Local dose analysis to predict acute and late urinary toxicities after prostate cancer radiotherapy: Assessment of cohort and method effects

Eugenia Mylona, Alessandro Cicchetti, Tiziana Rancati, Federica Palorini, Claudio Fiorino, Stephane Supiot, Nicolas Magne, Gilles Crehange, Riccardo Valdagni, Oscar Acosta, et al.

Local dose analysis to predict acute and late urinary toxicities after prostate cancer radiotherapy: assessment of cohort and method effects

Eugenia Mylona,MSc [1,*], Alessandro Cicchetti,MSc [2,3], Tiziana Rancati,MSc [2], Federica Palorini,PhD [2], Claudio Fiorino,MSc [4], Stephane Supiot MD,PhD [5], Nicolas Magne MD,PhD [6], Gilles Crehange, MD,PhD [7], Riccardo Valdagni,MD,PhD [2,8,9], Oscar Acosta,PhD [1], Renaud de Crevoisier,MD,PhD [1]

[1] *Univ Rennes, CLCC Eugène Marquis, INSERM, LTSI – UMR 1099, F-35000 Rennes, France.*

[2] Prostate Cancer Program, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

[3] *Università degli Studi di Pavia, Physics Department, Pavia, Italy.*

[4] *Department of Medical Physics, San Raffaele Scientific Institute, Milano, Italy.*

[5] *Department of Radiation Oncology, Centre Georges François Leclerc, Dijon, France.*

[6] *Radiotherapy Department, Lucien Neuwirth Cancer Institute, St Priest en Jarez, France.*

[7] *Medical Physics Department, Institut de Cancérologie de l'Ouest , Saint Herblain, France.*

[8] *Department of Oncology and Hemato-Oncology, Università degli Studi di Milano, Milan, Italy.*

[9] *Department of Radiation Oncology 1, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.*

Keywords: prostate cancer; radiotherapy; urinary toxicity; predictive model; dose-response relationship

Abbreviations: DSM: dose-surface map ; DVM: dose volume map ; Ssurf: sub-surface ; Svol: sub-volume

ABSTRACT

Purpose: To perform bladder dose-surface map (DSM) analysis for (1) identifying symptom-related sub-surfaces (Ssurf) and evaluating their prediction capability of urinary toxicity, (2) comparing DSM with dose-volume map (DVM) (method effect), and (3) assessing the reproducibility of DSM (cohort effect).

Methods and materials: Urinary toxicities were prospectively analyzed for 254 prostate cancer patients treated with IMRT/IGRT at 78/80Gy. DSMs were generated by unfolding bladder surfaces in a 2D plane. Pixel-by-pixel analysis was performed to identify symptom-related Ssurf. Likewise, DVM analysis was performed to identify sub-volumes (Svol). The prediction capability of Ssurf and Svol DVHs was assessed by logistic/Cox regression using the area under the ROC curve (AUC). The Ssurf localization and prediction capability were compared to (1) the Svol obtained by DVM analysis in the same cohort and (2) the Ssurf obtained from other DSM studies.

Results: Three Ssurf were identified in the bladder: posterior for acute retention (AUC =0.64), posterior-superior for late retention (AUC =0.68), and inferior-anterior-lateral for late dysuria (AUC=0.73). Five Svol were identified: one in the urethra for acute incontinence and four in the posterior bladder part for acute and late retention, late dysuria, and hematuria. The overlap between Ssurf and Svol was moderate for acute retention, good for late retention, and bad for late dysuria, and AUCs ranged from 0.62 to 0.81. The prediction capabilities of Ssurf and Svol models were not significantly different. Among five symptoms comparable between cohorts, common Ssurf was found only for late dysuria, with a good spatial agreement.

Conclusion: Spatial agreement between methods is relatively good although DVM identified more sub-regions. Reproducibility of identified Ssurf between cohorts is low.

INTRODUCTION

Urinary toxicity after prostate cancer radiotherapy (RT) is a limiting adverse effect. High inter-fraction bladder volume variations, the complexity of urinary symptoms, the potential simultaneous impact of dosimetric, clinical, and genetic factors and the possible involvement of the urethra, make urinary prediction particularly difficult. Predictive models of urinary toxicity commonly based on bladder dose-volume histograms (DVH) lack spatial information [1,2].

The hypothesis of heterogeneous intra-organ radiosensitivity brought a methodological evolution from the whole-organ-based philosophy toward more sophisticated predictive models that integrate local spatial descriptors of 3D dose distributions. Pixel and voxel approaches, based on dose-surface maps (DSM) and dose-volume maps (DVM), respectively, have been implemented aiming at investigating more localized dose-toxicity relationships and unveiling potential spatial signatures of radio-sensitivity in various sites, such as the lungs [3], heart [4], head and neck [5], rectum [6–9], and bladder [10–12].

Within the context of urinary toxicity, DSM analysis enables the evaluation of the dose on the bladder surface across the population via pixel-wise comparisons, while DVM analysis exploits the entire 3D dose distribution exploring differences not only within the organ but in the whole pelvis via voxel-wise comparisons. Bladder DSMs have recently been applied to urinary toxicity studies only by two research teams, providing evidence of a spatially variable dose-response relationship with respect to acute [8,9] and late [10] urinary symptoms. More recently, we showed the effectiveness of the DVM approach for identifying predictive symptom-related sub-regions in the bladder and the urethra [13]. When using such spatial dose-analysis methods, two main questions arise with regard to the localization and prediction capabilities of the identified subregions. The first concerns the reproducibility and generalizability of the results, and the second concerns the impact of using different methods (DSM and DVM).

The objective of this study was to perform a bladder dose-surface map (DSM) analysis to (1) identify specific symptom-related sub-surfaces (Ssurf) and evaluate their urinary toxicity prediction capability; (2) compare the results with DVM analysis in the same cohort of patients to estimate the impact of local dose analysis methods (method effect); and (3) compare the results with previous studies using the same method [10–12], thereby assessing their reproducibility (cohort effect). The predictive value of the whole bladder DVH and of clinical parameters was also analyzed.

MATERIALS AND METHODS

The workflow of the study is depicted in Figure 1. We first identified specific toxicity-related sub-surfaces using DSM and calculated their prediction capability in our cohort of patients. We secondly compared our results to those from the DVM analysis on the same cohort and to two previously published studies [10–12] using the same method (DSM).

Population dataset and urinary toxicity endpoints

The population dataset used in this analysis consists of 254 patients with localized prostate cancer treated with IMRT/IGRT at 78/80Gy (2Gy/fraction) from two prospective studies and has been described previously [13]. The one is called STIC-IGRT whose goal was to compare two prostate IGRT control frequencies (daily versus weekly) in terms of economic [14] as well as the clinical impact [15] and the other one called PROFIT aimed at evaluating the benefits of hypo-fractionation versus conventional fractionation in case of prostate cancer radiotherapy [16]. Patient- and treatment-related characteristics are described in Table A1. Patient characteristics include medical history and baseline symptoms. The median follow-up was 50 months (range: 6–102 months).

