From global to local spatial models for improving prediction of urinary toxicity following prostate cancer radiotherapy

Eugenia Mylona

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *AST – Signal, Image, Vision*

Par
Eugenia MYLONA

From global to local spatial models for improving prediction of urinary toxicity following prostate cancer radiotherapy

Thèse présentée et soutenue à Rennes, le 02 décembre 2019
Unité de recherche : LTSI-INSERM UMR 1099
Thèse N° :

Rapporteurs avant soutenance :

Laura Cella	Adjunct Professor, University Federico II Napoli, Italy
Javier Pascau	Associate Professor, Universidad Carlos III de Madrid, Spain

Composition du Jury :

Président :	Juliette Thariat	PU/PH, LPC, Université de Caen
Examinateurs :	Laura Cella	Adjunct Professor, University Federico II Napoli, Italy
	Javier Pascau	Associate Professor, Universidad Carlos III de Madrid, Spain
	Dimitri Lefkopoulos	Physicien Médical/HDR, Institut Gustave Roussy – Villejuif
	Oscar Acosta	MCU, LTSI, Université de Rennes 1
Dir. de thèse :	Renaud de Crevoisier	PU/PH, LTSI, Université de Rennes 1

Invité(s)

Tiziana Rancati	Medical physicist, INT Milan, Italy

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to every person that, in one way or another, contributed in the completion of this PhD thesis.

First and foremost, I thank my thesis supervisor, Renaud de Crevoisier, whose expertise was invaluable in formulating the research questions and methodology. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I cannot thank enough my mentor, Oscar Acosta, for all the guidance, time investment and his infinite support. From the first until the last day, he stood up for me, providing research assistantship, always encouraging me to work hard, but also he gave me moral support through the rough road to finish this thesis. He helped me press on even when I thought I couldn't achieve my goals, always believing in me and motivating me to give my best. He has been an constant source inspiration.

The work presented in this thesis has been critically assessed and approved by an outstanding committee to whom I am more than grateful: Juliette Thariat, Laura Cella, Javier Pascau and Dimitri Lefkopoulos. I feel proud and honored that they have accepted to be members of my jury committee.

I consider myself privileged to have completed my thesis in LTSI, not only because I collaborated with such talented and devoted researches but, above all, because I had the chance to meet and share moments with some wonderful people. From the director of LTSI, Lotfi Senhadji, to my colleagues in the office, particularly Axel, Jeremy, Carlos and Pierre, I am deeply grateful to everyone for their kindness and genuine support during various stages of my PhD.

From the bottom of my heart, many thanks to all my friends in the lab, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research. A very special thank you to Pablo, Karim, Elena, Helene and Gustavo as well as Kleo and Raphael, who have accompanied me through this journey, suffering and celebrating every step with me. More than friends, they have been a family to me.

As I was fortunate enough to work with exceptional scientists across the world, who actively participated in the research presented in this thesis, I want to express my deepest appreciation for the hospitality they offered me, the excellent collaboration we had and for all of the opportunities I was given to further my research. I am, therefore, indebted to Alessandro Cicchetti, a true friend, and Tiziana Rancati from Istituto Nazionale dei Tumori in Milan, Italy, as well as Martin Ebert and Angel Kennedy from Sir Charles Gairdner Hospital in Perth, Australia.

Finally, this wouldn't have been possible without the infinite support of my family. Thanks to my mother, whose strength and love inspires me every day.

Contents

Acronyms

3D conformal radiotherapy (3D-CRT)
Adaptive Synthetic oversampling (ADASYN)
androgen deprivation therapy (ADT)
area under the ROC curve (AUC)
area under the time-dependent ROC curve (tAUC/ t-AUC)
centerline distance (CLD)
common coordinate system (CCS)
common terminology criteria for adverse events (CTCAE)
computed tomography (CT)
cone beam computed tomography (CBCT)
Dice Similarity Coefficient (DSC)
digital rectal examination (DRE)
dose-organ overlap (DOO)
dose-surface map (DSM)
dose-volume histogram (DVH)
dose-volume map (DVM)
Edited Nearest Neighbor (ENN)
external beam radiotherapy (EBRT)
gastrointestinal (GI)
Generalized Linear Models with likelihood based boosting (GLMboost)
genitourinary (GU)
Groupe d'étude des tumeurs urogénitales (GETUG)
High Intensity Focused Ultrasound (HIFU)
image-guided radiotherapy (IGRT)
intensity-modulated radiotherapy (IMRT)
Kaplan-Meier (KM)
least absolute shrinkage and selection operator (LASSO)
Lyman–Kutcher–Burman (LKB) model
machine learning (ML)
Modified Hausdorff Distance (MHD)
Multi-Atlas-Based Urethra Segmentation (MABUS)
multileaf collimator (MLC)
normal tissue complication probability (NTCP)
organ at risk (OAR)
Partial Least Squares Discriminant Analysis (PLS-DA)
planning tumor volume (PTV)
prostate specific antigen (PSA)
quality-of-life (QoL)
Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC)
Radiation Therapy Oncology Group (RTOG)
Random Forests (RF)
Regularized Discriminant Analysis (RDA)
sub-surfaces (Ssurf)
sub-volume (Svol)
Support vector machines (SVM)
Synthetic Minority Oversampling Technique (SMOTE)
threshold-free cluster enhancement (TFCE)
transurethral resection of the prostate (TURP)
treatment planning system (TPS)
volumetric modulated arc therapy (VMAT)

List of figures

List of tables

Résumé en français

Amélioration de la prédiction de la toxicité urinaire après radiothérapie du cancer de la prostate à partir de modèles spatiaux multi-échelle de la dose: depuis les organes à risque aux sous-régions

Le cancer de la prostate est le deuxième cancer le plus fréquente chez l'homme et la cinquième cause de décès par cancer dans le monde. En 2018, cela représentait 1,3 million de nouveaux cas et 359 000 décès dans le monde. En France, le cancer de la prostate est le plus fréquent chez l'hommes. Le nombre de cas estimé pour 2018 est de 65 000 et le nombre de décès estimé à 9 000.

De nombreuses options thérapeutiques existent pour traiter cette maladie. Parmi elles, la radiothérapie externe (EBRT) est préconisée pour plus de deux tiers des patients, en l'associant ou non à une chirurgie et/ou à une chimiothérapie. Le but de ce traitement est de délivrer, sur plusieurs séances de traitement, une forte dose de rayonnement ionisant au volume cible, à savoir la prostate et les vésicules séminales. Le bénéfice d'une augmentation de la dose à la prostate, en termes de contrôle local et de survie globale, est atténué par la présence des organes à risque (OAR), tels que la vessie et le rectum, qui entourent le volume cible. Les effets secondaires dus à l'irradiation aux tissus sains peuvent entraîner des événements indésirables importants de nature urinaire, rectale ou sexuelle.

Les développements technologiques récents (systèmes de traitement, d'imagerie et de calcul) ont permis d'améliorer le rapport bénéfices/risques lié à la thérapie. L'introduction de techniques de traitement modernes, telles que la radiothérapie avec modulation d'intensité (IMRT) et la radiothérapie guidée par l'image (IGRT), ont permis une escalade de la dose. La radiothérapie avec modulation d'intensité constitue une avancée importante pour la délivrance d'une distribution de dose de haute précision. Elle permet notamment d'améliorer la conformation de l'irradiation à des formes tumorales complexes et concaves. La balistique de traitement est quant à elle calculée au moyen d'algorithmes permettant de simuler et d'optimiser la distribution de dose via une approche de planification inverse. L'objectif est d'optimiser le plan de traitement de façon à obtenir une dose de rayonnement élevée dans le volume tumoral mais aussi limitée que possible dans les organes à risque voisins.

Malgré les améliorations récentes dans la planification et la délivrance du plan du traitement, il n'est pas encore possible d'épargner totalement les organes à risques environnant la tumeur. Des effets secondaires liés à la toxicité du traitement sont toujours observés, en particulier dans les contextes d'escalade de dose. On distingue les effets secondaires qui se produisent pendant le traitement et les quelques semaines qui suivent, dits «aigus». Et les effets secondaires qui peuvent apparaître plusieurs mois/années après la fin du traitement, appelés effets « tardifs ». Cette différence d'apparition dans le temps est liée à la vitesse variable de prolifération des tissus. Les effets secondaires diffèrent largement d'une personne à l'autre selon la localisation et le volume irradié, la dose délivrée, la radiosensibilité

individuelle du patient et son état général. Les effets secondaires tardifs de la radiothérapie externe peuvent apparaître après la fin du traitement, au niveau urinaire, digestif ou sexuel. Au niveau digestif, il est possible que des saignements apparaissent dans les selles (on parle de rectorragies). Bien que la toxicité gastro-intestinale (GI) soit relativement réduite par l'introduction de techniques fortement conformationelles, les taux de toxicité génito-urinaire (GU) restent relativement stables. Comme l'indique le projet «Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) « une-réponse similaire à celle de la toxicité gastro-intestinale est loin d'être établie pour la toxicité GU.

La compréhension de la relation dose-volume avec la toxicité urinaire et le développement de modèles prédictifs fiables est d'une importance primordiale. La prédiction de toxicité peut être utile pour informer les patients sur les complications possibles liées au traitement et pour aider les cliniciens à mieux adapter le traitement aux caractéristiques du patient avec l'objectif de diminuer le risque d'effets secondaires. Cependant, la modélisation de la toxicité urinaire est un problème difficile, non seulement en raison de la variété des symptômes associés, mais aussi en raison des limites des descripteurs de dose et des difficultés à identifier les régions potentiellement responsables de ces symptômes. La vessie, par exemple, présente d'importantes variations inter fraction de forme, causant des incertitudes géométriques et de dose qui limitent la possibilité de modéliser avec précision la réponse dose-volume. Une autre structure potentiellement impliquée est l'urètre intra-prostatique qui reste, cependant, largement inexplorée dans la littérature en raison de la difficulté inhérente à être identifiée sur l'examen de planification (tomodensitométrie 3D ou scanner CT). Bien qu'il y ait des évidences d'implication de l'urètre dans la curiethérapie de cancer de la prostate ceci n'a pas encore été montré dans la radiothérapie externe. La quantification de la dose délivrée à l'urètre peut donc améliorer notre compréhension de la toxicité urinaire ou au moins certains des symptômes, mais nécessite l'identification de cet organe dans les images de planification.

Afin d'expliquer la toxicité urinaire, plusieurs études utilisent l'histogramme dose-volume (HDV) de la vessie entière sans un consensus clair. Les modèles prédictifs de toxicité généralement basés sur HDV, réduisent la distribution de dose 3D dans l'organe à une représentation unidimensionnelle de la relation dose-volume. L'information spatiale est donc perdue, ignorant la variabilité locale de la distribution de la dose 3D, ce qui peut limiter les capacités de prédiction. Les modèles HDV supposent que l'organe est homogène en termes de radiosensibilité et de statut fonctionnel. En fait, certaines manifestations individuelles de dysfonctionnement urinaire peuvent être liées à l'irradiation de régions spécifiques des voies urinaires inférieures, à savoir l'urètre et le col de la vessie, non expliquées par les méthodes traditionnelles HDV.

Cette hypothèse de radiosensibilité hétérogène intra-organe, a déclenché aujourd'hui une évolution méthodologique des approches de prédiction qui ne considèrent plus l'organe en entier, mais qui nécessitent l'analyse des sous-organes avec des modèles prédictifs plus sophistiqués qui intègrent des descripteurs des distributions de dose 3D. En ce qui concerne la toxicité urinaire, l'analyse de la distribution locale de la dose à des échelles spatiales plus fines, peut améliorer notre compréhension de la symptomatologie urinaire après RT. Jusqu'à présent, seulement un petit nombre d'études ont cherché à étudier l'effet local de dose-volume dans la vessie fournissant quelques évidences de radiosensibilité

intra-organe, quasi-hétérogène, tandis que le rôle de l'urètre reste à démontrer. Démêler la relation entre la dose locale à la vessie et l'urètre et la toxicité urinaire, peut être traduit en recommandations pour la planification du traitement dans la pratique clinique.

En partant des modèles globaux qui utilisent la dose aux organes entiers et en allant vers des méthodes plus locales, qui considèrent des sous-organes, cette thèse vise à améliorer notre compréhension de la toxicité urinaire après irradiation prostatique. Plus précisément, les objectifs de cette thèse sont les suivants :

1. Évaluer l'implication de l'urètre dans la toxicité urinaire.
2. Évaluer l'implication de sous-parties spécifiques de la vessie et de l'urètre dans la toxicité urinaire.
3. Étudier le potentiel d'une amélioration complémentaire de la prédiction en exploitant de nouvelles méthodes d'apprentissage automatique.

Afin d'évaluer la contribution de l'urètre à la toxicité urinaire, il a été nécessaire d'identifier cette structure sur la planification CT-images des patients traités par radiothérapie externe. Pour ce faire, nous avons proposé une approche de segmentation multi-atlas permettant d'identifier et segmenter avec précision l'urètre sur les images CT. L'atlas se compose d'un ensemble d'images CT de patients traités par curiethérapie, où l'urètre est visible grâce à l'utilisation d'une sonde urinaire. Pour un nouveau patient, les caractéristiques géométriques sont extraites de l'image à segmenter et comparées aux individus de la base de données. L'urètre pour ce nouveau patient est, alors, défini en combinant les urètres des atlas les plus semblables dans un processus de fusion pondérée.

Différentes approches peuvent être utilisées pour évaluer les relations dose-toxicité à des échelles spatiales plus fines. Tout d'abord, une analyse de la dose-sur la surface (DSM) de la vessie a été effectuée pour différents symptômes urinaires. Les résultats ont été comparés aux études précédentes qui utilisent les DSM afin d'évaluer l'impact des différentes cohortes dans l'identification des sous-régions radiosensibles. Cette analyse, limitée à la surface de l'organe, a été suivie par l'exploitation de l'ensemble de la distribution de la dose de planification 3D dans la vessie et l'urètre sans des hypothèses préalables quant à l'emplacement des régions à risque. Une nouvelle approche basée sur l'analyse de la carte 3D dose-volume (DVM) a ainsi été proposée permettant l'exploration de la relation dose locale-effet, au niveau du voxel. Cette méthode, validée rigoureusement, requiert plusieurs étapes : i) une normalisation spatiale des anatomies vers un référentiel commun, ii) la production d'une cartographie des doses à analyser et, iii) une méthodologie fiable pour effectuer l'analyse statistique locale.

La prédiction de la toxicité urinaire peut être améliorée non seulement en identifiant les descripteurs locaux de la dose qui seraient plus prédictifs que les descripteurs globaux (organe entier), mais aussi en analysant de façon fiable des données complémentaires disponibles, y compris dosimétriques, cliniques et biologiques. La majorité d'études existantes s'appuient sur des approches traditionnelles de régression (p. ex. régression logistique ou régression de Cox). Ces procédures, bien qu'elles soient couramment utilisées en raison de leur interprétablilté (valeurs p et intervalles de confiance), peuvent ne pas prendre en compte les interactions entre les variables, ce qui diminue leur capacité prédictive.

Des études récentes suggèrent que d'autres stratégies plus modernes peuvent donner des meilleurs résultats que les méthodes conventionnelles de modélisation prédictive. Les techniques d'apprentissage automatique, par exemple, peuvent potentiellement augmenter la prédiction de la toxicité car elles s'appuient sur des exemples informatifs antérieurs. Néanmoins, parmi la multitude de méthodes existantes il n'y a pas une évidence claire sur des méthodes plus performantes que d'autres, n'y a pas par ailleurs d'orientation claire sur leur utilisation dans ce contexte car leur application n'est pas simple Une étude comparative entre différentes techniques d'apprentissage automatique a été effectuée afin de fournir des éléments sur les avantages et inconvénients de différents classifieurs dans le cadre de la prédiction de la toxicité urinaire.

Cette thèse est structurée autour de plusieurs publications dans des revues et conférences internationales, des travaux soumis et des travaux en préparation, à savoir :

- Le chapitre 1 présente le contexte clinique et la modélisation de la prédiction de la toxicité urinaire après irradiation prostatique. Nous commençons par une description du cancer de la prostate et des options de traitement disponibles. Nous présenterons plus précisément les différentes techniques de radiothérapie et leurs limitations ainsi que les effets secondaires les plus fréquents en identifiant en particulier les symptômes urinaires. Ensuite, nous présentons les modèles traditionnels 1D DVH et identifiant leurs avantages et limites. Cette partie est suivie d'une description de deux nouvelles méthodologies qui permettent d'analyser les distributions de doses à des échelles spatiales plus fines, d'une part, basées sur la construction de cartographies 2D de la dose-(DSM) et, d'autre part, en exploitant la distribution de dose 3D pour une cartographie tridimensionnelle (DVM). Enfin, les objectifs de la thèse sont expliqués.
- Le chapitre 2 présente une méthodologie pour segmenter automatiquement l'urètre intra-prostatique sur l'image de planification CT. Tout d'abord, la méthodologie de segmentation multi-atlas est détaillée, y compris la construction des atlas et la segmentation finale de l'urètre. Ensuite, une évaluation de la méthode est réalisée en comparant à des méthodes existantes pour estimer la position de l'urètre. La dose à l'urètre est également calculée et comparé à la dose à la prostate.
- Le chapitre 3 décrit une méthodologie permettant de caractériser spatialement la distribution de la dose à la surface de la vessie à l'aide de DSM. Le premier objectif de ce chapitre est d'améliorer les approches existantes pour l'analyse des DSM et de proposer une méthodologie qui permettrait l'exploration de toute la surface de la vessie. Le deuxième objectif est d'identifier les sous-surfaces liées aux symptômes de la vessie qui sont potentiellement plus prédictives que l'ensemble de la vessie. Le troisième objectif est de comparer les résultats avec les études précédentes basées sur les DSM de la vessie et d'estimer le niveau d'accord avec nos observations.
- Le chapitre 4 propose une méthodologie basée sur les cartographies 3D de la dose (DVM) et qui permet une analyse statistique par voxel. Cette méthodologie appliquée à la région pelvienne a permis d'étudier la relation entre la dose locale et la toxicité dans la vessie et l'urètre. Tout d'abord, nous présentons la méthodologie de recalage rigide multi-organes qui permet la mise en correspondance anatomique de toute la population dans un seul espace commun de

référence. Ensuite les distributions des doses son déformées élastiquement, en suivant la transformation préalablement calculée, pour ensuite les comparer voxel par voxel permettant l'identification des sous-régions où des différences statistiquement significatives de la dose existent. Les sous-régions sont propagées à l'espace original de chaque patient et les respectifs DVH des sous-régions sont analysés et comparés aux DVH de la vessie. Enfin, la généralisation et la capacité prédictive de ces modèles ont été évaluées sur une population externe et indépendante.

- Le chapitre 5 évalue plusieurs algorithmes d'apprentissage automatique pour la prédiction de la toxicité urinaire en utilisant des données dosimétriques (DVH) et cliniques. Dans cette étude exploratoire, la performance de ces classifieurs a été évaluée sur la vessie en totalité afin de les comparer aux modèles traditionnels. Les problèmes de déséquilibre de la base de données ont été abordés à l'aide de quatre techniques de sur-échantillonnage synthétique différentes.
- Le chapitre 6 présente une discussion générale sur les principaux résultats et les contributions de cette thèse suivie des perspectives et des propositions sur l'orientation des études futures dans la prédiction de la toxicité urinaire après irradiation prostatique.

Les principales contributions de cette thèse sont donc :

- le développement d'une méthode de segmentation multi-atlas pour identifier l'urètre prostatique sur les images CT de planification
- une méthodologie pour cartographier avec précision les distributions de doses 3D à travers une population d'individus permettant d'effectuer des comparaisons par voxel des distributions de dose 3D
- l'identification et la validation de sous-régions spécifiques liées aux symptômes dans la perspective de réduire le risque de toxicité lors d'une planification spécifique-patient.
- l'ouverture méthodologique vers une amélioration de la prédiction de la toxicité urinaire à l'aide de méthodes d'apprentissage automatique et d'augmentation de données.

Introduction

Prostate cancer is the second most frequent malignancy among men and the fifth leading cause of cancer death worldwide. In 2018, this amounted to 1.3 million new cases and 359,000 deaths around the world from this disease. In France, prostate cancer is the most frequent type of cancer in men. The estimated number of cases for 2018 is 65,000 and the estimated number of deaths is 9,000.

A curative treatment modality for prostate cancer, which has emerged as a clinical standard, is external beam radiotherapy (EBRT). The aim of this technique is to deliver, over several treatment sessions, a high dose of ionizing radiation to the target volume, namely the prostate and the seminal vesicles. The proven benefit of increasing the dose to the prostate, in terms of local control and overall survival, is confined by the presence of healthy organs-at-risk (OAR), such as the bladder and the rectum. Radiation-induced damage to healthy tissues may result to significant adverse events of urinary, rectal or sexual nature.

Advances in technology and imaging, over the past decades, together with the introduction of modern treatment techniques, such as intensity-modulated radiotherapy (IMRT) and image-guided radiotherapy (IGRT), enabled dose escalation by geometrically shaping the dose beams around the target, whilst reducing the volume of normal tissues exposed to high radiation doses. While gastrointestinal (GI) toxicity has been relatively reduced by the introduction of highly conformal image-guided radiotherapy techniques, genitourinary (GU) toxicity rates remain relatively stable. As stated by the Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) project, a similar dose-response to that of GI toxicity is far from being established for GU toxicity.

Understanding the dose-volume relationship with urinary toxicity and developing reliable predictive models is of paramount importance. Predicting toxicity probability may be useful for informing the patients about possible treatment complications and helping clinicians to better tailor the treatment to patient's characteristics with the objective to decrease the risk of side effects. However, urinary toxicity modeling is a challenging issue, not only due to the variety of associated irritating or obstructive symptoms, but also owing to the limitations of dose descriptors and difficulties identifying the regions potentially responsible for those symptoms. The bladder, for example, presents important inter-fraction shape variations, causing geometric and dose uncertainties that limit the possibility of accurately modeling the dose-volume response concerning urinary toxicity. Another potentially critical structure is the intra-prostatic urethra which remains, however, largely unexplored in the literature due to inherent difficulty to be identified on the planning computed tomography (CT) scan. Although there is evidence in prostate cancer brachytherapy that some urinary symptoms are related to urethra damage this has not yet been shown in EBRT. Quantifying the delivered dose to the urethra may therefore improve our understanding of urinary toxicity or at least some of the related symptoms, , but requires the identification of this organ in the planning CT images.

Several studies used the whole bladder's dose-volume histogram (DVH) in an attempt to explain radiation-induced toxicities resulting in multiple DVH recommendations without any clear consensus. Predictive models of toxicity commonly based on DVH, reduce the 3D dose distribution within the organ to a unidimensional representation of the dose-volume relationship. The spatial information is, therefore, lost, ignoring the local variability of the 3D dose distribution, which may limit the prediction capabilities. DVH models assume that the organ is homogeneous in terms of radiosensitivity and functional status. In fact, some individual manifestations of urinary dysfunction may be linked to the irradiation of specific regions of the lower urinary tract, namely the urethra and the bladder neck, in which case DVH analysis of the entire bladder may not be sufficient to explain urinary toxicity.

This hypothesis of heterogeneous intra-organ radiosensitivity, in various anatomical sites, has induced a methodological evolution from the global, whole-organ-based philosophy towards more sophisticated predictive models that integrate local spatial descriptors of the dose distributions. With respect to urinary toxicity, analysis of the local dose distribution at lower spatial scales, may improve our understanding of urinary symptomatology after RT. To date only a small number of studies have sought to investigate the local dose-volume-effect in the bladder providing some evidence of quasi-heterogeneous intra-organ radiosensitivity while the role of urethra remains in obscurity. Unraveling the relationship between the local dose to the bladder and the urethra and urinary toxicity, may be translated into recommendations for treatment planning in clinical practice.

Going beyond the global, whole-organ-based models towards more local, sub-organ approaches, this thesis aims to improve our understanding of radiation-induced urinary side-effects and ameliorate the prediction of urinary toxicity following prostate cancer radiotherapy. More specifically, the objectives of this thesis are:

1. To assess the contribution of urethra damage to urinary toxicity.
2. To evaluate the involvement of specific bladder and urethra sub-parts in urinary toxicity.
3. To investigate the potential of further increasing prediction by exploiting data through new machine learning methods.

In order to assess the contribution of urethra damage to urinary toxicity, it was necessary to identify this structure on the planning CT-images of patients treated with EBRT. In this regard, we devised a multi-atlas segmentation approach to accurately identify and segment the otherwise undetectable urethra, on CT images. The atlas consists of a set of CT images of patients treated by brachytherapy, where the urethra is visible thanks to the use of an urinary probe. For a new patient, geometric features are extracted from the image to be segmented and compared to the features from the atlas database. The urethra for this new patient is, then, defined by combining the urethras of the most similar atlases in a weighed-fusion process.

Different approaches can be used to evaluate the dose-toxicity relationships in lower spatial scales with the objective of identifying anatomical regions correlated with urinary toxicity that are potentially more predictive than the dose to the whole bladder. First, a dose-surface map (DSM) analysis of the dose to

the surface of the bladder was performed to explore the local dose-response relationship for different urinary symptoms. The results were compared with previous DSM studies in order to evaluate the impact of different cohorts in the identification of radiosensitive subregions. This analysis, limited to the organ surface, was followed by the exploration of the entire 3D planning dose distribution in the bladder and the urethra without making any prior assumptions regarding the location of regions at risk. A novel approach based on 3D dose-volume map (DVM) analysis was thus proposed, allowing the investigation of the local dose-effect relationship at voxel level. The different pre-processing, including: i) an accurate spatial normalization to a single coordinate system, ii) the mapping of the doses to be analyzed and, iii) a reliable methodology to perform local statistical analysis, were thoroughly validated.

Urinary toxicity prediction may be improved not only by identifying local dose descriptors that are more predictive than global (whole-organ) dose descriptors but also by reliably analyzing the plethora of complex available data, including dosimetric, clinical and biological data. The vast majority of existing studies are based upon traditional regression approaches (e.g. logistic or Cox regression). These procedures, although commonly used due to their interpretability (P-values and confidence intervals), they may overlook the interactions among variables, resulting in modest prediction performance. Studies suggest that more contemporary strategies show promising results and perform better compared to conventional predictive modelling methods. Machine learning techniques can potentially increase toxicity prediction as they rely on previous informative examples. Nevertheless, a multitude of methods are emerging without a clear advantage of their use in this context as they application is not straightforward. A comparative study between different techniques was performed to provide some insights on the advantages and disadvantages of different classifiers in particular for urinary toxicity prediction.

This thesis is structured based on works published in international journals and conferences, works submitted and works in preparation for publication, which have all been co-authored. The thesis is divided in four parts and contains six chapters organised as follows:

Part I (Chapter 1)

- Chapter 1 presents the clinical context and the state-of-the-art predictive modeling for urinary toxicity following prostate radiotherapy. We start with a brief description of prostate cancer and the available treatment options. We will present more precisely the different techniques for radiotherapy and their limitations as well as the common side-effects emphasizing on the urinary symptoms. Then, we present the traditional 1D DVH-based models and we discuss their limitations. This is followed by a description of two emerging methodologies for analyzing the dose distributions at lower spatial scales, based on the construction of 2D dose-surface maps (DSM) and 3D dose-volume maps (DVM). Finally, the objectives of the thesis are explained.

Part II (Chapter 2)

- Chapter 2 presents a methodology for automatically segmenting the intra-prostatic urethra on the planning CT image of EBRT-treated patients. First, the entire framework of the multi-atlas-based segmentation method is detailed, from the atlas construction to the final urethra

segmentation. Then, the method's accuracy is evaluated with respect to the ground truth urethra and compared with existing surrogate models for estimating urethra's position. The dose to the urethra is also computed and compared to the dose delivered to the prostate.

Part III (Chapter 3 and Chapter 4)

- Chapter 3 describes a methodology for spatially characterizing the dose distribution on the bladder surface using DSMs. The first goal of this chapter is to improve the existent approaches for analyzing DSMs and propose a methodology that would allow the exploration of the entire bladder surface. The second goal is to identify symptom-related sub-surfaces of the bladder that are potentially more predictive than the whole bladder DVH. The third goal is to compare the results with previous bladder DSM studies and estimate the level of agreement between these studies and our observations.
- Chapter 4 presents a methodology for analyzing DVMs via voxel-wise comparisons, in order to investigate the local dose-effect relationship in the bladder and the urethra. First we present a robust multi-organ non-rigid registration strategy for anatomically aligning the population and propagating the dose distributions to a common space. Then, the DVMs of patients with and without toxicities were compared voxel-by-voxel allowing the identification of symptom-related subregions where statistically significant dose differences exist. The subregions are propagated to the native space of each patient where sub-region DVHs where computed and their discriminative power with respect to the DVH of the whole bladder was evaluated. Finally, the generalizability and the predictive capabilities of these models were assessed through external validation on a large, independent population.

Part IV (Chapter 5)

- Chapter 5 evaluates several machine learning algorithms for prediction of urinary toxicity using dosimetric and clinical data. The performance of these classifiers was evaluated on the original unbalanced dataset and also using four different synthetic oversampling techniques.

- Chapter 6 includes a general discussion on the main results and the contributions of this thesis, followed by the perspectives and suggestions on the direction of future studies in the prediction of urinary toxicity after prostate cancer radiotherapy.

The main contributions of this thesis are therefore:

- the development of a multi-atlas segmentation method for identifying the prostatic urethra on the planning CT images
- a methodology for accurately mapping the 3D dose distributions across a population of individuals allowing to perform voxel-by-voxel comparisons of the 3D dose distributions
- the identification and validation of specific symptom-related subregions with the perspective of reducing the risk of toxicity though patient-specific treatment planning.
- paving the way for further increasing the prediction of urinary toxicity using machine learning methods and data augmentation techniques.

Part I

Clinical context, state-of-the-art models and objectives

1 Clinical and scientific context and problem definition

This chapter presents the clinical and scientific context of this work. We will first discuss some general points about prostate cancer and the available therapeutic options and more specifically the external beam radiotherapy. We discuss its principle and limitations. We continue with a description of the possible side-effects emphasizing on the urinary symptoms. In the second part we describe the state-of-the-art predictive modeling for urinary toxicity following prostate radiotherapy. We first present the traditional whole organ DVH-based models and we discuss their limitations. Next, we describe two emerging methodologies for analyzing the dose distributions at lower spatial scales allowing to overcome some of the limitations of DVH-based models. These approaches are based on the construction of 2D dose-surface maps (DSM) coupled with pixel-wise comparisons and 3D dose-volume maps (DVMs) combined with voxel-wise comparisons. Their methodological challenges are also discussed. Finally, the motivation and objectives of this thesis are explained.

1.1 Prostate cancer

1.1.1 Prostate and prostate cancer

The prostate is a gland formed of both muscular and glandular tissues, surrounded by a 3-4 mm thick capsule. As part of the male reproductive system, prostate's most important function is the production of a fluid, containing several enzymes such as prostate specific antigen (PSA) that, together with sperm cells from the testicles and fluids from other glands, makes up semen. Prostate is located in the pelvic cavity, anterior to the rectum and at the base of the bladder, surrounding the prostatic urethra (the portion of the urethra that runs within the prostate), as shown in Figure 1.1. The seminal vesicles, two irregularly-shaped glands, are connected to the base of the prostate. Various nerves located around the prostate are implicated in the erection mechanism.

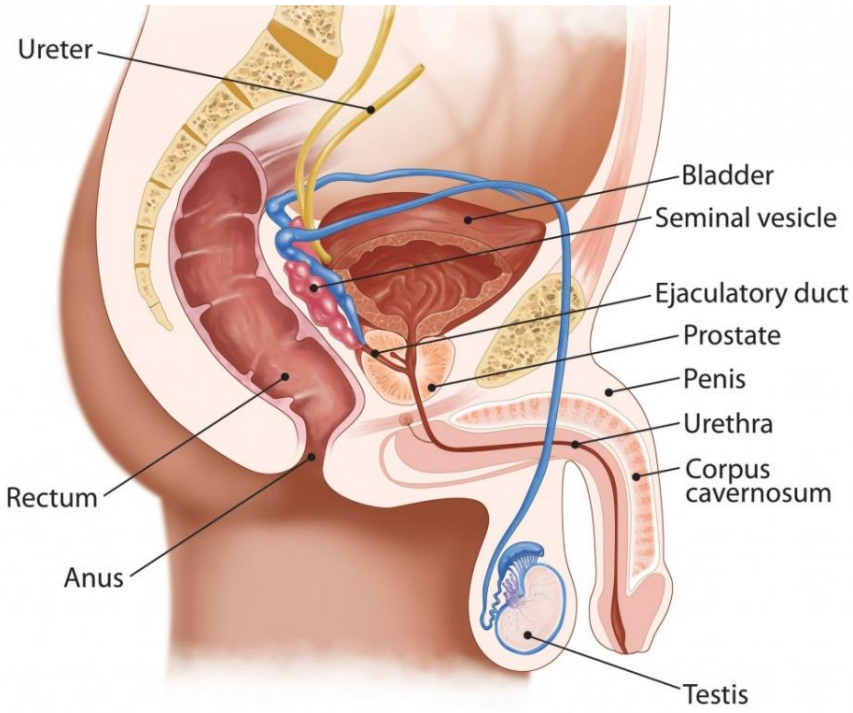


Figure 1.1 Illustration of the anatomy of the male reproductive system
(source: www.bladderclinic.com.au)

Prostate size changes through life, growing mostly during puberty, due to the rise of male hormones. For a healthy adult the average prostate dimensions are 4 cm in width and 3 cm in height and weight 20g. Around 80% of men over the age of 40 will undergo prostate hypertrophy, possibly compressing the urethra, leading to urinary dysfunctions. This hypertrophy, commonly called benign prostatic hyperplasia, is part of the natural evolution of prostate, but it can also be a sign of cancer.

Prostate cancer is the second most frequent type of cancer among men and the fifth leading cause of cancer death worldwide [1,2].In France, it is the most frequent type of male cancer with estimated

number of cases for 2018 being 65,000 and the estimated number of death 9,000. About 42% of fifty-year-old men have developed histological evidence of cancerous cells in the prostate; 9.5% of these will develop an aggressive form of cancer, and 2.9% will die from prostate cancer [3]. Medical treatments are therefore needed to prevent or delay the tumor from spreading outside the prostate. For a complete review on prostate cancer diagnosis, staging and treatments, see guidelines of the European Association of Urology (EAU) [4,5].

1.1.1.1 Diagnosis and staging of prostate cancer

Prostate cancer can be diagnosed through digital rectal examination (DRE) of the prostate and a blood test to measure the levels of PSA. An annual checkup is advised for men aged between 50 and 75 years. DRE is the standard way to define texture, shape, size and tenderness of the prostate gland. It is simple and complication-free, but subjective as it depends on the examiner [6]. PSA is a glycoprotein produced in the prostate and is used as marker for prostate cancer. Most healthy men have PSA levels below 4 ng/ml, and it usually goes above 4 when prostate cancer develops. However, DRE and PSA test have relatively low sensitivity, and they do not differentiate between aggressive and indolent disease [7]. Although the limited prognostic value of PSA, measurements of free-to-total (f/t) PSA ratio has improved the identification of patients with aggressive disease [8].

When a blood test shows high PSA levels or when an abnormal prostate or a lump is found through DRE examination, prostate biopsy might be performed. Definitive diagnosis is only confirmed after histopathological verification of adenocarcinoma in prostate biopsy specimens. The Gleason score is used to evaluate the prognosis of men with prostate cancer using samples from a prostate biopsy. Gleason grade tells how aggressive the cancer is, in other words, how likely it is to grow and spread outside the prostate capsule.

In the early stage, the cancer is located within the prostate but, with disease progression, it can expand to neighboring organs and tissues as well as more distant organs such as the lymph nodes and the bones. About 4% of prostate cancer patients already have metastases at the moment of diagnosis. Tumor extension is usually expressed as being at certain T stage using the tumor, nodes and metastasis (TNM) classification system according to the Union for International Cancer Control (UICC) [9]. There are four T stages with various subcategories indicating tumor size and location:

- T1: tumor present, but not detectable with imaging or clinically:
  - T1a: Tumor found in less than 5% of prostate tissue resected (for other reasons);
  - T1b: Tumor found in more than 5% of prostate tissue resected;
  - T1c: Tumor found in a needle biopsy performed due to an elevated PSA level;
- T2: tumor can be felt (palpated) on DRE examination, but it still appears to be confined to the prostate:
  - T2a: Tumor is in half or less than half of one of the prostate glands' two lobes;
  - T2b: Tumor is in more than half of one lobe, but not both;

  - T2c: Tumor is in both sides of the prostate lobes but still inside of the prostatic capsule.
- T3: tumor has spread throughout the prostatic capsule (if it is only half-way through, it is still classed under T2):
  - T3a: Tumor has spread through the capsule but not to the seminal vesicles;
  - T3b: Tumor has invaded one or both seminal vesicles.
- T4: the tumor has invaded other nearby structures.

