitis following 125I prostate brachytherapy using a rectal dose-volume histogram analysis*. Int J Radiat Oncol Biol Phys, 2001. **50**(2): p. 335-41.
- 280 8. Cox, B.W., et al., *Esophageal toxicity from high-dose, single-fraction paraspinal stereotactic radiosurgery*. Int J Radiat Oncol Biol Phys, 2012. **83**(5): p. e661-7.
- 285 9. Boersma, L.J., 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**(1): p. 83-92.
10. Koper, P.C., et al., *Impact of volume and location of irradiated rectum wall on rectal blood loss after radiotherapy of prostate cancer*. Int J Radiat Oncol Biol Phys, 2004. **58**(4): p. 1072-82.
- 290 11. Jackson, A., 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**(3): p. 685-698.
- 295 12. Mantel, N. and W. Haenszel, *Statistical aspects of the analysis of data from retrospective studies of disease*. Journal of the National Cancer Institute, 1959. **22**(4): p. 719-48.
13. Akimoto, T., et al., *Rectal bleeding after hypofractionated radiotherapy for prostate cancer: Correlation between clinical and dosimetric parameters and the incidence of grade 2 or worse rectal bleeding*. Int J Radiat Oncol Biol Phys, 2004. **60**(4): p. 1033-1039.
- 300 14. Altman, D.G., et al., *Dangers of using optimal cutpoints in the evaluation of prognostic factors* Journal of the National Cancer Institute, 1994. **86**(11): p. 829-835.
- 305 15. Miller, R. and D. Siegmund, *Maximally selected chi square statistics*. Biometrics, 1982. **38**(1011-1016).

16. Buettner, F., *On the relevance of the spatial distribution of dose in organs-at-risk for complications after radiotherapy*, in *Joint Department of Physics, Institute of Cancer Research and the Royal Marsden NHS Foundation Trust*. 2011, University of London.
- 310 17. Buettner, F., et al., *Assessing correlations between the spatial distribution of the dose to the rectal wall and late rectal toxicity after prostate radiotherapy: an analysis of data from the MRC RT01 trial (ISRCTN 47772397)*. *Phys Med Biol*, 2009. **54**(21): p. 6535-6548.
- 315 18. Huang, E.H., et al., *Late rectal toxicity: dose-volume effects of conformal radiotherapy for prostate cancer*. *Int J Radiat Oncol Biol Phys*, 2002. **54**(5): p. 1314-21.
19. Gulliford, S.L., et al., *A comparison of dose-volume constraints derived using peak and longitudinal definitions of late rectal toxicity*. *Radioth Oncol*, 2010. **94**(2): p. 241-247.
- 320 20. Kim, B., et al., *Development of the DVH management software for the biologically-guided evaluation of radiotherapy plan*. *Radiat. Oncol. J.*, 2012. **30**(1): p. 43-48.
21. van der Laan, H.P., et al., *Grading-system-dependent volume effects for late radiation-induced rectal toxicity after curative radiotherapy for prostate cancer*. *Int J Radiat Oncol Biol Phys*, 2008. **70**(4): p. 1138-1145.
- 325 22. Buettner, F., et al., *Using Bayesian logistic regression to evaluate a new type of dosimetric constraint for prostate radiotherapy treatment planning*. *Med Phys*, 2010. **37**(1768-1777).
23. Denham, J.W., et al., *Rectal and urinary dysfunction in the TROG 03.04 RADAR trial for locally advanced prostate cancer*. *Radioth Oncol*, 2012. **105**(2): p. 184-192.
- 330 24. Kearvell, R., et al., *Quality improvements in prostate radiotherapy: Outcomes and impact of comprehensive quality assurance during the TROG 03.04 'RADAR' trial*. *JMIRO*, 2013. **57**: p. 247-257.
- 335 25. *Lent soma scales for all anatomic sites*. *Int J Radiat Oncol Biol Phys*, 1995. **31**(5): p. 1049-1091.
26. Jackson, A., E.D. Yorke, and K.E. Rosenzweig, *The Atlas of Complication Incidence: A Proposal for a New Standard for Reporting the Results of Radiotherapy Protocols*. *Sem Radiat Oncol*, 2006. **16**(4): p. 260-268.
- 340 27. Gulliford, S.L., et al., *Dosimetric explanations of fatigue in head and neck radiotherapy: An analysis from the PARSPORT Phase III trial*. *Radioth Oncol*, 2012. **104**(2): p. 205-212.
28. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*. *Radiology*, 1982. **143**(1): p. 29-36.
- 345 29. Popescu, L.M., *Nonparametric ROC and LROC analysis*. *Med Phys*, 2007. **34**(5): p. 1556-64.
30. Zweig, M.H. and G. Campbell, *Receiver operator characteristic (ROC) plots; a fundamental evaluation tool in clinical medicine*. *Clin Chem*, 1993. **39**(561-577).
- 350 31. Colak, E., et al., *Comparison of semiparametric, parametric, and nonparametric ROC analysis for continuous diagnostic tests using a simulation study and acute coronary syndrome data*. *Comput Math Methods Med*, 2012. **2012**: p. 698320.
32. Youden, W.J., *Index for rating diagnostic tests*. *Cancer*, 1950. **3**(1): p. 32-5.

- 355 33. Perkins, N.J. and E.F. Schisterman, *The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve.* Am J Epidemiol, 2006. **163**: p. 670-675.
34. Laussen, B. and M. Schumacher, *Maximally selected rank statistics.* Biometrics, 1992. **48**: p. 73-85.
- 360 35. Westfall, P.H. and S.S. Young, *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* 1993, New york: Wiley.
36. Ebert, M.A., et al., *Detailed review and analysis of complex radiotherapy clinical trial planning data: Evaluation and initial experience with the SWAN software system.* Radioth Oncol, 2008. **86**: p. 200-210.

365

Figures

Figure 1: (a) Collated dose volume histogram (DVH) data from the RADAR trial for the anorectum, with colour scale according to peak LENT SOMA rectal bleeding grade over the entire late follow-up period (3months post radiotherapy to median 72 months). Note that very few grade 4 toxicities were reported. The median DVH is also shown. (b) Frequency histogram of volume and each bleeding grade by % volume at the specific EQD2 of 60 Gy (as indicated).

Figure 2: Using the data from Figure 1 (histograms greyed), (a) with complication dichotomised to $Y = 0$ (peak bleeding grade of 0) or 1 (peak bleeding grade of 1,2,3,4), the distributions of sensitivity, specificity and sensitivity - (1 - specificity) at 60 Gy EQD2, with volume split value. The corresponding AUC is 0.61, and the cut-point is shown corresponding ΔS_{max}^Y a relative volume of $c_{optimal} = 28.6\%$ (b) With inclusion of all grades ($Y = 0,1,2,3,4$), the corresponding T^* distribution and the optimal cut-point corresponding to T_{max}^{*Y} at a relative volume of $c_{optimal} = 29.0\%$.

Figure 3: Comparison of the two cut-point derivation methods for DVH data for the anorectum structure and for incidence of rectal bleeding. The cut-point values are shown as black circles (for $p > 0.05$) or white circles ($p \leq 0.05$). For (a), ROC derived points, for any grade above base-grade, are shown on the colour-scaled distribution of $\Delta S^Y(d, c)$, and in (b), the rank sum derived points are shown on the colour-scaled distribution of $T^{*Y}(d, c)$. Repeated analysis is shown in (c) and (d) with a random permutation applied to patient outcomes.