Radiotherapy and Oncology xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original article

Inter-institutional analysis demonstrates the importance of lower than previously anticipated dose regions to prevent late rectal bleeding following prostate radiotherapy

Maria Thor [a,*], Andrew Jackson [a], Michael J. Zelefsky [b], Gunnar Steineck [c], Asa Karlsdòttir [d], Morten Høyer [e], Mitchell Liu [f], Nicola J. Nasser [b], Stine E. Petersen [e], Vitali Moiseenko [g], Joseph O. Deasy [a]

[a] Dept of Medical Physics, Memorial Sloan Kettering Cancer Center; [b] Dept of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, USA; [c] Division of Clinical Cancer Epidemiology, Dept. of Oncology, Institute of Clinical Sciences, The Sahlgrenska Academy at the University of Gothenburg, Sweden; [d] Dept of Oncology, Haukeland University Hospital, Bergen, Norway; [e] Dept of Oncology, Aarhus University Hospital, Denmark; [f] British Columbia Cancer Agency, Vancouver Cancer Center, Canada; [g] Dept of Radiation, Medicine and Applied Sciences, University of California San Diego, La Jolla, USA

ARTICLE INFO

Article history:
Received 29 June 2017
Received in revised form 22 November 2017
Accepted 14 February 2018
Available online xxxx

Keywords:
Radiotherapy
Prostate cancer
Toxicity
GI
Rectal bleeding
Dose response

ABSTRACT

Purpose: To investigate whether inter-institutional cohort analysis uncovers more reliable dose–response relationships exemplified for late rectal bleeding (LRB) following prostate radiotherapy.
Material and methods: Data from five institutions were used. Rectal dose–volume histograms (DVHs) for 989 patients treated with 3DCRT or IMRT to 70–86.4 Gy@1.8–2.0 Gy/fraction were obtained, and corrected for fractionation effects ($\alpha/\beta = 3$ Gy). Cohorts with best-fit Lyman–Kutcher–Burman volume-effect parameter a were pooled after calibration adjustments of the available LRB definitions. In the pooled cohort, dose–response modeling (incorporating rectal dose and geometry, and patient characteristics) was conducted on a training cohort (70%) followed by final testing on the remaining 30%. Multivariate logistic regression was performed to build models with bootstrap stability.
Results: Two cohorts with low bleeding rates (2%) were judged to be inconsistent with the remaining data, and were excluded. In the remaining pooled cohorts ($n = 690$; LRB rate = 12%), an optimal model was generated for 3DCRT using the minimum rectal dose and the absolute rectal volume receiving less than 55 Gy (AUC = 0.67; $p = 0.0002$; Hosmer–Lemeshow p-value, $p_{HL} = 0.59$). The model performed nearly as well in the hold-out testing data (AUC = 0.71; $p < 0.0001$; $p_{HL} = 0.63$), indicating a logistically shaped dose–response.
Conclusion: We have demonstrated the importance of integrating datasets from multiple institutions, thereby reducing the impact of intra-institutional dose–volume parameters explicitly correlated with prescription dose levels. This uncovered an unexpected emphasis on sparing of the low to intermediate rectal dose range in the etiology of late rectal bleeding following prostate radiotherapy.
 Radiotherapy and Oncology xxx (2018) xxx–xxx

Most studies of normal tissue dose–response relationships use data from single institutions. Intra-institutional studies have only a limited variation of dose–volume variables. In essence, variables that can be identified as predictive are effectively restricted to those with sufficient variance in the investigated cohort [1], which is, consequently, closely related to the applied treatment technique, including prescription dose levels and beam arrangements.

Combining data from varied planning protocols has the potential to reduce statistical artifacts related to intra-institutional correlations among dose/volume variables. We hypothesize that combining data across institutions may shed new light on the dose tolerances for normal tissues due to a larger number of patients, and an increased variability in dose–volume histograms (DVHs) due to various treatment and delivery approaches [1]. While data-handling tools to facilitate pooled analyses are readily accessible [2], the feasibility of successfully modeling outcomes across institutions is potentially limited by differences in methods used to measure outcomes [3–6] or any unaccounted for properties of patient populations [5].