Among the most commonly used methods for risk stratification, the most commonly used is the grouping developed by D'Amico et al. [10] This score is based on Gleason score, T stage and PSA levels, to classify patients diagnosed with prostate cancer into low, intermediate and high risk categories as follows:

- Low risk: T1 to T2a stages, Gleason score ≤ 6, PSA ≤ 10 ng/mm.
- Intermediate risk: T2b stage or Gleason score = 7 or 10 ng/mm ≤ PSA ≤ 20 ng/mm.
- High risk: T3 to T4 stages or Gleason score ≥ 8 or PSA > 20 ng/mm.

Once the stage has been determined, different treatment options are considered.

1.1.2 Treatment options for prostate cancer

Different therapeutic options for localized prostate cancer exist that can be used in isolation or in combination. Depending on the age of the patient, his life expectancy, the aggressiveness of the cancer or the advantages /disadvantages of the therapies, the most appropriate treatment is chosen considering the therapeutic objective. This may include suppressing the tumor or metastases, reducing the risk of recurrence and slowing the tumor progression [11]. The main treatment modalities are active surveillance, surgery, radiotherapy (external or brachytherapy), and medication treatments such as hormonal therapy. More recent techniques like High Intensity Focused Ultrasound (HIFU) have also been developed. Hereinafter, the different therapeutic options are described according to their frequency of prescription.

- Radical prostatectomy (RP) is a surgical treatment for prostate cancer which consists of a total removal of the prostate gland and seminal vesicles, and may be accompanied by lymph node dissection. It is one of the standard treatments for localized prostate cancers with low and intermediate risk of recurrence. It can also be proposed in certain cases of localized high-risk, locally advanced cancers with lymph node involvement. Radiation therapy and / or hormonal therapy may supplement prostatectomy. The most common side effects of total prostatectomy are urinary incontinence, related to impaired functioning of the bladder and sphincter muscles, and erectile dysfunction. This treatment also implies a definite impossibility to ejaculate.
- Radiotherapy consists of delivering high doses of radiation to the cancer cells in order to prevent them from multiplying. Radiation can be administrated in two ways:
  - The source of radiation is located outside of the body (external), delivered in the form of photons (X-rays from linear accelerator machines), electrons, and more rarely other particles such as protons. It is one of the reference treatments for localized cancers at low and

intermediate risk. Coupled with hormone therapy, it is the recommended treatment modality for localized high-risk cancers and locally advanced cancers. It can also complete a prostatectomy to reduce the risk of recurrence. Irradiation of healthy tissues adjacent to the prostate can cause a variety of side-effects.

  - The source of radiation is located inside the body (brachytherapy) via the implantation of radioactive seeds within the prostate gland. These radioactive sources emit radiation that destroys the surrounding malignant cells. Because the gradient of dose drops sharply away from radioactive sources, brachytherapy is indicated for localized cancers with a low risk of recurrence. Side-effects are common due to prostate inflammation and urethra damage.

- Hormonal therapy, also called androgen deprivation therapy (ADT) acts by suppressing the levels of male hormones (androgens) which stimulate prostate cancer cells to grow. The main androgens are testosterone and dihydrotestosterone. Lowering androgen levels or stopping them from getting into prostate cancer cells often makes prostate cancers shrink or delay its' growth. However, hormonal therapy alone is not curative. Combined with radiation therapy, it is the gold standard for high-risk localized prostate cancers. It is also used in case of metastatic cancers. The most common side effects of hormonal therapy are hot flushes, erectile dysfunction, changes in physical appearance, and osteoporosis.
- Deferred treatment (active surveillance/watchful waiting): In some cases, treatment of localized prostate cancer may be deferred to avoid toxicity due to other treatment. There are two distinct strategies for conservative management that aim to reduce overtreatment: active surveillance and watchful waiting. The aim of active surveillance is to reduce overtreatment in patients with clinically confined, very-low-risk cancer, without renouncing curative treatment, as happens with watchful waiting [12]. Active surveillance is only proposed for highly selected low-risk patients. The surveillance of cancer evolution is carried out through frequent DRE, blood tests to measure the PSA levels, biopsies and MRI. Watchful waiting is recommended when cancer progresses slowly, or for older men with a high incidence of comorbidities and other causes of mortality [13].
- High-intensity focused ultrasound of the prostate (HIFU) consists of focused ultrasound waves to the prostate allowing high sound pressures to be delivered to a focal point in order to kill tumor cells via heating and cavitation [14]. The main treatment device is the Ablatherm®, which combines an ultrasound transducer and an echography probe introduced to the rectum during the operation. The goal of HIFU is to heat malignant tissues above 65°C so that they are destroyed by coagulative necrosis, under general or spinal anesthesia. The procedure is time-consuming, with about 10 g prostate tissue treated per hour. This technique is advised for patients aged over 70 years with low-risk cancers, but is mostly used in the case of local recurrence following radiotherapy [15].

In this thesis, we will focus on external beam radiotherapy for the treatment of prostate cancer. This modality is detailed in the following section.

1.2 External beam radiation therapy (EBRT)

Over the past few decades, external beam radiotherapy (EBRT) has advanced by leaps and bounds and is considered today a standard definitive treatment option for localized prostate cancer. Improvements in imaging and computing have led to a number of technical advances in planning and delivery of the treatment (Figure 1.2). These advances have permitted more precise and conformal delivery of doses of radiation to the prostate, thereby improving the therapeutic ratio.

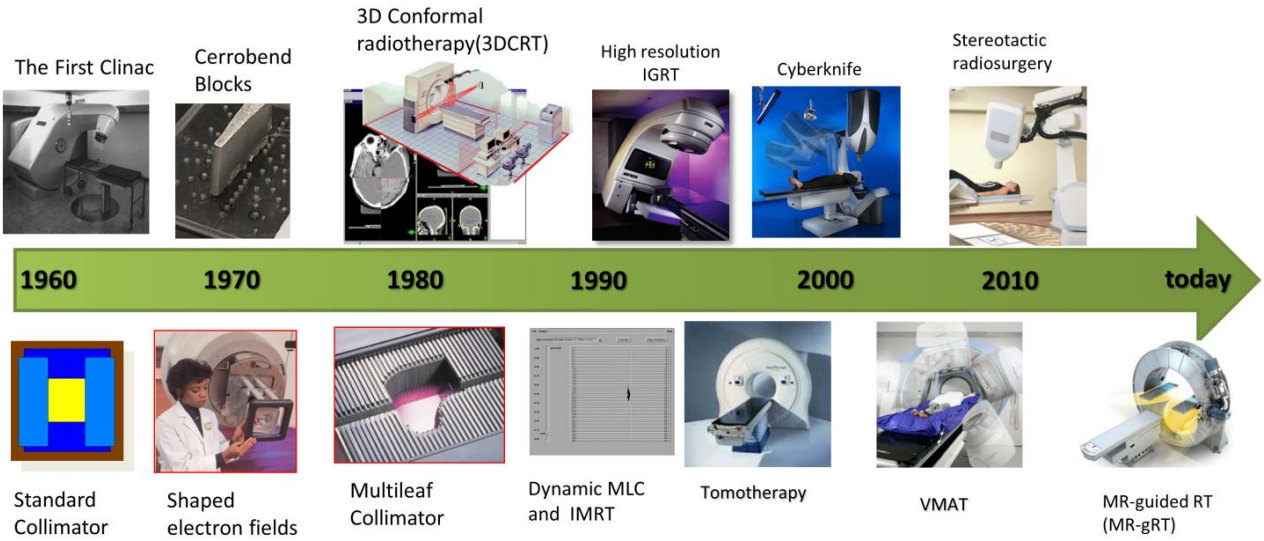


Figure 1.2 The evolution of EBRT

More than two thirds of the patients diagnosed with this prostate cancer are treated with EBRT, often combined with a concomitant treatment (e.g. surgery or hormonal therapy). During EBRT, ionizing rays charged with high energy photons are delivered to the tumor aiming to maximize local control whilst sparing neighboring organs (mainly the rectum and the bladder). An optimal irradiation dose is, therefore, the cornerstone of a successful treatment in terms of local control of the disease and the overall survival of the patient.

The radiation dose is expressed in Grays (Gy), which indicates the amount of radiation energy absorbed by 1 kilogram of human tissue (1 Jule/Kg). The total prescribed dose is delivered in several sessions in order to allow healthy tissue to recover and increase tumor damage by reoxygenation [16]. For example, if 80 Gy are prescribed to the prostate in a convenient treatment schedule, treatment can be completed in 40 sessions in a period of eight weeks, with a fractionation scheme of two Gray per fraction (2 Gy/fraction).

Before the treatment, the radiotherapy team will carefully plan the optimal radiation scheme (dose, fractionation and angles of the beams). This process, known as treatment planning, starts by acquiring a computer tomography (CT) scan of the anatomical region to be treated, namely the pelvic region for prostate cancer irradiation. The different structures shown on the CT scan are then delineated by an expert (e.g. prostate, seminal vesicles, bladder and rectum). This information is imported to a treatment

planning system (TPS), which is a software used to generate the irradiation scheme. The prescribed dose is a three-dimensional map that relates every point within patient with a level of dose. The following step is then to set up the linear accelerator to deliver the planning dose using the selected fractionation scheme, whereby each treatment fraction lasts only a few minutes. During the treatment the patient is placed on a table that is below the linear particle accelerator (Figure 1.3). Fiducial markers can be implanted on the patient's skin or prostate to align him at each treatment session.

1.2.1 Radiation techniques

For intermediate and high risk disease, radical EBRT is standard practice. There are two main types of irradiation techniques: three-dimensional conformal radiotherapy (3D-CRT) and intensity modulated radiotherapy (IMRT). IMRT with image-guided radiotherapy (IGRT) is nowadays considered as the gold standard for external radiotherapy.

Figure 1.3 Patient positioned on the linear accelerator
(source:www.cancer.net)

THREE-DIMENTIONAL CONFORMAL RADIOTHERAPY (3D-CRT)

In the past, radiation treatment matched the size of the tumor in a “box”-shaped region, meaning that adjacent healthy tissue was unavoidably included in the radiation field. Advances in imaging technology have made it possible to locate and treat the tumor more precisely. In case of prostate cancer, irradiation treatment plan is usually defined by five to nine convergent beams, which conform the target volume via the modulation of a multileaf collimator (MLC) located at the linear accelerator output (Figure 1.4). This form of irradiation is known as 3D conformal radiotherapy (3D-CRT). 3D-CRT uses a planning CT to focus precisely on the tumor region, while trying to spare the healthy surrounding tissue. This exact targeting makes it possible to use higher levels of radiation in treatment, which are more effective in shrinking and killing tumors.

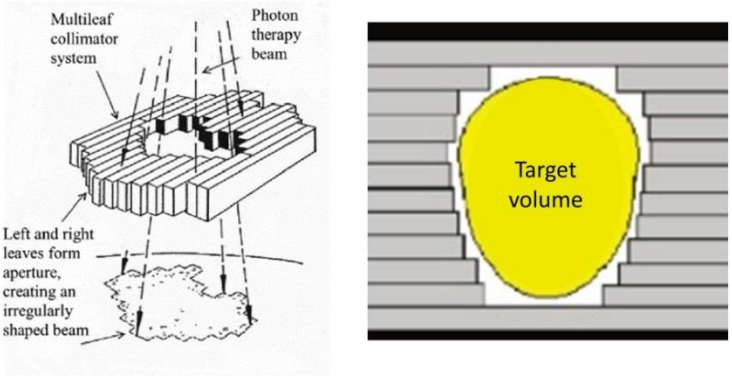


Figure 1.4 Schematic of multileaf collimator (source:[17])**; Individually positioned leaflets shape the beam to the target**

INTENSITY MODULATED EXTERNAL-BEAM RADIOTHERAPY (IMRT)

Intensity-modulated radiation therapy (IMRT) has been considered the most successful development in radiation oncology since the introduction of CT to treatment planning [18,19]. Over the past two decades, IMRT has supplanted 3D-CRT thanks to its high conformity which facilitates dose escalation and improves local control without significantly increasing the risk of morbidity [20,21]. These features make it particularly suitable for the treatment of diseases that involve high rates of local recurrence and complications [22]. A comparison of planning treatments using 3D-CRT and IMRT is shown in Figure 1.5.

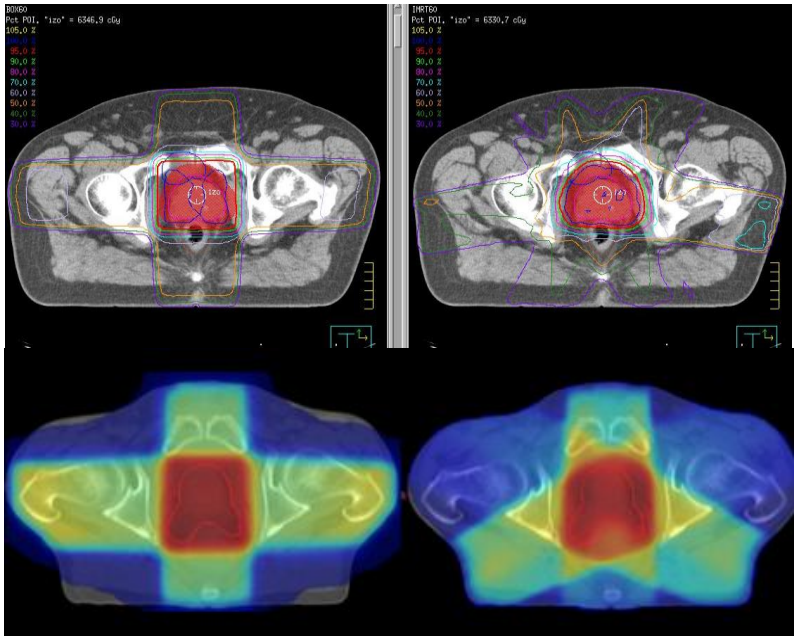


Figure 1.5 Treatment plan for prostate cancer: 3D-CRT (left side) and IMRT (right side)

On the top is illustrated an example of CT image and the design of the irradiation field surrounding target volume and on the bottom the resulting planning 3D dose distributions for the given irradiation field (source: www.mistir.info).

During the irradiation process with IMRT, the fluence (photon amount per surface unit) is no longer homogeneous but is modulated using a MLC, by continuously adapting the beams to the shape of the target volume [23]. This allows for a more conformational dose distributions to be delivered within the

treatment field, which may fit complex structures. This technique is particularly useful to spare neighboring organs at risk. The dose is derived by an inverse-planning software that starts from the end-product (i.e. desired dose) and ends with the input (i.e. fluence profile) [24]. The desired dose is represented as a mathematical cost function to be optimized with some constraints. The input parameters of this function include leaf positions, some weights, and the fluence matrix.

IMAGE-GUIDED RADIOTHERAPY (IGRT)

Different anatomical references, such as skin markers, can be used to allow for reproducible patient positioning through the treatment sessions with respect to the irradiation field. However, as the patient might experience anatomical changes, such as weight loss, the effectiveness of these markers is limited and may lead to setup errors [25].

Portal imaging systems have developed with the introduction of linear accelerators [26]. A digitally-reconstructed radiograph (DRR) (based on the planning CT) is generated by simulating irradiation in a process that mimics the geometry of the treatment. This image allows for bone structure alignment and for patient repositioning, which is intended to decrease the amount and frequency of setup errors. However, repositioning the patient according to the bone structures does not completely solve the problem of geometric uncertainties. Indeed, from one irradiation session to another or even during a session in some cases, anatomical variations can occur even in a fixed bone reference system. These variations may correspond to displacements / deformations of the target or the organs at risk. In the first case, the risk is to under-irradiate the tumor and thus reduce local control. In the second, organs at risk can be over-irradiated, increasing the risk of complications [27].

Because of these targeting uncertainties during treatment, the need to precisely locate the clinical target and / or OARs, has led to the appearance of new imaging devices integrated into the accelerator. In prostate cancer radiotherapy, the term IGRT generally implies the use of imaging that allows tumor localization, as opposed to the use of imaging that allows visualization of bone structures only. The localization of the tumor can be direct, via 3D imaging showing the soft tissues, or indirect when using markers implanted in the prostate [28]. The most common imaging modality used for this purpose is cone beam computed tomography (CBCT) and is present on all modern accelerators. This is based on an X-ray source of energy kV (more rarely MV) and a 2D detector mounted on the arm of the linear accelerator opposite the source. The system rotates around the patient, and the acquisition of a large number of 2D projections allows the reconstruction of a 3D image that can be compared to the CT. More recently, there has been a shift towards the yield of real-time motion data using non-ionizing radiation modalities such as electromagnetic transponders (EMT) and four-dimension (4D) transperineal ultrasound (TPUS) [29].

IGRT has played an important role in the evolution of radiotherapy by reducing the uncertainty of the exact position of tumor and OARs, and improving the precision of the delivered dose. IGRT has also enabled the development of adaptive radiotherapy based on the assessment of information obtained from daily images [25].

EVOLVING TREATMENT TECHNIQUES

Volumetric modulated arc therapy (VMAT), is a novel type of IMRT technique, which can achieve highly conformal dose distributions with improved target volume coverage and sparing of normal tissues, compared with conventional radiotherapy techniques [30]. In contrast to static field IMRT, with VMAT the radiotherapy machine rotates around the patient in an arc shape. The machine continuously reshapes and changes the intensity of the radiation beam as it moves around the body. Giving the radiotherapy in this way makes it more accurate and shortens the treatment time.

Another technique, the CyberKnife Robotic Radiosurgery System is a form of targeted radiation therapy known as stereotactic body radiotherapy (SBRT). The CyberKnife machine has a robotic arm that moves around the treatment couch to deliver doses of radiation from different angles. It continuously identifies the exact location of the prostate and makes active corrections for any movement of the prostate throughout the course of the treatment by tracking fiducial markers into the organ [31]. The CyberKnife may work on a moving target - the prostate - without harming surrounding areas. As a result, the procedure may be more comfortable for the patient and radiation delivered may be more accurate. It also allows to treat multiple tumor sites at the same time, making it useful for areas of cancer spread.

Magnetic resonance imaging (MRI) machines and linear accelerators have been used separately in the care of cancer patients for years. A new era in image-guided technology is rapidly evolving with the integration of an onboard MRI with a radiotherapy treatment system, with the emergence of magnetic resonance-guided radiation therapy (MRgRT). The integration of these two powerful technologies into one machine allows radiation oncologists to track and monitor the movement of tumors during radiation delivery, and potentially track radiation response in real-time, without any added radiation dose to the patient. We expect significantly decreased target margin and increased target dosage by using online adaptive MRI-based linac in the future [32].

Although most of the current practice of clinical radiotherapy utilizes photon beams, particle therapy and, in particular proton therapy, has recently gained interest for the treatment of prostate cancer. Given the physical properties of photons, normal tissues surrounding the target volume still receive a substantial amount of unwanted dose. The theoretical advantage of proton therapy in reducing radiation dose to normal tissue is based on its intrinsic radiation properties. A heavy, charged particle such as a proton deposits most of its dose at a prescribed depth in the body with a rapid dose fall-off beyond this point. This peak of energy delivery is commonly referred to as the Bragg peak. This means that proton therapy can reduce radiation dose delivered to tissues beyond the target compared to photon radiation. Although proton and photon particles are deemed to be similarly effective in prostate cancer treatment, the interest in proton treatment stems from its potential for lower toxicity and therefore further dose escalation [33]. Current evidence has not clearly demonstrated that this theoretical difference translates to improved patient outcomes compared to IMRT. However, as proton technology continues to evolve, this is an active area of research.

1.2.2 Dose-effect relationships in local control

There is a clear relationship between prostate dose and tumor control [21,34–43]. Tumor control can be expressed in terms of two main types of recurrence, namely biochemical and clinical recurrence. As shown in Table 1.1 [44], six large randomized trials have demonstrated that increasing the dose to the prostate to 74-80 Gy with standard fractionation (1.8–2 Gy), may improve biochemical recurrence-free survival and disease-specific survival [21,34–36,42,43].

Table 1.1 Randomized controlled trials evaluating the efficacy of radiation dose escalation for prostate cancer
(source: [44])

Trial	Number of patients	Dose comparison (Gy)	Outcome
MD Aderson [36]	301	70 versus 78	78% versus 59% freedom from biochemical or clinical failure
PROG 95-09 [42]	393	70.2 versus 79.2	32% versus 17% 10-year biochemical failure
MRC RT01 [21]	843	64 versus 74	43% versus 55% 10-year biochemical recurrence-free survival
Dutch [43]	664	68 versus 78	54% versus 64% freedom from failure
GETUG 06 [34]	306	70 versus 80	39% versus 28% biochemical failure
RTOG 0126 [35]	1532	70.2 versus 79.2	35% versus 20% 8-year biochemical failure

According to the International Commission on Radiation Units and Measurements (ICRU), there are five target volumes defined in radiotherapy: gross tumor volume (GTV), clinical target volume (CTV), planning tumor volume (PTV), and treated and irradiated volume. CTV represents the main target volume to be irradiated in radiotherapy, receiving the prescribed dose. The CTV contains the primary tumor (called GTV) and/or sub-clinical malignant tissue that have to be eradicated in order to control the tumor. 3D safety margins around CTV define the planning target volume (PTV). This target volume is to ensure CTV dose coverage in case of inter- or intra-fraction geometric variations.

Planned dose distributions are usually evaluated by means of dose-volume histograms (DVH). The DVH, as show in Figure 1.6, is a graph representing the volume of a structure receiving a dose equal to higher than a given value (cumulative DVH). It allows volumetric quantification of the dose distribution but it does not provide any information about the spatial distribution of the dose within the organ or interest.

Regardless of the techniques and their degree of sophistication, side effects related to the healthy tissue irradiation during the treatment, are always observed, in particular in a context of dose escalation. Understanding the toxicity is therefore crucial to improve the reliability of the treatment. In order to achieve local control and spare organs at risk, various recommendations have been proposed by different cooperative groups [45]. The constraints recommended by the French group, Groupe d'étude des tumeurs urogénitales (GETUG), are summarized in Table 1.2.

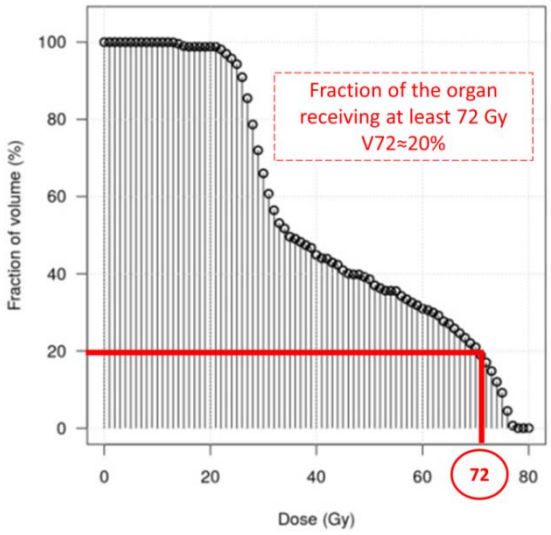


Figure 1.6 Representation of a Dose-Volume Histogram (DVH)

Table 1.2 GETUG dose-volume constraint recommendations for PTV and OARs in prostate cancer radiotherapy

Volume	Notation	Definition
Target volume (PTV)	$D_{min} > 90\%$	Minimum dose to PTV must be higher than 90% of the prescribed dose
	$V_{95} > 95\%$	The volume receiving at least 95% of the prescribed dose must be higher than 90% of the total volume
Bladder wall (7 *mm*)	$D_{max} < 80\ Gy$	The average dose to 1.8 cm^3must be always lower than 80 *Gy*
	$V_{70} < 50\%$	The volume receiving at least 70 *Gy* must be lower than 50%
Rectal wall (7 mm)	$D_{max} < 76\ Gy$	The average dose to 1.8 cm^3 must always be lower than 76 *Gy*
	$V_{72} < 25\%$	The volume receiving at least 72 *Gy* must be lower than 25%
Femoral heads	$V_{55} < 5\%$	For each femoral head, the volume receiving at least 55 *Gy* must be lower than 5%

1.3 Radiation toxicity following prostate cancer radiotherapy

Radiation triggers its therapeutic effect by damaging the DNA of actively dividing cells, causing division delay, reproductive failure and interphase arrest [46]. These consequences are more frequently encountered in rapidly dividing cells. However, it may not only affect malignant cells, but also adjacent normal tissue. Side-effects are usually secondary to fibrosis and progressive endarteritis that take place in poorly oxygenated submucosal and muscular tissues, which may lead to further tissue scarring [46].

Prostate cancer has a remarkably high incidence-to-mortality ratio, meaning a large part of men diagnosed with prostate cancer will die of other, unrelated, causes. Radiation therapy does, however, cause a wide range of side effects, that can be severe and cause temporary or permanent damage to the patient. As most patients survive early-stage prostate cancer after treatment, quality-of-life (QoL) outcome has emerged as an important factor to consider in treatment decisions. QoL refers to the

impact of disease and treatment on patients' well-being and physical, emotional and social functioning [47].

Radiation toxicity can manifest in a number of different ways, as summarized in Table 1.3, impairing the urinary, bowel and sexual functioning. The terms "side-effects" and "toxicity" are considered equal and exchangeable in this thesis. Side-effects from radiation therapy are classified as acute or late. Acute side-effects occur during treatment or up to six months after its completion, and usually resolve within four to six weeks. Late side-effects are observed from six months up to several years after completion of treatment and may be permanent [48].

Table 1.3 Side-effects of prostate cancer radiotherapy classified according to their frequency

Very Likely	Less likely but serious
• Tanning or redness of skin in treatment area • Rash, itching or peeling of skin • Temporary hair loss in the treatment area • Temporary fatigue, nausea or diarrhea • Abdominal cramps • Bladder irritation with a stinging sensation	• Injury to the bladder, urethra, bowel, or other tissues in the pelvis or abdomen • Intestinal or urinary obstruction • Erectile dysfunctions • Rarely, rectal bleeding that requires medication or burning/cutting of tissues to stop

In prospective clinical trials, physicians record patient symptoms at each follow-up visit. Afterwards, the toxicity events are graded using standard grading scales. These scoring systems include late effects normal tissues (LENT) / subjective, objective management (SOMA) (LENT/SOMA); Radiation Therapy Oncology Group and the European Organization for Research and Treatment of Cancer (RTOG/EORTC); and the common terminology criteria for adverse events (CTCAE). An additional effort must be made when using retrospectively collected data to match records from physicians' reports to a common terminology. For this reason, prospective studies are the basis of almost all clinical guidelines. Grade refers to the severity of the side-effect. The systems display Grades 1 to 5 with unique clinical descriptions of severity for each side-effect based on this general guideline:

- Grade 1: Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated;
- Grade 2: Moderate; minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental activities of daily living (*);
- Grade 3: Severe or medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care activities of daily living (**);
- Grade 4: Life-threatening consequences; urgent intervention indicated;
- Grade 5: Death related to adverse event.

(*) refers to preparing meals, shopping for groceries/clothes, using the telephone, managing money, etc; (**) refers to bathing, dressing and undressing, self-feeding, using the toilet, taking medications, and not bedridden.

Toxicity can also be rated by the patients, and is important because physicians often underestimate the impact of disease and treatment on patients' life [49]. Side-effects and their severity are measured using standardized questionnaires, which provide a more objective assessment of general and disease-specific domains [50] including well-being, vitality, fatigue, pain, general health status, global QoL, and life satisfaction [51].

As mentioned, EBRT may cause urinary and gastro-intestinal toxicities, as well as, sexual dysfunction. Sexual function after RT for prostate cancer is an important topic with serious impact on patient QoL. Although the specific mechanism by which radiation therapy reduces erections is uncertain, it has been suggested that radiation therapy does not damage the corporal nerves, but rather it causes vascular damage [35]. Small bowel or rectal irritation can manifest as abdominal cramping, diarrhea, fecal incontinence, proctitis (urgency, tenesmus), or rectal bleeding. Urogenital complications include urinary frequency/urgency, dysuria, urinary obstruction/retention, hematuria, urinary tract infection, and incontinence.

Understanding the dose-toxicity relationship is a central question for improving treatment reliability of radiotherapy treatment. This thesis work is exclusively devoted to investigating urinary toxicity and it is described in the following paragraph.

1.3.1 Genitourinary (GU) toxicity

Bladder, urethra, and the urinary sphincter are sub-parts of the lower urinary track (Figure 1.7) and their injury can result to manifestation of urinary symptoms. Radiation-induced injury of the urinary tract is a complex and debilitating complication and can lead to significant morbidity for the patient. The symptoms may be mild or severe, self-limiting, or progressive and may develop gradually or suddenly. The pathophysiology of urinary radiation injury is still not completely understood. The lesions are diverse pathologically including fibrosis, necrosis, atrophy, and vascular damage. The consequences may include a contracted and defunctionalized bladder, radiation cystitis with hematuria due to breakdown of the mucosa secondary to loss of supporting submucosal blood supply, ureteral and urethral strictures, sphincteric deficiency, as well as urinary fistulae[46,52,53].

In the literature, the rectum is an extensively studied OAR, and dose-volume predictors of late GI toxicity are established. These are commonly incorporated into radiotherapy protocols in clinical practice. However, for urinary toxicity, there is a significant paucity of knowledge [54]. On top of that, the incidence of moderate/severe urinary toxicities, which play a major role in QoL, increased, as a consequence of more aggressive treatments [45,55]. With the introduction of high-dose IGRT, the incidence of gastrointestinal toxicity has reduced to approximately 1% compared to the increase about 10% for genitourinary toxicity [56]. This is speculated to be related to the fact that in almost all cases the bladder neck and prostatic urethra are inevitably included in the high-dose region [57–60].

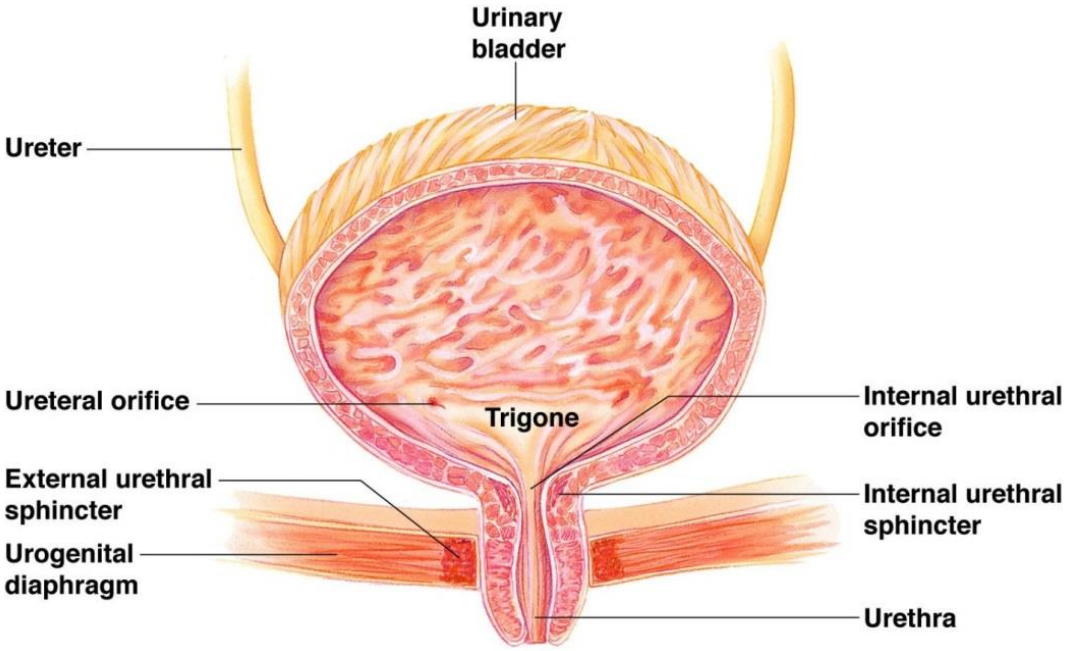


Figure 1.7 Anatomy of urinary bladder
(source: http://droualb.faculty.mjc.edu)

Indeed, existing evidence support that regions of the bladder neck, in particular the trigone, are associated with urinary toxicity. Another critical structure is the urethra, as it passes through the prostate and it is unavoidably irradiated with the total prescribed dose. In the literature of prostate cancer brachytherapy, there is strong evidence that some urinary symptoms are related to urethra damage [61–63]. However, such an effect has never been investigated in external-beam radiotherapy since this structure is not visible on CT images. Indeed, as it is shown in Figure 1.8, the prostatic urethra and the surrounding prostatic tissue share similar physical characteristics and the low contrast of the image does not allow the distinction of the two structures. However, quantifying the delivered dose to the urethra is crucial as it may improve our understanding of urinary toxicity.

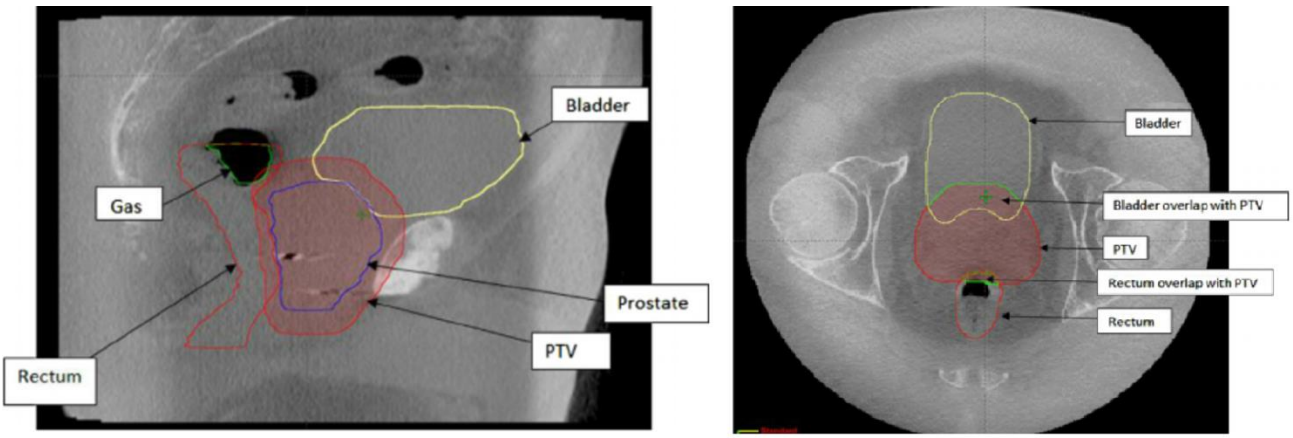


Figure 1.8 CT scan in sagittal (left) and axial (right) view and organ deligneations

Overall, genitourinary toxicity is currently perceived as the most relevant long-term problem following prostate radiotherapy and a major factor determining post-treatment QoL [64]. Given its growing significance, urinary toxicity has been poorly investigated. Unraveling the underlying dose-volume effect relationship and identifying patients at higher risk of toxicity, appears as a cornerstone in further definitions of constraints for personalized treatment planning.

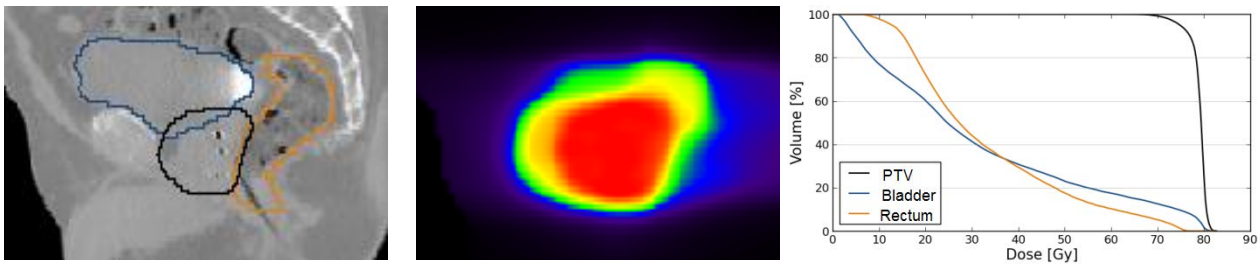
In the next section we will present the different methodologies that allow the exploitation of dosimetric data and have been employed in this thesis for developing predictive models of urinary toxicity.

1.4 Predictive models of toxicity: state of the art

Research groups around the world have attempted to study the potential of treatment optimization and individualization through the use of knowledge of the associations of treatment, clinical and dose factors with specific side-effects. It is generally acknowledged in these studies that radiation-induced side-effects are associated with a large number of factors that differ for each individual patient, with a great difficulty of inferring these complex models directly from current radiobiological knowledge.