Follow-up evaluation was scheduled at 3 months after the beginning of the radiation and every 6 months thereafter. It included digital rectal examination, PSA evaluation, and assessment of genitourinary (GU) morbidity. Acute (≤3 months from RT start) and late (>3 months) urinary toxicity was scored, for the STIC-IGRT trial, using the CTCAE v.3.0 (scoring of reference) and LENT/SOMA and for the PROFIT trial, using the RTOG/EORTC and CTCAE v.3.0 (used only for the French cohort). In total, 20 endpoints were considered: 5 symptoms (incontinence, retention, dysuria, hematuria, and frequency), grade ≥1 and grade ≥2, and acute and late urinary toxicity. Table A2 displays the rates of grade ≥1 and grade ≥2 acute and late urinary toxicity by symptom. Acute grade ≥1 and grade ≥2 toxicity rates ranged from 3 to 80% and 0 to 26%, respectively. Late 5-year grade ≥1 and grade ≥2 toxicity rates ranged from 10 to 55% and 1 to 11%, respectively. Six endpoints with event rate <5% were excluded from the study due to the low number of toxicity events. The remaining 14 toxicity endpoints were analyzed: 9 corresponding to grade≥1 ranging from 7% to 80% and 5 corresponding to grade≥2 ranging from 9% to 26%. The percentage of the follow-up period for which the patients remain symptomatic is also given in the Table A2, ranging

from 38% to 53% and from 21% to 46% for grade ≥1 and grade ≥2 toxicity, respectively. All of the patients provided informed consent. The trials were approved by the French Institutional Review Board and are registered in ClinicalTrials.gov (NCT00433706 for the STIC-IGRT trial, NCT00304759 for the PROFIT trial). The ethics clearance number of the study was 2006-A00524-47.

Details of the radiation treatment are provided in Supplementary material (Appendix A).

2D dose-surface map (DSM) and pixel-wise analysis to identify sub-surfaces related to toxicity

DSMs were generated from the planning CT delineations and dose distributions using dedicated software (VODCA, MSS Medical Software Solutions GmbH, Hagendorn, Switzerland). The workflow is shown in Figure A1. For each patient, the bladder surface was cut anteriorly at the points of intersection with the sagittal plane passing through its center of mass and virtually unfolded in a 2D plane. The dose distribution was transposed accordingly (step 1) [10]. Each dose map was first normalized in the axial direction (step 2). After aligning all of the maps in the population to the most inferior-central point of the bladder base (step 3), they were normalized to the DSM template in the cranial-caudal direction (step 4). The smallest vertical bladder extension present in the cohort (29 mm above the bladder base) was selected as the reference plane (DSM template). This affine transformation allowed the representation of the entire bladder surface of each patient on the same 2D plane. Pixel-wise comparisons between patients with/without toxicity were then performed for each endpoint using the Mann-Whitney U test. Average dose maps for each group and the corresponding dose differences and *p*-value maps were generated. *P*-value maps were thresholded at $p \leq 0.01$, thereby yielding sub-surfaces (Ssurf) of the bladder with significant dose differences. Permutation tests were also performed to account for false positives due to multiple comparisons [17] (Appendix B).

3D dose-volume maps (DVM) and voxel-wise analysis to identify sub-volumes related to toxicity

The same dataset was used to perform DVM analysis yielding 3D sub-volumes (Svol) in the bladder and urethra. The urethra was segmented on the CT image of each individual using a multi-atlas-based approach [18]. The DVM method is extensively described in [13] and detailed in Supplementary Material (Appendix C, Figure A2). In brief, DVMs were generated by first non-

rigidly registering the organs (bladder, prostate, and urethra) to a common coordinate system and then propagating the 3D planning dose distribution according to the transformation obtained beforehand.

Comparison of sub-surfaces and sub-volumes in the same cohort

The comparison between Ssurf and Svol was performed both visually and by computing their spatial overlap after propagating the 3D Svol to a common 2D plane (Supplementary Material; Figure A3). For this purpose, the bladder surface was progressively eroded up to 6 mm according to GETUG recommendations for defining the bladder wall thickness [19]. Thus, four consecutive surfaces (at 0, 2, 4, and 6 mm from the bladder surface) intersecting the Svol were unfolded using VODCA. The union of these propagated regions defined the 2D Svol. The overlap between Ssurf and 2D Svol was computed using the Dice score.

Toxicity prediction capability of the dose in sub-surfaces, sub-volumes, and the whole bladder in univariate analysis

DVHs were calculated in the Ssurf, the Svol, and the whole bladder and univariate analyses were performed bin-wise (1Gy increment) to identify the most predictive dose bin and the range of significant dose bins, for each toxicity endpoint. Logistic regression was used for acute urinary toxicity, and the discriminative performance was assessed with the area under the ROC curve (AUC). Cox regression was used for late urinary toxicity. The five-year discriminative performance was measured with the area under the time-dependent ROC curve (tAUC) as described in [20], which accounts for censoring in survival analysis. The AUC/ tAUC and 95% confidence intervals (CI) were computed using 1000 bootstrap replicates and the dose bins with the highest significant AUC (AUCmax) were selected for further analysis.

Toxicity prediction capability of dosimetric and clinical parameters in multivariate analysis

Multivariate logistic/Cox models were constructed including clinical parameters and the pre-selected dosimetric variables (dose bins) from the univariate analysis. These models were constructed using the least absolute shrinkage and selection operator (LASSO) method [21,22], which enables the simultaneous analysis of the correlations between the features and also prevents overfitting [23,24] (Appendix D). The AUC/ tAUC and 95% CIs from 1000 bootstrap replicates

were used to evaluate the models' discriminative performance. For the comparison between AUCs of Ssurf and Svol, a method accounting for the number of events/non-events in the dataset was used [25]. Overfit-corrected calibration curves were generated via bootstrapping (n=500). Multivariate analysis with backward elimination was also performed to assign model uncertainties in terms of hypothesis testing (*p*-values).

Assessment of the DSM reproducibility with comparison of the sub-surfaces between cohorts

The reproducibility of the DSM results was assessed through comparisons between the results obtained in our cohort and the results published by Palorini et al. for acute toxicity [10,11] and by Yahya et al. for late toxicity [12]. Table 1 summarizes the population and treatment characteristics of these DSM studies [10–12]. Table A3 shows the approximate correspondences for all the symptoms and grades between different scoring systems used across the studies. Overall, five symptoms with similar inter-study definitions were considered for comparison (acute frequency and retention, late dysuria, incontinence, and hematuria). The concordance between cohorts was assessed in terms of localization of identified Ssurf. The localization of the Ssurf was visually defined, first in our cohort and then retrospectively in the other cohorts, with respect to the cranio-caudal, antero-posterior, and lateral axes of the bladder. The inter-cohort agreement of the Ssurf overlap was then categorized as good, moderate, or bad.

RESULTS

Symptom-related sub-surfaces: localization and prediction capability

Figure 2 shows the localization of identified Ssurf, the average dose received, and the dose differences between patients with/without toxicities. Three Ssurf were identified only for grade≥1 toxicity endpoints, located at the posterior bladder part for acute retention, posterior-superior part for late retention, and inferior-anterior and lateral for late dysuria.

Table 2 shows the prediction capabilities of the DVH for the three Ssurf in univariate analysis. The most predictive dose bin and the corresponding AUC as well as the range on the dose bins with significant AUC ($p \leq 0.05$) are given. The maximum AUCs were 0.64 for acute retention, 0.68 for late retention, and 0.73 for late dysuria. The prediction capabilities of clinical variables in univariate analysis are given in the Supplementary material (Tables A4 and A5). For both acute and late toxicities, out of the 12 explored clinical variables 9 were significantly predictive.

Table 3 details the results of multivariate analysis using LASSO. The variables' units are given in Supplementary material (Appendix E). The AUCs were 0.70 for acute retention, 0.72 for late retention, and 0.73 for late dysuria. Table A6 (Supplementary material) shows the results of multivariate analysis using backward elimination. The AUCs were 0.70 for acute retention, 0.73 for late retention, and 0.78 for late dysuria.