To test our hypothesis we combined six datasets from five institutions ($n = 989$) and asked if a generalizable dose–response relationship can be established for late rectal bleeding (LRB) after RT for localized prostate cancer. Late rectal bleeding has the potential to negatively impact quality of life [7]. Previous dose–response

* Corresponding author at: Dept of Medical Physics, Memorial Sloan Kettering Cancer Center, 485 Lexington Ave, NYC, NY 10012, USA.
E-mail address: thorm@mskcc.org (M. Thor).

https://doi.org/10.1016/j.radonc.2018.02.020
0167-8140/

efforts for LRB have used data from single cohorts and institutions, or synthesized dose–volume cut points from individual studies into a combined plot [3]. In this study, we first addressed whether data are fundamentally similar enough to be pooled. We then generated a dose–response relationship incorporating patient and treatment characteristics.

Methods and materials

Cohort-specific information

Six cohorts were initially identified for this pooled dose–response analysis of LRB. These cohorts comprised 989 patients treated with primary external-beam RT for localized prostate cancer in 1991–2007 to 70–86.4 Gy@1.8–2.0 Gy/fraction (Tables 1, S1 and S2). Institutions included the British Columbia Cancer Agency, Canada (Cohort 1 [8]), Aarhus University Hospital, Denmark (Cohort 2 [9]), Memorial Sloan Kettering Cancer Center, USA (Cohorts 3 and 4 [10,11]), Haukeland University Hospital, Norway (Cohort 5 [12]), and Sahlgrenska University Hospital, Sweden (Cohort 6 [13]). Treatment was typically 3D Conformal Radiotherapy (3DCRT), except in one cohort where intensity-modulated RT (IMRT) had been used (Table S1). Dose was prescribed to the isocenter except in Cohorts 3 and 4, where the prescription dose was given as the minimum isodose surface encompassing the planning target volume. Only in Cohort 2 was image-guidance routinely performed, which consisted of multiple era-specific procedures [9]. Cohorts 3 and 4 included dose/volume data for all treated patients that experienced LRB (cases), but only a subset of the patients that did not (controls): three controls were matched per case based on RT technique and year of RT, (as proposed by Jackson et al. [10]), resulting in 72 patients chosen from 369 in Cohort 3, and 68 patients chosen from 601 patients in Cohort 4. In all conducted analyses, each control was, therefore, weighted by the inverse of the sampling frequency (accounting for both RT technique and treatment year). In what follows, quoted LRB rates reflect the rates observed in the complete cohorts.

To exclude uncertainties in rectal definition, the rectum was manually re-defined in all patients to be the volume within the outer rectal contour (including contents) from the slice below the recto-sigmoid junction to the slice above the anal canal. Pre-treatment rectal preparation protocols were not used on a routine basis.

Assessment of LRB after RT had been performed by patients in two cohorts, and by physicians in four cohorts, using a total of five scoring systems (*cf*. Table S2 for a complete overview of all LRB assessments being used) [9,13–16]. The minimum follow-up time criterion was three months (Table S2). For each scoring system, LRB was defined as the maximum-recorded LRB grade within an individual's follow-up time. Across all cohorts, the median follow-up time for LRB was 3.0–7.3 year. For physician-assessed scores, LRB was defined as ≥Grade 2 (denoted $LRB_{\geq 2}$). For patient-assessed scores, there were three candidate LRB definitions (≥monthly, ≥weekly, and ≥daily occurrence of LRB, denoted $LRB_{\geq m}$, $LRB_{\geq w}$, $LRB_{\geq d}$, respectively), and we, therefore, investigated each of these three candidate definitions.