1.4.1 DVH-based models: Global Organ Analysis of Dose-Effect Relationships

The availability of individual, 3-dimensional, dosimetric information permits the quantitative assessment of dose-volume relations for specific endpoints by investigating the correlations between dose-volume data and toxicity data. Parameters extracted from DVHs at the treatment planning stage include dose that a tissue received and volume irradiated. As explained before, the DVH is represented as a non-increasing function of the dose that matches any given dose value with the fraction of the organ volume receiving at least that dose. DVHs are essential decision-support tools for the evaluation of radiotherapy treatment plans (Figure 1.9).



(a) (b) (c)

Figure 1.9 Illustrations of DVH (c) calculated for PTV, bladder and rectum from manual organ delineations (a) and planning dose distribution (b)

Normal tissue complication probability (NTCP) models attempt to condense the dose-volume information into a number that expresses the risk of a certain toxicity. Most NTCP models are phenomenological and have the advantage of being characterized by few parameters (typically 3). There exist many different approaches to model NTCP in the context of radiotherapy outcomes, with Lyman–Kutcher–Burman (LKB) model [65,66] being the most commonly employed:

$$NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

where

$$t = \frac{gEUD - D_{50}}{m \cdot D_{50}}$$

$$gEUD = \left(\sum_i v_i \cdot D_i^{\frac{1}{n}} \right)^n$$

where i is the bin of the DVH corresponding to relative volume v_i that receives dose D_i. D_{50} is the value of dose corresponding to the 50% probability to induce normal tissue complication, the parameter m is inversely proportional to the slope of the dose-response curve and the parameter *n* accounts for the magnitude of volume effects and takes values between 0 and 1. For serial organs, such as the spinal cord or the rectum, resulting in $n \ll 1$, means that toxicity is mainly linked to high doses or hot spots. For parallel organs, such as parotid, lung, liver and kidney, $n \sim 1$ means that toxicity is associated to the mean dose to the organ.

Even if the NTCP-based models are common practice to evaluate the best plan, the values of the parameters used by the model are often questionable. For instance, the *n* value of the NTCP model, although relatively well identified for the rectal toxicity (range: 0.06–0.24), the *n* value for the bladder is still far from being established. Several studies suggest bladder as a serial organ [57,58,67–72], since side-effects, mainly urgency and obstruction, were related to the high dose delivered in specific bladder parts; primarily the trigone region. On the other hand, some studies found bladder as a mixed serial–parallel organ because severe urinary toxicity was related to both low and high doses delivered to the whole bladder [73–75]. The QUANTEC (QUAntitative Normal TissuE models in the Clinic) investigation [76] concluded prudently that both maximum dose and a relatively large irradiated bladder volume (50%) may correlate with bladder toxicity (Grade≥ 3 late RTOG).

Given the inconsistency of available data concerning radiation-induced toxicity, it is crucial to develop robust models with superior predictive performance. Dimensionality reduction, feature extraction strategies together with machine-learning methodologies, aiming at exploiting more available multimodal data, have emerged to overcome some of these issues, exhibiting promising prediction capabilities. Among them, Principal Component Analysis (PCA) was proposed to reduce the dimensionality of the DVH data and quantify the variability of DVH shapes [77]; functional data analysis [78] enabled the representation of the DVH as a curve rather than discrete measurements. Other machine-learning methods such as artificial neural networks [79,80], together with genetic algorithms and comparison with support vector machines [81] or random forest [82] have also been investigated, reporting competitive predictive results. In the context of urinary toxicity prediction, a review and comparison of six different machine learning techniques was presented by Yahya et al. [83], nevertheless without any consensus on the best algorithms, as the predictive power of the models was modest and endpoint-dependent.

1.4.2 Beyond DVH: Local Analysis of Dose-Effect Relationships

The DVH reduces the 3D dose distribution within an organ to a unidimensional and discrete representation of the dose-volume relationship. Several limitations arise when only scalar values or DVH are used within predictive models: i) A DVH is limited to a single organ, ii) 3D dose distributions may lead to the same DVH, iii) the information on the spatial distribution of dose is lost by merely considering the organ volume thus ignoring the local variations and the potentially heterogeneous intra-organ radiosensitivity, iv) correlation may exist between adjacent DVH bins. More broadly, urinary toxicity is a complex multiparametric phenomenon that may involve structures at different scales, from sub-organ parcels to large structures or regions whose response may additionally depend on individual radiosensitivity. These factors may explain the limited prediction capability of DVH-based models [84].

Although these models are continuously being improved and are bringing new insights into the understanding of dose volume effects from DVHs, the data collected through the whole radiotherapy treatment (clinical history, tumor stage, multimodal imaging, organ delineations, 3D dose distribution, intra-individuals changes, etc.) are very rich and on most occasions not thoroughly exploited. The planning dose distribution, for instance, can be considered as a 3D function presenting a large variability across the treated populations and is strongly linked to the individuals anatomy which in addition may change between-fractions sometimes at the expense of the prediction [85,86]. With the steadily increasing computational capabilities, exploiting information from more available data within integrative approaches becomes nowadays feasible.

Going further beyond the concept of whole-organ DVH, recent approaches aim at investigating more localized dose-toxicity relationships by analyzing the dose at lower spatial scales. For example, by analyzing dosimetric parameters of the lower GI anatomy, it has been shown that dose on some sub-regions correlated better with specific toxicities. This is the case of the works undertaken by Stenmark et al. [87] dividing the rectum in three different regions, or other authors by demonstrating anatomical dependence of specific GI toxicities [88–90]. Likewise, for urinary toxicity [58] by separately analyzing DVHs of the whole bladder, bladder wall, urethra, and bladder trigone. Other models have also sought to geometrically represent the 3D dose distribution in a single coordinate system via a spatial normalization for a joint analysis of dose at the lowest sampling scales (pixel and voxel levels). Different studies appear in the literature, either by building a parametric mapping via an intermediate spherical /cylindrical coordinate system such as the DSM in 2D [91], in 3D [57,92] or via 3D anatomical non-rigid registration (DVM) [93,94].

One of the major advantages of the pixel- and voxel-wise analyses is that no prior assumptions are made regarding the location of regions correlating with toxicity. Although these methodologies are still emerging, low spatial scale analyses of the dose distribution in different organs, have allowed the unravelling of the local dose-effect relationship across a population at each single pixel / voxel. With this kind of analysis, the implication of multiple structures was identified, such as in the head and neck area [95], or in the heart in patients treated for lung cancer [96].

These methods can potentially increase the prediction capabilities and improve the patient-specific treatment planning [94,97]. New planning systems are steadily moving from global DVH-based constraints, applied as suggested by international recommendations and available on the commercial TPS, towards the definition of spatially localized 3D patient-specific constraints as part of the TPS optimization. In the following sections we describe the methodological details for the construction of DSMs and DVMs and we discuss the challenges that these methods are facing.

1.4.2.1 2D Dose-suface maps (DSM): a 2D to 3D mapping

A DSM is a mapping of the 3D dose to a 2D representation of an organ wall and is constructed by virtually unfolding the organ in a slice wise manner. Different algorithms exist for generating DSMs from the 3D dose distribution [91,98,99]. To build the map, a 2D image is constructed and a parametric mapping is established between the 3D coordinate system of the organ wall and the 2D image. Thus, each pixel in the 2D image corresponds to a portion of the organ wall with the local dose computed, for instance, by interpolation at that 3D point. By construction, these DSMs reflect the dose of the organ surface. A dedicated module of VODCA (MSS Medical Software Solutions, Hagendorn, Switzerland) allows the generation of DSMs from organ contours and the calculated dose distributions. A representative output of this software is shown in Figure 1.10Figure 1.10

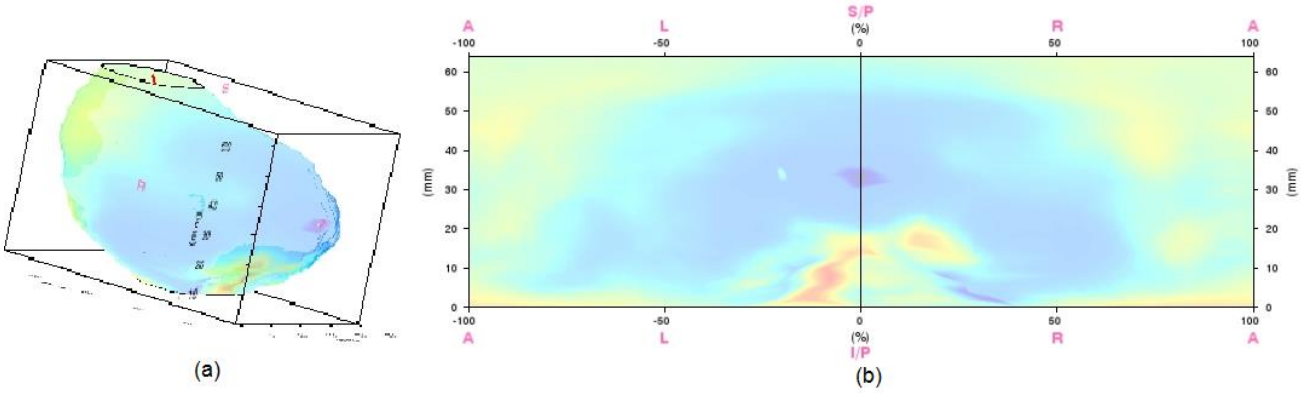


Figure 1.10 Bladder dose-surface map (DSM) extracted from VODCA software in 3D (a) and unfolded in a 2D plane (b)

The DSMs have been exploited in several ways to show relationships between toxicity and 2D local dose distributions. Via extraction of geometric features from isodose curves [100–103] or through direct pixel-wise comparisons [60,104–107], several studies have identified regions that better discriminate patients with and without toxicity. More recently, deep learning was also applied to study dosimetric effects, based on a convolutional neural network model to exploit rectal dose distribution on DSMs [107].

1.4.2.2 3D Dose-volume maps (DVM): a 3D to 3D mapping

The 3D DVM stands upon the 3D inter-individual spatial alignment, allowing for the subsequent dose propagation to a common coordinate system. The spatial alignment may be obtained via a parametric

representation of the anatomy in a spherical or cylindrical coordinate system [57,92] or can be more precisely computed through non-rigid registration methods [93,108] or tailored to a particular anatomy [109]. After this inter-individual normalization, voxel-wise statistical tests are performed on the 3D dose maps resulting in the localization of regions where statistically meaningful differences between or within groups may exist. Hence, organ subregions are computed as the clusters of voxels within the organ, where significant dose differences have been found. An example of DVM is shown in Figure 1.11 with color scales representing the dose distribution ranging from the high dose (red) to low/zero dose (blue).

Voxel-wise comparisons in a common frame of reference represent a reliable strategy to reveal local differences across individuals within a whole volume at low spatial scales. These methods are inspired by the voxel-based morphometry [110]. Applied to toxicity studies, the works undertaken in this field have allowed the identification of more predictive sub-regions within the organs in several locations, such as the rectum for gastrointestinal toxicity in prostate cancer [93,94], in the heart and lung for corresponding toxicities in thoracic cancer [96,111], and in the cricopharyngeus muscle and cervical esophagus for dysphagia in head and neck cancer [95]. A 3D voxel-based approach has never been applied to explore dosimetric patterns associated with urinary toxicity.

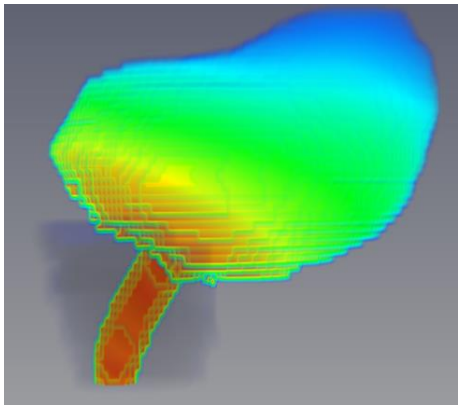


Figure 1.11 A 3D dose-volume map (DVM) of the bladder and the urethra

Table 1.4 summarizes some representative works in this field where both DSM and DVMs have been used in several clinical locations [112]. As mentioned before once the dose distribution is normalized to a single coordinate system, which is central to this methodology, a comparison of extracted features or a pixel or voxel wise analysis can be performed.

1.4.3 2D DSM and 3D DVM Methodological Challenges

Pixel/voxel-based methods share several methodological aspects as they require the dose to be mapped to a single coordinate system and thus different steps must be performed for the comparisons to be anatomically meaningful. These steps, as illustrated in Figure 1.12, include: (1) the spatial normalization of a population of individuals in terms of their anatomy to a common coordinate system; (2) the mapping of dose distributions according to the anatomical transformation obtained; (3) a reliable methodology to perform statistical analysis of the local dose-volume-effect relationship [112].

Table 1.4 Summary of works using 2D Dose-surface maps (DSM) and 3D Dose-volume maps (DVM) for toxicity prediction
H&N (Head and Neck); NRR (Non-Rigid Registration); GI(Gastrointestinal Toxicity); GU (Genitourinary Toxicity); studies on GU toxicities are shown highlighted; adjusted from [112]**.**

Reference	Model	Cancer location	Organ/ toxicity	Spatial Normalization / Dose mapping	Pixel/voxel-wise analysis
(Munbodh et al. 2008)	2D DSM	Prostate	GI	Conformal mapping	Geometric features
(Buettner et al. 2011, Buettner et al. 2012)		Prostate	GI	Geometric Dose surface map	Geometric features
(Buettner et al. 2009)		Prostate	GI	Geometric Dose Surface map	Geometric features/ Neural networks
(Palorini et al. 2014)		Prostate	GU	Geometric Dose Surface map	Geometric features
(Wortel et al. 2015)		Prostate	GI	Geometric Dose Surface map	Dose surface features
(Palorini et al. 2016)		Prostate	GU	Geometric Dose Surface map	Pixel-wise comparison
(Improta et al. 2016)		Prostate	GU	Geometric Dose Surface map	Pixel-wise comparison/ spatial descriptors
(Calyn et al. 2017)		Prostate	GI	Geometric Dose Surface map	Spatial features
(Yahya et al. 2017)		Prostate	GU	Geometric Dose Surface map	Pixel-wise comparison/ spatial descriptors
(Xin et al. 2017)		Cervix	GI	Geometric Dose Surface map	Deep learning
(Heemsbergen et al. 2010)	3D DVM	Prostate	GU	Geometric 3D mapping	Voxel-wise comparisons
(Ziad et al. 2012, Rao et al. 2012)		H&N	Trismus	NRR	Voxel-wise comparisons
(Coloigner et al. 2015)		Prostate	GI	NRR	ICA for classification
(Chen et al. 2011) (Fargeas et al. 2013)		Prostate	GI	NRR	PCA for feature extraction and classification
(Fargeas et al. 2015)		Prostate	GI	NRR	Tensor decompositions
(Ospina et al. 2013)		Prostate	GI	NRR	Tensor value decomposition for subregion identification
(Liu et al. 2015)		Prostate	GI	NRR	Non-negative matrix factorization for Classification
(Acosta et al. 2013)		Prostate	GI	NRR	Voxel-wise Comparisons
(Drean et al. 2016)		Prostate	GI	NRR on different templates	Voxel-wise comparisons, definition of a generic 3D patient-specific region
(McWilliam et al. 2016) (McWilliam et al. 2017)		Lung	Heart	NRR	Voxel-wise
(Palma et al. 2016)		Lung	Lung	NRR	Voxel-wise differences
(Monti et al. 2017)		H&N	Acute dysphagia	NRR	Voxel-wise comparisons
(Avanzo et al. 2017)		Lung	Radiation induced lung injury	NRR	Voxel based longitudinal comparison of CT density & dose

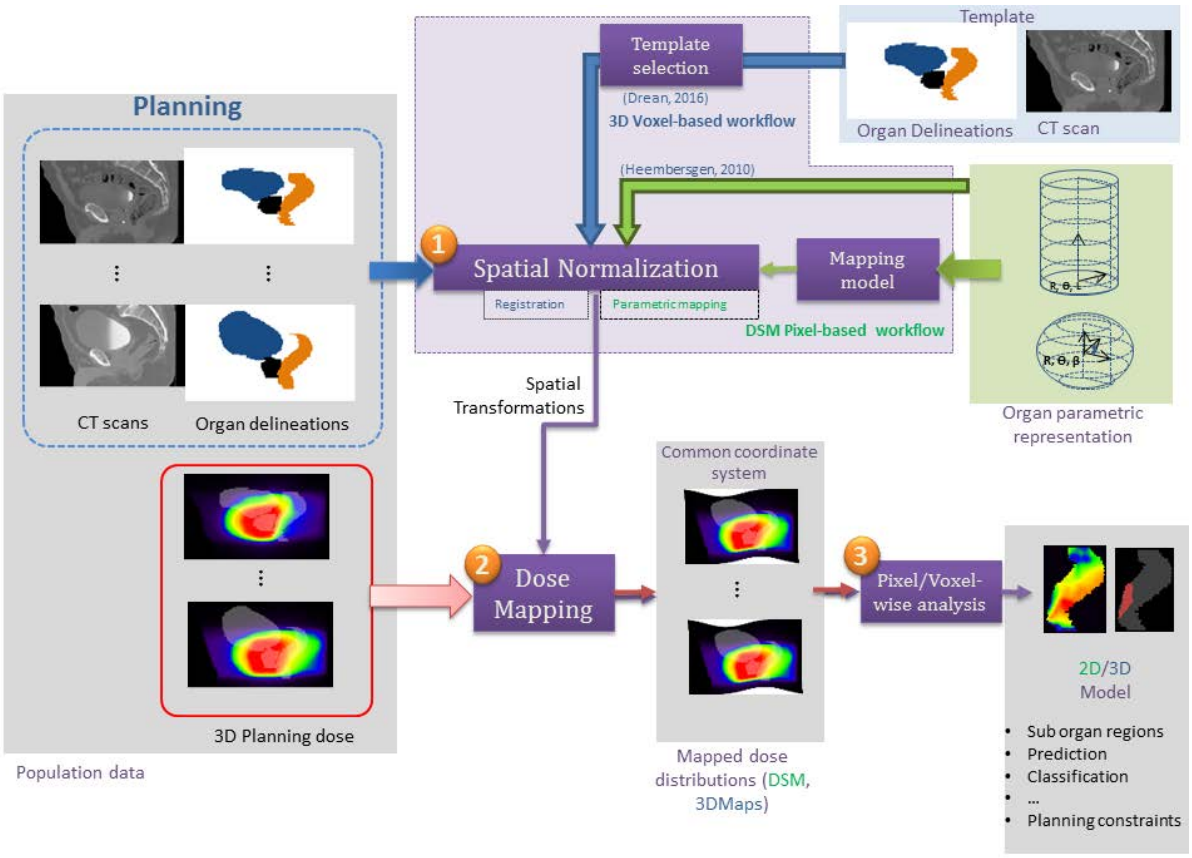


Figure 1.12 Methodological aspects of voxel/pixel based models

(source: [112])

1.4.3.1 Spatial normalization and dose propagation to a common coordinate system

Spatial normalization is the process of obtaining a transformation between the native individual's coordinate system and the common coordinate system leading to meaningful correspondences across the population. This is a key step in pixel/voxel wise analysis since dose comparison results rely on anatomical alignment accuracy. In the case of DSMs, the mapping is generated by the direct relationship between a 3D coordinate system and the 2D map. After the 3D–2D relationships are obtained, the dose is propagated and interpolated, yielding a 2D image of dose on the unfolded organ.

The geometric correspondences may be extended to 3D by simply including the third axis (i.e., R in cylindrical or spherical coordinates) to build a 3D dose map. This was done in Heemsbergen et al. and Witte et al. [57,92] where the dose mapping relies on a parametric representation of the anatomy in a spherical coordinate system and mapped back again to a single anatomy to perform voxel-wise analysis. Spatial normalization may also be performed by non-rigid registration between the population data and individual template. In that case several questions arise, such as the selection of the most representative template and the most reliable inter-individual registration method. This appears as particularly difficult given the high inter-individual anatomical variability (organ volume, artefacts, presence of gas, air, etc.) and the low contrast of soft tissues if CT scans are used for registration.

In the setting of population analysis, a mapping error may lead to invalid results in statistical analysis. In the presence of a high dose gradient, which is the case for most OARs, small shifts in registration may result in large differences in dose on the reference template. Evaluating the accuracy of inter-individual mapping is particularly complex because of the lack of ground truth. In the literature, different measures have been proposed to estimate intra-individual anatomical mapping accuracy and transformation validity [113,114]. Dice similarity coefficient and Hausdorff distance are among the most commonly used evaluation metrics. However, they only reflect overall geometric overlap between transformed structures and do not show local mapping errors within the structure. Given that a high overlap score does not necessarily imply good point-to-point mapping [115], Drean et al. [109] assessed the mapping accuracy not only from an anatomical view point but also from a dosimetric one, by introducing a new metric to estimate the organ overlap relative to dose distribution (dose-organ overlap, DOO). This metric measures the ratio between dose distributions on the intersection and union of the considered region. The DOO is a value between 0, when structures have no voxels in common, and 1, when the dose to the structures is identical. In practical terms, the score penalizes anatomical difference by taking into account the dose that would be mapped onto the reference structure. Dice score, Hausdorff distance, and DOO score have also been used to assess spatial normalization in recent voxel-based toxicity studies [95,111].

1.4.3.2 Pixel/Voxel-wise analysis in a common coordinate system

After 3D doses are spatially normalized, several comparisons can be performed in the common coordinate system. Very often the incorporation of spatial descriptors improves discrimination of multi-variable models including dosimetric and clinical parameters [94,101,105]. The descriptors may be computed after extracting isodose curves for instance [100]. Several studies found that toxicity is related to the shape of isodoses as well as dose coverage [103]. Comparisons of dose average at the intragroup level (with and without toxicity) are frequently investigated with additional tests seeking for statistically significant differences. Either parametric or non-parametric voxel-wise hypothesis tests can be performed, depending on the data. This is the most frequently found case, i.e., where pixel/voxel two-sided t-tests and the resulting p-values map (p-value < threshold value) were used for delimiting the regions better discriminating between groups [60,93,94,116].

Caution must be taken however because of the multiple comparison problem, arising when performing thousands of simultaneous tests that may be correlated which is the case in voxel-wise methods. Those issues have been largely treated in voxel-based morphometry studies. Several correction techniques exist such as Bonferroni, false discovery rate [110,117], threshold-free cluster enhancement (TFCE) methods [118], or permutation tests [119], which have been implemented in several of the aforementioned toxicity studies [106,111,120]. Permutation tests allow inferences while taking into account the multiplicity of tests as described in [121]. The TFCE offers an interesting spatial characteristic as it takes advantage of neighborhood information to increase the belief in contiguous areas of the considered signal introducing spatial coherence to the findings. Permutation testing can also be coupled

to TFCE. In this thesis, the multiple comparison problem was addressed by implementing the method proposed by Chen et al. [119], as will be described in Chapter 4.

In the next section we will present the motivation and objectives of the thesis.

1.5 Motivation and thesis objectives

As already explained, although the dose is planned and delivered to the target following recommendations aimed at maximizing control and diminishing toxicity, the surrounding healthy organs (rectum, bladder ...) are often impaired from irradiation and present adverse events. To date, urinary side-effects are not well understood, in particular how they relate to underlying dose-volume characteristics. In a review by Fiorino et. al. [45], it was highlighted that 3D dose-volume-response data for the bladder are still lacking.

In the past, the whole bladder's DVH has been used in an attempt to explain radiation-induced toxicities. Although some studies found a relationship between urinary toxicity and bladder dose [57,58,72,73,122,123], others did not find any significant correlation [124,125]. No general consensus exists among the studies reporting positive findings. This lack of homogeneity in the reported relationship between urinary toxicity and bladder dose can be attributed to several factors [126], such as:

- The high variability of bladder volume limits the capability to determine the actual dose received by the bladder during the treatment. Consequently, the planning dose may not be representative of the actual dose delivered because of the high interfraction bladder volume variations [127,128].
- Urinary symptoms are multifactorial and depend on patient-specific clinical parameters [57,129–134], individual biological patterns [129,135–137] and dosimetric parameters.
- Urinary toxicity events may occur late after RT, in contrast to late gastro-intestinal toxicity which generally reaches a plateau after 3 years, suggesting that a longer follow-up is required to properly estimate late urinary toxicity [133].
- Specific sub-regions of the bladder may present different radiobiological behaviors/sensitivities, which may have different impacts on distinct side effects [57,58].

Above all these, it is important to stress that, currently, all the studies of urinary toxicity following EBRT have been focused on the dose to the whole bladder or parts of it, while the potential involvement of urethra damage due to radiation remains largely unexplored. The main reason for this is the lack of tools to identify this structure on the planning CT images.

Going beyond the global, whole-organ-based models towards more local, sub-organ approaches, this thesis aims to improve our understanding of radiation-induced urinary side-effects and ameliorate the prediction of urinary toxicity following prostate cancer radiotherapy. More specifically, the objectives of this thesis are:

1. To assess the contribution of urethra damage to urinary toxicity.
2. To evaluate the involvement of specific bladder and urethra sub-parts in urinary toxicity.
3. To investigate the potential of further increasing prediction by exploiting data through new machine learning methods.

An overview of the thesis structure is illustrated in Figure 1.13.

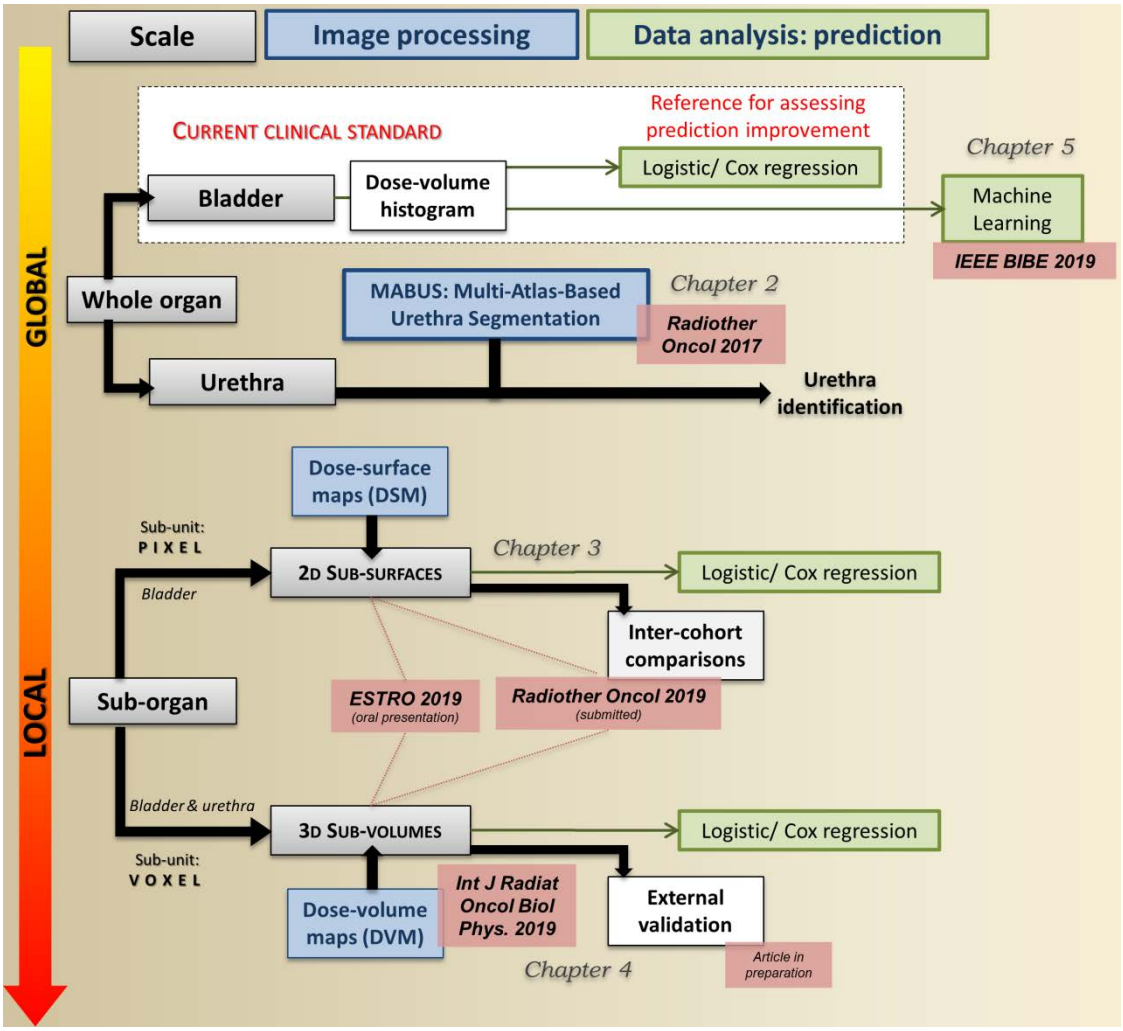


Figure 1.13 Schematic overview of the thesis structure

Part II

Quantifying the dose to the prostatic urethra

2 Segmentation of intra-prostatic urethra from planning CT images to quantify dose distribution in external radiotherapy

In this chapter a methodology is proposed for automatically segmenting the intra-prostatic urethra on the planning CT images of EBRT-treated patients with the objective to assess the involvement of this structure in urinary toxicity. First, the entire framework of the multi-atlas-based segmentation method is detailed, from the atlas construction to the final urethra segmentation. Then, the method's accuracy is evaluated with respect to the ground truth urethra and compared with existing surrogate models for estimating urethra's position. Finally, the DVH in the urethra is computed in a population dataset in order to quantify the dose received by this structure.

The content of this chapter has been published in the journal Radiotherapy and Oncology.

Acosta O, **Mylona E**, Le Dain M, Voisin C, Lizee T, Rigaud B, Lafond C, Gnep K, de Crevoisier R. *Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy*. Radiother. Oncol. 2017;125:492–499.

2.1 Introduction

Although there is evidence in prostate cancer brachytherapy that some urinary symptoms are related to urethra damage [61–63] this has not yet been shown in external-beam radiotherapy. Quantifying the delivered dose to the urethra may therefore improve our understanding of urinary toxicity or at least some of the related symptoms if we can accurately identify the organ from the planning CT. Thus, segmenting the urethra from the planning CT in order to assess the dose it receives, would pave the way for further studies on urinary toxicity prediction considering both the bladder and urethra.

To our knowledge, a formal segmentation of the intra-prostatic urethra from CT images has not been addressed yet. However, indirect surrogate models for estimating the urethra position have been previously proposed [138,139]. They are nevertheless based on empirical considerations with respect to the prostate midplane. Segmenting the urethra from CT scans is fairly challenging. Not only is there already poor contrast between soft tissues like the prostate, bladder, and rectum, thus rendering segmentation difficult for planning, but the intra-prostatic urethra itself is completely invisible. These issues restrict the use of classic intensity-based segmentation methods. Atlas-based approaches, widely discussed in the literature [140–146], are common methods for organ segmentation. In atlas-based methods, precomputed segmentation in a template space is propagated onto the image to be segmented via spatial normalization (registration) as depicted in Figure 2.1. Several individuals from a population can be used to constitute the atlas (multi-atlas). This allows to overcome the inter-individual variability and registration issues. Previous works have shown the benefits of combining multiple atlases in improving segmentation accuracy [140,142,143,147–149].

In this paper, we propose a weighted Multi-Atlas-Based Urethra Segmentation strategy, herein called MABUS, from planning CTs. Our goal is to provide a method that could be applied to a different set of patients receiving external-beam radiotherapy in order to assess the dose to the urethra and relate with toxicity effects.

2.2 Material and methods

This study is divided into three main parts: i) a description of MABUS, the multi-atlas-based urethra segmentation method which illustrates the whole implemented framework, from the atlas construction to the final urethra segmentation, ii) the evaluation of the method's accuracy with respect to the urethra ground truth in a leave one out cross validation framework, and comparison with the existing surrogates proposed by Bucci [138] and Waterman [139] and finally iii) the computation of the dose received by the urethra in a different series of patients with prostate cancer IMRT which aims to introduce the way in which the method may be used in toxicity studies.

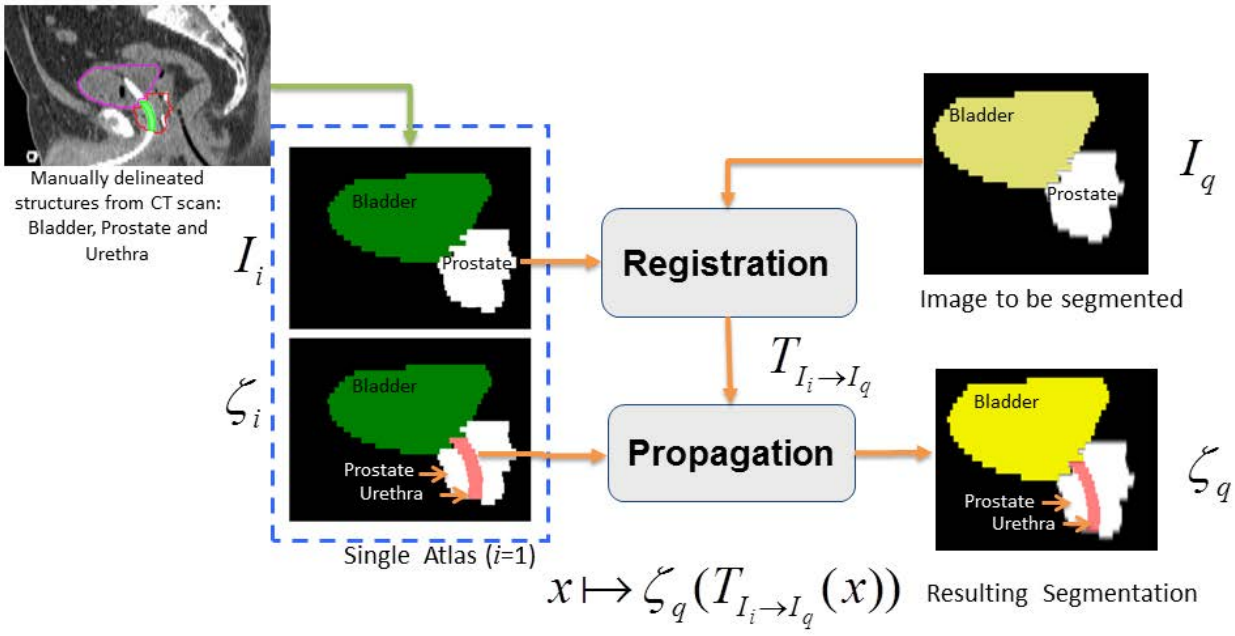


Figure 2.1 Atlas-based segmentation framework

2.2.1 Multi-atlas based urethra segmentation (MABUS) method description

In general, as depicted in Figure 2.1, atlas-based segmentation relies on the registration of a template I_i to the query image I_q, in order to obtain a transformation $T_{I_i \to I_q}$, which maps a set of generated labels ζ_i onto I_q. If the mapping is anatomically correct, the yielded segmentation is accurate and anatomically meaningful. Multi-Atlas based segmentation builds upon this idea by extending the number of atlases thereby reducing the interindividual variability issues.

Following this multi-atlas idea, the proposed MABUS was devised and can be divided into seven steps as depicted in Figure 2.2. In summary, an atlas dataset was first built from manually-delineated CTs including the urethra, thanks to the presence of an urinary probe (Step 1). The query image to be segmented I_q was then rigidly aligned with the same template I_T as the atlas database (Step 2) and features were extracted $F_q = \{f_{1q},..,f_{5q}\}$(Step 3). By comparing the features, the atlases were ranked according to their similarity to the query image (Step 4). The labels $\zeta_i = \{urethra\}$ from the top *n*=10 ranked atlases were then propagated to the query image using an accurate non-rigid registration method (Step 5) designed to match the prostate anatomies. Finally, the urethra segmentation was obtained by combining different labels in a weighted-fusion process (Step 6), followed by centerline detection (Step 7). The image segmentation methods were developed in C++ using the Insight Toolkit libraries (ITK) [150] and python open source technologies.

ATLAS BUILDING FROM TRAINING DATA (STEP 1)

For the atlas building, we used an initial series of CT scans (512×512 0.63×0.63mm axial pixels and 3mm slices) from 55 patients treated for localized prostate cancer with Iodine-125 brachytherapy. All the patients were fitted with urinary catheters, enabling urethra segmentation. The prostate, bladder, and urethra were delineated for each by the same radiation oncologist, constituting the set of atlases I_i, with the label $\zeta_i = \{urethra\}$.