Symptom-related sub-volumes: localization and prediction capability

Figure A4 shows the localization of symptom-related Svol, the average dose received, and the dose differences between patients with/without toxicities. Five Svol were identified only for grade ≥1 toxicity endpoints, located in the urethra for acute incontinence, at the posterior part of the bladder for acute retention, at the posterior part for late retention and late dysuria, and at the superior part for late hematuria.

Table 2 shows the prediction capabilities of the DVH for the five Svol in univariate analysis. The maximum AUCs were 0.73 for acute incontinence, 0.62 for acute retention, 0.70 for late retention, 0.81 for late dysuria, and 0.67 for late hematuria.

Table 3 details the results of multivariate analysis using LASSO. The AUCs were 0.73 for acute incontinence, 0.71 for acute retention, 0.79 for late retention, 0.82 for late dysuria, and 0.68 for late hematuria. Table A6 (Supplementary material) shows the results of multivariate analysis using backward elimination. The AUCs were 0.74 for acute incontinence, 0.71 for acute retention, 0.78 for late retention, 0.82 for late dysuria, and 0.67 for late hematuria.

Comparison between methods (DSM and DVM): sub-region localization and prediction capability

A total of three Ssurf and five Svol were identified. Spatial comparison between Ssurf and Svol could be performed for three symptoms. By visual inspection, the overlap agreement was considered moderate for acute retention, good for late retention, and bad for late dysuria. The corresponding dice scores are given in Figure 3. The discriminative performance of the three Ssurf and Svol in multivariate analysis was not significantly different. Calibration curves of the multivariate models are provided in the Figure A5.

Whole bladder DVH prediction capability

The prediction capabilities of the whole bladder's DVH in univariate analysis are given in Table 2. DVHs were significantly predictive for four symptoms, with AUC ranging from 0.64 to 0.72.

Comparison between cohorts: sub-surface localization and prediction capability

Table 1 summarizes the results of DSM analysis for all the cohorts [10–12]. Only five of the urinary symptoms analyzed in our cohort may be comparable with previous studies: acute frequency and retention and late incontinence, dysuria and hematuria. Among these symptoms, four Ssurf were identified in other cohorts, and two in our cohort. Only for one symptom, late dysuria, was Ssurf found in both our study and another [12], with good spatial agreement (inferior-anterior-lateral). Inter-cohort comparison of the prediction capabilities was not feasible since in Yahya et al. [12] such analysis was not performed.

DISCUSSION

Two local dose-analysis methods (DSM and DVM) were used to identify symptom-related sub-regions (Ssurf and Svol) predictive for urinary toxicity after prostate cancer radiotherapy. Results were compared with those from previous bladder DSM studies. The goals were to compare the DSM and DVM methods and to assess the reproducibility of the DSM method. Compared to previous DSM studies, the local dose-effect relationship was confirmed in our population only for one symptom out of the 5 symptoms in common, suggesting a strong cohort effect. When comparing DSM and DVM methods in the same cohort, the method effect was less pronounced, although more sub-regions were identified using DVMs.

We found a weak reproducibility (1 out of 5) of the Ssurf between cohorts, suggesting that DSM results are strongly dependent on cohort characteristics. The cohort effect may be related to population and statistics (cohort size, toxicity rates, endpoint definition) and treatment-related factors (total dose, fractionation, and technique). Indeed, Table 1 shows the diversity between cohorts. One study included only 72 patients [10]. Across the studies, prescribed doses ranged from 66Gy to 80Gy and both standard fractionation and hypo-fractionation were used. The treatment techniques were either IMRT or 3D-CRT. Toxicity rates were also different between cohorts, mostly concerning acute toxicities, namely 26% [10] versus 42% in our study, for acute retention. Nevertheless, there was one symptom (late dysuria) for which a sub-region was identified and

confirmed in two independent cohorts (anterior-inferior and lateral bladder surface, receiving 40–60Gy).

With respect to the method effect, using DVM analysis enabled the identification of five Svol in the bladder and urethra, in contrast to the DSM analysis, with which only three Ssurf were found in the bladder. This difference may be related to the fact that DVMs enable the simultaneous exploration of multiple 3D anatomical structures (e.g., the bladder and the urethra), whereas DSMs are limited to a single organ surface. For example, for acute incontinence, one Svol was found in the prostatic urethra with no evidence of dose-volume effect in the bladder, strengthening the assumption of urethra involvement in urinary toxicity [26,27]. For two symptoms (acute and late retention), both methods identified similar sub-regions in the bladder (posterior part of the bladder including the bladder trigone) corresponding to intermediate-high doses.

Concerning the prediction capability of the dose to the sub-regions, all the sub-regions identified by the two methods appear more predictive than the whole bladder, suggesting the advantage of using such sub-regions for toxicity prediction. An additional benefit of using DVMs in particular, is the possibility of performing personalized treatment planning by back-propagating the identified Svol from the template to the patient's native space and adding specific constraints to each region. However, the applicability of these methods in the clinical routine is yet to be demonstrated for the bladder, as previously done for the rectum [28]. Whereby, sparing sub-regions in the treatment planning may reduce the risk of specific side effects. Nevertheless, this planning perspective assumes that the local dose-toxicity relationship can be translated into a causality relationship, which is not necessarily true. External validation is required to confirm the generalizability of the models.

As it is acknowledged in the literature, several clinical/patient-related risk factors were found to impact on urinary toxicity, such as baseline urinary function [29–35], prior TURP [29,30,34,36,37], age [32,34,35,38–41], diabetes [33,35,41] and the intake of certain medications like anti-coagulants [35,41,42] and anti-hypertensives [31,38,39,43]. For some symptoms, such as urinary frequency, clinical parameters proved to be particularly predictive (AUC up to 0.81).

One of the limitations of our study is that we analyzed the dose distribution obtained at the planning step, while in reality, bladder volume during treatment may significantly vary from its initial planning volume [44]. These interfraction anatomical variations may have an impact on the actual delivered dose. In a series of 24 patients with the mean planning dose in the bladder being

41.4 Gy ,Nassef et al. [45] showed that differences between planned and cumulated dose could be up to 18 Gy. Nevertheless, dosimetric uncertainties are less pronounced in the bladder base compared to the cranial and anterior/lateral part [46,47]. Moreover, different scoring systems have been used across the studies to assess urinary toxicity and, in contrast to previous studies, patient-reported outcomes were not available for our population dataset. This might have obscured our ability to properly evaluate the reproducibility of the identified Ssurf. No associations were found between dose and grade ≥2 symptoms possibly due to the low number of toxicity events or because of the lack of association. Yet, we were able to demonstrate an association between grade ≥2 symptoms and several clinical parameters, stressing the importance of including clinical parameters in predictive models. Also, given the reversibility of urinary symptoms (Table A2), the Cox model might not be the optimal choice for analyzing late urinary toxicity. Finally, the frequency of 6-month follow-up might be too long to detect minor endpoints arising in between.

In conclusion, DSM and DVM analyses can unveil the heterogeneous intra-organ radiosensitivity by identifying symptom-specific sub-regions that are more predictive than the whole bladder DVH/DSH. However, the reproducibility of identified symptom related Ssurf between cohorts is weak in our study, potentially due to a cohort effect related to the heterogeneity of the compared populations. On the other hand, spatial agreement of the identified sub-regions from the two methods is relatively good, although DVM identifies more sub-regions. These results should be clearly confirmed by other prospective analyses exploring both methods before being used in clinical practice.

REFERENCES

[1] Fiorino C, Rancati T, Valdagni R. Predictive models of toxicity in external radiotherapy: Dosimetric issues. Cancer 2009;115:3135–40. doi:10.1002/cncr.24354.