Pooling approach

Our approach was to consider whether all datasets were consistent enough to justify pooling, as commonly performed in meta-analyses. As a measure of comparability, we used the commonly reported Lyman–Kutcher–Burman (LKB) model that essentially weights different regions of the DVH according to a power-law [17,18]. Within the LKB formalism, large heterogeneities in the volume-effect parameter a indicate distinctively different volume-effects: a high value of a indicates that the highest doses in the DVH drive the complication probability, whereas a value of a near 1 indicates the mean dose drives the probability of a complication [3]. The LKB model further includes two additional parameters: the probability of a 50% complication rate (D_{50}), and the slope of the dose–response curve (m). Since both D_{50} and m depend on the a value of the investigated organ, we assumed that pooling feasibility is primarily determined by the a value rather than focusing on either D_{50} or m individually.

Prior to DVH extraction and to adjust for differences in fractionation schemes, the dose–distribution for each patient was converted into equivalent doses as if all doses were delivered in 2 Gy fractions, assuming $\alpha/\beta = 3$ Gy [3,19].

Best-fit LKB parameters (a, D_{50}, and m) for LRB were initially assessed from rectal DVHs in each cohort using Maximum Likelihood estimation with a grid search (grid size: $a = 0.001{:}100$ on a logarithmic scale in 55 steps; $D_{50} = 25{:}250$ in 2 Gy steps; $m = 0.01{:}1.1$ in 0.02 steps). For the best-fit a in each cohort, 95% confidence intervals were estimated ($95\%CI_{BP}$) from 95th percentiles of the fitted values from 1000 bootstrap sample populations [20]. The heterogeneity index I^2 [21,22] was then calculated for the cohort-specific a relative to the $95\%CI_{BP}$ of a in the other cohorts. The I^2 statistic describes the percentage of total variation across studies that is due to heterogeneity rather than chance, and ranges from 0 to 100%, with lower values indicating no observed heterogeneity [22]. We calculated the I^2 statistic after omitting each cohort in turn. An I^2 statistic close to 0% among the remaining cohorts, therefore, indicates no residual heterogeneity, and that the omitted cohort was fundamentally different from the remaining cohorts, and should not be pooled.

Subsequently, best-fit LKB parameters were assessed for the remaining pooled cohort. The area under the receiving-operating characteristics curve (AUC) of the related generalized equivalent uniform dose, gEUD [23], was compared to that of the gEUD using the QUANTEC recommended a value of 11 [3]. The AUCs had to be within the 95%Cs (AUC95%CI) of each other for the models to be considered to have same predictive ability [24].

Dose–response modeling

For the pooled cohort, more general multivariate dose–response modeling was performed based on including variables related to rectal dose and geometry, as well as patient characteristics. Dose for each patient was represented by a total of 104 variables (including also gEUD with the best-fit a), geometry by three, and patient characteristics by five variables (Table S3). All analyses were conducted in MATLABv. R2016a, and extraction of dose data was performed in the computational environment for radiotherapy research, CERR [25].

Overall, the modeling approach followed that of the Transparent Reporting of a multivariate prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement, and further details can be found in [26]. The pooled cohort was randomly split into 70% and 30%; the former was used for model training, and the latter for model testing. Dose–response modeling was based on logistic regression. Within the model building process (training), univariate and multivariate analysis (UVA, MVA) was applied with bootstrap resampling using 1000 sample populations. A backward-forward stepwise selection was used in the MVA with the objective of minimizing the Akaike Information Criterion. A variable was considered a candidate predictor for MVA if presenting with an average p-value < 0.20 across all bootstrap samples on UVA. Candidate predictors were then eliminated until no variable had a Spearman's rank correlation coefficient ($|R_s|$) ≥0.70 with any other selected variable. In MVA, a model was considered a candidate

Please cite this article in press as: Thor M et al. Inter-institutional analysis demonstrates the importance of lower than previously anticipated dose regions to prevent late rectal bleeding following prostate radiotherapy. Radiother Oncol (2018), https://doi.org/10.1016/j.radonc.2018.02.020