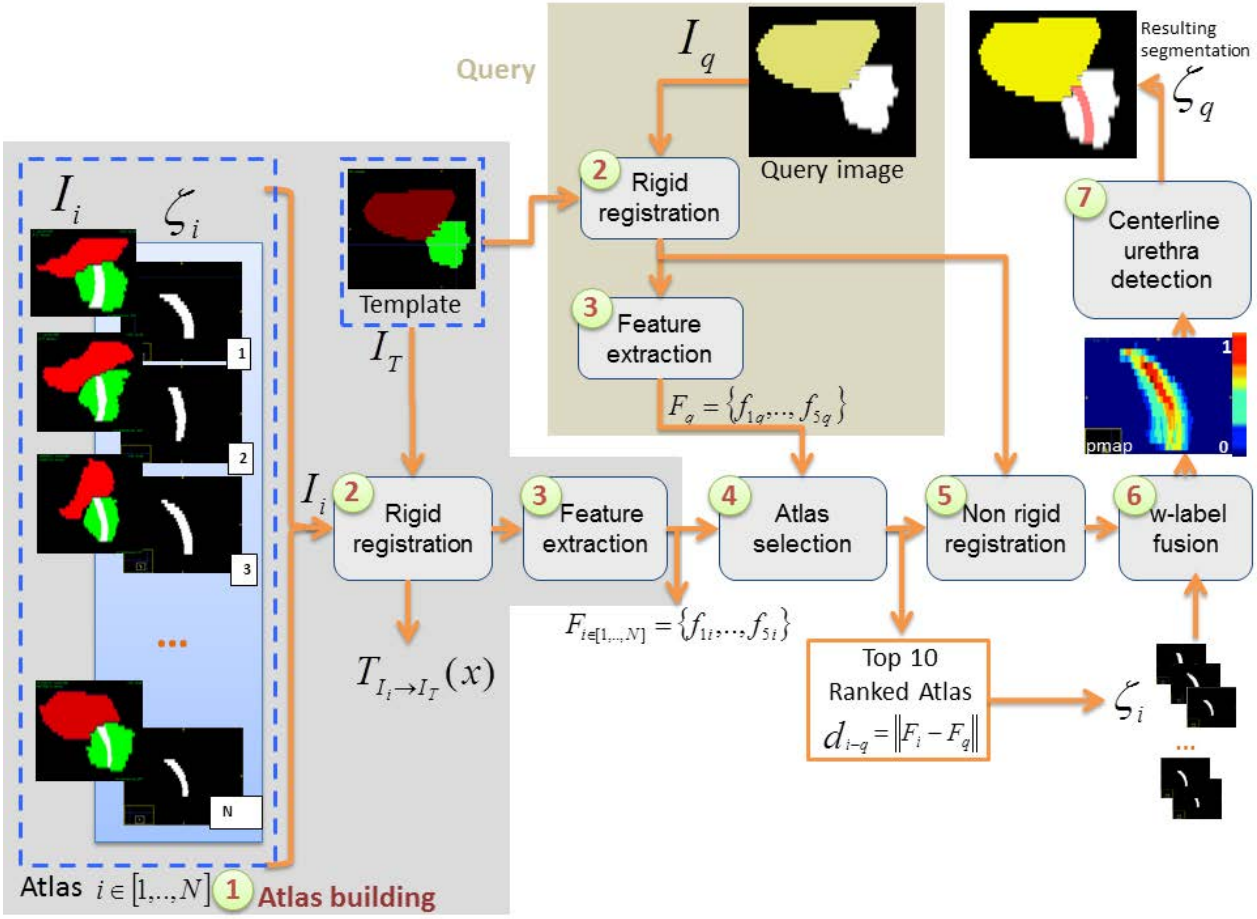


Figure 2.2 Overall proposed framework of multi-atlas based urethra segmentation (MABUS) method

TEMPLATE SELECTION AND RIGID REGISTRATION (STEP 2)

A first average patient I_T was selected as a common coordinate system. This patient was the closest to all others in terms of prostate volume. The whole population was then rigidly registered to this patient by aligning the prostate centroids followed by a fine alignment of bony anatomy. This enabled geometrical descriptors to be generated and compared in the same common space. The central lines of the manually delineated catheters were computed by extracting their centroid at each slice.

FEATURE EXTRACTION (STEP 3)

A simplified geometrical description of the anatomy (prostate/bladder) was generated to characterize each individual. Hence, the obtained vector $F_{i\in[1,..,N]} = \{f_{1i},..,f_{5i}\}$ describes the individuals in terms of i) prostate volume ii) distance between prostate and bladder centroids, iii) the extension of the bladder in the anterior posterior direction, and iv) the orientation of the bladder with respect to the prostate centroid, regarding two angles (ϕ) and (θ), which respectively describe the anteroposterior and lateral directions. The descriptors were then normalized across the population with a z-score.

QUERY IMAGE AND ATLAS SELECTION (STEP 4)

For an image I_q to be segmented, the two previously described steps (2 and 3) were applied as for the training database. Thus, rigid registration to the common template I_T and characterization yield for I_q the vector $F_q = \{f_{1q},..,f_{5q}\}$, exhibiting similar features as computed for the atlas dataset. Following z-score normalization, the Euclidean distances $d_{i-q} = \|F_i - F_q\|$ between features enabled the individuals from the atlas I_i to be ranked in terms of similarity to the query image I_q. In a multi-atlas strategy, only the top (*n*=10) ranked atlases were selected as the closest to the query image I_q, with all remaining atlases discarded. Since their configuration is similar to that of the query image, the urethra is expected to lie inside the prostate in a similar position. The number of atlases (*n*=10) was selected as a tradeoff between computational time and optimized results in a leave-one-out segmentation process in which the top 1,2,..,n atlases were tested.

NON-RIGID REGISTRATION (STEP 5)

In this step, the labels ζ_i from the *n* most suitable previously-selected atlases were non-rigidly propagated to the prostate of the query image. To this end, the prostates from the best atlases were non-rigidly registered to the prostate of the query image I_q. We applied a Laplacian-based registration method, based on a previous work [109], but here only considering the prostate. In our implementation, instead of using the central line, we selected the centroid Cp of the prostate for computing a scalar field $u(x,y,z)$ by applying Laplace's equation inside the prostate volume, demarcated by an external boundary F_{Ext}, here the prostate surface, and an internal boundary, here the prostate centroid Cp, as:

$$\Delta u(x,y,z) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0 \quad (1)$$

where $u(x,y,z) = u_{Ext}$, if $(x,y,z) \in F_{Ext}$ and $u(x,y,z) = u_{Cp}$, if $(x,y,z) \in Cp$, where $\{u_{Ext}, u_{Cp}\} \in \mathbb{R}$. This scalar field *u* provides a structural and normalized descriptor to be exploited in a Demons-based non-rigid registration framework. This step yields a 3D deformation field (DF) as a set of 3D vectors describing the voxel-wise deformation of each of the atlas images to I_q. The binary labels $\zeta_i = \{urethra\}$ from the selected atlases were then propagated to the query image space using the calculated transformation $T = DF \circ T_{Rigid}$ and nearest-neighbor interpolation in order to preserve the binary nature of the propagated labels.

WEIGHTED-LABEL FUSION (STEP 6)

Once the labels were propagated to the same coordinate system, namely the prostate of the query image I_q, the raised question was how to fuse all the warped labels $\zeta_i = \{urethra\}$ to yield the best segmentation result. Different decision rules may be applied, such as a simple voting-rule [151], a weighted decision based on similarity [152] or a Bayesian approach, such as the simultaneous truth and performance level estimation (STAPLE) [149]. We opted for the weighted-label fusion approach, resulting in the following probability map:

$$p_{\zeta_i} = \frac{1}{N}\sum_{i=1}^{N} w_i \zeta_i \tag{2}$$

with the weights *w* derived as a function of the Euclidean distance between geometric features, $d_{i-q} = \|F_i - F_q\|$, giving

$$w_i = \frac{e^{-d_{i-q}}}{\sum_{i=1}^{N} d_{i-q}} \tag{3}$$

Thus, the contribution of each non-rigidly propagated label to this map heavily depended on the similarity between the atlas and the query image.

CENTERLINE AND URETHRA DETECTION (STEP 7)

A 50% threshold was applied to the probability map p_{ζ_i}, followed by a centerline computation as a 3D cubic spline curve of equidistant points. This resulted in a smooth path γ_q describing the urethra within the prostate. Considering the urethra to be a tubular-like structure, the final urethral region ζ_q was obtained as a 5mm zone around γ_q.

2.2.2 Evaluation of the accuracy of the segmentation method

Leave-one-out cross validation was performed to evaluate the accuracy of the proposed method. Thus, each of the individuals from the atlas database was iteratively selected as query image I_q and the resulting segmentations compared with the catheter positions. For our purposes, the central path described by the catheter was considered the ground truth.

Two metrics were used to evaluate the method's accuracy. A score based on the centerline distance (CLD) was devised to assess the differences between the obtained discretized urethra path γ_q and the catheter central line ρ_{GT} as follows:

$$CLD(\gamma_q, \rho_{GT}) = \frac{1}{N}\sum_{k=1}^{N} dist(\gamma_q^k, \rho_{GT}) \tag{4}$$

where γ_q^k represents the k^{th} point of the urethra central line, *dist* (γ_q^k, ρ_{GT}) is the Euclidean distance of this point to the catheter central line, and *N* is the number of points. Figure 2.3 illustrates the way in

which the CLD was computed from this distance map. This score was computed for the whole segmented path γ_c , as well as for each region produced after splitting the central path ground truth into three equivalent segments (from the apex to base). The percentage of the points from γ_c lying inside a region around the centerline ground truth (PWR) within 3.5 mm and 5mm radius were also quantified. For comparison with the existing surrogates, we implemented the methods proposed by Bucci [138] (deviated surrogate) and Waterman [139] (centered surrogate) which are based on the geometrical center of the axial midplane.

2.2.3 Dosimetric study: assessment of the dose received by the urethra in prostate cancer IMRT

A second series of 95 patients having received 78Gy IMRT for prostate cancer between July 2012 and June 2015 were analyzed. The target volume included the prostate, sparing the pelvic lymph nodes. Target volumes and organs at risks (bladder, prostate) were delineated on CT slices according to the French GETUG group recommendations [153]. IMRT combined with image-guided radiation therapy (IGRT) were used to deliver a total dose of 78 Gy for cone-beam CT or 80 Gy for fiducials to the prostate over eight weeks at 2Gy/fraction. The pelvic lymph nodes were not irradiated. Bladder dose-volume histograms (DVHs) complied with GETUG recommendations, namely V70 < 50%.

The proposed segmentation method was applied to patients treated with IMRT. Manual delineations of the prostate and bladder were used to automatically segment the urethra within a 5mm diameter along the centerline. The dose within the urethra was assessed from the 3D planning dose distribution and compared to the dose to the prostate. A comparison between the prostate and urethra DVHs was performed using a Wilcoxon non parametric test.

2.3 Results

2.3.1 Atlas construction

The inter-individual variability found in terms of feature descriptors $F_{i\in[1,..,N]} = \{f_{1i},..,f_{5i}\}$ was very high. The features represented in average 62.6 ± 15 cc for the prostate volume, 43.1 ± 15 mm for the distance between prostate and bladder centroids, 19.3 ± 19 mm for the bladder extension in the horizontal/y-axis and 58.9 ± 19.5 (ϕ) and 120.5 ± 61.5 (θ) for the angles describing respectively the antero-posterior and lateral directions of the bladder with respect to the prostate centroid. Figure 2.3 displays the distribution of the interindividual normalized similarity across individuals in the z-score space.

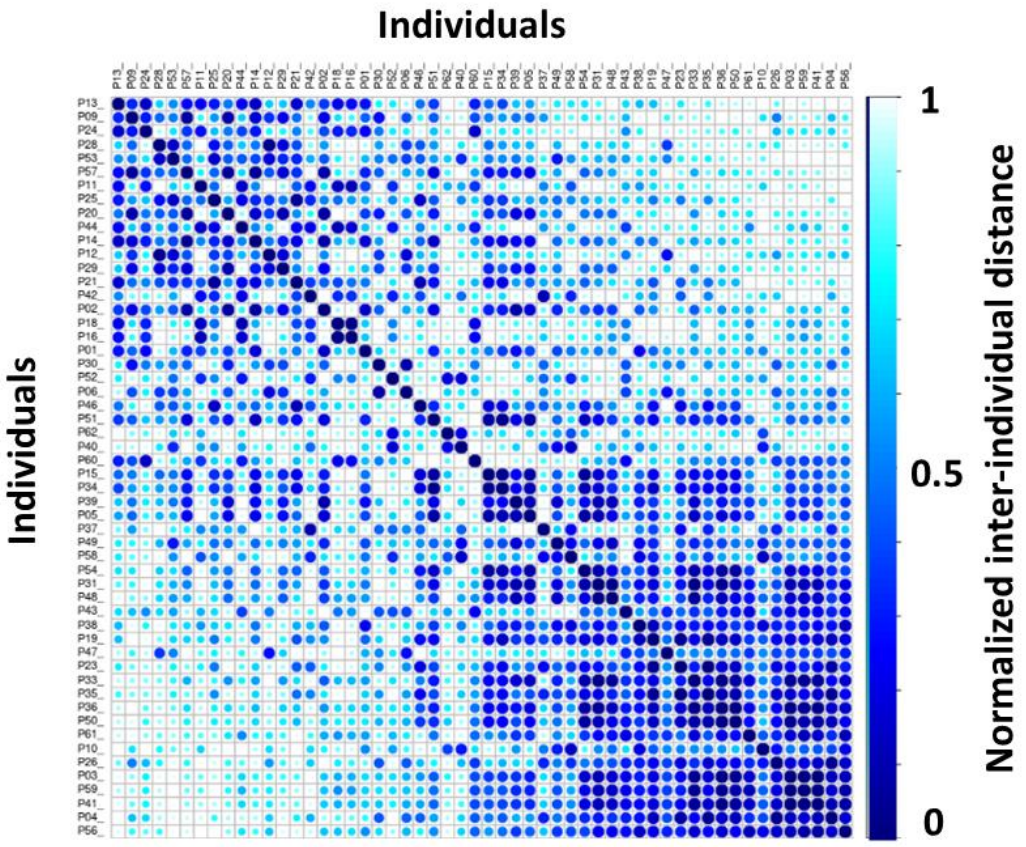


Figure 2.3 Inter-individual distance map

2.3.2 Urethra segmentation accuracy

Leave one out results and comparisons with Waterman's and Bucci's are shown in Figure 2.4. In average MABUS outperformed the other two surrogate models globally and by thirds. Global CLD was 3.25±1.2 with MABUS, while 6.11±1.96 with Waterman and 3.91±1.46 with Bucci's method ($p<0.001$). In addition by thirds with MABUS the computed CLD scores were 3.67 ±1.66 mm, 2.52 ±1.58 mm and 3.01 ±1.76 mm for the top middle and bottom thirds respectively. Likewise for the portion of the central line (PWR) within the 3.5mm and 5mm radius regions: 0.53±0.29 and 0.83±0.18 respectively for MABUS with a maximum of 1 in both cases. With Waterman we obtained 0.32±0.19 and 0.53±0.17 ($p<0.001$) and with Bucci 0.51±0.22 ($p=0.5$) and 0.74±0.21 ($p < 0.01$). Figure 2.5 depicts those results. There were some individuals for whom the segmentation was not as good as expected. Those outliers appear as dissimilar to the remaining individuals as shown in Figure 2.3. Only 3 segmentations out of 54 presented a score inferior to 0.5 in the 5mm region.

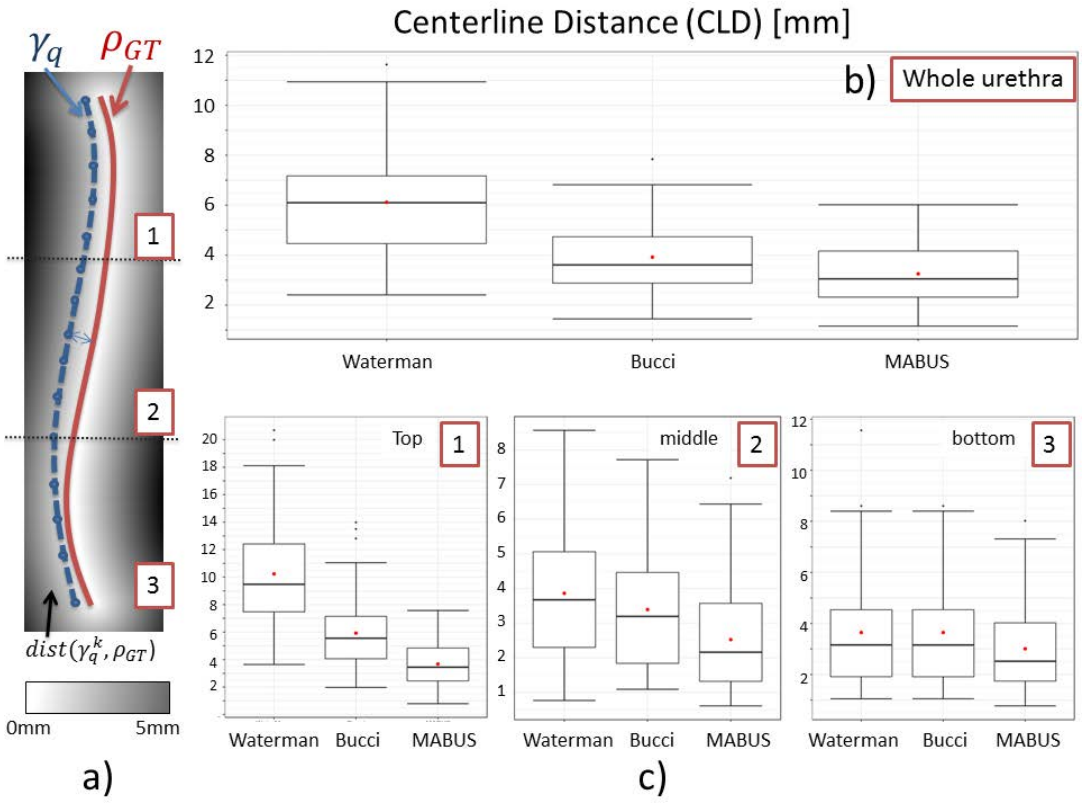


Figure 2.4 Centerline distance (CLD) definition and leave one out validation results of the proposed method compared with previously proposed surrogate models

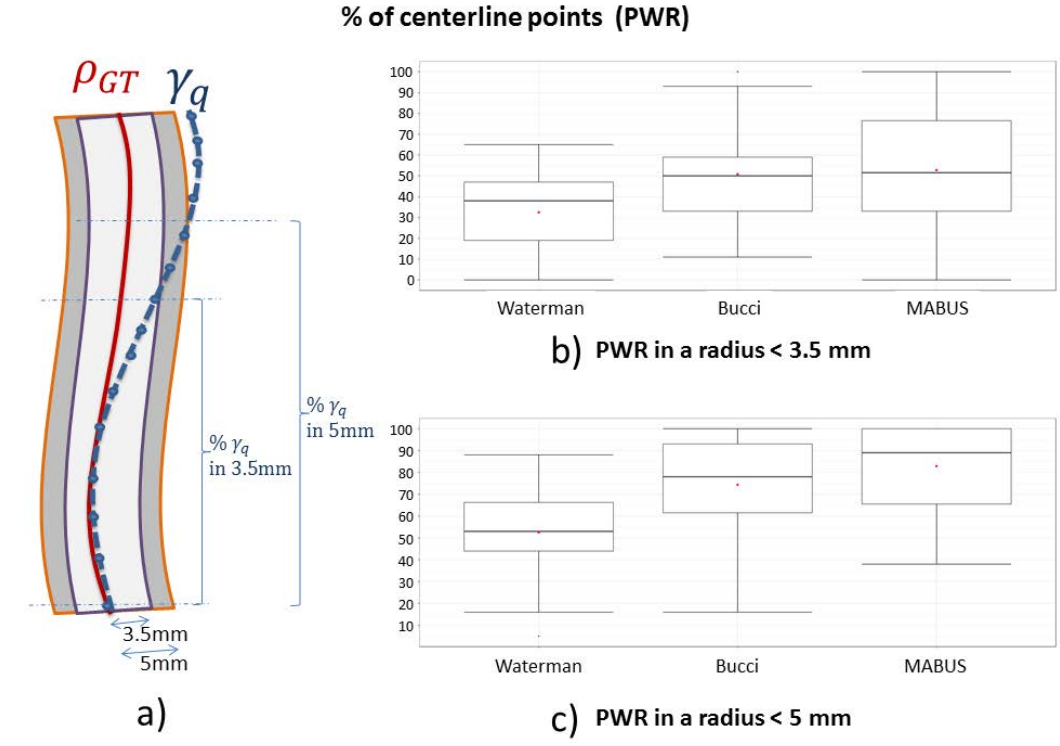


Figure 2.5 Percentage of segmented centerline (γ_q) lying within a radius (PWR) of 3.5 or 5mm around the ground truth centerline (γ_q)

Figure 2.6 shows six examples of resulting segmentations overlaid on the manually delineated catheters. Top row illustrates a comparison of the proposed method with the surrogate models Figure 2.6 a) Waterman, b) Bucci and c) MABUS. Low row shows different results with MABUS on different individuals. It can be observed the complimentary information brought by the two different scores. For instance, in case of Figure 2.6 d) the CLD was 3.4mm and the whole centerline was within the first 5mm region (PWR). For c) CLD=2.8mm. Although only half of the points where within the first 3.5mm, the whole centerline was within the first 5mm.

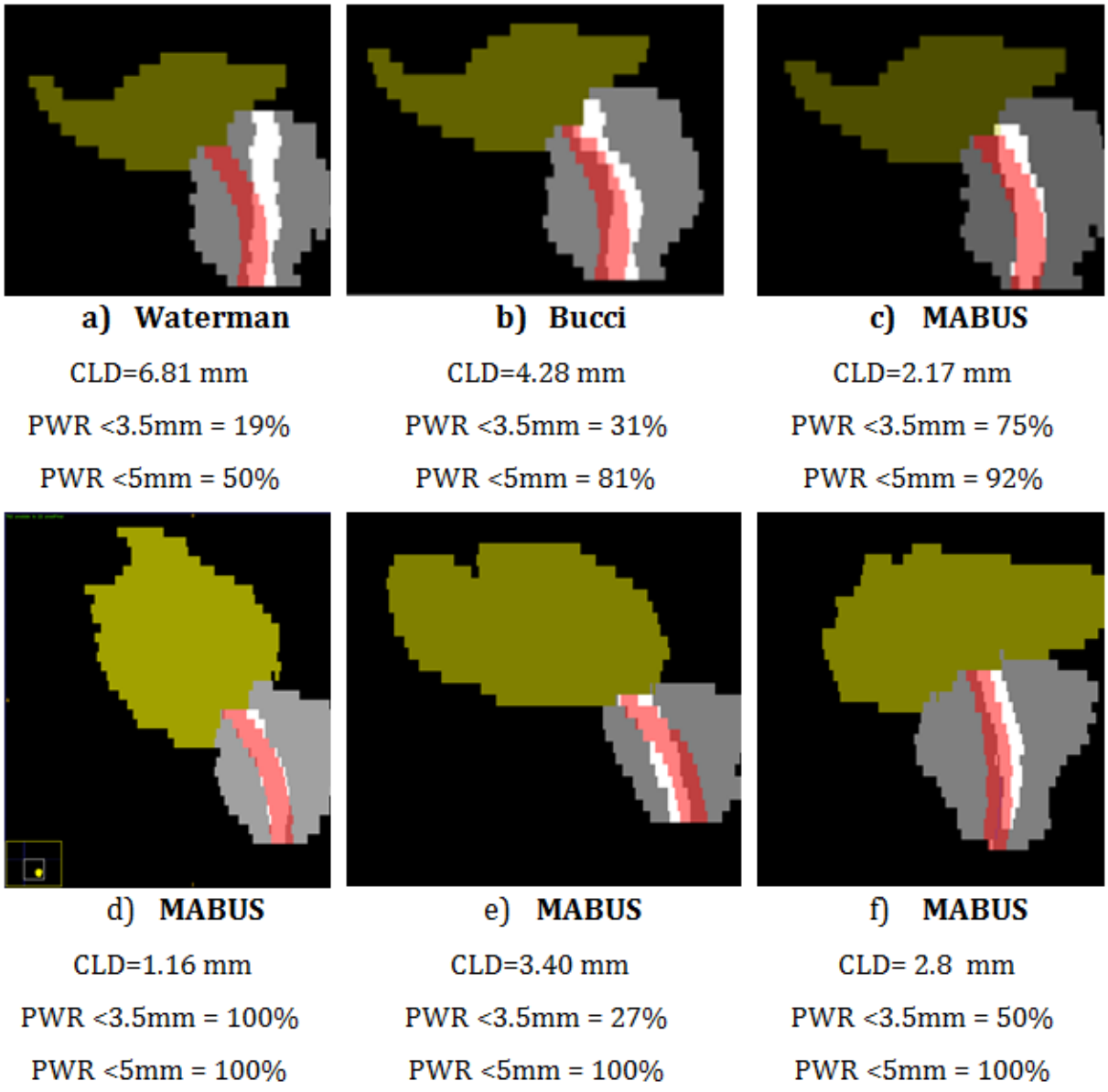


Figure 2.6 Examples of urethra segmentations (white) overlaid on the actual urinary catheter (red).
Top row: Comparison of MABUS with two surrogate models. Low row: .different results for individual configurations with the proposed methodology (MABUS) only

2.3.3 Dosimetric study

Urethra and prostate DVHs were significantly different. The bin-wise DVH comparison show that the volume (%) receiving a dose between D74Gy to D79Gy by the urethra was significantly higher than in the prostate ($p < 0.01$). Figure 2.7 highlights the prostate and urethra DVH bins, where statistically significant differences are represented by the red circles.

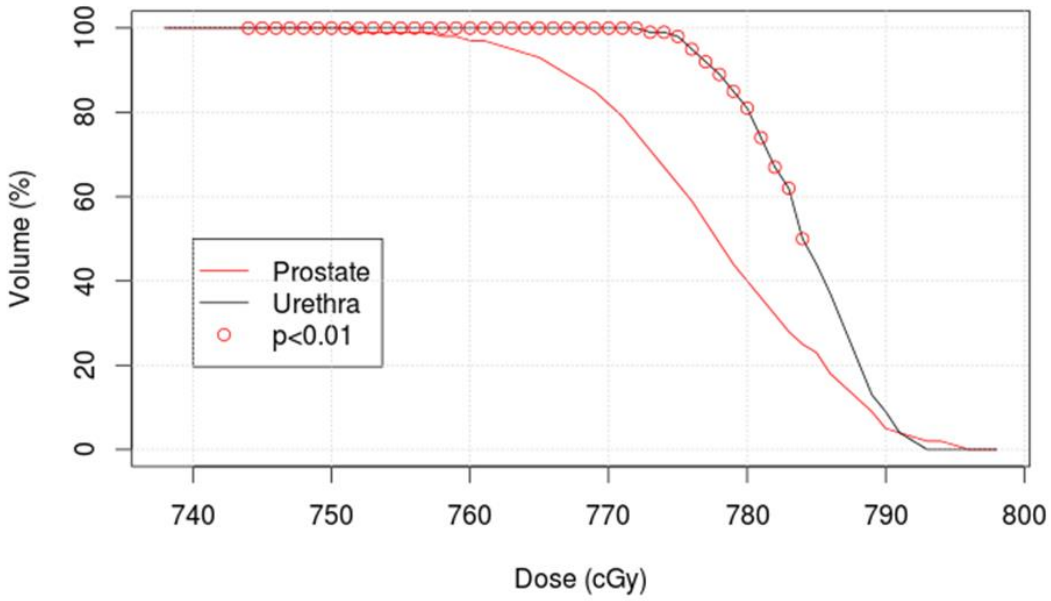


Figure 2.7 DVH differences between urethra and prostate in case of IMRT delivering 80Gy to the prostate

Figure 2.8 shows an example of 3D dose distribution within the manually segmented volumes for planning together with the automatically segmented urethra. Considering the urethra, the prostate and the PTV, it can be seen that high doses appear in the urethra (isodose curves and DVH values > V70). During the 3D dose optimization, the constraints to the PTV and to the organs at risk, bladder and rectum, will introduce a high dose gradient close to the PTV. Although the PTV is well covered it will receive 90% of the dose, whereas the urethra being central to the prostate likely receives the higher doses.

2.4 Discussion

We proposed a weighted multi-atlas based method to segment the intra-prostatic urethra from planning CT and compared with two previously surrogate models based on the central axis of the prostate, Waterman et al. [139] (centered surrogate) and Bucci et al. [138] (deviated surrogate). The method does not need any catheter to estimate the position of the urethra as it is based on the combination of similar cases contained in a large data set of atlases, which are weighted to achieve an accurate segmentation

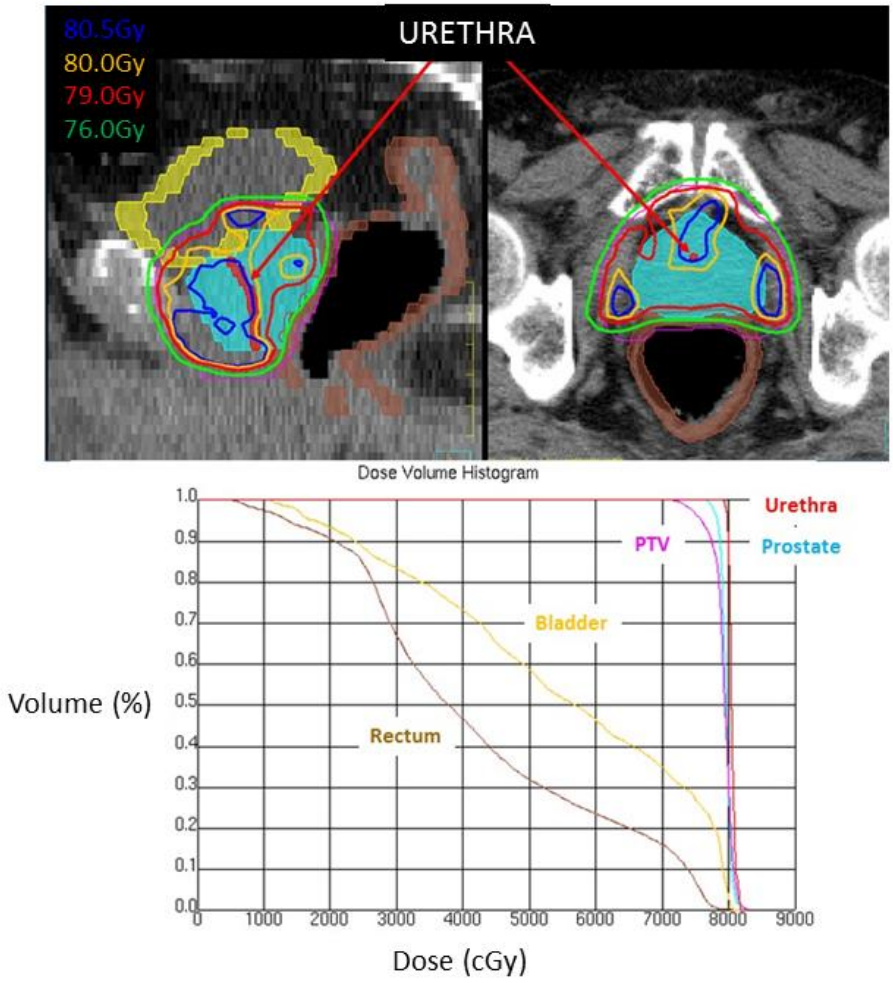


Figure 2.8 Example of urethra segmentation and dose distribution in IMRT planning delivering 80 Gy to the prostate

The obtained accuracy of the urethra segmentation considering the CLD (in average 3.25 mm), computed in a leave one out cross-validation enabled to assess dose to the urethra in a different IMRT database. With our method, the measured dose received by the urethra appears slightly higher than the dose received by the whole prostate. This is likely due to the position of the urethra rather central within the prostate, relatively far from the rectum where a gradient of dose appears (Figure 2.8). Such findings support the use of our urethra segmentation method to potentially improve urinary toxicity prediction by considering both the dose received by the urethra and the prostate.

To our knowledge there is no evidence in the literature of any method for explicitly segmenting the urethra in the planning CTs. An atlas based urethra segmentation method in MRI was proposed [154], within a SBRT perspective, but without a formal segmentation propagation towards the planning CT. The first method proposed by Waterman to estimate the dose to the urethra from CT scan appears in brachytherapy [139]. As mentioned before, in that study, the urethra was estimated as a geometric surrogate based upon the prostate centerline. They found a good correspondence of the urethral doses

(D10, D25 and D50) between this model and the urinary catheter. This model was latterly evaluated in brachytherapy patients with visible catheter [155]. They showed that a surrogate defined at the geometric center of the prostate may significantly overestimate the dose to the urethra. A surrogate urethra model by considering a slight deviation of 30 degrees anteriorly with respect to the central axis was proposed by Bucci et al. [138]. This deviation provided a better dose estimate than Waterman [139]. Here, we evaluated these two urethral surrogates with the same dataset. Compared to them our method performed better in the same regions (overall CLD=3.25± 1.2mm with MABUS vs 3.91±1.46 with the deviated surrogate and 6.11±1.96 mm with the centered surrogate). The difference with the central axis is less pronounced in the bottom third, suggesting that a good approximation of the urethra in this region may be achieved. However in the upper part, although the slight deviation of 30 degrees offers a good approximation for some individuals, this is not the case for all the patients. Our multi-atlas approach enabled to devise a strategy aimed at finding the n most suitable atlas within the dataset by defining a similarity metric based on simple prostate and bladder geometric features. The prostate and bladder segmentations were used as they are generated during the dose planning clinical protocol.

Considering the proposed features, the large interindividual variability was captured as shown in the similarity map (Figure 2.3). Thus, with the exception of some outliers, one can find good candidates in the atlas selection procedure for each considered individual. The atlas selection step is indeed crucial to accurately segment the urethra as demonstrated in the leave one out experiments. With the proposed features, it has been shown that a trend arises when correlating similarity metric (distance) and segmentation outcome. Then, adding multiple atlases improves the accuracy in a voting strategy when fusing labels from the closest atlas [148]. Other global or local features based on CT intensity or shape descriptors may have been proposed (mutual information, cross correlation, SPHARM, etc.), but with a limitation concerning accuracy assessment due to the presence of the catheter. The fusion step takes into account the interindividual similarity by weighting the contribution to the probability map via an exponential function, which has been shown to be more performant than simple averaging [152]. This strategy led us to limit to 10, the number of selected atlas, as by adding more their contribution is vanished in an exponential function. Highly contributive is the prostate non-rigid registration based on the Laplacian scalar field [109]. Indeed, the main feature brought by the Laplacian is the computation of a normalized structural description comparable across individuals, as opposed to classical distance maps.

We used for evaluation the scores based on distance centerlines as proposed by [155] as the urethra is considered as a path within a tiny tube-like structure. Other scores based on volume overlap (Dice, Jaccard) are not suited here for assessing segmentation accuracy. The obtained scores enable the method to be used within a perspective of assessing dose to the urethra within an acceptable margin. However, improvements may be done within a perspective of IMRT planning and dose escalation. The atlas segmentation method may exhibit some limitations if we consider the hypothesis of that the urethra keeps a similar shape with and without a catheter. One way to demonstrate that hypothesis will be the MRI where the urethra might be visible. However, in brachytherapy studies the urethra position is only given by the catheter [61–63]. Further limitations concern the patient variability in IMRT-treated patients compared to the atlas database as prostate and bladder volumes may be higher because of the patient selection and the presence of a catheter.

2.5 Conclusion

When applying the proposed method to an independent data set of patients treated with IMRT, a large difference was found in the dose to the urethra (high dose range) compared to the prostate. These findings suggest further studies to be performed on urinary toxicity by quantifying the dose to the urethra as reported in brachytherapy [15-17]. New multimodal models combining also dose to the bladder with clinical factors, biological parameters and other multimodal data within a radiomics framework could provide new insights into the urinary toxicity.

In this chapter, we developed a methodology for segmenting the prostatic urethra on the planning CT images of patients treated with EBRT. Segmenting the urethra on the planning CT images paves the way for future studies for investigating the involvement of this structure on urinary side-effects and potentially improving our understanding of urinary toxicity after prostate cancer radiotherapy. Furthermore, by analyzing the dose at fine scales and integrating spatial descriptors of the dose to both the bladder and the urethra might provide new insights into the urinary toxicity and enable the development of more reliable predictive models.

The next part of this thesis explores two methodologies for performing population analysis using 3D images. The purpose is to characterize spatial dose patterns at sub-organ level and anatomical regions implied in toxicity following prostate cancer radiotherapy. Going beyond the whole bladder DVH to image subunit scales, we attempt to identify local spatial dose descriptors and develop predictive models of urinary toxicity, using bladder DSMs (Chapter 3) and urethro-vesical DVMs (Chapter 4).