[2] Viswanathan AN, Yorke ED, Marks LB, Eifel PJ, Shipley WU. Radiation Dose-Volume Effects of the Urinary Bladder. Int J Radiat Oncol Biol Phys 2010;76:116–22. doi:10.1016/j.ijrobp.2009.02.090.

[3] Palma G, Monti S, D'Avino V, Conson M, Liuzzi R, Pressello MC, et al. A Voxel-Based Approach to Explore Local Dose Differences Associated With Radiation-Induced Lung Damage. Int J Radiat Oncol Biol Phys 2016;96:127–33. doi:10.1016/j.ijrobp.2016.04.033.

[4] McWilliam A, Kennedy J, Hodgson C, Vasquez Osorio E, Faivre-Finn C, van Herk M. Radiation dose to heart base linked with poorer survival in lung cancer patients. Eur J Cancer 2017;85:106–13. doi:10.1016/J.EJCA.2017.07.053.

[5] Monti S, Palma G, D'Avino V, Gerardi M, Marvaso G, Ciardo D, et al. Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients. Sci Rep 2017;7:7220. doi:10.1038/s41598-017-07586-x.

[6] Dréan G, Acosta O, Ospina JD, Fargeas A, Lafond C, Corrégé G, et al. Identification of a rectal subregion highly predictive of rectal bleeding in prostate cancer IMRTRectal subregion involved in bleeding. Radiother Oncol 2016. doi:10.1016/j.radonc.2016.04.023.

[7] Buettner F, Gulliford SL, Webb S, Partridge M. Using dose-surface maps to predict radiation-induced rectal bleeding: a neural network approach. Phys Med Biol 2009;54:5139–53. doi:10.1088/0031-9155/54/17/005.

[8] Scaife JE, Thomas SJ, Harrison K, Romanchikova M, F Sutcliffe MP, Forman JR, et al. Accumulated dose to the rectum, measured using dose-volume histograms and dose-surface maps, is different from planned dose in all patients treated with radiotherapy for prostate cancer. Br J Radiol 2015;88. doi:10.1259/bjr.20150243.

[9] Acosta O, Drean G, Ospina JD, Simon A, Haigron P, Lafond C, et al. Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. Phys Med Biol 2013;58:2581–95. doi:10.1088/0031-9155/58/8/2581.

[10] Palorini F, Cozzarini C, Gianolini S, Botti A, Carillo V, Iotti C, et al. First application of a pixel-wise analysis on bladder dose-surface maps in prostate cancer radiotherapy. Radiother Oncol 2016;119:123–8. doi:10.1016/j.radonc.2016.02.025.

[11] Improta I, Palorini F, Cozzarini C, Rancati T, Avuzzi B, Franco P, et al. Bladder spatial-dose descriptors correlate with acute urinary toxicity after radiation therapy for prostate cancer. Phys Medica 2016. doi:10.1016/j.ejmp.2016.08.013.

[12] Yahya N, Ebert MA, House MJ, Kennedy A, Matthews J, Joseph DJ, et al. Modeling Urinary Dysfunction After External Beam Radiation Therapy of the Prostate Using Bladder Dose-Surface Maps: Evidence of Spatially Variable Response of the Bladder Surface. Int J Radiat Oncol 2017;97:420–6. doi:10.1016/j.ijrobp.2016.10.024.

[13] Mylona E, Acosta O, Lizee T, Lafond C, Crehange G, Magné N, et al. Voxel-Based Analysis for Identification of Urethrovesical Subregions Predicting Urinary Toxicity After Prostate Cancer Radiation Therapy. Int J Radiat Oncol 2019;104:343–54. doi:10.1016/J.IJROBP.2019.01.088.

[14] Perrier L, Morelle M, Pommier P, Lagrange JL, Laplanche A, Dudouet P, et al. Cost of prostate image-guided radiation therapy: Results of a randomized trial. Radiother Oncol 2013;106:50–8. doi:10.1016/j.radonc.2012.11.011.

[15] de Crevoisier R, Bayar MA, Pommier P, Muracciole X, Pêne F, Dudouet P, et al. Daily Versus Weekly Prostate Cancer Image Guided Radiation Therapy: Phase 3 Multicenter Randomized Trial. Int J Radiat Oncol 2018;102:1420–9. doi:10.1016/j.ijrobp.2018.07.2006.

[16] Catton CN, Lukka H, Gu C-S, Martin JM, Supiot S, Chung PWM, et al. Randomized Trial of a Hypofractionated Radiation Regimen for the Treatment of Localized Prostate Cancer. J Clin Oncol 2017;35:1884–90. doi:10.1200/JCO.2016.71.7397.

[17] Chen C, Witte M, Heemsbergen W, van Herk M. Multiple comparisons permutation test for image based data mining in radiotherapy. Radiat Oncol 2013;8:293. doi:10.1186/1748-717X-8-293.

[18] Acosta O, Mylona E, Le Dain M, Voisin C, Lizee T, Rigaud B, et al. Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy. Radiother Oncol 2017;125:492–9. doi:10.1016/j.radonc.2017.09.015.

[19] Beckendorf V, Guerif S, Le Prisé E, Cosset J-M, Bougnoux A, Chauvet B, et al. 70 Gy Versus 80 Gy in Localized Prostate Cancer: 5-Year Results of GETUG 06 Randomized Trial. Int J Radiat Oncol 2011;80:1056–63. doi:10.1016/j.ijrobp.2010.03.049.

[20] Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Med 2006;25:3474–86. doi:10.1002/sim.2299.

[21] Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B 1996;58:267–88. doi:10.2307/2346178.

[22] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385–95.

[23] Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1–22.

[24] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics n.d.

[25] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36. doi:10.1148/radiology.143.1.7063747.

[26] Jung J, Ahn HK, Huh Y. Clinical and functional anatomy of the urethral sphincter. Int Neurourol J 2012;16:102–6. doi:10.5213/inj.2012.16.3.102.

[27] Nitti VW. The prevalence of urinary incontinence. Rev Urol 2001;3 Suppl 1:S2-6.

[28] Lafond C, N'Guessan J, Dréan G, Perichon N, Delaby N, Acosta O, et al. PO-0841: Feasibility of dose decrease in a rectal subregion predictive of bleeding in prostate radiotherapy. Radiother Oncol 2017;123:S454–5. doi:10.1016/S0167-8140(17)31278-1.

[29] Peeters STH, Heemsbergen WD, Van Putten WLJ, Slot A, Tabak H, Mens JW, et al. Acute and late complications after radiotherapy for prostate cancer: Results of a multicenter randomized trial comparing 68 Gy to 78 Gy. Int J Radiat Oncol Biol Phys 2005;61:1019–34. doi:10.1016/j.ijrobp.2004.07.715.

[30] Heemsbergen WD, Al-Mamgani A, Witte MG, Van Herk M, Pos FJ, Lebesque J V. Urinary obstruction in prostate cancer patients from the dutch trial (68 Gy vs. 78 Gy): Relationships with local dose, acute effects, and baseline characteristics. Int J Radiat Oncol Biol Phys 2010;78:19–25. doi:10.1016/j.ijrobp.2009.07.1680.

[31] Barnett GC, De Meerleer G, Gulliford SL, Sydes MR, Elliott RM, Dearnaley DP. The Impact of Clinical Factors on the Development of Late Radiation Toxicity: Results from the Medical Research Council RT01 Trial (ISRCTN47772397). Clin Oncol 2011;23:613–24. doi:10.1016/j.clon.2011.03.001.

[32] Wortel RC, Incrocci L, Pos FJ, van der Heide UA, Lebesque J V., Aluwini S, et al. Late Side Effects After Image Guided Intensity Modulated Radiation Therapy Compared to 3D-Conformal Radiation Therapy for Prostate Cancer: Results From 2 Prospective Cohorts. Int J Radiat Oncol 2016;95:680–9. doi:10.1016/j.ijrobp.2016.01.031.

[33] Jolnerovski M, Salleron J, Beckendorf V, Peiffert D, Baumann A-S, Bernier V, et al. Intensity-modulated radiation therapy from 70Gy to 80Gy in prostate cancer: six- year outcomes and predictors of late toxicity. Radiat Oncol 2017;12:99. doi:10.1186/s13014-017-0839-3.

[34] Cozzarini C, Rancati T, Palorini F, Avuzzi B, Garibaldi E, Balestrini D, et al. Patient-reported urinary incontinence after radiotherapy for prostate cancer: Quantifying the dose–effect. Radiother Oncol 2017;125:101–6. doi:10.1016/j.radonc.2017.07.029.

[35] Yahya N, Ebert MA, Bulsara M, Haworth A, Kennedy A, Joseph DJ, et al. Dosimetry, clinical factors and medication intake influencing urinary symptoms after prostate radiotherapy: An analysis of data from the RADAR prostate radiotherapy trial. Radiother Oncol 2015;116:112–8. doi:10.1016/j.radonc.2015.06.011.

[36] De Langhe S, De Meerleer G, De Ruyck K, Ost P, Fonteyne V, De Neve W, et al. Integrated models for the prediction of late genitourinary complaints after high-dose intensity modulated radiotherapy for prostate cancer: Making informed decisions. Radiother Oncol 2014;112:95–9. doi:10.1016/j.radonc.2014.04.005.

[37] Fonteyne V, Villeirs G, Lumen N, De Meerleer G. Urinary toxicity after high dose intensity modulated radiotherapy as primary therapy for prostate cancer. Radiother Oncol

2009;92:42–7. doi:10.1016/j.radonc.2009.03.013.

[38] Cozzarini C, Rancati T, Carillo V, Civardi F, Garibaldi E, Franco P, et al. Multi-variable models predicting specific patient-reported acute urinary symptoms after radiotherapy for prostate cancer: Results of a cohort study. Radiother Oncol 2015;116:185–91.

[39] Palorini F, Rancati T, Cozzarini C, Improta I, Carillo V, Avuzzi B, et al. Multi-variable models of large International Prostate Symptom Score worsening at the end of therapy in prostate cancer radiotherapy. Radiother Oncol 2016;118:92–8. doi:10.1016/j.radonc.2015.11.036.

[40] Ahmed AA, Egleston B, Alcantara P, Li L, Pollack A, Horwitz EM, et al. A novel method for predicting late genitourinary toxicity after prostate radiation therapy and the need for age-based risk-adapted dose constraints. Int J Radiat Oncol Biol Phys 2013;86:709–15. doi:10.1016/j.ijrobp.2013.03.020.

[41] Mathieu R, Arango JDO, Beckendorf V, Delobel JB, Messai T, Chira C, et al. Nomograms to predict late urinary toxicity after prostate cancer radiotherapy. World J Urol 2014;32:743–51. doi:10.1007/s00345-013-1146-8.

[42] Schaake W, van der Schaaf A, van Dijk L V., van den Bergh ACM, Langendijk JA. Development of a prediction model for late urinary incontinence, hematuria, pain and voiding frequency among irradiated prostate cancer patients. PLoS One 2018;13:e0197757. doi:10.1371/journal.pone.0197757.

[43] Carillo V, Cozzarini C, Rancati T, Avuzzi B, Botti A, Borca VC, et al. Relationships between bladder dose-volume/surface histograms and acute urinary toxicity after radiotherapy for prostate cancer. Radiother Oncol 2014;111:100–5. doi:10.1016/j.radonc.2014.02.006.

[44] Fuchs F, Habl G, Devečka M, Kampfer S, Combs SE, Kessel KA. Interfraction variation and dosimetric changes during image-guided radiation therapy in prostate cancer patients. Radiat Oncol J 2019;37:127. doi:10.3857/ROJ.2018.00514.

[45] Nassef M, Simon A, Cazoulat G, Duménil A, Blay C, Lafond C, et al. Quantification of dose uncertainties in cumulated dose estimation compared to planned dose in prostate IMRT. Radiother Oncol 2016;119:129–36. doi:10.1016/J.RADONC.2016.03.007.

[46] Palorini F, Botti A, Carillo V, Gianolini S, Improta I, Iotti C, et al. Bladder dose-surface maps and urinary toxicity: Robustness with respect to motion in assessing local dose effects. Phys Medica 2016;32:506–11. doi:10.1016/j.ejmp.2016.03.006.

[47] Casares-Magaz O, Moiseenko V, Hopper A, Pettersson NJ, Thor M, Knopp R, et al. Associations between volume changes and spatial dose metrics for the urinary bladder during local versus pelvic irradiation for prostate cancer. Acta Oncol (Madr) 2017;56:884–90. doi:10.1080/0284186X.2017.1312014.

Tables and Figures

A. ACUTE TOXICITY

Method	Study	N	Treatment technique	PD (fraction)	Toxicity scoring	Global toxicity			Frequency			Incontinence			Retention*			Dysuria			Hematuria			Predictive clinical factors
						Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	
						Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	
DSM	Palorini et al. 2016	72	IMRT/ IGRT	70-74 Gy (2.5-2.65 Gy)	IPSS	Inf- lat	0.67	0.62	Post	0.71	0.61	Not studied			Not found	-	-	Not studied			Not studied			Age, vascular comorbidities, cardiovascular drugs, antihypertensive drugs
DSM	Improta et al. 2016	375	IMRT/ IGRT	76 Gy (1.8-2 / 2.2-2.7 Gy)	IPSS	Inf-post	0.70	0.66	Not studied			Not studied			Not studied			Not studied			Not studied			Age, BMI, HT, 5-ARIs, prostate volume, cardiovascular drugs, hypercholesterolemia treatment, hypertension treatment
DSM	Current study	254	IMRT/ IGRT	78-80 Gy (2Gy)	RTOG/ CTCAE v3	Not studied			Not found	-	NS	Not found		NS	Post	0.64	0.60	Not found	-	NS	Not studied			Age, TURP, previous surgery, anticoagulant drugs, antihypertensive drugs, prostate volume, baseline symptoms
DVM	Mylona et al. 2019								Not found	-		Urethra	0.73		Inf-post	0.62		Not found	-		(No events)			

B. LATE TOXICITY

Method	Study	N	Treatment technique	PD (fraction)	Toxicity scoring	Global toxicity			Frequency			Incontinence			Retention			Dysuria			Hematuria			Predictive clinical factors
						Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	Subregion		Bladder	
						Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	Location	AUC	AUC	
DSM	Yahya et al. 2017	754	3D-CRT	66-74 Gy (1.8-2.2 Gy)	IPSS (global) LENT-SOMA	Inf-ant	Not given	Not given	Not studied			Inf-post	Not given	Not given	Not studied			Inf-ant & lat	Not given	Not given	Sup-post & Inf-ant	Not given	Not given	Cerebrovascular condition, bowel disorder, respiratory disorder, dyslipidaemia,

					(symptoms)																	smoking, alcohol intake, baseline symptoms
	Current study	254	IMRT/ IGRT	78-80 Gy (2Gy)	RTOG/ CTCAE v3	Not studied			Not found	-	NS	Not found	-	NS	Sup-post	0.68	0.67	Inf-ant & lat	0.74	0.72	Not found	-	0.65	Age, TURP, previous surgery, HT, diabetes, anticoagulant drugs, antihypertensive drugs, prostate volume, baseline symptoms
DVM	Mylona et al. 2019								Not found	-		Not found	-		Sup-post	0.70		Post	0.81		Sup-post	0.67	