Part III

Going beyond bladder dose- volume histograms: novel methods for spatially analyzing local dose distribution

3 Urinary toxicity prediction using dose-surface maps

This third chapter describes a methodology for spatially characterizing the dose distribution on the bladder surface using DSMs. The first goal of this chapter was to improve the existent approaches for analyzing DSMs and propose a methodology that would allow the exploration of the entire bladder surface. The second goal, was to identify symptom-related sub-surfaces of the bladder that are potentially more predictive that the whole bladder DVH. The third goal was to compare the results with previous bladder DSM studies and assess the reproducibility of the results.

Part of the work presented in this Chapter was conducted at the Istituto Tumori di Milano, in Italy, in collaboration with Tiziana Rancati and Alessandro Chicchetti. This work was presented at the ESTRO 38 conference in Milano as an oral communication and has been submitted for publication to the journal Radiotherapy and Oncology. For the shake of coherence of the thesis manuscript, the original paper has been divided in two parts: the one is presented in this chapter and the other part in the Chapter 6, section 6.1.

Mylona E, Cicchetti A, Rancati T, Palorini F, Fiorino C, Supiot S, Magne N, Creange G, Valdagni R, Acosta O, de Crevoisier R. *Local dose analysis to predict acute and late urinary toxicities after prostate cancer radiotherapy : assessment of cohort and method effects.* Radiother. Oncol. (Submitted to Radiotherapy and Oncology)

Mylona E, Cicchetti A, Rancati T, Palorini F, Supiot S, Magne N, Creange G, Acosta O, de Crevoisier R. *Predicting urinary toxicity via 2D and 3D dose map analyses in prostate cancer radiotherapy.* Radiother. Oncol. 2019;133:326. (Presented at the ESTRO 38 as an oral communication)

3.1 Introduction

Dose-surface maps (DSM) of the bladder have recently been applied to urinary toxicity studies only by two research teams, providing evidence of spatially variable dose-response relationship with respect to acute [8,9] and late [10] urinary symptoms. These studies, however, were limited to the global urinary toxicity [156], included a small number of patients [60], or the discriminative power of the models was disregarded [106]. More importantly, since spatial dosimetric patterns can be surrogates of underlying population characteristics, the generalization of these observations is not established. In addition, in all the studies, absolute DSMs were used which restricted their analysis to only a few centimeters form the bladder base because of the variable bladder extension across the population.

In the following sections we will present a methodology for analyzing the whole bladder surface using normalized DSMs. Then we compare the DSMs pixel-wise between patients with and without toxicities in order to identify sub-surfaces (Ssurf) of the bladder where significant dose differences exist. To answer the question whether the impact of dose to specific subregions is associated specific symptoms we compare our results with those from previous studies.

3.2 Population data set, treatment, and urinary toxicity

The study included a total of 272 patients with localized prostate cancer treated with intensity-modulated radiation therapy/image-guided radiation therapy (IMRT/IGRT) between May 2008 and July 2018 within two multicentric prospective phase III trials (STIC-IGRT) [157] and (PROFIT, normofractionated arm) [158]. The mean age of the patients was 70 years (range: 52-81). The target volume included the prostate and seminal vesicles (SV), avoiding the pelvic lymph nodes. Target volume and organs at risks (bladder, rectum, and femoral heads) were delineated on computed tomography (CT) slices according to the GETUG and PROFIT recommendations. IMRT combined with IGRT was used to deliver a total dose of 78 Gy (in case of cone-beam CT [CBCT]) or 80 Gy (in case of fiducials) to the prostate over 8 weeks and 46 Gy to the SV over 4.6 weeks, at 2 Gy/fraction. Bladder dose-volume histograms (DVHs) were compiled either with GETUG recommendations as maximum dose <80 Gy or with PROFIT recommendations as 50% and 70% of the bladder wall receiving less than 53 Gy and 71 Gy, respectively. Patient and treatment characteristics are described in Table 3.1. All 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 median follow-up was 50 months (range: 6–102 months). Acute (≤3 months from RT start date) and late (> 3months) urinary toxicity was scored using the CTCAE v.3.0 (210 patients from the STIC-IGRT trial) or the RTOG (62 patients from the PROFIT trial) radiation morbidity-scoring schema. We considered 20 endpoints: 5 symptoms (incontinence, retention, dysuria, hematuria, frequency), for both grade $\geq$ 1 and grade $\geq$ 2, and for acute and late urinary toxicity. The 5-year toxicity rates were estimated using Kaplan-Meier (KM) analysis. Table 3.2 displays the number of events and rates of acute and 5-year late urinary toxicity (grade ≥1 and grade ≥2) by symptom as calculated by KM. Among the 20 endpoints, 4 symptoms

(acute grade≥2 incontinence and hematuria, late grade≥2 dysuria and hematuria) were excluded from the analysis due to the absence or very low number of events.

Table 3.1 Patient and treatment characteristics of the population

Characteristics		Values
Age (years, mean ±SD)		69 ± 6
TURP		7 %
Previous abdominal surgery		20 %
Diabetes		12 %
Anti-coagulant treatment		25 %
Antihypertensive treatment		33 %
Hypercholesterolemia treatment		19 %
Baseline symptoms	Hematuria	0%
	Dysuria	<1%
	Retention	8%
	Incontinence	<1%
	Frequency	14%
ADT		28 %
Prescribed dose	80 Gy	32 %
	78 Gy	68 %
Prostate volume* (cc, mean ± SD)		52.1 ± 21.3
Bladder volume* (cc, mean ± SD)		227.6 ± 145.5

SD: standard deviation; TURP: transurethral resection of the prostate; ADT: Androgen deprivation therapy.: calculated from the planning CT*

Table 3.2 Rates of grade ≥ 1 and grade ≥ 2 acute and late urinary toxicity (by symptom)

Endpoints	Acute toxicity (No. of cases and rates)			
	Grade ≥ 1		Grade ≥ 2	
Incontinence	17 (7%)		0	
Frequency	202 (80%)		67 (26%)	
Retention	106 (42%)		49 (19%)	
Dysuria / Pain	113 (45%)		19 (7%)	
Hematuria	7 (3%)		0	
Endpoints	**Late 5-year toxicity (No. of cases and rates)**			
	Grade ≥ 1		**Grade ≥ 2**	
	No. of cases	Kaplan-Meier (95% CI)	No.of cases	Kaplan-Meier (95% CI)
Incontinence	33	15% (10-20%)	8	4% (1-7%)
Frequency	125	55% (48-62%)	27	11% (7-15%)
Retention	40	34% (26-42%)	23	10% (5-15%)
Dysuria / Pain	31	13% (8-18%)	2	1% (0-2%)
Hematuria	23	10% (6-14%)	5	2% (0-4%)

3.3 Materials and methods

3.3.1 DSM construction and pixel-wise analysis to identify sub-surfaces

Absolute DSMs were generated from the planning CT delineations and dose distributions using a dedicated software (VODCA, MSS Medical Software Solutions GmbH, Hagendorn, Switzerland). The workflow is shown in Figure 3.1. 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) [60]. 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). As opposed to previous DSM analyses, 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 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. The p-value maps were threshold at p ≤0.01 to identify sub-surfaces (Ssurf) of the bladder with significant dose differences.

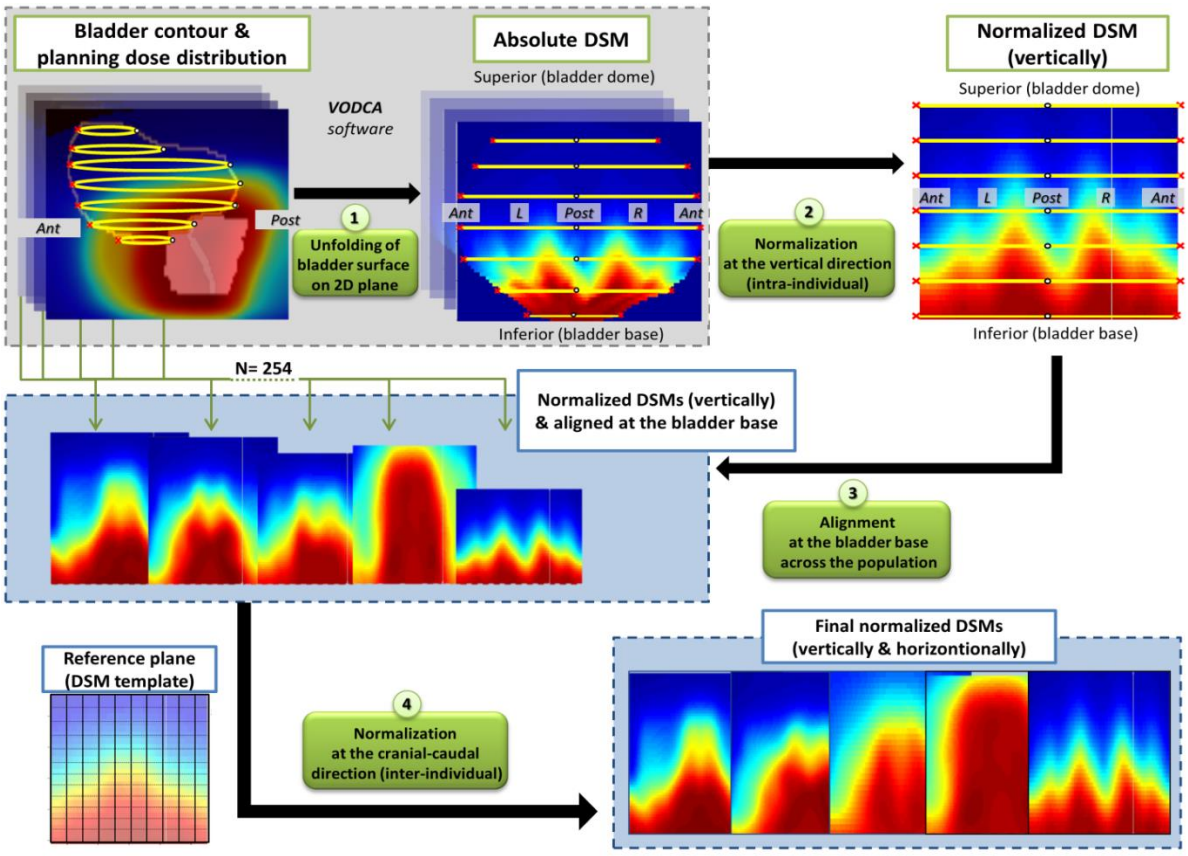


Figure 3.1 Workflow of dose-surface map (DSM) construction
Ant=anterior; Post=posterior; R=right; L=left

3.3.2 Prediction capability of dosimetric and clinical parameters

UNIVARIATE ANALYSIS

Univariate analysis was performed independently for each Ssurf and the whole bladder to identify the most predictive dose bins of the DVH/ DSH. Logistic regression was used for acute toxicity, and the discriminative performance was assessed with the area under the ROC curve (AUC). Cox regression was used for late toxicity. The 5-year discriminative performance was measured with the area under the time-dependent ROC curve (tAUC) as described in [159], 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.

MULTIVARIATE ANALYSIS

Multivariate logistic/Cox models were constructed including clinical parameters and pre-selected dosimetric variables (dose bins) from the univariate analysis. Models were constructed using two methods: i) backward elimination and ii) the least absolute shrinkage and selection operator (LASSO) method [160,161], which enables the simultaneous analysis of the correlations between the features and also prevents overfitting [162,163], as described below. The AUC/ tAUC and 95% CIs from 1000 bootstrap replicates were used to evaluate the models' discriminative performance.

Variable selection using LASSO

Penalized regression methods, such as the Least Absolute Shrinkage Selection Operator (LASSO) [160], allow the generation of parsimonious models that balance accuracy and model complexity. Regularization by LASSO-type penalties usually leads to sparse solutions and performs both variable and model selection. LASSO can be applied to many types of regression, including logistic and Cox regression.

LASSO parameters are estimated by maximizing a log-likelihood criterion in which an L1-norm penalizes large absolute values of the parameter estimates. Many regression coefficients are shrunk to zero and only a few others remain in the model. In this way, coefficient estimates of redundant features can be forced to be exactly zero, resulting in an optimal sparse model. The degree of shrinkage depends on the tuning parameter lambda (λ). The choice of the shrinkage parameter, thus, drives the model selection. Moreover, the shrinkage introduces bias in the parameter estimates, which reduces the variance of the parameter estimates. This phenomenon is known as the bias-variance trade-off, which could improve the model's prediction performance.

To avoid dealing with high dimensional data in the regression analysis, the set of dosimetric variables was reduced prior to LASSO by selecting a single dosimetric variable per DVH/DSH (the bin corresponding to AUCmax). As a result, no more than 16 variables were used as input for feature selection. Standardization of all of the variables was also performed prior the LASSO feature selection to guarantee that the penalty term would treat the variables in a comparable way. The number of selected features

and their coefficients were tuned with the parameter lambda (λ) optimized using 5-fold cross-validation (CV) [162]. This was performed using the cv.glmnet routine in the *glmnet* package in R. At each iteration, the data set was divided into 5 sub-sets (folds). Four were used to train the model and the fifth set was used for validation. The process was repeated 5 times so that each sub-set was used once as a validation set. For the logistic regression models, the minimum deviance was selected as the criterion for finding the tuning parameter λ, while for the Cox regression, the log partial-likelihood was used. The cross-validated LASSO method was repeated 100 times, resulting in 500 models in total to identify different combination of features and coefficients yielding minimum deviance values. We then applied the "one standard error" rule [163] to select the most sparse model (the model with the fewest non-zero parameters) whose deviance is no more than one standard error greater than the deviance of the "best" model (the model among the 500 with minimum deviance).

3.4 Results

Three subregions were identified on the bladder surface (only for grade ≥1 toxicity endpoints) as predictive of specific urinary symptoms: acute and late retention and late dysuria. For acute and late retention, the two subregions where found on the posterior bladder surface corresponding to the intermediate dose region with dose differences 7.9±2.9 Gy and 10.9±1.9 Gy, respectively. For late dysuria a subregion was identified on the anterior-lateral and mostly inferior bladder surface corresponding to intermediate-high dose region with dose differences 8.3±2.7 Gy.

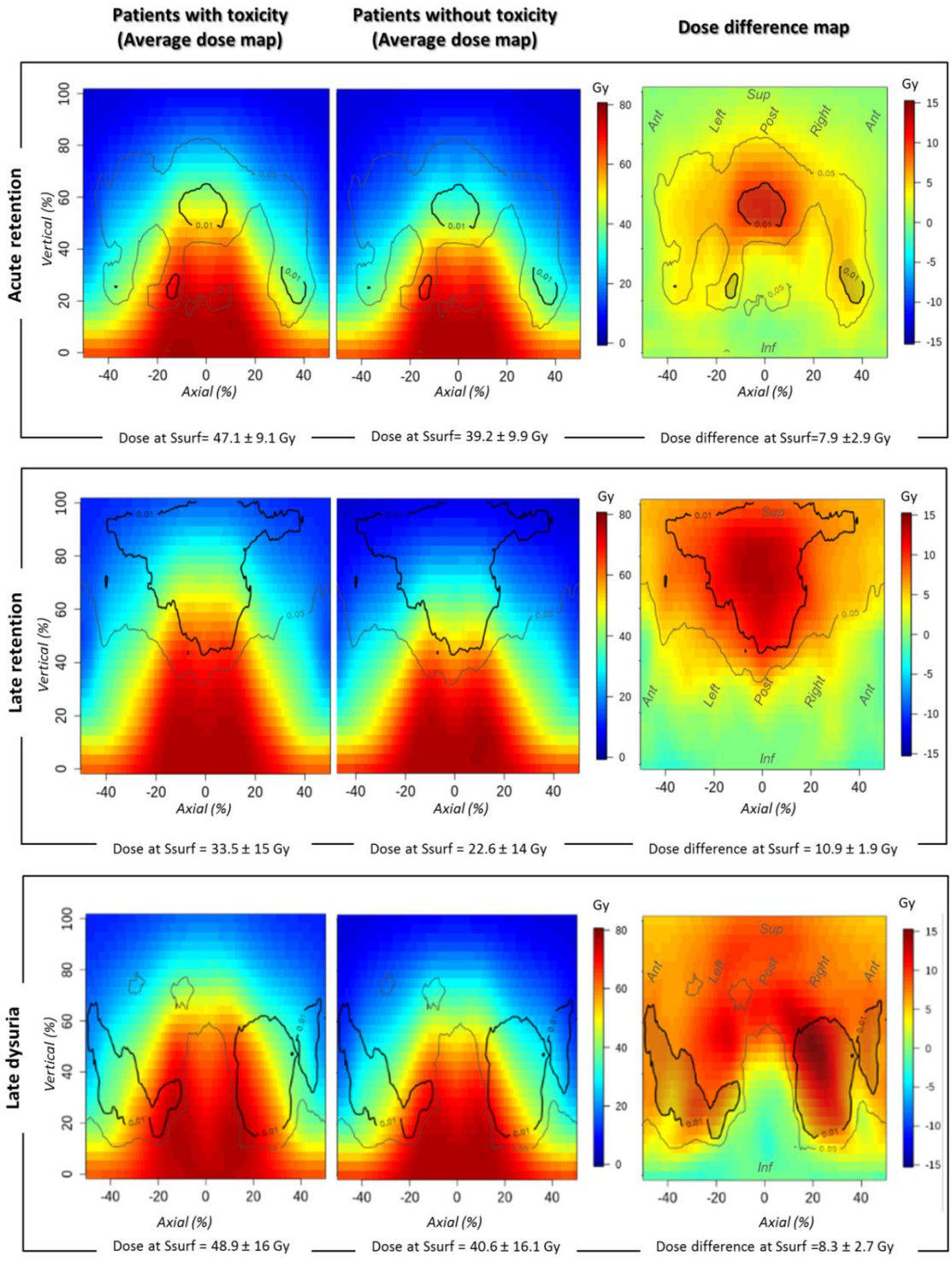


Figure 3.2 Symptom-related sub-surfaces (Ssurf) of statistically significant dose differences between patients with/without toxicity from DSM analysis

Figure 3.2 shows the DSMs of the average dose distribution for patients with (left) and without (middle) urinary toxicities and the corresponding dose difference maps (right), for the 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 (for the region corresponding to p-value ≤0.01).

UNIVARIATE ANALYSIS

Table 3.3 displays the association between urinary endpoints and the DSHs of the Ssurf and the DVH of the whole bladder. For the three Ssurf, AUC was 0.64 for acute retention, 0.68 for late retention and 0.74 for late dysuria. For the whole bladder the AUCs was 0.60 for acute retention, 0.67 for late retention, 0.72 for late dysuria and 0.65 for late hematuria.

Table 3.3 Univariate analysis of the DVH/DSH for the whole bladder and the sub-surfaces (Ssurf)

A. Acute toxicity (Grade ≥1)					
Symptom	**Region**	**Most predictive DVH/DSH bin (range of predictive bins)***	**p value**	**OR (95% CI)**	**AUC (95% CI)**
Retention	Whole bladder	V79 (V77-V79)	0.05	1.06 (1.01-1.13)	0.60 (0.51-0.67)
	Ssurf	S42 (S15-S60)	<0.01	1.01(1-1.02)	0.64 (0.56-0.72)
B. Late toxicity (Grade ≥1)					
Symptom	**Region**	**Most predictive DVH/DSH bin (range of predictive bins)***	**p value**	**HR (95% CI)**	**tAUC at 5 years (95% CI)**
Retention	Whole bladder	V19 (V4-V78)	<0.01	1.02 (1-1.04)	0.67 (0.59-0.75)
	Ssurf	S38 (S3-S79)	<0.01	1.02 (1.01-1.03)	0.68 (0.60-0.75)
Dysuria	Whole bladder	V67 (V5- V78)	<0.01	1.03 (1.01-1.05)	0.72 (0.63-0.82)
	Ssurf	S70 (S8-S80)	<0.01	1.07 (1.03-1.1)	0.74 (0.64-0.83)
Hematuria	Whole bladder	V7 (V6-V18)	0.04	1.04 (1-1.07)	0.65 (0.55-0.75)
	Ssurf	NS	-	-	-

*Ssurf: Sub-surface; OR: Odds ratio; HR: Hazard ratio; CI: Confidence interval; AUC: Area under the ROC curve; tAUC=time dependent AUC; NS: not significant (p > 0.05); *DVH bin with the highest AUC and range of bins with statistically significant p-value (≤ 0.05)*

Clinical parameters significantly associated with acute urinary symptoms were age, transurethral resection of the prostate (TURP), previous abdominal surgery, anti-coagulant treatment, anti-hypertensive treatment, prostate volume, and baseline symptoms, as shown in Table 3.4. The impact of hematuria, dysuria and incontinence baseline symptoms on toxicity has not been assessed due to the very low occurrence of these symptoms (<1%). Clinical parameters significantly associated with late urinary symptoms were age, transurethral resection of the prostate (TURP) previous abdominal surgery, diabetes, anti-coagulant treatment, anti-hypertensive treatment, androgen-deprivation therapy (ADT), prostate volume, and baseline symptoms, as shown in
Table 3.5. The impact of hematuria, dysuria and incontinence baseline symptoms on toxicity has not been assessed due to the very low occurrence of these symptoms (<1%).

Table 3.4 Univariate analysis testing the impact of patient/ treatment characteristics on acute urinary toxicity

Parameters	Grade ≥ 1 toxicity										Grade ≥ 2 toxicity					
	Hematuria		Dysuria		Retention		Frequency		Incontinence		Dysuria		Retention		Frequency	
	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)
Age	0.39		**0.01**	**0.94 (0.9-0.99)**	0.87		0.44		0.8		0.68		0.31		0.72	
TURP	**0.04**	**6.9 (1.07-4.8)**	0.41		**0.04**	**0.27 (0.07-0.96)**	**<0.01**	**0.19 (0.07-0.54)**	1		0.47		0.15		0.16	
Previous abdominal surgery	**0.04**	**10.2 (1.12-84)**	0.38		0.96		0.67		0.54		0.47		0.58		0.43	
Diabetes	0.99		0.63		0.5		0.33		0.1		0.99		0.52		0.32	
Anti-coagulant treatment	**0.01**	**8.1 (1.6-42.8)**	0.92		0.3		0.8		0.08		0.71		0.17		0.43	
Antihypertensive treatment	0.17		0.41		0.56		0.57		1		0.04	**4.1 (1.09-15.5)**	0.46		0.78	
Hypercholesterolemia treatment	0.73		0.76		0.93		0.09		0.47		0.33		0.5		0.86	
ADT	0.44		0.07		0.75		0.7		0.81		0.39		0.44		0.08	
Prostate volume	0.21		0.12		0.75		0.07		0.56		0.01	**1.03 (1.01-1.05)**	0.42		0.27	
Bladder volume	0.97		0.39		0.31		0.06		0.14		0.33		0.59		0.19	
Baseline retention	0.99		0.62		**0.04**	**3(1.03-0.76)**	0.19		1		0.99		**0.01**	**3.53 (1.3-9.6)**	**<0.01**	**4.4 (1.6-12.3)**
Baseline frequency	0.38		0.19		0.39		**0.04**	**3.17 (1.03-9.76)**	1		0.6		0.24		**<0.01**	**3.2 (1.4-7.3)**

Table 3.5 Univariate analysis testing the impact of patient/ treatment characteristics on late urinary toxicity

Parameters	Grade ≥ 1 toxicity										Grade ≥ 2 toxicity					
	Hematuria		Dysuria		Retention		Frequency		Incontinence		Retention		Frequency		Incontinence	
	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)
Age	0.22		0.81		**0.01**	**0.95 (0.91-0.99)**	**0.05**	**1.03 (1-1.06)**	**0.03**	**1.07 (1-1.13)**	**0.02**	**0.92 (0.86-0.98)**	0.28		0.49	
TURP	0.86		0.24		0.44		0.51		**<0.01**	**5.72 (2.02-16.22)**	0.66		0.13		**<0.01**	**17.7 (2.95-76.26)**
Previous abdominal surgery	0.12		0.06		**<0.01**	**2.17 (1.25-3.77)**	0.4		0.45		**0.05**	**2.62 (1-6.78)**	0.3		0.06	
Diabetes	0.74		0.98		0.29		0.86		**0.03**	**2.58 (1.11-5.97)**	1		**0.01**	**3.08 (1.3-7.28)**	0.82	
Anti-coagulant treatment	0.26		0.72		0.24		0.24		**0.02**	**2.32 (1.16-4.63)**	0.74		0.68		0.92	
Antihypertensive treatment	0.91		0.71		0.63		**0.04**	**1.58 (1.01-2.47)**	0.39		**0.05**	**0.33 (0.11-1.02)**	0.74		0.53	
Hypercholesterolemia treatment	0.22		0.42		0.68		0.69		0.55		0.68		0.2		0.68	
ADT	0.7		0.11		0.23		**<0.01**	**1.91 (1.3-2.74)**	0.91		0.1		0.08		0.69	
Prostate volume	0.76		0.67		0.61		0.46		0.43		0.16		0.75		0.85	
Bladder volume	0.13		0.08		0.06		0.06		0.08		0.34		0.37		0.52	
Baseline retention	0.4		0.94		**<0.01**	**2.59 (1.54-4.35)**	0.08		0.85		**<0.01**	**3.64 (1.66-7.96)**	0.95		1	
Baseline frequency	0.4		0.64		0.12		**<0.01**	**3.78 (2.4-6.05)**	0.84		0.19		**<0.01**	**7.19 (2.4-21.62)**	0.53	

MULTIVARIATE ANALYSIS

Table 3.6 displaces the results of the multivariate analysis with backward elimination. Table 3.7 shows the detailed results of the LASSO multivariable regression for the different symptoms. The two methods resulted in similar models.

Table 3.6 Parameters affecting acute and late urinary toxicity in multivariate analysis using backward elimination

A. Acute toxicity (logistic regression)					
Grade ≥1					
Endpoints	**Parameters**	**p-value**	**OR (95% CI)**	**Model p-value**	**AUC (95% CI)**
Retention	Ssurf (S42)	0.02*	1.01 (1-1.02)	<0.01	0.70 (0.63-0.78)
	TURP	0.06	0.27 (0.07-1.03)		
	Baseline retention	<0.01*	4.39 (1.48-13.0)		
Dysuria	Age	0.01*	0.94 (0.9-0.99)	<0.01	0.63 (0.54-0.71)
	ADT	0.07	0.52 (0.25-1.07)		
Frequency	TURP	<0.01*	0.07 (0.01-0.37)	<0.01	0.77 (0.68-0.85)
	Hypercholesterolemia treatment	0.01*	0.3 (0.11-0.78)		
	Baseline frequency	0.02*	5.75 (1.4-23.6)		
Grade ≥2					
Retention	TURP	0.17	0.23 (0.03-1.85)	<0.01	0.64 (0.55-0.72)
	Anticoagulant treatment	0.09	0.36(0.11-1.17)		
	Baseline retention	<0.01*	3.69 (1.43-9.5)		
Dysuria	Antihypertensive treatment	0.04*	3.58 (1.39-15.65)	<0.01	0.72 (0.55-0.88)
	Prostate volume	0.02*	1.03 (1.01-1.05)		
Frequency	TURP	0.2	0.36 (0.08-1.71)	<0.01	0.61 (0.54-0.68)
	Baseline frequency	0.01*	3.17 (1.31-7.64)		
B. Late toxicity (Cox regression)					
Grade ≥1					
Endpoints	**Parameters**	**p-value**	**HR (95% CI)**	**Model p-value**	**tAUC (95% CI)**
Retention	Ssurf (S52)	<0.01*	1.02 (1-1.03)	<0.01	0.73 (0.68-0.78)
	Age	0.02*	0.95 (0.91-0.99)		
	Previous abdominal surgery	0.07	1.69 (0.96-2.97)		
	Baseline retention	<0.01*	3.17 (1.66-5.05)		
Incontinence	Age	0.06	1.09 (0.99-1.19)	<0.01	0.81 (0.71-0.88)
	TURP	<0.01*	5.06 (1.73-14.76)		
	Diabetes	<0.01*	5.59 (2.05-15.21)		
Hematuria	-	-	-	-	-
Dysuria	Ssurf (S70)	<0.01*	1.06 (1.03-1.10)	<0.01	0.78 (0.69-0.85)
	ADT	0.1	2.27 (0.89-5.76)		
Frequency	Bladder volume	0.08	1.02 (0.99-1.03)	<0.01	0.8 (0.75-0.87)
	Antihypertensive treatment	0.02*	1.76 (1.1-2.83)		
	ADT	<0.01*	2.92 (1.72-4.95)		
	Baseline frequency	<0.01*	4.28 (2.55-7.17)		
Grade ≥2					
Retention	Age	0.12	0.95 (0.89-1.03)	<0.01	0.75 (0.63-0.84)
	Previous abdominal surgery	0.11	2.2 (0.8-6)		
	Antihypertensive treatment	0.1	0.37 (0.12-1.21)		
	Baseline retention	<0.01*	4.49 (1.64-12.26)		
Frequency	Hypercholesterolemia treatment	0.09	0.18 (0.02-1.38))	<0.01	0.77 (0.66-0.86)
	Baseline frequency	<0.01*	5.29 (1.8-15.6)		
Incontinence	TURP	<0.01	17.7 (2.95-76.26)	<0.01	0.74 (0.62-0.87)

*TURP: transurethral resection of the prostate; ADT: Androgen deprivation therapy; OR: Odds ratio; HR: Hazard ratio; CI: Confidence interval; AUC: Area under the ROC curve; tAUC: time-dependent area under the ROC curve; *significant variables (p≤0.05); Model p-value: Global statistical significance of the model as defined by the Likelihood ratio test.*

Table 3.7 Parameters affecting acute and late urinary toxicity in multivariate analysis using LASSO

A. Acute toxicity (logistic regression)				
Grade ≥1				
Endpoints	**Parameters**	**beta**	**OR (e^{beta})**	**AUC (95% CI)**
Retention	Ssurf (S42)	0.0043	1.01	0.70 (0.62-0.78)
	TURP	-0.2190	0.80	
	Baseline retention	0.7763	2.17	
Dysuria	Age	-0.0318	0.96	0.63 (0.55-0.72)
	ADT	-0.1531	0.85	
Frequency	Bladder volume	-0.0002	0.99	0.79 (0.7-0.88)
	TURP	-1.2162	0.30	
	Hypercholesterolemia treatment	-0.4177	0.66	
	Baseline frequency	0.5518	1.73	
Grade ≥2				
Retention	TURP	-0.2528	0.78	0.63 (0.54-0.72)
	Anticoagulant treatment	-0.2320	0.79	
	Baseline retention	0.774	2.17	
Dysuria	Antihypertensive treatment	0.3210	1.38	0.72 (0.55-0.88)
	Prostate volume	0.0127	1.01	
Frequency	Baseline frequency	0.5908	1.81	0.59 (0.52-0.65)
B. Late toxicity (Cox regression)				
Grade > 1				
Endpoints	**Parameters**	**beta**	**HR (e^{beta})**	**tAUC (95% CI)**
Retention	Ssurf (S52)	0.0089	1.01	0.72 (0.67-0.77)
	Age	-0.0285	0.97	
	Previous abdominal surgery	0.3268	1.39	
	Baseline Retention	0.8131	2.26	
Incontinence	Age	0.0355	1.04	0.81 (0.71-0.88)
	TURP	1.2058	3.34	
	Diabetes	1.2925	3.64	
Dysuria	Ssurf (S70)	0.0411	1.04	0.73 (0.63-0.81)
Frequency	Age	0.0072	1.01	0.81 (0.75-0.86)
	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	Age	-0.0414	0.96	0.77 (0.67-0.85)
	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	Diabetes	0.4673	1.60	0.81 (0.72-0.90)
	Hypercholesterolemia treatment	-0.4663	0.63	
	Baseline frequency	1.1170	3.06	
Incontinence	TURP	2.0501	7.77	0.82 (0.65-0.97)
	Previous abdominal surgery	0.6994	2.01	

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

3.5 Discussion

An affine transformation of the DSMs was implemented to represent the entire population on a common reference frame while preserving the spatial information of the dose distribution to the entire surface of the bladder. Subsequent pixel-wise comparisons allowed us to investigate the local dose-effect relationship for acute and 5-years late urinary toxicity after prostate cancer RT. This fine-scale analysis of the dose distribution indicated the presence of a spatially variable dose-effect relationship on the surface of the bladder. The doses to three Ssurf located in the inferior hemisphere of the bladder were found to be good predictors for three grade ≥ 1 symptoms (acute retention, late retention and late dysuria).

For late dysuria, a large surface of the bladder seems to be involved including the anterior-inferior and lateral region. For acute retention, one Ssurf was found on the posterior bladder region, while for late retention the Ssurf was located on the posterior-superior bladder. Although the relationship between bladder dose and obstructive symptoms has not been explored in previous DSM studies, our results are in line with other urinary toxicity studies [54,57,58], suggesting that high/intermediate dose might be impacting on the manifestation of obstructive effects.

Concerning the discriminative power of the models in univariate analysis, the doses to the Ssurf were slightly more predictive than the dose to the whole bladder. For acute retention the Ssurf was moderately predictive (AUC=0.64) but outperformed the whole bladder model (AUC=0.60, p-value=0.05). For late toxicities, the predictive capabilities of Ssurf and the whole bladder were marginally different. More specifically, the dose to the Ssurf for late retention resulted to AUC=0.68 while the dose to whole bladder resulted to AUC=0.67. Similarly, for late dysuria the Ssurf AUC was 0.74 while the bladder AUC was 0.72.

3.6 Comparison with previous studies to estimate the "cohort-effect"

To date, bladder DSMs coupled with pixel-wise comparisons have been investigated by two research teams, providing evidence of spatially variable dose-response relationship with respect to acute [60,156] and late [106] urinary symptoms, as shown in
Table 3.8. Concerning acute toxicity analyzed by DSMs, urinary frequency/urgency was significantly related to the dose at 5–12 mm posterior to the bladder base, corresponding to trigone [60]. Concerning late toxicity, another DSM analysis [106] found incontinence to be related to the region lateral to trigone. They also reported that the dose to the anterior-inferior and posterior-superior bladder surface was associated with worsening of symptoms as measured by IPSS, dysuria, and hematuria. These studies, however, were limited to the global urinary toxicity [156], included a small number of patients [60], or the discriminative power of the models was disregarded [106]. More importantly, since spatial dosimetric patterns can be surrogates of underlying population characteristics, the generalizability of these observations is not established. Confirmation in independent cohorts is essential to answer the question whether the impact of dose to specific subregions is associated specific symptoms.

The reproducibility of the DSM results was assessed through comparison between the results obtained in our cohort and the results from the aforementioned cohorts.
Table 3.8 summarizes the population and treatment characteristics of these three DSM studies used for comparison. 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.

Overall, five symptoms with similar inter-study definitions were considered for comparison (acute frequency and retention, late dysuria, incontinence, and hematuria). Among these five symptoms, four Ssurf were identified in other cohorts, and two in our cohort. Only for one symptom, late dysuria, was Ssurf was found in both our study and another study [106], with good spatial agreement (inferior-anterior-lateral).

Table 3.8 Overview of existing bladder DSM studies

Acute toxicity																							
Method	**Study**	**N**	**Treatment technique**	**PD (fraction)**	**Toxicity scoring**	**Global toxicity**			**Frequency**			**Incontinence**			**Retention***			**Dysuria**			**Hematuria**		
						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		
	Improta et al. 2016-	375		76 Gy (1.8-2 / 2.2-2.7 Gy)		Inf-post	0.70	0.66	Not studied			Not studied			Not studied			Not studied			Not studied		
	Current study	272	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 (No events)		
Late toxicity																							
Method	**Study**	**N**	**Treatment technique**	**PD (fraction)**	**Toxicity scoring**	**Global toxicity**			**Frequency**			**Incontinence**			**Retention**			**Dysuria**			**Hematuria**		
						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 (symptoms)	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
	Current study	272	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

*"weak stream" as defined by IPSS was considered to be similar to retention

Color shading represents the symptoms that are comparable across DSM studies

The reproducibility of identified symptom related Ssurf between cohorts is weak in our study, potentially due to the heterogeneity of the compared populations The local dose-effect relationship was confirmed in our population for late dysuria but remains unclear for other symptoms, 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 3.8 shows large differences between cohorts. Statistical differences between the studied populations are expected to be strongly influential. For example, the first pilot study of bladder DSM analysis for acute urinary toxicity [60] included only 72 patients. 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% [60] versus 42% in our study, for acute retention. Nevertheless, there was one symptom (late dysuria) for which Ssurf was identified and confirmed in two independent cohorts (anterior-inferior-lateral bladder surface, receiving 40–60Gy).

3.7 Conclusion

Dose-surface maps can unveil the heterogeneous intra-organ radiosensitivity by identifying symptom-specific subregions of the bladder surface that might be more predictive than the dose to the whole bladder. This has already been proposed by others. However, this study demonstrated that spatial dosimetric patterns can be surrogates of underlying population characteristics since, compared to previous DSM studies, we were not able to confirm a local dose-toxicity relationship but for one symptom. Consequently the local dose-toxicity relationship is not necessarily translated into a causality relationship. Therefore, generalization of these observations is not a foregone conclusion. Careful assessment and external validation of such models is indispensable to establish clinically meaningful dosimetric constraints on the bladder surface. In addition, further investigation is needed to prove the improvement on prediction capabilities brought by these models compared to the whole bladder DVH.