Abbreviations: DSM=dose-surface map; DVM=dose-volume map; N=number of patients, PD=prescribed dose; AUC=area under the ROC curve; NS=not significant; TURP=transurethral resection of the prostate; HT= Hormonal therapy, BMI=body-mass index; Inf=inferior; Sup=superior; Ant=anterior; Post=posterior; Lat=lateral

*We consider "weak stream" in IPSS as the corresponding symptom for retention

Table 1. Summary of DSM and DVM studies for identification of urinary toxicity sub-regions

A. Acute toxicity (Grade ≥1)					
Endpoints	**Region**	**Most predictive DVH/DSH bin (range of predictive**	**p value**	**OR (95% CI)**	**AUC (95% CI)**
Incontinence	Whole bladder	NS	-	-	-
	Ssurf	NA	-	-	-
	Svol	V80 (V80-V83)	0.037	1.02 (1.01-1.04)	0.73 (0.64-0.81)
Retention	Whole bladder	V79 (V77-V79)	0.052	1.06 (1.01-1.13)	0.60 (0.51-0.67)
	Ssurf	S42 (S15-S60)	<0.005	1.01(1-1.02)	0.64 (0.56-0.72)
	Svol	V72 (V63-V79)	0.011	1.02 (1.01-1.04)	0.62 (0.55-0.68)
B. Late toxicity (Grade ≥1)					
Symptom	**Region**	**Most predictive DVH bin (range of predictive bins)***	**p value**	**HR (95% CI)**	**tAUC at 5 years (95%**
Retention	Whole bladder	V19 (V4-V78)	<0.005	1.02 (1-1.04)	0.67 (0.59-0.75)
	Ssurf	S38 (S3-S79)	<0.005	1.02 (1.01-1.03)	0.68 (0.60-0.75)
	Svol	V35 (V7-V76)	<0.005	1.02 (1.01-1.02)	0.70 (0.62-0.77)
Dysuria	Whole bladder	V67 (V5- V78)	<0.005	1.03 (1.01-1.05)	0.72 (0.63-0.82)
	Ssurf	S70 (S8-S80)	<0.005	1.07 (1.03-1.10)	0.74 (0.64-0.83)
	Svol	V52 (V32 - V76)	<0.005	1.05 (1.02-1.08)	0.81 (0.72-0.90)
Hematuria	Whole bladder	V7 (V6-V18)	0.044	1.04 (1-1.07)	0.65 (0.55-0.75)
	Ssurf	NA	-	-	-
	Svol	V17 (V5-V25)	<0.005	1.02 (1.01-1.04)	0.67 (0.56-0.77)

Svol: Sub-volume; Ssurf: Sub-surface; OR: Odds ratio; HR: Hazard ratio; CI: Confidence interval; AUC: Area under the ROC curve; tAUC=time dependent AUC; NA: not applicable (not identified any sub-surface); NS: not significant (p > 0.05);

**DVH bin with the highest AUC and range of bins with statistically significant p-value (≤ 0.05);*

Table 2. Univariate analysis of the DVH for the whole bladder, the sub-volumes (Svol), and the sub-surfaces (Ssurf)

No sub-regions were found for grade ≥2 toxicities; grade ≥1 acute dysuria, frequency, and hematuria; or grade ≥1 late incontinence and frequency. Ssurf (S*x*) and Svol (V*x*) parameters represent the portion of the sub-surface or sub-volume, respectively, receiving at least *x* Gy of dose. The increment of DVH and DSH bins are the % of the volume and surface, respectively.

A. Acute toxicity (logistic regression)						
Grade >1						
Endpoints	**Model**	**Parameters**	**beta**	**OR (e^{beta})**	**AUC (95% CI)**	**p-value of AUC**
Retention	Ssurf	Ssurf (S42)	0.0043	1.01	0.70 (0.62-0.78)	0.9
		TURP	-0.2190	0.80		
		Baseline retention	0.7763	2.17		
	Svol	Svol (V72)	0.0098	1.01	0.71 (0.63-0.78)	
		TURP	-0.4981	0.62		
		Baseline retention	1.0355	2.81		
Incontinence	Svol	Svol (V80)	0.00891	1.01	0.73 (0.61-	NA
Dysuria	Clinical only	Age	-0.0318	0.96	0.63 (0.55-0.72)	NA
		ADT	-0.1531	0.85		
Frequency	Clinical only	Bladder volume	-0.0002	0.99	0.79 (0.7-0.88)	NA
		TURP	-1.2162	0.30		
		Hypercholesterolemia treatment	-0.4177	0.66		
		Baseline frequency	0.5518	1.73		
Grade >2						
Retention	Clinical only	TURP	-0.2528	0.78	0.63 (0.54-0.72)	NA
		Anticoagulant treatment	-0.2320	0.79		
		Baseline retention	0.774	2.17		
Dysuria	Clinical only	Antihypertensive treatment	0.3210	1.38	0.72 (0.55-0.88)	NA
		Prostate volume	0.0127	1.01		
Frequency	Clinical only	Baseline frequency	0.5908	1.81	0.59 (0.52-	NA
B. Late toxicity (Cox regression)						
Grade > 1						
Endpoints	**Model**	**Parameters**	**beta**	**HR (e^{beta})**	**tAUC (95% CI)**	**p-value of AUC**
Retention	Ssurf	Ssurf (S52)	0.0089	1.01	0.72 (0.67-0.77)	0.08
		Age	-0.0285	0.97		
		Previous abdominal surgery	0.3268	1.39		
		Baseline Retention	0.8131	2.26		
		Svol (V35)	0.0114	1.01		

	Svol	Age	-0.0222	0.98	0.79 (0.72-0.85)	
		Previous abdominal surgery	0.3486	1.41		
		Baseline Retention	0.8857	2.34		
Incontinence	Clinical only	Age	0.0355	1.04	0.81 (0.71-0.88)	NA
		TURP	1.2058	3.34		
		Diabetes	1.2925	3.64		
Hematuria	Svol	Svol (V17)	0.0105	1.02	0.68 (0.57-0.78)	NA
		Anticoagulant treatment	0.0841	1.93		
Dysuria	Ssurf	Ssurf (S70)	0.0411	1.04	0.73 (0.63-	0.1
	Svol	Svol (V52)	0.0171	1.02	0.82 (0.72-0.90)	
		Previous abdominal surgery	0.1427	1.15		
Frequency	Clinical only	Age	0.0072	1.01	0.81 (0.75-0.86)	NA
		Bladder volume	-0.00038	0.99		
		Antihypertensive treatment	0.3003	1.35		
		ADT	0.5538	1.74		
		Baseline frequency	1.0396	2.82		
Grade > 2						
Retention	Clinical only	Age	-0.0414	0.96	0.77 (0.67-0.85)	NA
		Previous abdominal surgery	0.3828	1.47		
		Diabetes	-0.3501	0.70		
		Antihypertensive treatment	-0.4946	0.61		
		ADT	0.2863	1.33		
		Baseline retention	1.0885	2.96		
Frequency	Clinical only	Diabetes	0.4673	1.60	0.81 (0.72-0.90)	NA
		Hypercholesterolemia treatment	-0.4663	0.63		
		Baseline frequency	1.1170	3.06		

Svol: Sub-volume; Ssurf: Sub-surface, TURP: transurethral resection of the prostate; ADT: Androgen deprivation therapy; OR: Odds ratio; HR: Hazard ratio; AUC: Area under the ROC curve; tAUC: time-dependent area under the ROC curve;
**statistical significance of the differences in AUCs/tAUCs between Ssurf models and Svol models*

Table 3. Multivariate predictive models for acute and late urinary toxicity using LASSO

The analysis was not performed for acute hematuria (grade ≥1 and ≥2), late dysuria and hematuria (grade ≥2), because of the low number of events. Ssurf (S*x*) and Svol (V*x*) parameters represent the portion of the sub-surface or sub-volume, respectively, receiving at least *x* Gy of dose. The units for the variables are given in Appendix E.

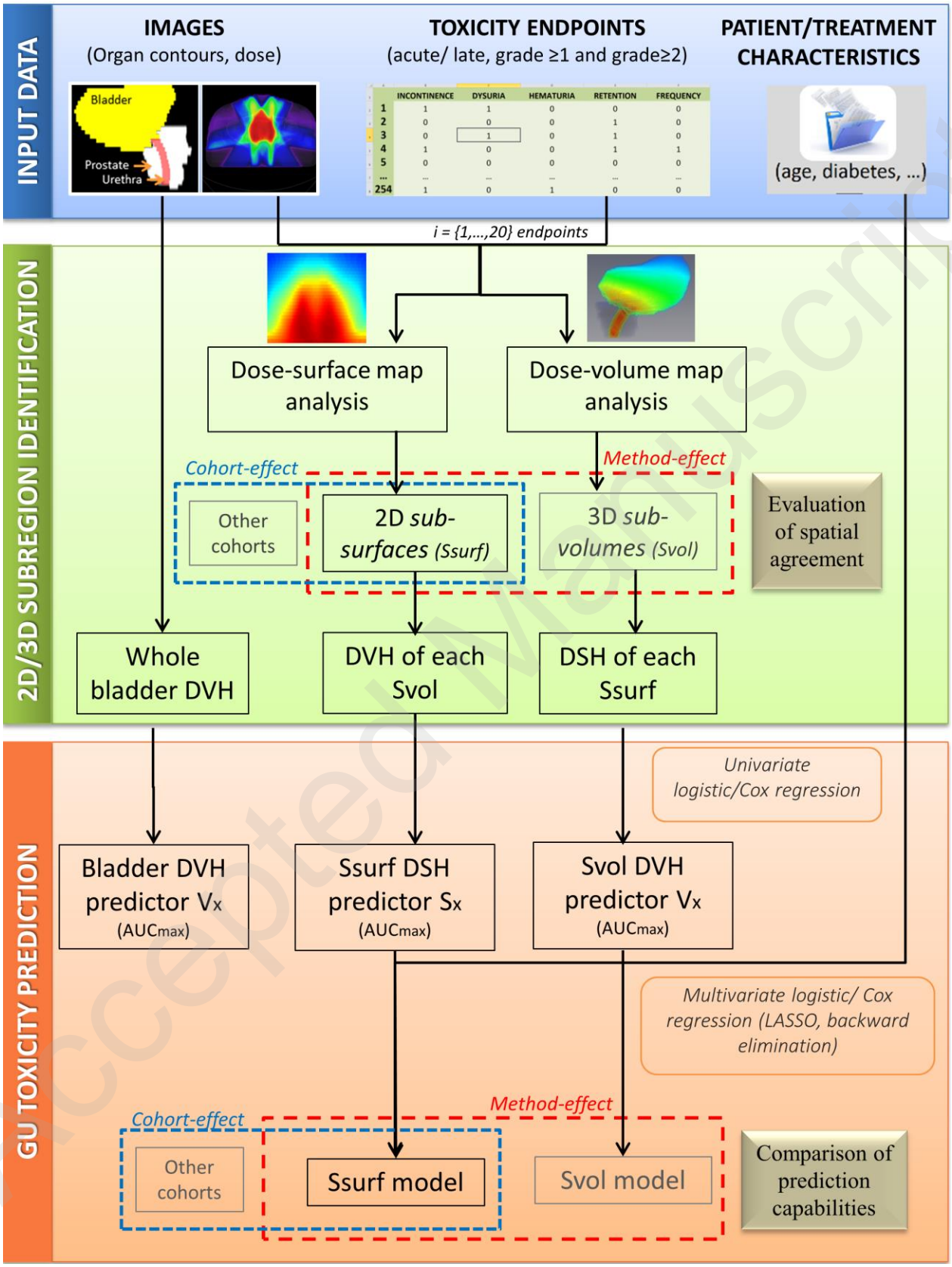


Figure 1. Study workflow

We first identified specific symptom-related sub-surfaces using DSM and calculated their prediction capabilities in our population dataset. Then, we compared our results, in terms of spatial localization and prediction capabilities to those from DVM analysis using the same cohort (assessment of method effect) and to those from other studies [10–12] using the same method (assessment of cohort effect).

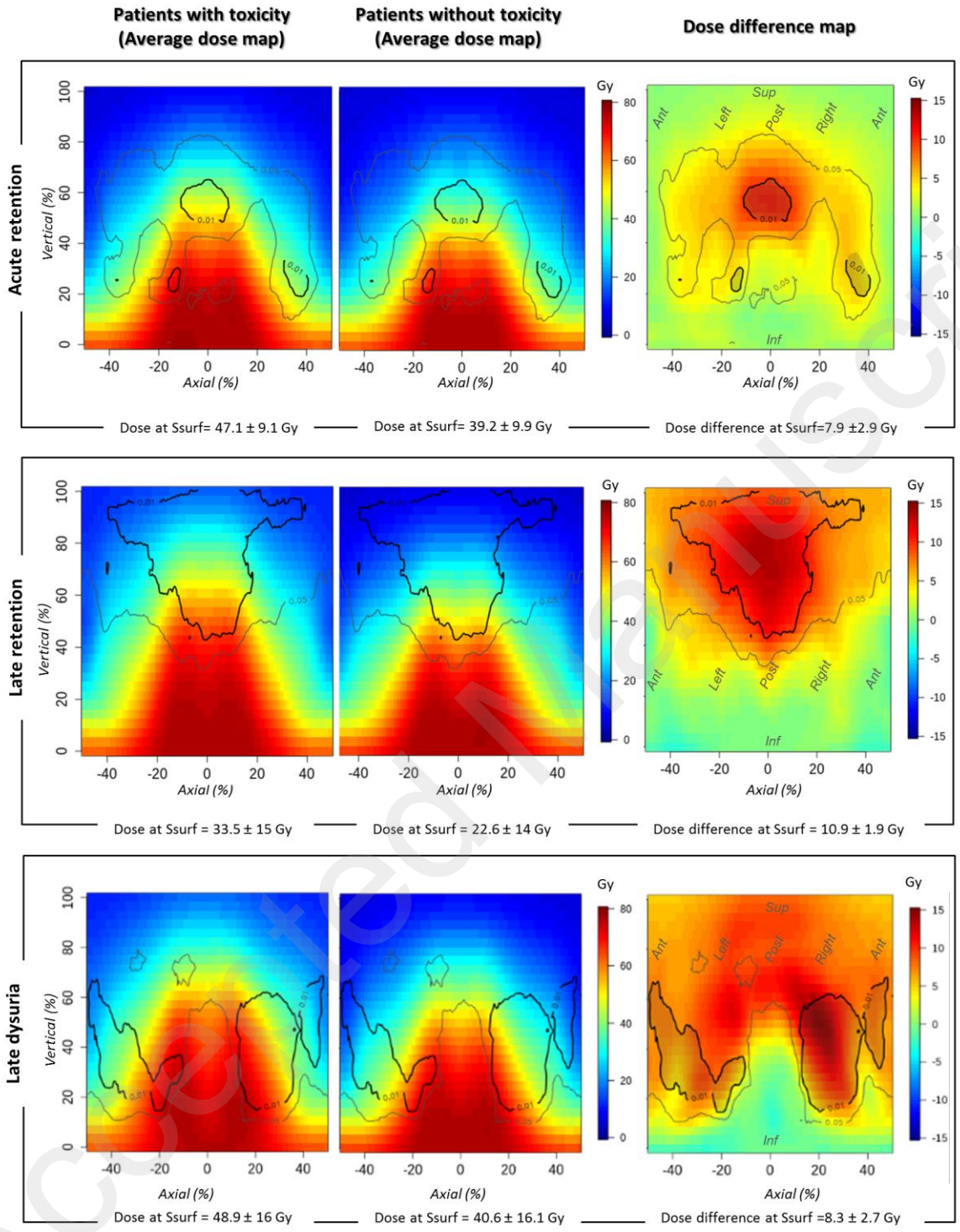


Figure 2. Symptom-related sub-surfaces (Ssurf) of statistically significant dose differences between patients with/without toxicity from DSM analysis