In this chapter, we tackled the problem of characterizing dose patterns on the bladder surface possibly correlated with toxicity, using DSMs. Indeed, DSM remains an attractive method for investigating the dose-response relationship at a pixel level. It is, nevertheless, limited by two main factors. First, the two dimensional nature of DSMs is unavoidably restricted to the surface of the organ, assuming that a dose-response relationship exist only on the bladder surface which is not necessarily true. Second, DSMs are limited to a single organ, while in reality, more than one structures can be contributing to urinary toxicity. This is particularly true for urinary toxicity, as symptom manifestation might be associated not only with the dose to the bladder but also with the dose to the urethra or other sub-parts of these structures. To overcome these limitations with the premise to improve prediction of urinary toxicity, in the following chapter we propose a new methodology to perform population analysis of the entire 3D

dose distribution in the bladder and the urethra, using DVMs, which aims at identifying 3D anatomical sub-regions implied in urinary toxicity following prostate cancer radiotherapy.

4 Urinary toxicity prediction using dose-volume maps

In this chapter we present a methodology for analyzing dose-volume maps via voxel-wise comparisons. The proposed framework allows to investigate the local dose-effect relationship in the bladder and the urethra with the objective to identify sub-regions that are potentially more predictive of urinary toxicity than the whole bladder DVH. First we describe an original and robust multi-organ non-rigid registration strategy for anatomically aligning the population and propagating the dose distributions to a common space. Then, the dose-volume maps of patients with and without toxicities were compared voxel-by-voxel allowing the identification of symptom-related subregions where statistically significant dose differences exist. The subregions where propagated to the native space of each patient where sub-region DVHs were computed and their discriminative power with respect to the DVH of the whole bladder was evaluated. Finally, the generalizability of these models was assessed through external validation on a large, independent population.

The content of this chapter has been published to the International Journal of Radiation Oncology, Biology, Physics (with exception paragraph 4.4). The section 4.4 contains preliminary (unpublished) results of a work realized in Sir Charles Gairdner Hostital, Western Australia in collaboration with Martin Ebert and Angel Kennedy.

Mylona E, Acosta O, Lizee T, Lafond C, Crehange G, Magné N, Chiavassa S, Supiot S, Ospina JD, Campillo-Gimenez B, Castelli J, de Crevoisier R. *Voxel-Based Analysis for Identification of Urethrovesical Subregions Predicting Urinary Toxicity After Prostate Cancer Radiation Therapy*. Int J Radiat Oncol Biol Phys, 2019. 104 (2): 343-54.

Introduction

Increasing evidence has recently arisen showing that localized anatomical subregions of the bladder are correlated with acute and late urinary toxicities [58,105,106] . Therefore, there is a clear need to exploit the rich information of the planning three-dimensional (3D) dose distribution, together with the individual's anatomy, to accurately identify subregions of the bladder/urethra at high risk of damage for toxicity prediction and more accurate organ-dose constraint definition. To this end, voxel-based methods for dose distribution analysis via dose-volume maps (DVM) may help unravel the complexity of toxicity and local dose-volume relationships by identifying simultaneous involvement of different radiosensitive structures [84,93,94]. The DVM approach is based on non-rigid registration to align patient anatomies and map dose distributions to a single reference [109]. A subsequent voxel-wise statistical analysis is performed to test for local dose differences between patients with/without toxicity. Hence, organ subregions are computed as the clusters of voxels within the organ, where significant dose differences have been found. This methodology has been recently applied to identify regional dose differences in the rectum for gastrointestinal toxicity in prostate cancer [93,94], in the heart and lung for corresponding toxicities in thoracic cancer [96,111], and in the cricopharyngeus muscle and cervical esophagus for dysphagia in head and neck cancer [95]. To our knowledge, a 3D voxel-based approach has never been applied to explore dosimetric patterns associated with urinary toxicity.

The objective of this study was to identify bladder and urethra sub-volumes (Svol) associated with urinary toxicity after prostate cancer RT. The proposed framework combines urethra segmentation, an accurate anatomical non-rigid registration approach for mapping the population 3D dose distributions to a single coordinate system, a voxel-wise analysis with respect to toxicity in the common space, and a further subregional analysis in the individual's space.

4.1 Materials and Methods

The population dataset and treatment characteristics of the patients included in this study have been thoroughly described in the previous chapter. In brief, it consists of 272 patients with localized prostate cancer treated with IMRT/IGRT at 78/80Gy (2Gy/fr), in a multicenter setting (STIC-IGRT and PROFIT) [157,158]. The median follow-up was 50 months (range: 6-102 months). Acute (≤3 months from RT start) and late (>3 months) urinary toxicity was scored using the CTCAE v.3.0 or RTOG radiation morbidity-scoring schema. 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.

The workflow of the study is divided into 7 steps, as presented in Figure 4.1.

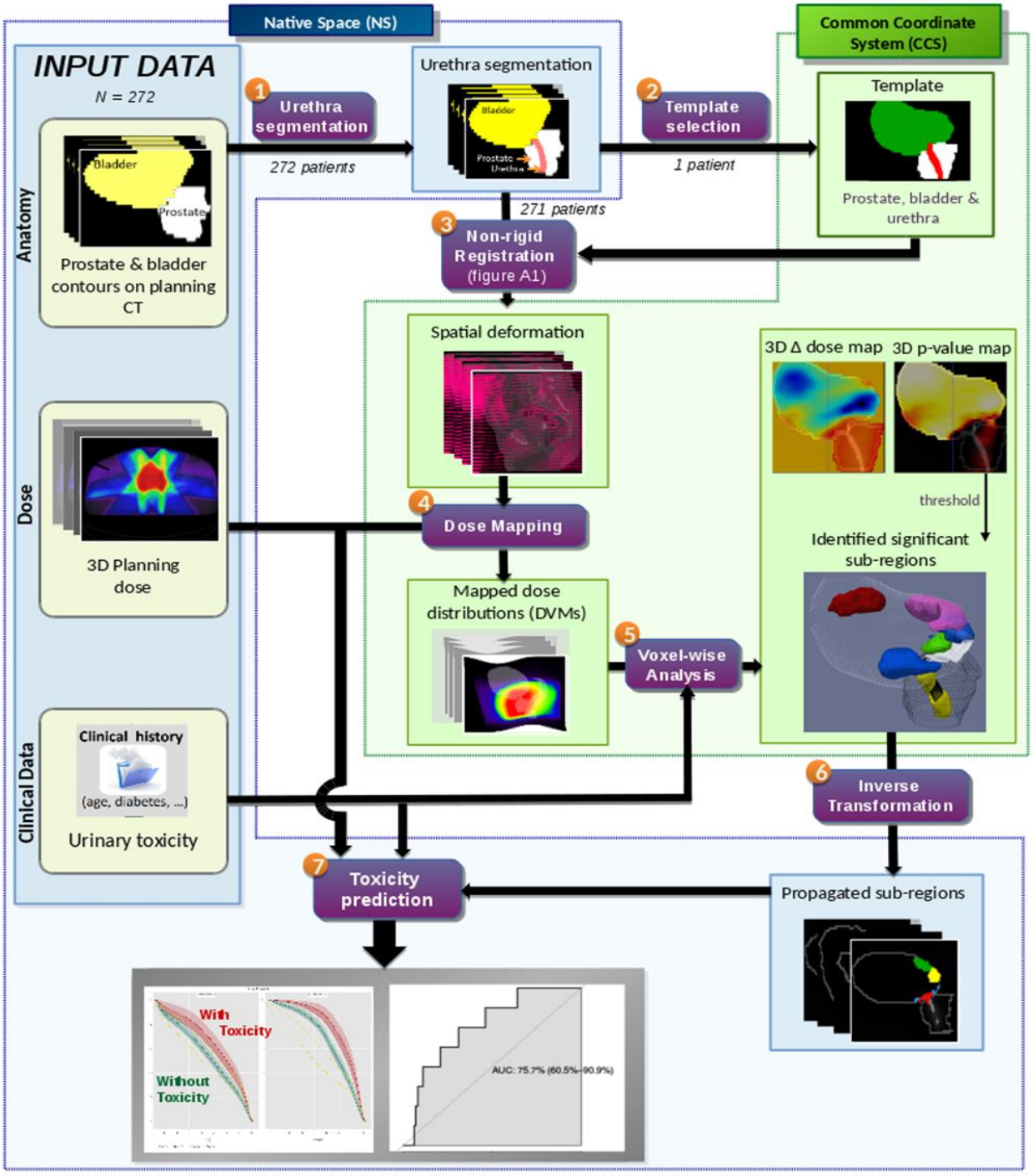


Figure 4.1 Workflow of the study in 7 steps

4.1.1 Automatic urethra segmentation on planning CT images (step 1)

The urethra was automatically segmented on the CT image of each individual using the MABUS approach described in Chapter 2 (Figure 4.1, step 1).

4.1.2 Template selection, registration of population to the template, and dose propagation (steps 2, 3, and 4)

An average patient, close to the whole population in terms of prostate volume, bladder volume, and urethral length, was selected as a common coordinate system (CCS) for aligning the entire population (Figure 4.1, step 2). A customized algorithm was devised for non-rigidly registering the structural description of the organs (bladder, urethra, and prostate; Figure 4.1, step 3) [109]. Figure 4.2 depicts the overall workflow of the non-rigid registration method used for aligning the whole population to a single template yielding anatomical correspondences for dose comparisons. The method exploits organ delineations (prostate and bladder) as obtained during the treatment planning as well as the urethra segmentation [164] (Figure 4.2, step 1).

A structural description of the organs is first obtained by combining different scalar maps into a single 3D image which represents the whole pelvis. For the structural description of the bladder, a map with the Euclidean distance of each voxel inside the volume to the surface of the bladder was computed (Figure 4.2, step 2). For the prostate a scalar field was computed by applying the Laplacian equation inside the prostate volume considering both limit conditions at the urethra (u=0) and the prostate surface (u=1000) as described in (12) (Figure 4.2, step 3). The Laplacian field provides a normalized distance map centered on the urethra, which allows alignment of intra-prostatic structures together. Finally the bladder Euclidean distance map and prostate Laplacian scalar field are combined to produce a global structural description of the organs to be registered (Figure 4.2, step 4). By doing so, the structures are comparable across individuals and therefore may be aligned in 3D by devising a cost function based on the sum of squared differences of the scalar fields inside and outside the organs. The approach was embedded in a multi-resolution framework where the mean square metric of the scalar fields was adopted as cost function.

Once the structural description of the organs is computed, the non-rigid registration workflow was implemented in three steps using the Elastix framework [165]. Firstly, a rigid registration of prostate was performed to roughly align the urethra and prostate volumes (Figure 4.2, step 5). Non-rigid registration was then applied to the bladder only (Figure 4.2, step 6) to cope with large interindividual deformations. This step was followed by a non-rigid registration of the whole set of structures (bladder, urethra and prostate), which together steer the registration towards a global alignment (Figure 4.2, step 7). This approach allowed to deal with the high inter-individual variability in terms of bladder volume and shape, while at the same time, achieving a precise registration of the urethra. Non-rigid transformations were optimized with the Adaptive Stochastic Gradient Descent. The obtained 3D deformation fields from each step are finally combined to propagate the planning dose distributions from the native space of each patient to the CCS, using trilinear interpolation. The final transformation computed as a composite of T1, T2 and T3, is used later to propagate back the identified Svol from the template coordinate system to the native space, as described in section 5.2.4.

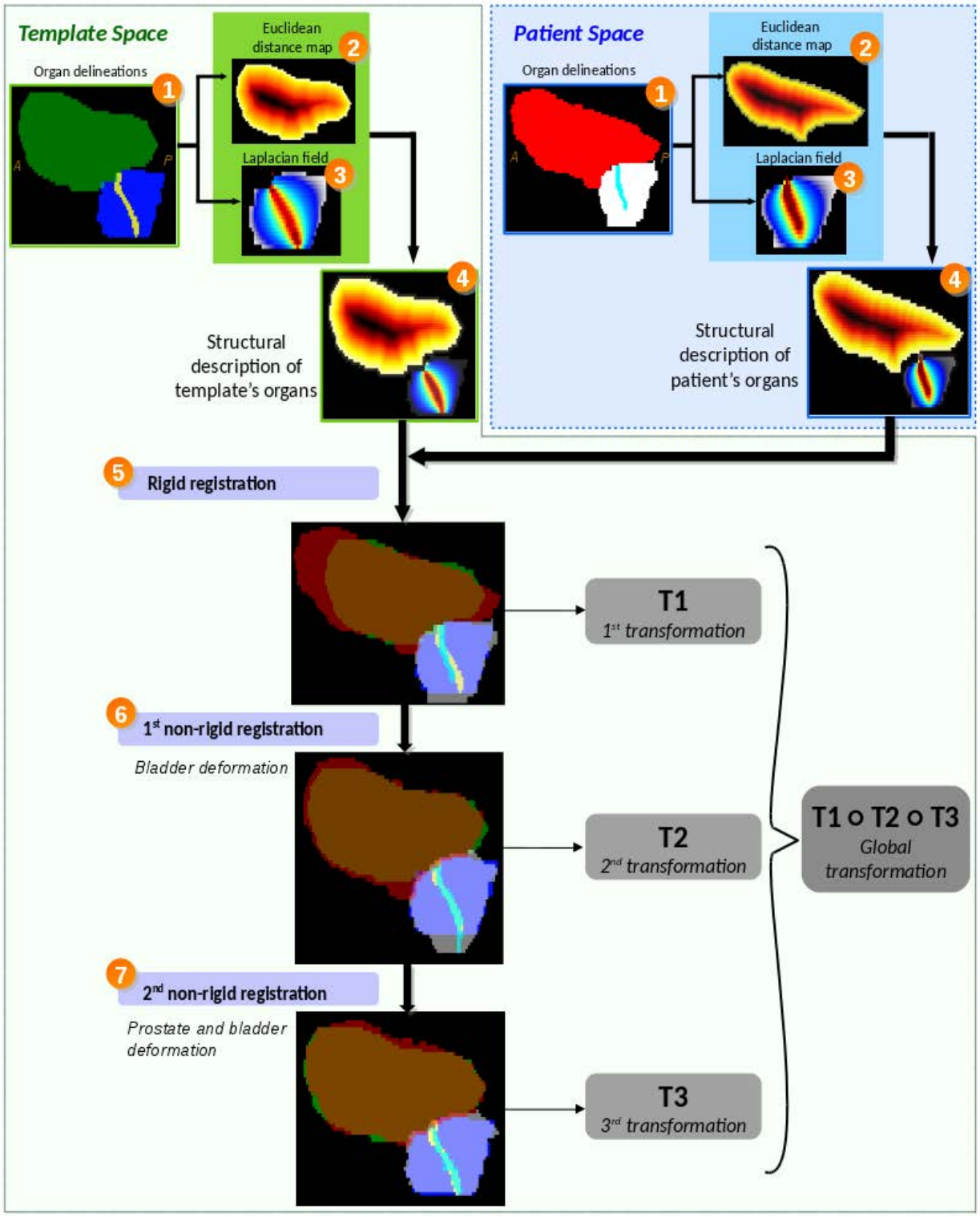


Figure 4.2 Workflow of registration via structural description of the bladder, prostate and urethra

Geometric and dosimetric scores were used to quantitatively assess the registration accuracy. Thus, the centerline distance (CLD) was used for the urethra alignment [164] and the Dice Similarity Coefficient (DSC) and the Modified Hausdorff Distance (MHD) [166] for the prostate and the bladder. In addition,

the dose-organ overlap (DOO) was computed to evaluate the dose-warping accuracy as in [94]. The DOO score measures the coincidence of both the organs and dose distribution in the common space by penalizing the overlap errors within the higher dose gradients. Finally, the 3D registration uncertainty was estimated by computing the standard deviation of the center of mass coordinates of rigidly registered bladder contours. Subsequently, dose distributions for each patient were smoothed accordingly with a Gaussian kernel with corresponding width, to overcome registration uncertainties.

4.1.3 DVM construction and definition of sub-volumes on the template (step 5)

Once the 3D dose distributions are propagated to the CCS, voxel-wise analysis allowed the identification of symptom-related Svol by generating for each symptom a DVM representing the dose differences between patients with and without toxicity. Only accurately registered patient data were included. Unilateral Mann-Whitney *U* tests were voxel-wise performed between the two populations for each endpoint. The alternative hypothesis was that patients with toxicity received a higher dose. This step resulted in 3D maps for both dose differences and *P*-values (threshold at $P = 0.01$; Figure 4.1, step 5), thereby characterizing each subregion in terms of average dose difference.

To cope with the multiple comparisons problem, arising when performing multiple statistical tests (voxel-wise), a nonparametric permutation test [119] was performed, which allowed the description of the entire DVM with a single adjusted *P*-value. At each voxel, the average dose difference was normalized to the standard deviation computed over all random samples generated from 1000 permutations on the urinary toxicity labels (yes vs. no). The normalized maximum dose difference was selected as test statistic summarizing the discrepancy between the two groups (with/without toxicity) and therefore avoiding a voxel-wise test and a consequent multiple comparison problem. After each permutation, we obtain a distribution of test statistic (Mann-Whitney U test), from which the adjusted p-value can be computed as the probability of having a test statistic greater than the test statistic in the observed sample and compared with a significance level of 5%, 10%, 15%, and 20%. The corresponding percentile of the normalized maximum dose difference value possibly determines a voxel region with a statistically significant dose difference.

The resulting subregions were compared with those generated by the voxel-wise Mann-Whitney *U* test (uncorrected). The overlapping regions (Mann Whitney U test region and permutation test region) were finally considered in order to define the Svol. The absolute volume (in cc) of each subregion in the CCS was also calculated.

4.1.4 Inverse mapping of sub-volumes to the native patient space and toxicity prediction (steps 6 and 7)

All segmented Svol were propagated from the CCS back to each individual native space by applying the inverse of the previously computed 3D deformation field (Figure 4.1, step 6). Subsequently, DVHs and mean doses were computed for the Svol and the bladder and compared across the patients with and

without toxicity. Unilateral Mann-Whitney *U* test was used to compare the mean dose and each DVH bin between the two groups of patients. The prediction capability of the dose (mean and bin-wise) in the identified Svol and the bladder was evaluated for the corresponding toxicity endpoints in the native space of the patients. Additionally, we tested, by symptom, whether the doses in the Svol identified for the grade≥1 toxicity endpoints were also predictive for the corresponding grade≥2 toxicities. For acute toxicities, the predictive performances of the Svol and the bladder were estimated and compared using the logistic regression. For late toxicities, the Cox proportional hazard model was used to compute risk estimates. The prediction capability of the Cox model at 5-years was evaluated with the area under the time-dependent ROC curve (t-AUC), based on the approach proposed by Chambless and Diao [159], which accounts for censoring in survival analysis (also defined as cumulative/dynamic AUC). The AUC and t-AUC and 95% confidence intervals (CI) were computed using 1000 bootstrap replicates. The impact of patient and treatment characteristics on each toxicity endpoint was also assessed in a univariate analysis.

Multivariate logistic/Cox models were constructed including clinical parameters and pre-selected dosimetric variables (dose bins) from the univariate analysis. Models were constructed using two methods: i) backward elimination and ii) the least absolute shrinkage and selection operator (LASSO) method [160,161], as described in Chapter 3. The AUC and t-AUC and 95% CIs from 1000 bootstrap replicates were used to evaluate the models' discriminative performance.
Statistical analysis and graphics were performed in Python and R.

4.2 Results

4.2.1 Accuracy of spatial normalization

In total, 18 patients out of 272 were excluded from the study because of either low prostate (≤0.88; n = 6) or bladder (≤0.85; n = 12) Dice scores. The cut-off points where empirically chosen after visual inspection of the registered images, considering both organ overlap and urethra alignment. For the remaining 254 patients, the computed mean values and standard deviations of DSC, MHD, and DOO scores for the prostate and the bladder and the CLD score for the urethra are reported in Table 4.1. The standard deviations of center of mass coordinates after the non-rigid registration were 2.4 mm in the left-right, 5.3 mm in the anterior-posterior and 6.2 mm in the cranial-caudal direction.

Table 4.1 Co-registration scores after non-rigid registration for the prostate and the bladder

Organs	DI	MHD (mm)	DOO	CLD (mm)
Prostate	0.92 ± 0.02	0.18 ± 0.17	0.87 ± 0.03	-
Bladder	0.88 ± 0.07	0.68 ± 0.88	0.87 ± 0.08	-
Urethra	-	-	-	1.82 ± 0.80

DI= Dice score Index, MHD= Modified Hausdorff Distance[166]*, DOO= Dose-Organ Overlap, CLD: Centerline Distance*[164].

4.2.2 Identification of the symptom-related Subvolumes with significant dose differences between patients with/without toxicity, in the template

Figure 4.3 shows the Svol where statistically significant dose differences between patients with/without toxicity appear in the common coordinate system (CCS). On the left is shown the dose distribution to the bladder and the urethra for patients with/without toxicity. The dose displayed in each voxel corresponds to the mean dose of each population. The mean dose (± standard deviation) received by the whole Svol in each population is given in the figure. On the right is shown the mean dose difference to the bladder and the urethra (sagittal and coronal views). The black rectangle indicates the region of the template's trigone. Svol were identified for a total of 5 grade ≥1 symptoms: in the prostatic urethra for acute incontinence, in the bladder trigone for acute retention, late retention and in the posterior part of the bladder for dysuria, and in the superior part of the bladder for late hematuria. These volumes ranged from 2.4 to 16.9 cc. Figure 4.4 shows a 3D representation of these Svol.

The dose differences across the patients (with and without toxicity) in acute and late retention Svol were not statistically significant for grade ≥2 toxicity. For the remaining Svol, namely acute incontinence, late hematuria and late dysuria, the low number of grade ≥ 2 events did not allow to perform a statistical analysis.

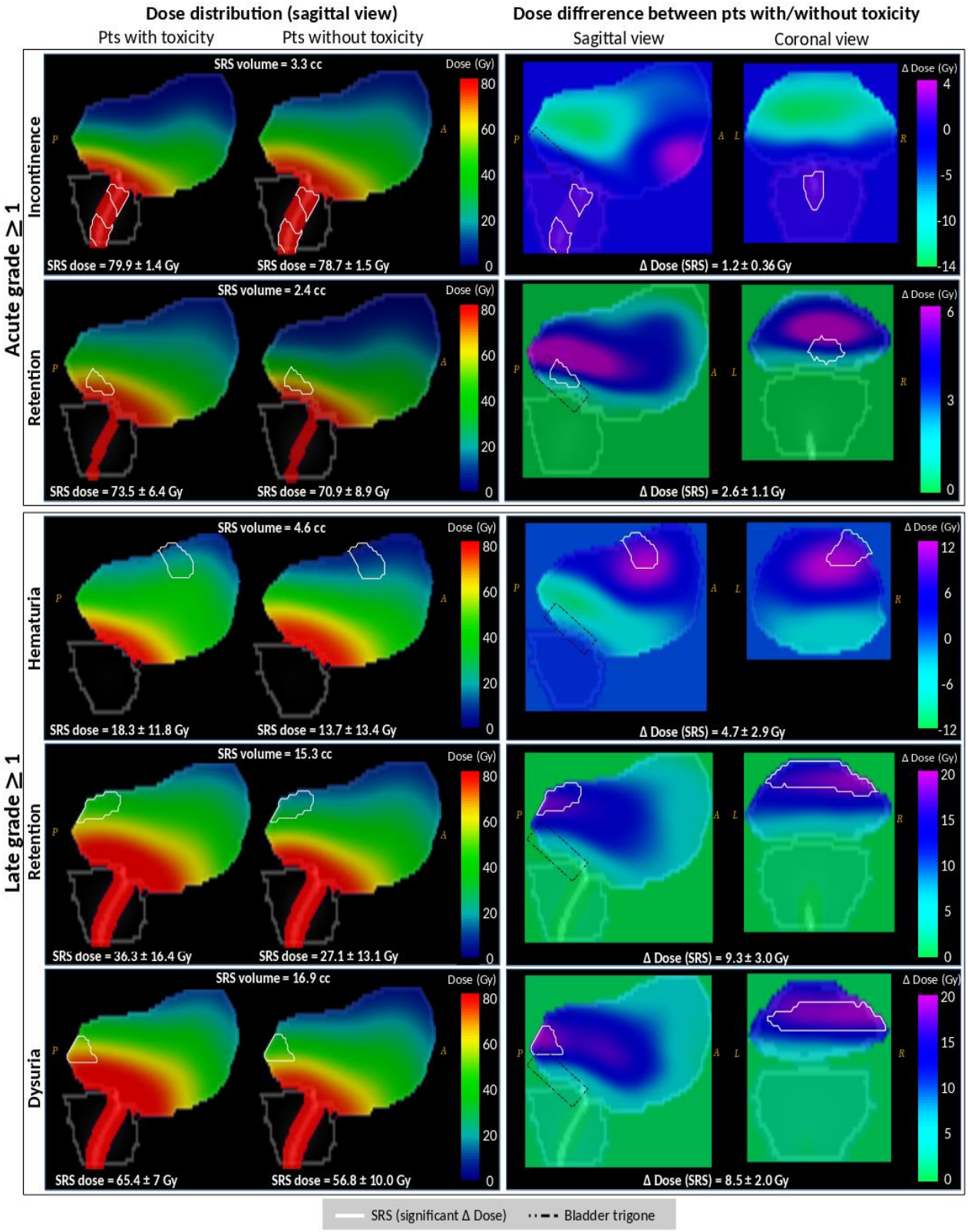


Figure 4.3 Symptom-related subregions (SRS) of statistically significant dose differences between patients with/without toxicity in the common coordinate system (CCS)

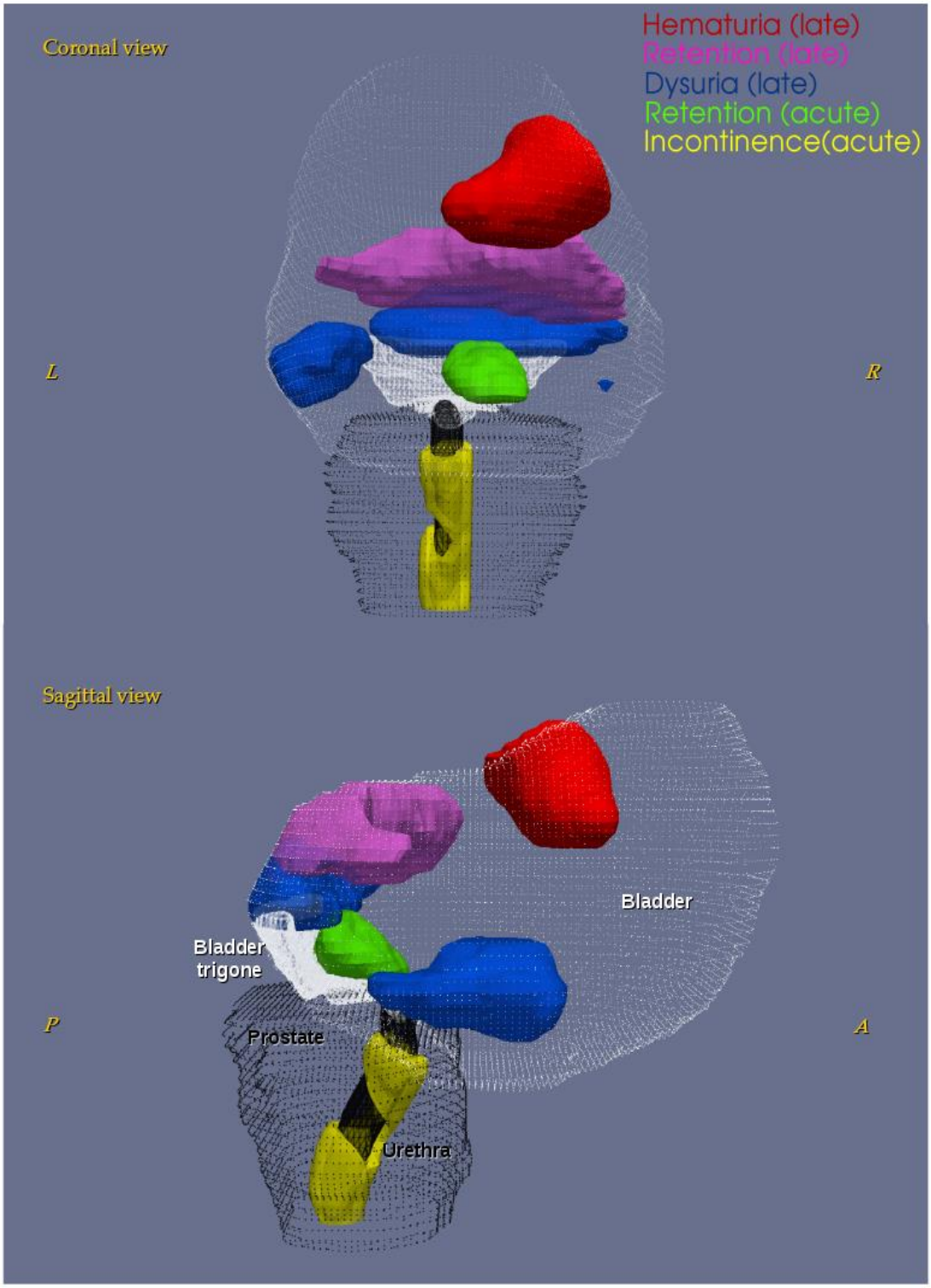


Figure 4.4 Identified sub-volumes (Svol) in the in the common coordinate system

4.2.3 Dose differences in the sub-volumes and in the whole bladder between patients with/without toxicity in the native patient space

The averaged dose differences between patients with/without toxicity in the whole bladder, the prostatic urethra and the five Svol after propagation of each Svol from the template to the native space are reported in Table 4.2. These differences, ranging from 1.2 to 9.3 Gy, were significant for the five Svol ($p<0.01$). For the whole bladder, the dose differences were significant for acute retention (4.4 Gy), late retention (4.4 Gy) and late dysuria (9.3 Gy).

Figure 4.5 depicts the DVHs of the Svol and the whole bladder for the two groups of patients. The dose bins with significant differences among the two populations are identified. The red and green curves represent the average DVHs of the patient with/without toxicity, respectively, in the Svol (continuous lines) and in the whole bladder (dashed lines). The shadowed region indicates the dose bins where the dose for the group with toxicity is significantly higher than the group without toxicity (Man-Whitney U test; $p \leq 0.05$).

Table 4.2 Dose differences between patients with/without acute (A) and late (B) urinary toxicity in the whole bladder, the urethra and in the sub-volumes (Svol)

A. Acute urinary toxicity

Grade≥ 1					
Endpoints	**Regions**	**Dose of pts with toxicity (mean ± SD ,**	**Dose of pts without toxicity (mean ± SD , Gy)**	**Dose difference (mean ± SE , Gy)**	**p value**
Hematuria	Whole bladder	39.5 ± 15.8	35.3 ± 14.4	4.2 ± 6.05	NS
	Urethra	77.9 ± 0.4	78.5 ± 0.5	-0.6 ± 0.2	NS
	Svol	-	-	-	-
Dysuria	Whole bladder	37.5 ± 15.4	38.1 ± 14.9	-0.6 ± 2.3	NS
	Urethra	78.3 ± 0.5	78.5 ± 0.5	-0.3 ± 0.1	NS
	Svol	-	-	-	-
Retention	Whole bladder	40.4 ± 14.6	36 ± 15.3	4.4 ± 2.3	0.03
	Urethra	78.4 ± 0.3	78.4 ± 0.4	-0.05 ± 0.1	NS
	Svol	73.5 ± 6.4	70.9 ± 8.9	2.6 ± 1.2	<0.01
Frequency	Whole bladder	35.6 ± 14.8	34.7 ± 12.8	0.9 ± 2.1	NS
	Urethra	78.5 ± 0.7	78.5 ± 0.7	0	NS
	Svol	-	-	-	-
Incontinence	Whole bladder	31.3 ± 6.5	35.7 ± 14.7	-4.4 ± 1.9	NS
	Urethra	79.3 ± 1.2	78.5 ± 0.9	0.8 ± 0.5	0.06
	Svol	79.9 ± 1.4	78.7 ± 1.5	1.2 ± 0.3	<0.01
Grade≥ 2					
Dysuria	Whole bladder	35.5 ± 17.3	38 ± 15	-2.5 ± 4.9	NS
	Prostatic	78.5 ± 0.5	78.2 ± 0.4	0.3 ± 0.2	NS
	Svol	-	-	-	-
Retention	Whole bladder	38.8 ± 15.8	37.6 ± 15	1.2 ± 2.9	NS
	Urethra	78.2 ± 0.3	78.5 ± 0.4	-0.3 ± 0.1	NS
	Svol	-	-	-	-
Frequency	Whole bladder	36.6 ± 14.7	35 ± 14.3	1.6 ± 2	NS
	Urethra	78.6 ± 0.8	78.4 ± 0.7	0.2 ± 0.1	NS
	Svol	-	-	-	-

B. Late urinary toxicity (at 5 years)

Grade≥ 1					
Endpoints	**Regions**	**Dose of pts with toxicity (mean ± SD , Gy)**	**Dose of pts without toxicity (mean ± SD , Gy)**	**Dose difference (mean ± SE , Gy)**	**p value**
Hematuria	Whole bladder	38.7 ± 11.5	39.8 ± 13.4	-1.1 ± 2.8	NS
	Urethra	77.1 ± 1.1	78.1 ± 0.6	-0.4 ± 0.7	NS
	Svol	18.3 ± 11.8	13.7 ± 13.4	4.69 ± 2.9	0.04
Dysuria	Whole bladder	50.1 ± 10.1	40.8 ± 12.9	9.3 ± 2.8	<0.01
	Urethra	78.0 ±0.3	77.9 ± 0.5	0.1 ± 0.4	NS
	Svol	65.4 ± 7	56.8 ± 10.0	8.5 ± 2	<0.01
Retention	Whole bladder	46 ± 13.9	41.6 ± 12.5	4.4 ± 2.7	0.05
	Urethra	78.1 ± 0.4	78.0 ± 0.6	0.06 ± 0.1	NS
	Svol	36.3 ± 16.4	27.1 ± 13.1	9.3 ± 3	<0.01
Frequency	Whole bladder	38.9 ± 13.6	40.4 ± 13.7	-1.4 ± 2.4	NS
	Urethra	78.2 ± 0.7	78.2 ± 0.6	0.02 ± 0.1	NS
	Svol	-	-	-	-
Incontinence	Whole bladder	38 ± 12.6	40.4 ± 13.3	-2.4 ± 2.6	NS
	Urethra	78.2 ± 1	78.0 ± 0.5	- 0.2 ± 0.5	NS
	Svol	-	-	-	-
Grade≥ 2					
Retention	Whole bladder	42.2 ± 13.3	39.6 ± 13.1	2.6 ± 3.5	NS
	Urethra	77.9 ± 0.5	78.1 ± 0.6	-0.3 ± 0.3	NS
	Svol	-	-	-	-
Frequency	Whole bladder	38.6 ± 12	40.4 ± 13.51	-1.8 ± 2.8	NS
	Urethra	78.5 ± 0.6	78.0 ± 0.6	0.5 ± 0.2	NS
	Svol	-	-	-	-
Incontinence	Whole bladder	36.2 ± 13.4	40.3 ± 13.1	-4.1 ± 4.9	NS
	Urethra	78.8 ± 0.6	77.9 ± 0.6	0.9 ± 0.2	0.07
	Svol	-	-	-	-

Svol:sub-volume; SD: Standard deviation, SE: Standard error; NS: not significant ($p > 0.05$)

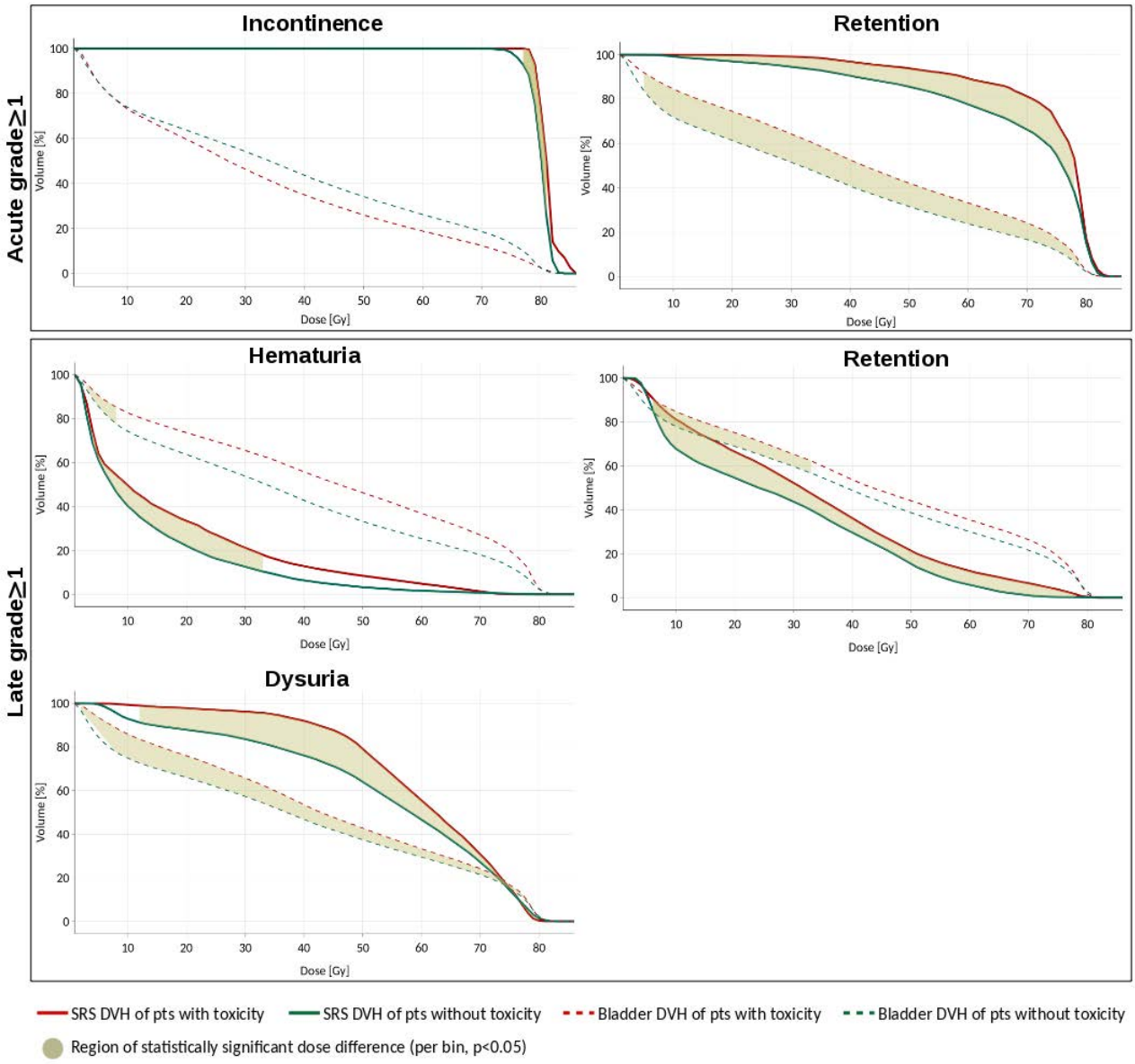


Figure 4.5 DVHs of patients with and without urinary toxicity for the whole bladder and the Svol in the native space

4.2.4 Predictive capabilities of the Svol and whole bladder DVHs in the native spaces

UNIVARIATE ANALYSIS

The predictive capabilities of the mean dose and the DVHs for the five Svol and the whole bladder are shown in Table 4.3A for acute toxicity and Table 4.3B for late toxicity. The dose bins with the highest significant AUC/ t-AUC are reported. The AUC calculated from the doses in the Svol ranged from 0.62 to 0.81 for the five symptoms (grade ≥ 1 acute incontinence and retention, and late dysuria and retention), while the doses to the whole bladder were predictive only for late dysuria (highest AUC=0.75) and late retention (highest AUC=0.71). The doses delivered in the Svol were not predictive of grade≥2 toxicity.