Ant=anterior; Post=posterior; Inf= inferior; Sup=Superior

Dose surface maps (DSMs) of the average dose distribution for the two groups (left and middle) and the corresponding dose difference (right) for three symptoms (grade ≥1). Contours show the regions with statistically significant dose differences corresponding to *p*-value ≤0.01 (bold) and ≤0.05 (light) using the pixel-wise Mann-Whitney test. The mean dose (±SD) to the Ssurf for each group and the mean dose differences are given below each DSM (only for the region corresponding to *p*-value ≤0.01).

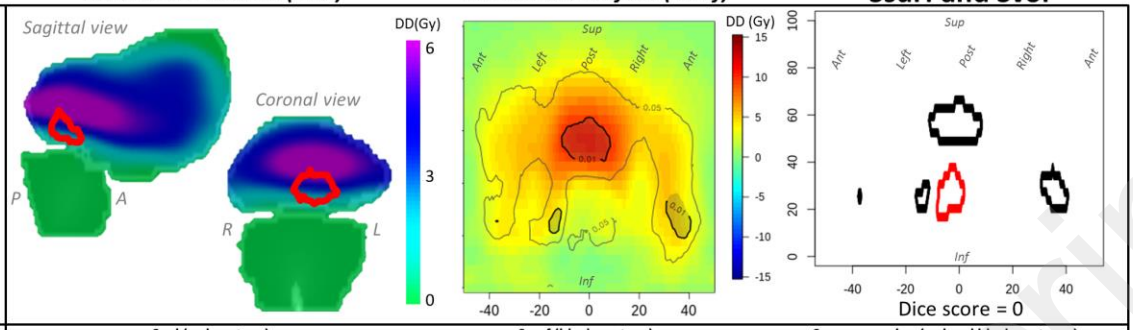
Dose difference maps

DVM & Sub-volume (Svol)

DSM & Sub-surface (Ssurf)

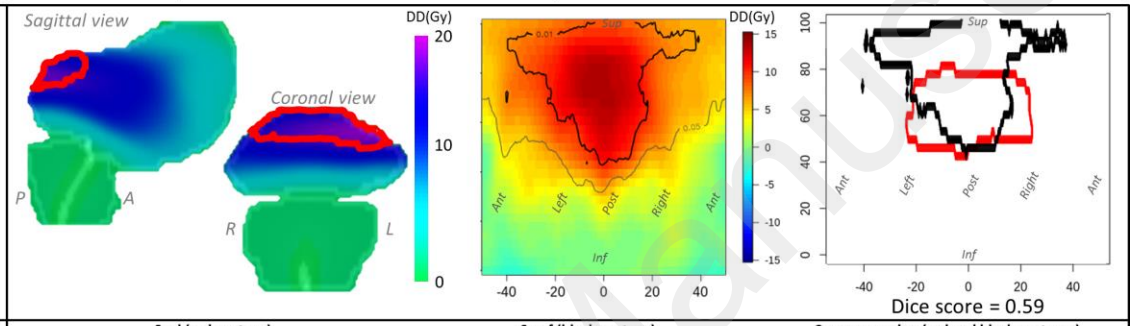
Overlap of Ssurf and Svol

Acute retention



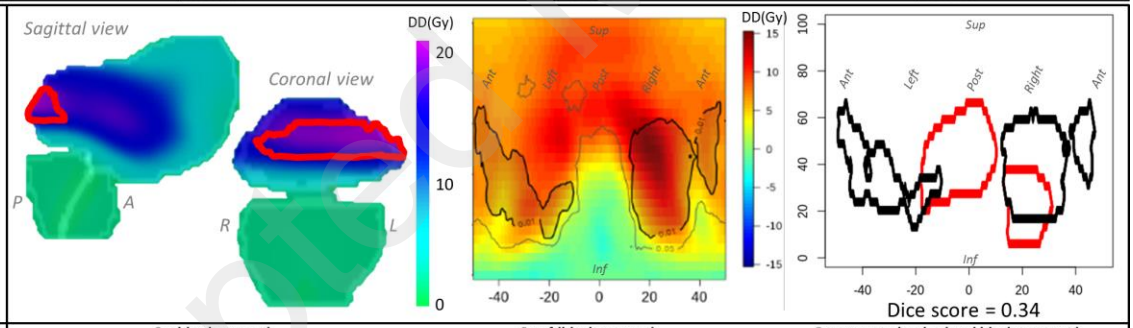
Region	Svol (red contour)	Ssurf (black contour)	Common region (red and black contours)
sup-inf	inferior	middle	-
right-left	middle	middle	-
ant-post	posterior	posterior	-

Late retention



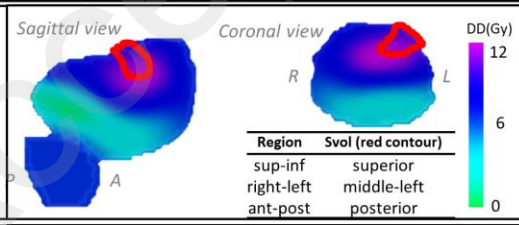
Region	Svol (red contour)	Ssurf (black contour)	Common region (red and black contours)
sup-inf	middle-superior	superior	middle-superior
right-left	middle	middle	middle
ant-post	posterior	posterior	posterior

Late dysuria



Region	Svol (red contour)	Ssurf (black contour)	Common region (red and black contours)
sup-inf	middle-inferior	middle-inferior	inferior
right-left	middle-left	right & left	right & left
ant-post	posterior	anterior	middle

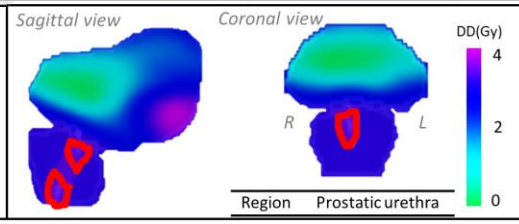
Late hematuria



Region	Svol (red contour)
sup-inf	superior
right-left	middle-left
ant-post	posterior

No sub-surface found

Acute incontinence



Region	Prostatic urethra

No sub-surface found

Figure 3. Spatial overlap between sub-surfaces (Ssurf) and sub-volumes (Svol)

Ant=anterior; Post=posterior; Sup=superior; Inf=inferior; DD=dose difference

Left: 3D dose-difference maps (DVMs) between patients with/without toxicity in sagittal and coronal views for five symptoms (grade ≥1) and the identified Svol (red contours). Middle: 2D dose-difference maps (DSMs) between patients with/without toxicity for three symptoms (grade ≥1) and the identified Ssurf (black contours). Right: Overlap of the Ssurf (black) and 2D Svol (red) and the corresponding Dice scores. The relative location of each identified sub-region is provided below each figure.