Table 4.3 Urinary toxicity prediction capability of the mean dose and the DVH for the whole bladder and the identified symptom-related sub-volumes (Svol) in the native space of the patients

A. Acute toxicity (grade≥1)

Endpoints	Regions	Mean Dose			DVH			
		p value	OR (95% CI)	AUC (95% CI)	Most predictive DVH bin (range of bins)*	p value	OR (95% CI)	AUC (95% CI)
Retention	Whole bladder	NS	-	-	V79 (V77-V79)	0.052	1.06 (1.01-1.13)	0.60 (0.51-0.67)
	Svol	0.04	1.06 (1.01-1.2)	0.62 (0.56-0.68)	V72 (V63-V79)	0.01	1.02 (1.01-1.04)	0.62 (0.55-0.68)
Incontinence	Whole bladder	NS	-	-	-	NS	-	-
	Svol	<0.01	2.1 (1.27-3.43)	0.74 (0.66-0.81)	V80 (V80-V83)	0.04	1.02 (1.01-1.04)	0.73 (0.67-0.81)

B. Late toxicity (grade≥1)

Endpoints	Regions	Mean Dose			DVH			
		p value	HR (95% CI)	t-AUC (95% CI)	Most predictive DVH bin (range of bins)*	p value	HR (95% CI)	t-AUC (95% CI)
Dysuria	Whole bladder	<0.01	1.05 (1.01-1.09)	0.75 (0.66-0.84)	V67 (V5- V78)	<0.01	1.03 (1.01-1.05)	0.72(0.63-0.82)
	Svol	<0.01	1.1 (1.04-1.17)	0.8 (0.71-0.88)	V52 (V32 - V76)	<0.01	1.05 (1.02-1.08)	0.81(0.72-0.90)
Retention	Whole bladder	<0.01	1.03 (1.01-1.06)	0.66 (0.58-0.75)	V19 (V4-V78)	<0.01	1.02 (1-1.04)	0.67(0.59-0.75)
	Svol	<0.01	1.04 (1.02-1.05)	0.71 (0.63-0.78)	V35 (V7-V76)	<0.01	1.02 (1.01-1.02)	0.70(0.62-0.77)
Hematuria	Whole bladder	NS	-	-	V7 (V6-V18)	0.04	1.04 (1-1.07)	0.65(0.55-0.75)
	Svol	0.04	1.02 (1.01-1.04)	0.64 (0.55-0.73)	V17 (V5-V25)	<0.01	1.02 (1.01-1.04)	0.67(0.56-0.77)

MULTIVARIATE ANALYSIS

Table 4.4 presents the significant dosimetric and/or clinical parameters impacting on acute (Table 4.4 A) and late (Table 4.4 B) urinary toxicity in multivariate analysis selected using backward elimination. The corresponding models resulted from LASSO multivariate analyses are reported in Table 4.5. The dosimetric impact of the doses to the Svol is confirmed.

Table 4.4 Parameters impacting on acute and late urinary toxicity in multivariate analysis using backward elimination

A. Acute toxicity (logistic regression)

Grade ≥1					
Endpoints	**Parameters**	**p-value**	**OR (95% CI)**	**Model p-value**	**AUC (95% CI)**
Retention	Svol (V72)	0.01*	1.19 (1.04 – 1.4)	0.01	0.70 (0.60-0.80)
	TURP	0.06	0.12 (0.01-1.05)		
	Baseline retention	0.09	2.64 (0.86-8.1)		
Incontinence	Svol (V80)	0.03*	1.8 (1.05-3.05)	<0.01	0.74 (0.60-0.87)
	Prescribed dose	0.11	1.6(0.9-2.84)		
Hematuria	TURP	0.08	6.9 (0.8-59)	<0.01	0.86 (0.75-0.97)
	Previous abdominal surgery	0.07	8.5 (0.85-85.3)		
	Anti-coagulant treatment	0.04*	9.4 (1.2 – 71.2)		
Dysuria	Age	0.01	0.94 (0.9-0.99)	0.01	0.6 (0.52-0.69)
Frequency	TURP	<0.01*	0.1 (0.02-0.5)	<0.01	0.7 (0.62-0.77)
	Baseline frequency	0.01*	6 (1.5 – 24.8)		
Grade ≥2					
Retention	Baseline retention	0.01	3.53(1.29-9.63)	0.01	0.59 (0.5-0.68)
Dysuria	Antihypertensive treatment	0.06	3.6 (0.94-13.8)	<0.01	0.72 (0.56-0.88)
	Prostate volume	0.02*	1.03 (1.01-1.05)		
Frequency	Baseline retention	<0.01*	4.9 (1.57-15.26)	<0.01	0.7 (0.6-0.8)
	Baseline frequency	0.01*	3.2 (1.3-7.7)		

*OR: Odds ratio; HR: Hazard ratio; CI: Confidence interval; AUC: Area under the ROC curve; t-AUC: time-dependent area under the ROC curve; * significant parameters (p≤0.05); Model p-value: Global statistical significance of the model as defined by the Likelihood ratio test*

B. Late toxicity (Cox regression)

Grade ≥1					
Endpoints	**Parameters**	**p-value**	**HR (95% CI)**	**Model p-value**	**t-AUC (95% CI)**
Retention	Svol (V35)	<0.01*	1.03 (1.01-1.05)	<0.01	0.74 (0.66-0.84)
	Age	0.1	0.96 (0.92-1.01)		
	Previous abdominal surgery	0.06	1.65 (0.92-2.96)		
	Baseline retention	<0.01*	3.7 (1.9-7.1)		
Incontinence	Age	0.07	1.09 (0.99-1.19)	<0.01	0.77 (0.65-0.88)
	TURP	<0.01*	4.9 (1.66-14.4)		
	Diabetes	<0.01*	5.1 (1.67-15.5)		
	Anti-coagulant treatment	0.7	1.2 (0.4-3.6)		
Hematuria	Svol (V17)	0.03*	1.04 (1.01-1.07)	0.01	0.67 (0.54-0.78)
	Baseline retention	0.2	2.6 (0.7-9.9)		
Dysuria	Svol (V52)	<0.01	1.1(1.04-1.17)	<0.01	0.80 (0.71-0.88)
Frequency	Age	0.15	1.03 (0.99-1.07)	<0.01	0.76 (0.68-0.84)
	Antihypertensive treatment	0.06	1.58 (0.97-2.57)		
	ADT	<0.01*	2.44 (1.5-3.98)		
	Baseline frequency	<0.01*	3.79 (2.32-6.2)		
Grade ≥2					
Retention	Age	0.2	0.95 (0.89-1.03)	<0.01	0.74 (0.59-0.9)
	Previous abdominal surgery	0.12	2.2 (0.8-6)		
	Antihypertensive treatment	0.14	0.4 (0.12-1.34)		
	Baseline retention	<0.01*	4.9 (1.76-13.6)		
Frequency	Diabetes	0.09	3 (0.83-10.9)	<0.01	0.74 (0.62-0.88)
	Baseline frequency	0.02*	5.29 (1.8-15.6)		
Incontinence	TURP	<0.01	17.7 (2.95-76.26)	<0.01	0.74 (0.62-0.87)

*OR: Odds ratio; HR: Hazard ratio; CI: Confidence interval; AUC: Area under the ROC curve; t-AUC: time-dependent area under the ROC curve; * significant parameters (p≤0.05); Model p-value: Global statistical significance of the model as defined by the Likelihood ratio test*

Table 4.5 Parameters impacting on acute and late urinary toxicity in multivariate analysis using LASSO

A. Acute toxicity (logistic regression)

Grade ≥1				
Endpoints	**Parameters**	**beta**	**OR (e^{beta})**	**AUC (95% CI)**
Retention	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 (V80)	0.00891	1.01	0.73 (0.61-0.85)
Dysuria	Age	-0.0318	0.96	0.63 (0.55-0.72)
	ADT	-0.1531	0.85	
Frequency	Bladder volume	-0.0002	0.99	0.79 (0.7-0.88)
	TURP	-1.2162	0.30	
	Hypercholesterolemia treatment	-0.4177	0.66	
	Baseline frequency	0.5518	1.73	
Grade ≥2				
Retention	TURP	-0.2528	0.78	0.63 (0.54-0.72)
	Anticoagulant treatment	-0.2320	0.79	
	Baseline retention	0.774	2.17	
Dysuria	Antihypertensive treatment	0.3210	1.38	0.72 (0.55-0.88)
	Prostate volume	0.0127	1.01	
Frequency	Baseline frequency	0.5908	1.81	0.59 (0.52-0.65)
OR: Odds ratio; CI: Confidence interval; AUC: Area under the ROC curve;				

B. Late toxicity (Cox regression)

Grade ≥1				
Endpoints	**Parameters**	**beta**	**HR (e^{beta})**	**tAUC (95% CI)**
Retention	Svol (V35)	0.0114	1.01	0.79 (0.72-0.85)
	Age	-0.0222	0.98	
	Previous abdominal surgery	0.3486	1.41	
	Baseline Retention	0.8857	2.34	
Incontinence	Age	0.0355	1.04	0.81 (0.71-0.88)
	TURP	1.2058	3.34	
	Diabetes	1.2925	3.64	
Hematuria	Svol (V17)	0.0105	1.02	0.68 (0.57-0.78)
	Anticoagulant treatment	0.0841	1.93	
Dysuria	Svol (V52)	0.0171	1.02	0.82 (0.72-0.90)
	Previous abdominal surgery	0.1427	1.15	
Frequency	Age	0.0072	1.01	0.81 (0.75-0.86)
	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	Age	-0.0414	0.96	0.77 (0.67-0.85)
	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	Diabetes	0.4673	1.60	0.81 (0.72-0.90)
	Hypercholesterolemia treatment	-0.4663	0.63	
	Baseline frequency	1.1170	3.06	
Incontinence	TURP	2.0501	7.77	0.82 (0.65-0.97)
	Previous abdominal surgery	0.6994	2.01	

HR: Hazard ratio; CI: Confidence interval; t-AUC: time-dependent area under the ROC curve

4.3 Discussion

A robust non-rigid registration strategy, coupled with voxel-wise comparisons, allowed us to investigate the local dose-effect relationship for acute and 5-years late urinary toxicity after prostate cancer RT. The doses on five Svol, located in the urethra and the bladder, were identified as good predictors for five grade ≥ 1 symptoms (acute incontinence and retention, late retention, dysuria and hematuria). The predictive capabilities of these Svol outperformed the predictive capabilities of the whole bladder. To our knowledge, this is the first study to explicitly correlate the dose to specific subregions of the bladder and the urethra with urinary toxicity, within a voxel-based framework.

This study is the first to explicitly correlate the 3D dose to the urethra with urinary toxicity following external beam radiation therapy. Acute and late retention Svol were found in the trigone and posterior part of the bladder. Indeed, retention can attributed to reduced bladder contractility (detrusor muscle). Relationship between the dose to the bladder trigone or high bladder-dose and obstructive side effects, has also been reported in the literature [54,57,58]. Dysuria (painful urination or difficulty urinating) Svol is also mainly located in the posterior part of the bladder, partially in the trigone. Surprisingly, a Svol for hematuria was found superiorly, at the bladder dome (a region which receives a relatively low dose, < 20 Gy), while bleeding classically related to high dose [54,129,167], and telangiectasia are mainly observed in the bladder neck/trigone at the cystoscopy. Incontinence appears related to the dose delivered to the prostatic urethra. Indeed, incontinence may result from malfunctioning of the urethral sphincter [168]. A dosimetric association with urinary frequency could not be demonstrated in the present study, although this symptom was strongly predicted by the baseline frequency.

One of the major advantages of the voxel-wise analysis is that the whole 3D volume can be explored and compared without any prior knowledge of regions correlating with toxicity. However, our approach relies on a multi-organ deformable image registration, which appears particularly challenging given the high interindividual variability. Conventional interindividual CT registration methods are not accurate enough for reliable anatomical mapping because of the low soft-tissue contrast. To cope with this issue, we combined a structural description of the pelvic region using the contours obtained within the clinical protocol. Each step was thoroughly validated, including the automatic segmentation of the urethra, thus increasing the reliability of dose mapping. We propagated the Svol found in the template back to each individual's space and, as such RT planning may be modified to spare specific subregions. The feasibility of reducing the dose in the Svol, while preserving the dose to the PTV, still requires demonstration. This goal seems achievable when dealing with Svol distant from the PTV (such as hematuria Svol) but much more difficult when Svol are located inside or close to the PTV (such as incontinence Svol).

Our study presents some limitations. The results must be carefully interpreted, as a correlation between local dose and toxicity does not necessarily mean causality, especially given the paucity of events for certain endpoints. Second, we failed to find a spatial dose-effect relationship for some symptoms, in particular for grade ≥2 toxicity. Although this may imply the absence of specific Svol for these symptoms, it can also be due to the limited follow-up or to the patient-specific parameters that affect toxicity and

were not taken into account or to the limited follow-up. Furthermore, the predictive performance of our model has been internally validated (via bootstrapping) on the same subjects used to construct the model. External cohorts are therefore required to validate the results. Finally, statistical analyses were performed on the planning dose distribution, which can differ from the actual delivered dose [127,128]. Daily 3D imaging, such as CBCT, could be used to compute cumulated doses, thereby confirming some of the Svols' implication in toxicity [169].

4.4 External validation to estimate the "cohort-effect"

4.4.1 Population dataset, treatment and toxicity

For the external validation of the five previously identified Svol we used patients from the Randomised Androgen Deprivation and Radiotherapy (RADAR, Trans-Tasman Radiation Oncology Group 03.04) trial, which examined the influence of the duration of androgen deprivation therapy with or without bisphosphonate treatment, adjuvant with radiotherapy. Data were collected from 23 centres in Australia and New Zealand between 2003 and 2008. The RADAR trial is a phase 3 trial with a 2 × 2 factorial design in patients with non- metastatic adenocarcinoma of the prostate (stage T2b–4 or T2a, Gleason score ≥7, and baseline prostate-specific antigen concentration [PSA] ≥10 μg/L). This trial is registered with number NCT00193856 in ClinicalTrials.gov.

Data collection, protocol requirements and QA have been summarized in multiple publications [134,167,170–179]. All participants received center-nominated radiotherapy, where 813 had EBRT (without a brachytherapy boost) to 66, 70 or 74 Gy of dose, delivered in up to 2 treatment phases. Delineation of the bladder was not mandatory for RADAR and no specific dose constraints were applied for the bladder. Centers were free to prescribe different bladder filling protocols; 701 patients were prescribed to full bladder, 34 empty and 19 with no or missing protocol [167]. Dose-surface histograms (DSH) of the bladder wall were calculated independently using a plan review software (SWAN) [174] to ensure consistency across datasets submitted from different centers [180]. More complete descriptions of the study protocol and treatment technique specifications can be found elsewhere [177].

In total, 476 were treated with conventional fractionation (1.8-2Gy/fraction) and had sufficient dosimetric, clinical and imaging data available for inclusion in the analysis. Rates by symptom of grade ≥ 1 and grade ≥ 2 acute and late urinary toxicity are given in Table 4.6.

HARMONISATION OF ENDPOINTS

The urinary function of RADAR patients used as the validation cohort was assessed at baseline and at the end of radiotherapy using physician-assessed LENT-SOMA and the International Prostate Symptom Score (IPSS) questionnaire. Patients were routinely followed up every three months for 18 months, then six-monthly up to five years and then annually where urinary symptoms were assessed using LENT-SOMA. Patients were asked to complete the International Prostate Symptom Score (IPSS) questionnaire at 12,

18, 24, 36 and 60-month follow-up post-randomization. The median follow-up was 72 months. Urinary symptom endpoint definitions were in concordance with the definitions used to develop the predictive models for dysuria, hematuria and incontinence (physician assessed; CTCAE v.3.0 or RTOG). The symptom of retention is not described by the LENT-SOMA scoring and, thus, for the validation of the corresponding models equivalent endpoints were derived from the IPSS. Retention was defined by aggregating the scoring of the questions describing an obstructive effect (q1, q3, q5,q6) and herein called IPSS-O.

Table 4.6 Rates of grade ≥ 1 and grade ≥ 2 acute and late urinary toxicity (by symptom)

Endpoints	**Acute toxicity**			
	Grade ≥ 1 (No. of cases and %)		**Grade ≥ 2 (No. of cases and %)**	
Incontinence	67 (15% [12-19%])		29 (7% [4-9%])	
Dysuria / Pain	221 (51% [46-55%])		81 (18% [15-22%])	
Hematuria	12 (2% [1-4%])		2 (0% [0-1%])	
Retention	155 (40% [33-43%])		65 (14% [10-16%])	
	Late 5-year toxicity			
Endpoints	**Grade ≥ 1**		**Grade ≥ 2**	
	No. of cases	**Kaplan-Meier**	**No. cases**	**Kaplan-Meier**
Incontinence	128 (35%)	32% (27-37%)	49 (15%)	12% (9-16%)
Dysuria / Pain	111 (31%)	28% (23-33%)	32 (10 %)	8% (5-11%)
Hematuria	52 (15%)	13% (9-16%)	7 (2 %)	2% (0-3%)
Retention	60 (16%)	13% (10-18%)	27 (9 %)	7% (5-10%)

4.4.2 DVM construction and statistical analysis

The same template patient that was previously selected for Svol identification (from the French cohort) was used to align the population in a common coordinate system. Non-rigid registration was performed as described in section 1.2.2 and a deformation field was generated. From the 476 patients, the registration accuracy, as evaluated in terms of Dice score, was considered as acceptable for 454 of them (≤0.88 for the prostate and ≤0.85 for the bladder) and they were included in the analysis.

For each successfully registered individual, the five previously identified Svol were back-propagated from the template space to the native space of the patient by applying the inverse of the transformation obtained during the registration process. Subsequently, DVHs and mean doses were computed for the Svol, the bladder and the urethra and compared across the patients with and without toxicity. Unilateral Mann-Whitney *U* test was used to compare the mean dose and each DVH bin between the two groups of patients. For acute toxicities, the predictive performances of the Svol, the bladder and the urethra were

estimated and compared using the logistic regression. For late toxicities, the Cox proportional hazard model was used to compute risk estimates. The prediction capability of the Cox model at 5-years was evaluated with the area under the time-dependent ROC curve (t-AUC) [159]. The AUC and t-AUC and 95% confidence intervals (CI) were computed using 1000 bootstrap replicates.

4.4.3 Results

The predictive capabilities of the mean dose and the DVHs for the five Svol, the whole bladder and the urethra are shown in Table 4.7A for acute toxicity and Table 4.7B for late toxicity. The dose bins with the highest significant AUC/ t-AUC are reported. Subregions, confirmed in the RADAR cohort where acute incontinence, late retention and late dysuria. The AUC calculated from the doses in the Svol ranged from 0.65 to 0.70 for the symptoms. The dose to the whole bladder was predictive only for acute incontinence (AUC=0.65). The dose to the urethra was predictive only for acute incontinence (AUC=0.72).

Table 4.7 Urinary toxicity prediction capability of the mean dose and the DVH for the whole bladder and the identified sub-volumes (Svol) in the native space of the patients

A. Acute toxicity (grade≥1)

Symptom	Region	Mean Dose			DVH			
		p value	OR (95% CI)	AUC	Most predictive DVH bin (range of bins)	p value	OR (95% CI)	AUC
Retention	Whole bladder	NS	-	-	-	NS	-	-
	Urethra	NS	-	-	-	NS	-	-
	Svol	NS	-	-	-	NS	-	-
Incontinence	Whole bladder	NS	-	-	V70 (V63-V73)	<0.01	1.05 (1.01-1.09)	0.65
	Urethra	<0.01	1.4 (1.11-1.74)	0.72	V71 (V71-V75)	0.01	1.03 (1.01-1.05)	0.71
	Svol	0.01	1.07 (1.01-1.13)	0.70	V71 (V71-V75)	0.01	1.03 (1.01-1.05)	0.70

B. Late toxicity (grade≥1)

Symptom	Region	Mean Dose			DVH			
		p value	HR (95% CI)	t-AUC	Most predictive DVH bin (range of bins)	p value	HR (95% CI)	t-AUC
Dysuria	Whole bladder	NS	-	-	-	NS	-	-
	Urethra	NS	-	-	-	NS	-	-
	Svol	0.05	1.04 (1-1.07)	0.60	V67 (V55-V73)	<0.01	1.02 (1-1.03)	0.66
Retention	Whole bladder	NS	-	-	-	NS	-	-
	Urethra	NS	-	-	-	NS	-	-
	Svol	<0.01	1.03 (1.01-1.06)	0.65	V38 (V13-V64)	<0.01	1.02 (1.01-1.04)	0.70
Hematuria	Whole bladder	NS	-	-	-	NS	-	-
	Urethra	NS	-	-	-	NS	-	-
	Svol	NS	-	-	-	NS	-	-

The Svol for acute incontinence, which included the inferior and posterior part of the prostatic urethra was found to be predictive (AUC=0.70), although the mean dose to the whole prostatic urethra showed the best performance (AUC=0.72). For late dysuria, the dose to the Svol identified on the posterior bladder part, was confirmed with moderate performance (AUC=0.66). Late retention Svol, located at the posterior bladder, was confirmed with good predictive performance (AUC=0.70). The dose to the whole bladder was predictive only for one symptom, acute incontinence, with modest performance (AUC=0.65). Acute retention and late hematuria models were not confirmed on the RADAR population.

4.5 Conclusion

A voxel-wise analysis allowed the identification of urethro-vesical Svol, whose irradiation appears highly correlated to specific urinary side effects after prostate cancer RT. The dose received by these Svol was more predictive than the dose to the whole bladder. External validation on a large independent population confirms three out of the five subregions as predictive for specific symptoms, suggesting that the posterior part of the bladder and the prostatic urethra are particularly involved in urinary toxicity. The therapeutic benefit of adding dosimetric constraints to these sub-regions needs to be demonstrated.

Using DVMs we were able to identify and validate specific sub-parts of the bladder as predictive of specific urinary side-effects. Although the prediction capabilities of current models exploiting available are improved compared to the whole bladder DVH models, they might be limited by cohort characteristics (size, toxicity rates, etc…) and other methodological issues.

Along with traditional regression, more advanced machine learning strategies have been used and tested in the development of predictive models. A plethora of algorithms arise as promising tools for improving the prediction capabilities but their performance can significantly vary depending on the learning task and the specific data characteristics. In the next chapter we implement and compare some of the most popular machine learning techniques aiming to detect those that are better suited for radiotherapy outcome prediction using common dosimetric and clinical data. Further improvement of their performance was also attempted through data augmentation techniques. For simplicity, comparisons were performed for one symptom only and using the DVH of the whole ladder.

Part IV

Future directions for improving urinary toxicity prediction: A machine learning approach

5 Comparison of machine learning algorithms and oversampling techniques for urinary toxicity prediction after prostate cancer radiotherapy

In the present chapter we assess the robustness of several machine learning algorithms for prediction of urinary toxicity following prostate cancer radiotherapy using dosimetric and clinical data. For this explanatory study, the performance of the classifiers was evaluated using the whole bladder DVH, which is considered today the clinical standard. Moreover, to cope with the problem of low number of events on the unbalanced dataset the implemented machine learning strategies included four different synthetic minority class oversampling techniques.

The work presented in this chapter has been accepted for publication at the peer-reviewed IEEE conference: International Conference on Bioinformatics and Bioengineering (IEEE BIBE 2019), in Athens Oct 28-30 2019.

Mylona E, Lebreton C, Fontaine C, Crehange G, Magné N, Supiot S, de Crevoisier R, Acosta O. *Comparison of machine learning algorithms and oversampling techniques for urinary toxicity prediction after prostate cancer radiotherapy*. IEEE proceedings, Athens Oct 28-30 2019.

5.1 Introduction

Given the inconsistency of available data concerning radiation-induced toxicity, it is crucial to develop robust models with superior predictive performance in order to perform tailored treatments. Machine Learning techniques emerge as appealing in this context, nevertheless without any consensus on the best algorithms to be used. This work proposes a comparison of several machine-learning strategies together with different minority class oversampling techniques for prediction of urinary toxicity following prostate cancer radiotherapy using dosimetric and clinical data. The performance of these classifiers was evaluated on the original dataset and using four different synthetic oversampling techniques.

Urinary toxicity modeling has been addressed in several studies without a clear consensus [126]. Although many important clinical factors have been identified, the relationship between urinary toxicity and bladder dose is far from reaching unanimity. Indeed, the vast majority of existing studies are based upon traditional regression approaches (e.g. logistic or Cox regression) to identify the most important predictors. Although these approaches are often preferred over Machine Learning (ML) techniques due to their interpretability, the prediction power of such models is yet modest. Machine learning techniques can potentially increase toxicity prediction after RT as they rely on previous informative examples. Plethora of methods are emerging without a clear advantage of its use in this context as they application is not straightforward.

One of the main issues that may arise when dealing with real-world data, is the class imbalance problem [181]. That is, in the simplest case when modeling a binary outcome, the majority of the individuals belong to a specific class (majority class) and far less number of individuals are assigned to the counterpart class (minority class). Commonly, the minority class is also the class of interest to model. Within the context of urinary toxicity this phenomenon is particularly evident due to the paucity of events for certain endpoints. Training a classifier on highly imbalanced data can be particularly misleading since the minority class has minimal effect on overall accuracy. Consequently, even if the algorithm classifies all the samples in the majority class, this may result in seemingly high performance.

Within the context of urinary toxicity prediction after RT, we attempt to answer three research questions:

- Is there a superior classifier in terms of discriminative performance?
- When using machine learning techniques on an imbalanced dataset, can we increase classifier's performance by handling the class imbalance?
- Which oversampling technique is more robust for coping with class imbalance issues?

5.2 Materials and Methods

We analyzed 30 dosimetric parameters extracted from the whole bladder's DVH, together with 10 clinical- and patient-related parameters, in a training/test pipeline using different classification methods and oversampling techniques.

5.2.1 Population dataset

The population dataset and treatment characteristics of the patients included in this study have been thoroughly described in the previous chapter. It consists of 254 patients with localized prostate cancer treated with IMRT/IGRT at 78/80Gy (2Gy/fr), in a multicenter setting (STIC-IGRT and PROFIT) [157,158]. Twenty urinary toxicity outcomes were available corresponding to different urinary symptoms, toxicity grades and time of symptom manifestation. Due to the loss-to-follow up, which can occur any time during the follow-up period, some datasets consist of subsamples with fewer/more patients and with lower/higher class imbalance. For the purpose of this study, we chose to evaluate late urinary retention as the corresponding dataset was moderately imbalanced (1:2). The dataset for 5-years late retention consisted of 122 patients. 82 of them (66%) remained asymptomatic by that time (majority class) and 40 of them (33%) developed urinary toxicity within the 5-years period (minority class).

5.2.2 Classifiers

Eight common classifiers were selected and implemented using the R package, caret. The selection includes classifiers frequently used in medical data analysis:

- Least Absolute Shrinkage and Selection Operator (LASSO) [160] is a regularized form of logistic regression that effectively performs model selection. The added shrinkage regularization (i.e. feature selection) makes it is suitable for datasets with many features while maintaining the interpretability of a standard logistic regression.
- Generalized Linear Models with likelihood based boosting (GLMboost) [182] is a machine learning method for optimizing prediction accuracy and for obtaining statistical model estimates via gradient descent techniques. A key feature of the method is that it carries out variable selection during the fitting process without relying on stepwise variable selection.
- Naive-Bayes is a probabilistic classifier based on Bayes' theorem [183] with the assumption of independence between features. It detects the class type based on the maximum probability obtained for the given tuple to a particular class.
- Decision Trees C5.0 [184]. A decision tree iteratively subdivides the training set by selecting feature cutoffs. Decision trees can model nonlinear effects and are easily interpretable.
- Random Forests (RF) generate a large number of decision trees based on random subsamples of the training set while also randomly varying the features used in the trees. Random forests allow modeling nonlinear effects. A random forest model is an ensemble of many decision trees and is, therefore, more difficult to interpret.

- Support vector machines (SVM) is a non-probabilistic, linear, binary classifier used for classifying data by learning a hyperplane separating the data. Linear SVMs is defining the separating hyperplane in the original feature space and thus is more intuitive than the radical kernel SVM.
- Partial Least Squares Discriminant Analysis (PLS-DA) [185] is a dimensionality reduction technique based on PLS regression that is used when the response variable is categorical. Instead of finding hyperplanes of maximum variance between the response and independent variables PLS finds a linear regression model by projecting the predicted variables and the observed variables into a new space. PLS-DA can provide good insight into the causes of discrimination via weights and loadings.
- Regularized Discriminant Analysis (RDA) [186] builds a classification rule using regularized group covariance matrices that are supposed to be more robust in presence of multicollinear data. It lacks, however, interpretability and is impractical for high-dimensional datasets.

More details about these classifiers can be found in Machine Learning textbooks [163].

5.2.3 Oversampling techniques

Random oversampling often results in overfitting while undersampling may weaken the classifiers performance. Thus, the method of choice for handling the class imbalance problem was the Synthetic Minority Oversampling Technique (SMOTE) [187] and some of its' variations. We implemented four commonly used techniques for synthetically balancing the data by oversampling the minority class:

- SMOTE produces new minority observations based on weighted average of the k-nearest neighbors of the same class. These synthetically generated minority class instances make the class distributions more balanced.
- Borderline (BD) SMOTE [188] oversamples the minority examples only near the borderline with the majority class. Compared to regular SMOTE, borderline-SMOTE does not create synthetic examples for noise instances, but concentrates its effort near the borderline, which in turn helps the decision function to create better boundaries between classes.
- SMOTE+ENN [189] copes with the issue of SMOTE where artificial minority class examples are produced too deeply in the majority class space. Inducing a classifier under such a situation can lead to overfitting. In this study Wilson's Edited Nearest Neighbor Rule (ENN) [190] was used to remove noisy SMOTE examples while leaving the original data unchanged.
- Adaptive Synthetic oversampling (ADASYN) [191] uses a weighted distribution for the minority class objects according to their level of difficulty of learning. In comparison to borderline-SMOTE, ADASYN creates different synthetic samples for the minority class depending on its distribution and not just for the borderline instances. The synthetic samples are created based on the majority nearest neighbors via the k-NN method. One drawback of this approach is that it does not identify noisy instances, and thus becomes susceptible to outliers.

5.2.4 *Experimental Design*

The performance of the classifiers was initially computed on the original dataset (before oversampling), which served as the reference dataset for the analysis. Given the small number of available examples, splitting beforehand into training and testing sets was not attempted. Thus, a 5-fold cross-validation (CV) was performed, repeated 20 times, resulting in 100 models. The performance of each classifier was estimated according to the maximum AUC obtained through the 5-fold CV. This is a common strategy to build and validate a model internally when the amount of data is not large enough to split in training/test sets. The results were then used to evaluate the impact of the different oversampling techniques on classifiers' performance.

When the oversampling was performed, a nested CV was implemented whereby the number of both inner and outer folds was set to 5. The experimental design for the nested CV is depicted in Figure 5.1.

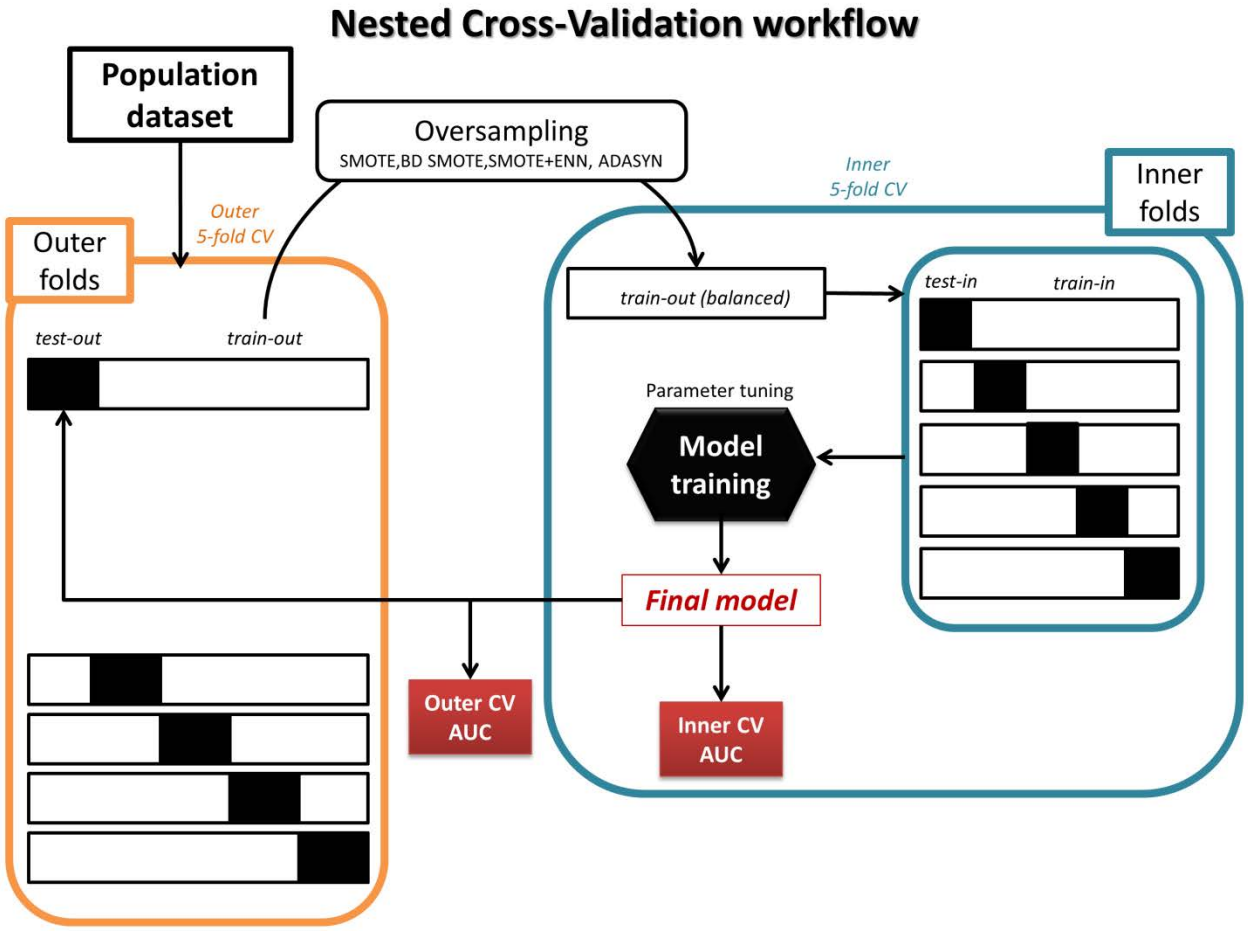


Figure 5.1 Workflow of the nested cross-validation

Each dataset was split into five random subsamples stratified for outcome classes. Hence, each subsample was used as a test-set (test-out) and the remaining observations as training set (train-out). The oversampling was then performed on the minority class of the train-out to balance the two classes.

Features were first centered and then re-scaled and the same transformation was applied to the corresponding test-out. The models were trained on the train-out and applied on the test-out to compute the performance metrics resulting in five estimates per performance metric (1 per test-out).
In the inner CV, the train-out was split again into five subsamples, where each subsample served once as a test set (test-in) and the remaining observations as testing set (train-in). This inner 5-fold CV was repeated 20 times and the models with different tuning parameters were compared. The best tuning parameters were selected according to the maximum AUC of an inner 5-fold CV (test-in). For each repetition different randomization seeds were used to make the process reproducible and ensure that exactly the same splits would be performed across the different oversampling techniques.

To provide meaningful comparisons of the performance of each classifier on the reference dataset (without oversampling) and the datasets after oversampling, the test-in results of the CV were used. For this purpose the area under the ROC curve (AUC) was used to compare the models. The actual performance of the classifiers was then evaluated based on their scores on the test-out sets. The AUC and the F-measure were computed.

The analysis was implemented using various open-source R packages interfaced with the R package caret.

5.3 Results

Table 5.1 shows the best discriminative performance of each classifier on the original dataset (without oversampling) and on the oversampled datasets, after repeated 5-fold CV. Before oversampling, the performance of the classifiers ranged from 0.41 (RF) to 0.65 (PLS-DA). After oversampling the best performance was achieved on the SMOTE+ENN dataset with the RDA classifier (AUC=0.89) and the lowest performance on the SMOTE dataset using the Naive-Bayes classifier (AUC=0.57).

Table 5.1 The AUC for each classifier after repeated 5-fold CV

CLASSIFIERS	No_oversampling	SMOTE	BD_SMOTE	SMOTE_ENN	ADASYN
RF	0.41	0.69	0.70	0.77	0.69
Decision trees	0.53	0.77	0.78	0.83	0.76
SVM	0.62	0.78	0.73	0.79	0.77
PLS-DA	0.65	0.78	0.75	0.84	0.76
GLMboost	0.56	0.66	0.63	0.75	0.60
LASSO	0.56	0.66	0.62	0.73	0.60
Naive-Bayes	0.53	0.57	0.58	0.65	0.61
RDA	0.58	0.85	0.82	0.89	0.83

Figure 5.2 shows the results of the pairwise comparisons between the datasets averaged over all the classifiers (one-sided Wilcoxon signed-rank test). The alternative hypothesis was that the oversampling techniques 1 performed better than the oversampling techniques 2. The numbers in the plot indicate the p-values. The color indicates whether the increased AUCs by oversampling technique 1 are statistically

significant, or close to significant (p-value <0.1), than the oversampling technique 2. Consistently higher was the performance on the SMOTE+ENN dataset. However, all the oversampling techniques showed improved classification performance compared to the original dataset.

Figure 5.2 Pairwise comparisons between classifiers

Figure 5.3 shows the results of the pairwise comparisons between the classifiers averaged over all the datasets (one-sided Wilcoxon signed-rank test). The alternative hypothesis was that the Classifiers 1 performed better than the Classifiers 2. The numbers in the plot indicate the p-values. The color indicates whether the increased AUCs by the classifiers 1 are statistically significant, or close to significant (p-value <0.1), than the classifiers 2. The best performing classifier was the RDA followed by the PLS-DA. The least performing classifier was Naive-Bayes. Overall, RDA, PLS-DA, SVM and Decision Trees outperformed RF, LASSO, GLMboost and Naive-Bayes.

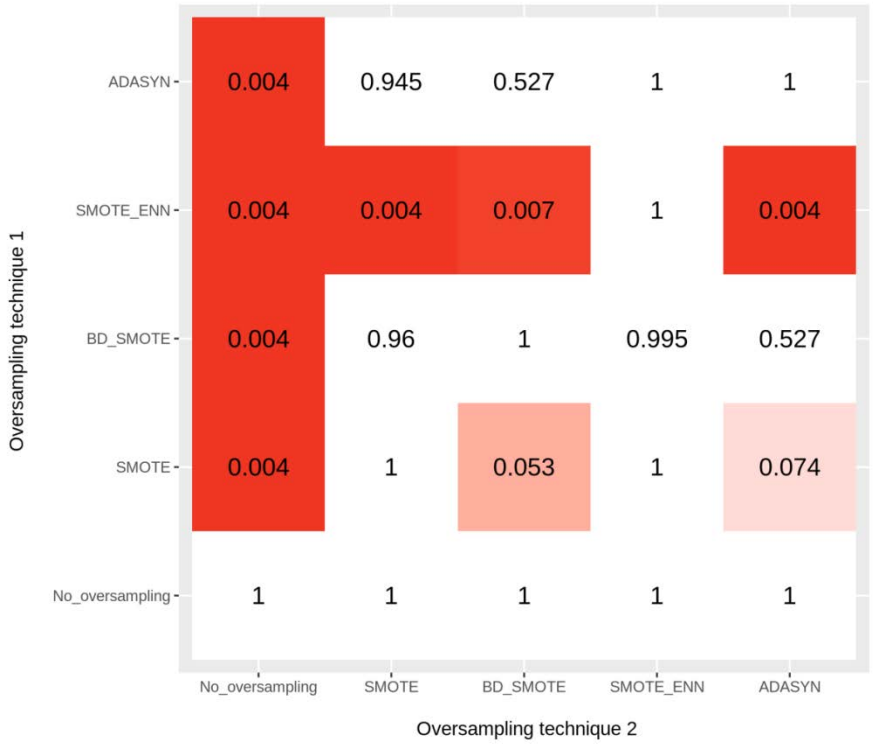


Figure 5.3 Pairwise comparisons between oversampling techniques

Table 5.2 shows the actual performance of each classifier and for each dataset on the "never-seen" test-sets (test-out). The highest performing classifier across all the datasets was again the RDA with average AUC=0.69. Second scored the PLS-DA with AUC=0.65. Similar were the results for the F-measure as shown in the Table 3. The highest F-measure=0.70 was obtain with RDA followed by the PLS-DA algorithm (F-measure=0.66).

Table 5.2 The average AUC of the test-out for each classifier and resampling technique

CLASSIFIERS	AUC				
	SMOTE	*BD SMOTE*	*SMOTE + ENN*	*ADASYN*	average
RF	0.60	0.63	0.63	0.61	0.62
Decision trees	0.57	0.55	0.62	0.58	0.58
SVM	0.64	0.61	0.61	0.62	0.62
PLS-DA	0.65	0.62	0.70	0.61	0.65
GLMboost	0.57	0.58	0.58	0.58	0.58
LASSO	0.59	0.57	0.53	0.58	0.57
Naive-Bayes	0.59	0.58	0.6	0.61	0.60
RDA	0.68	0.69	0.71	0.67	**0.69**
average	0.61	0.61	**0.63**	0.61	

Concerning the different oversampling techniques, there were no significant differences between SMOTE, Borderline SMOTE and ADASYN while SMOTE+ENN performed slightly higher. Overall the highest performance was obtained with RDA on the SMOTE+ENN dataset (AUC=0.71, F-measure=0.72).

Table 3 The average F-measure of the test-out for each classifier and resampling technique

CLASSIFIERS	F-measure				
	SMOTE	*BD SMOTE*	*SMOTE + ENN*	*ADASYN*	average
RF	0.53	0.55	0.59	0.51	0.55
Decision trees	0.61	0.62	0.64	0.65	0.63
SVM	0.63	0.63	0.65	0.60	0.63
PLS-DA	0.61	0.67	0.70	0.65	0.66
GLMboost	0.62	0.63	0.61	0.63	0.62
LASSO	0.62	0.64	0.60	0.61	0.62
Naive-Bayes	0.49	0.54	0.50	0.52	0.51
RDA	0.69	0.69	0.72	0.68	**0.70**
average	0.60	0.62	**0.63**	0.61	

5.4 Discussion

The purpose of this study was to compare the discriminative performance of eight binary classifiers in predicting urinary toxicity after prostate cancer radiotherapy. In order to deal with the class imbalance problem, the performance of these classifiers was evaluated using four different synthetic oversampling techniques. Both dosimetric and clinical features were included.

Our results suggest that there is indeed an overall ranking of classifiers, with the two types of discriminant analysis, RDA and PLS-DA, performing the best. We also observed that all the oversampling techniques significantly increased the performance of the models, with the SMOTE+ENN providing the best results. Interestingly, there was no strong dependence of the classifier performance on the oversampling technique as the RDA scored the highest AUC across all the datasets, including the original non-resampled one.

Although it is expected that reducing the number of features in the model, might increase performance, we intentionally performed the analysis without applying any previous feature selection procedure, given that in real world situations, RT data usually include a large number of highly correlated features. This can possibly explain the reason why certain ML algorithms performed better than others in our dataset. A common limitation of many ML algorithms is their poor ability to properly handle strongly dependent variables. Conversely both RDA and PLS-DA have been reported on the literature as robust techniques under such situations. For instance, Dumancas and Bello [192] compared the predictive performance of 12 machine learning algorithms on a dataset with the presence of high multi-colinearity. The objective was to use lipid profile data to predict 5-year mortality. They showed that the highest scoring classifier among the 12, as measured by the AUC, was the PLS-DA.

With respect to synthetically oversampling methods, SMOTE [187] was the first technique which introduced new samples by using the feature space rather than the data space and is today, a well-established tool for oversampling. Although SMOTE seems to work well with low dimensional data, it is less effective when applied on high-dimensional data. This is due to the fact that SMOTE is not able to manage the bias in the majority class for the classifier where the data is high dimensional. Another drawback of the SMOTE algorithm is over generalization of the minority class space. Thus, a plethora of algorithms based on this concept have been proposed in order to overcome the limitations of SMOTE.

Over the past decades, several studies have investigated the potential of SMOTE and SMOTE-variations to overcome the class imbalance problem. Batista et al. [189] performed a study on 13 datasets with different imbalance levels and showed oversampling methods to perform well on datasets with few positive examples. They proposed two methods, SMOTE+ENN and SMOTE+Tomek, and analyzed their behavior against other resampling techniques for dealing with class imbalance. Lopez et al. [193] compared SMOTE, Borderline-SMOTE, ADASYN and some other methods using Decision Trees, Support Vector Machines and k Nearest Neighbors classifiers and evaluated their performance in terms of AUC on 66 datasets. They reported SMOTE and SMOTE+ENN as the top methods with Borderline-SMOTE and ADASYN being also competitive. More et al. [194] performed a survey of different resampling techniques including Random Oversampling, SMOTE, Borderline-SMOTE etc. and concluded SMOTE+ENN to be the best approach in terms of Precision for the majority class and Recall for the minority class. Our results also suggest that SMOTE+ENN is the most efficient approach for dealing with the class imbalance problem. However, we were not able to confirm any significant improvement on the classifiers' performance when using the ADASYN or Borderline-SMOTE over the traditional SMOTE technique. Nevertheless, all the oversampling techniques significantly improved the average classification performance of the ML algorithms compared to the original dataset (without oversampling). These findings demonstrate the importance of handling class imbalance and highlight their efficacy on RT datasets. We have to stress out that complexity of data, level of imbalance, evaluation criteria and choice of classifier, all play crucial role in the evaluation process.

Another common approach to perform class imbalance is undersampling which consists in down-sizing the majority class by randomly removing observations until the dataset is balanced. Despite the popularity of undersampling, this technique was intentionally disregarding in the present study because of the risk of removing relevant observations from the dataset, since the process is performed in an unsupervised manner. Consequently, the more imbalanced the dataset, the more samples will be discarded when undersampling. A study from Pozzolo et al. [195] showed that the beneficial impact of undersampling is strongly dependent on the nature of the classification task (degree of unbalancedness and nonseparability) and on the variance of the classifier and, as a consequence, it is extremely dependent on the specific test point. Although our study is "optimistic" in terms of class imbalance (minority-majority ratio was 1:2), it is particularly common in toxicity prediction studies that the minority class consist of only of a few dozens of patients, in which case undersampling would result to a balanced dataset with only a few observations.

One of the most critical points in performance evaluation is the choice of the correct metric. A metric may correspond to some expected loss over different operating conditions. In the taxonomy proposed by Ferri et al. [196] three families of metrics were recognized: performance metrics which account for the quality of classification (such as accuracy), performance metrics which account for a ranking quality (such as AUC), and performance metrics which evaluate the quality of scores or how well the model does in terms of probability estimation(such as the Brier score or log-loss). The most widespread, but also controversial, measure to evaluate a classifier's performance is accuracy (or error rate which is defined as 1 minus accuracy). As it simply measures the number of correctly predicted samples over the total number of samples, this metric should be used only if the assumptions of balanced class distribution and equal cost of misclassification errors hold true, which is not usually the case for real-world applications.

In this study we used the F-measure and the AUC to evaluate models' performance. The F-measure (also called F1 score) conveys the balance between the precision and the recall and is recommended for imbalanced datasets. It is assumed, however, that precision and recall are equally important. The use of AUC is well-accepted for evaluating classifiers performance, especially for medical applications. The ROC compares the classifiers' performance across the entire range of class distribution and error costs and, in contrast to accuracy, does not require the choice of a single threshold value. Ling et al. [197] were the first to establish that AUC is statistically consistent, more discriminating and an overall better measure than accuracy in evaluating and comparing classification learning algorithms. Nevertheless, AUC also suffers from some limitations. For example, when comparing two classifiers, if ROC curves cross, then it is possible that one curve has a higher AUC (and so is apparently better) even though the alternative may show superior performance over almost the entire range of values of the classification threshold for which the curve will be used [198].

Finally, for the classifiers to be clinically useful, model interpretability is arguably a major requirement. Although all the implemented classifiers yield a variable importance, which allows the identification of the most pertinent features in the model, they are not all as easily interpretable. Unfortunately, our study shows that the most intuitive classifiers (such as LASSO and GLMboost) were among the least performing.

Future work may include the combination of the best classifier in a majority voting scheme in order to take advantage of the best of their individual performance. This can be done by weighting the importance (or credibility) of each classifier based on the training outcomes. Survival machine learning techniques should also be investigated as they take into account right-censoring.

5.5 Conclusion

Oversampling of imbalanced datasets coupled with machine learning models in the present work, offers a benchmark for predicting radiation-induced side-effects following prostate radiotherapy. The results suggest that properly handling class imbalance can significantly improve classifiers' performance, paving

the way for the development of more accurate predictive models for toxicity prediction after RT. Overall, RDA performed better than other methods. However, more research is needed to understand the advantages and limitations of machine learning methods for the prediction of different urinary symptoms.

The next and final chapter contains a general discussion of the main results, the contributions and the limitations of this thesis. Recommendations are also made for future work in prediction of urinary toxicity following prostate cancer radiotherapy.

6 General discussion, conclusions and perspectives

The work presented in this thesis aimed at investigating spatial correlations between dose and side effects by taking into consideration the spatial dose distribution, with the objective to improve urinary toxicity prediction following prostate cancer radiotherapy. This problem was addressed in a population analysis framework where the dose distribution was explored in different spatial scales, from the traditional whole-organ-based models towards sub-organ models.

Historical observations have suggested that there is no clear dose-volume relationship for the bladder. Only recently, extensive studies of the impact of dose distributions across the bladder have been undertaken. These studies, including our own works in this thesis, suggest that there is a significant association between dose distribution and urinary symptom manifestation, with this outcome potentially supplanting previous hypotheses of homogeneous bladder radiosensitivity.

Beyond bladder dose, the assessment of urethra involvement in urinary toxicity was of major importance. Because the urethra is not visible on a CT scan, contouring and assigning constraints to this structure during planning, is not feasible using conventional approaches. To cope with this, we developed a multi-atlas-based methodology for segmenting the prostatic urethra on CT images, as detailed in Chapter 2. This approach for segmenting the urethra is clinically relevant and paves the way for assessing urethral toxicity potentially improving the overall urinary toxicity prediction by considering both the dose received by the bladder and the urethra. Indeed, our work allowed for the first time to correlate urethra dose with urinary toxicity after EBRT.

In the context of local dose analysis, two different methods were implemented and thoroughly described in Chapters 3 and 4. On the one hand, DSMs were generated by unfolding bladder surfaces and normalizing not only to lateral direction, as had been done in previous studies, but also to the cranio-caudal direction allowing to evaluate the dose on the entire bladder surface. On the other hand, for the DVMs, a robust multi-organ registration strategy was devised in order to accurately align the organs (bladder, prostate, and urethra) to a common coordinate system while copying with the high inter-individual anatomical variations. As explained, the proposed DVM models strongly depend on the validity of dose mapping and, hence, on the reliability of the alignment of the whole population to a single coordinate system (template). Taking into consideration this important source of error, we applied a relatively strict threshold for excluding patients from the analysis depending on the organ alignment obtained after non-rigid registration (in terms of Dice score). On top of that, we smoothed the propagated dose distributions based on the registration accuracy.

The generalizability and reproducibility of our observations was established for specific symptoms, through inter-cohort comparisons (DSM models) and external validation (DVM models). This is a crucial step for developing reliable predictive models as an apparent correlation between dose and toxicity does not necessarily mean causality. The identification of subregions may also depend on specific cohort characteristics such as the total number of patients, the toxicity rates, the treatment modality, the total

delivered dose and the dose per fraction. Indeed, the comparison between a DSM analysis on our cohort and other DSM studies in Chapter 3, suggests that the results are strongly dependent on cohort characteristics as only one subregion was confirmed in our cohort (late dysuria). The external validation of our DVM models in Chapter 4, was successful for three out of five subregions highlighting the role of urethra and posterior bladder region in urinary toxicity.

Given the recent shift from classical statistical analysis methods, towards machine learning algorithmic and data mining techniques, as powerful tools to extract information from existing recorded data, this thesis may also lay the foundations for predicting radiation-induced side-effects using machine learning strategies. Our observations suggest that discriminant analysis algorithms, such as RDA and PLS-DA, may be suitable for analyzing highly correlated, structured data, like DVH bins. In the case of class imbalance, as commonly observed in real-world data, we recommend to synthetically create new patient data with the objective to balance the ratio between patients with and without toxicities, which we showed to effectively improve outcome prediction capabilities of machine learning algorithms. In the clinical perspective, such approaches could be used concomitantly to create computer-based decision support tools for treatment optimization.

6.1 Comparison of 2D DSM and 3D DVM methods

Two local dose-analysis methods (DSM and DVM) were used to identify predictive symptom-related sub-regions (Ssurf and Svol). Both 2D DSMs and 3D DVMs analyses allowed the identification of sub-regions predictive of distinct acute and late urinary symptoms. Given that DSM and DVM analyses were applied to the same population dataset, a direct comparison of the two methods, in terms of sub-region identification and predictive performance, is meaningful.

Using DVMs five Svol in the bladder and urethra were identified, in contrast to the DSM analysis which allowed the identification of only three Ssurf 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 to urinary toxicity [168,199]. Nevertheless, for two symptoms (acute and late retention), both methods identified a quite similar sub-region in the bladder (posterior part of the bladder including the bladder trigone) corresponding to intermediate-high doses. Figure 6.1 summarizes the number and the localization of the identified subregions with the two methods and depicts the spatial overlap between Ssurf and Svol.

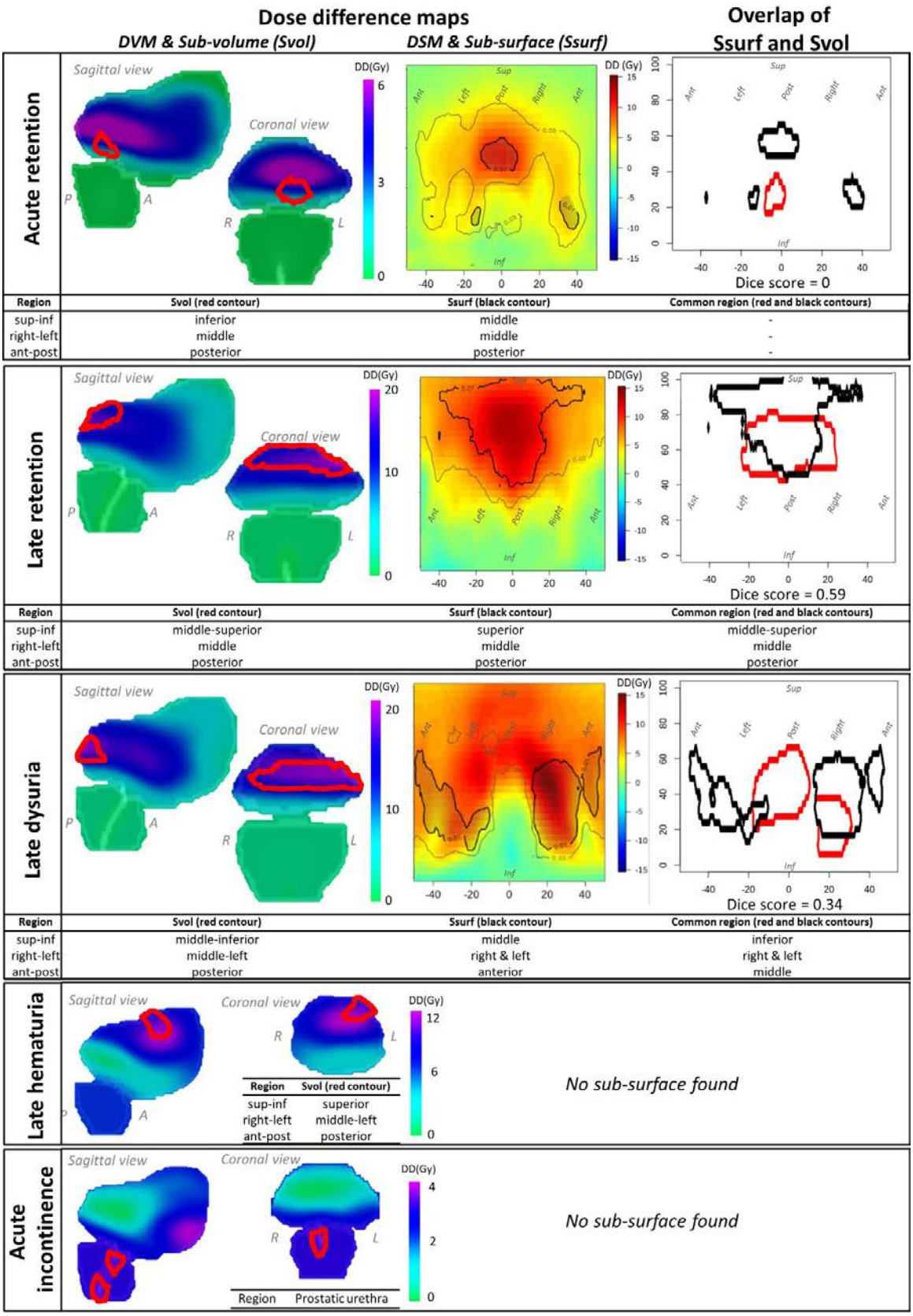


Figure 6.1 Spatial overlap between sub-surfaces (Ssurf) and sub-volumes (Svol)

The predictive capabilities of the dose to the identified sub-volumes were evaluated with respect to the current clinical standard, which is the dose received by the whole bladder. Indeed, the dose to the whole bladder was found to be informative to some extent. Nevertheless, the subregions identified by any of the two methods, were consistently more predictive than the dose to the whole bladder. This trend is in line with previous DSM studies. In our dataset, the AUCs ranged from 0.60 to 0.72 for the whole bladder, from 0.64 to 0.74 for the Ssurf and from 0.62 to 0.81 for the Svol, in univariate analysis.

Each of the two methods presents advantages and disadvantages with respect to the data they require, the information they provide, and the flexibility of implementation. DVM construction requires a multi-organ non-rigid registration with large number of degrees of freedom and uncertainties, although reduced with a regularized deformation field and a dose smoothing process, they are not eradicated. DSM registration uncertainties are less pronounced since it is based upon an affine parametric transformation which implies less degrees of freedom. DVM methods have, however, the advantage that they can be used to explore the entire pelvic region without any prior information while DSM methods assume that a subregion is located exclusively on the bladder surface. Considering the planning dose distribution only on the bladder surface is subject to high dosimetric uncertainties due to the high inter-fraction bladder variations. A 3D Svol close to the bladder surface might better represent the actual dose delivered to the surface, although bladder is a hollow organ, because it implicitly accounts for inter-fraction bladder-surface motion within this region.

More generally speaking, DSM methods are intrinsically restricted to the surface of the organ, and hence, exclusively applicable to hallow organs. However, most organs are solid, in which case DVM analyses, taking into considerations the entire 3D dose distribution within the organ, are undeniably more appropriate.

6.2 Limitations of the work

One limitation to develop and apply statistical methodologies is the data availability and in our studies we considered a relatively small number of patients (n = 254). Also, the limited number of patients who experience grade≥2 toxicities hindered our ability to model these clinically relevant endpoints. Although, clinician-only reporting can result in under-reporting of lower grade morbidity and the downgrading of symptom severity [200–203], predictive models developed based on solely grade≥1 urinary toxicity may be suboptimal.

Another limitation of our work is the exploitation of information from the planning step. Because of the anatomical variations that may occur during the treatment, the dose actually received by the bladder may be different [86,127,204,205], although IGRT was used for our population. Any discrepancy between planning and actual delivered dose may be the origin of uncertainties in toxicity prediction. In fact bladder base remains relatively stable with minor or no day-to-day variations. The bladder dome, on the

other hand, presents the highest extensibility raising the question of whether the planning dose distribution at this region can be considered as the actual delivered dose.

In general the prediction capabilities of the proposed models, including clinical and dosimetric variables, were not remarkable. This may be due to the fact that for some patients the relationship between dose and side effect is not directly established. Indeed, there may exist individual specificities related with the occurrence of side-effects such as individual radio-sensitivity or other factors determined by genetic and/or epigenetic mechanisms [129,206,207], that were not investigated in this thesis.

A potential source of error is the choice of template for aligning the whole population. In our analyses, we assumed that the selected template is representative of the given population. However, the remarkably high inter-individual bladder volume variations raise questions concerning the validity of this assumption. Since the impact of different anatomical references for spatially aligning the population was not evaluated, the possibility of identifying different subregions on different templates cannot be excluded.

Concerning the performance of different machine learning strategies, side-by-side comparisons were performed using only one endpoint. Concluding the superiority of a strategy based on only one endpoint may not be optimal for deriving a solid conclusion. Findings have less chance of being fortuitous if they are repeated across endpoints and datasets.

6.3 Perspectives

From methodological point of view, the work presented in this thesis paves the way for in-depth studies, or new ones, aiming to improve our understanding of urinary symptom manifestation and increase urinary toxicity prediction following prostate cancer radiotherapy. From clinical point of view, a potential benefit of using DVMs is the possibility of performing personalized treatment planning by back-propagating the identified sub-regions from the template to the patient's native space and adding specific dosimetric constraints. Sparing sub-regions in the treatment planning system, as previously done for the rectum [97], may prevent specific side effects.

More precisely, among numerous perspectives we can mention:

1. The possibility to decrease the dose in specific sub-volumes by defining 3D patient-specific constraints as part of the TPS optimization, while preserving the dose in the prostate, needs to be demonstrated. As such, a randomized clinical study on the therapeutic benefit of adding dosimetric constraints to the sub-regions during treatment planning could be considered.
2. The availability of strong features is key in the construction of robust predictive models. Indeed, we found that including spatial descriptors of the dose distributions can increase model performance but these models still require new features. By including individual biological parameters is expected to achieve better predictive capabilities.

3. By combining DSM and DVM analyses, the impact of dosimetric and geometric uncertainties can be potentially eliminated, thereby allowing the development of more reliable models. Thus it might be useful to evaluate the complementarity of the two methods.
4. Future studies should also focus on studying the impact of dosimetric uncertainties produced by inter-fraction variations of the bladder on the prediction of urinary toxicity in prostate cancer radiotherapy by considering the cumulative dose distribution during treatment. Some works in this direction have been previously performed in order to use machine learning strategies to predict inter-fraction organ deformations [208,209].
5. In the same context, in order to account for bladder motion, DVM studies should be also extended to areas outside of the bladder and even to the whole pelvic region if accurate alignment of the population is feasible.
6. Apart from prostate cancer radiotherapy, our proposed DVM approach can be extended to other organs for investigating toxicities following radiotherapy at different anatomical sites. Through multi-organ non-rigid registration the simultaneous exploration of the dose to multiple organs is also feasible.
7. Our 3D DVM approach can still be improved. In particular different registration pipelines and correction techniques should be tested in order to maximize registration accuracy and minimize potential bias of subregion identification due to misalignments.
8. Selection of the optimal template to be representative of a given population need to be further investigated. Considering the high interindividual variability, a study of the influence of the template selection on the identification of risk regions should also be considered.
9. Future research studies should focus on grade ≥2 urinary toxicity rather than grade ≥1, as it is a more clinically relevant endpoints with significant impact on the QoL for the patient. Given the lack of events for these types of toxicities in our dataset, a potential solution would be to apply oversampling techniques such as SMOTE in order to allow the exploration of these endpoints.
10. The best performing machine learning algorithms can also be ensembled, using the majority voting technique, in order to increase the classification and prediction accuracy. Such elaborate models will offer new perspectives in predicting and preventing radiation-induced urinary complications.
11. New Machine Learning strategies, such as Deep Learning can be also considered. These emerging techniques would be able to train neural networks, fully exploiting the 3D dose distributions, and potentially increase the prediction without requiring previous registrations.

The overview of the thesis structure, as well as the contributions and the perspectives are illustrated in Figure 6.2.

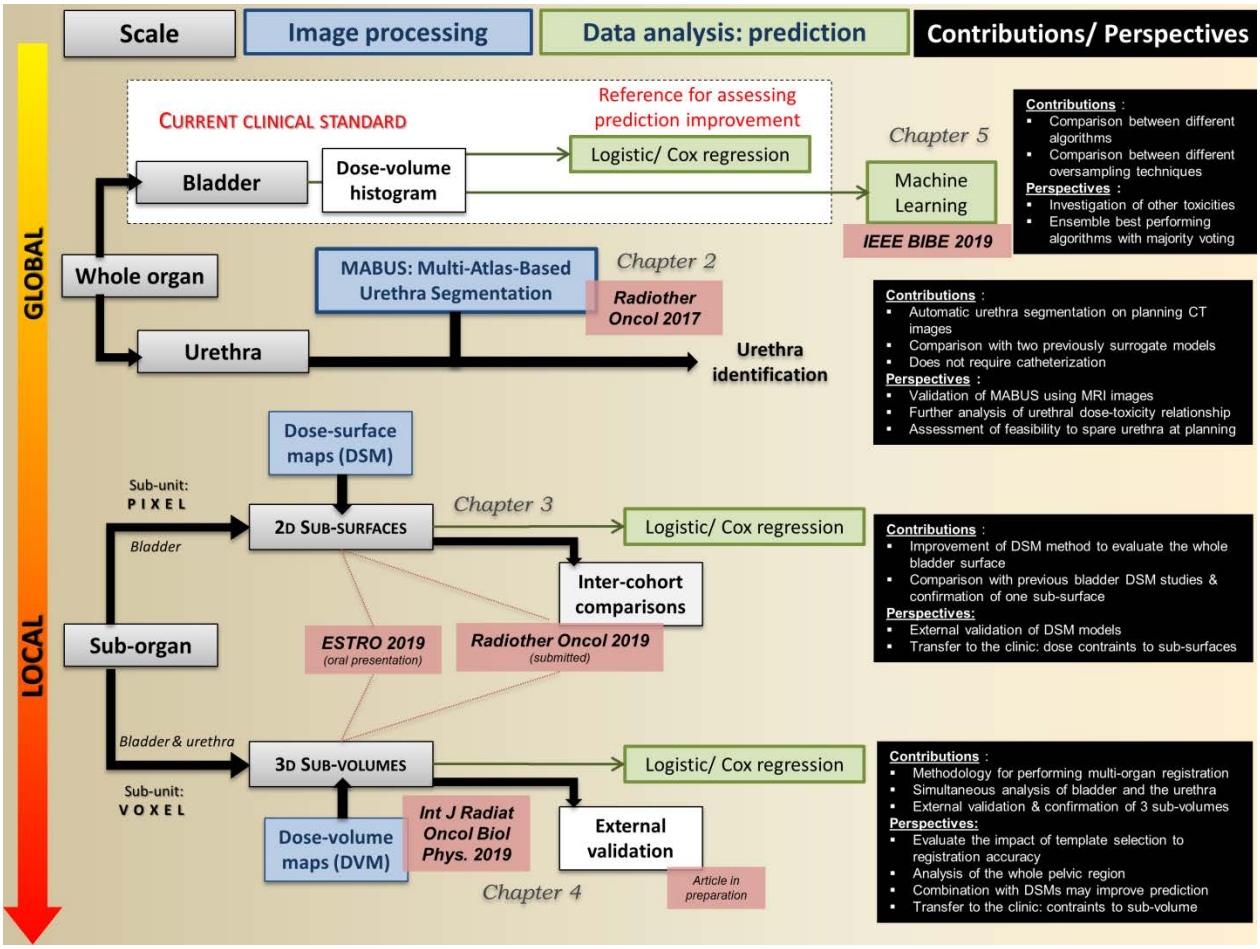


Figure 6.2 Schematic overview of the thesis structure, contributions and perspectives

List of Publications

The work presented in this thesis have been already published or submitted for publication in the following journals and communications:

International journal papers

- ✓ Acosta O, **Mylona E**, Le Dain M, Voisin C, Lizee T, Rigaud B, Lafond C, Gnep K, de Crevoisier R. *Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy*. Radiother. Oncol. 2017;125:492–499.

- ✓ **Mylona E**, Acosta O, Lizee T, Lafond C, Crehange G, Magné N, Chiavassa S, Supiot S, Ospina JD, Campillo-Gimenez B, Castelli J, de Crevoisier R. *Voxel-Based Analysis for Identification of Urethrovesical Subregions Predicting Urinary Toxicity After Prostate Cancer Radiation Therapy.* Int J Radiat Oncol Biol Phys, 2019. 104 (2): 343-54.

- ✓ Largent A, Barateau A, Nunes JC, **Mylona E**, Castelli J, Lafond C, Greer P, Dowling J, Baxter J, Saint-Jalmes H, Acosta O, de Crevoisier R*. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning.* Int J Radiat Oncol 2019. doi:10.1016/J.IJROBP.2019.08.049.

- ✓ **Mylona E**, Cicchetti A, Rancati T, Palorini F, Fiorino C, Supiot S, Magne N, Creange G, Valdagni R, Acosta O, de Crevoisier R. *Local dose analysis to predict acute and late urinary toxicities after prostate cancer radiotherapy : assessment of cohort and method effects.* Radiother. Oncol. (Submitted)

Peer-reviewed conference papers

- ✓ **Mylona E**, Lebreton C, Fontaine C, Crehange G, Magné N, Supiot S, de Crevoisier R, Acosta O. *Comparison of machine learning algorithms and oversampling techniques for urinary toxicity prediction after prostate cancer radiotherapy.* IEEE proceedings (BIBE 2019, 28-30/10/2019, Athens, Greece)

Communications in international conferences

- ✓ Lizee T, **Mylona E**, Lafond C, Supiot S, Acosta O, de Crevoisier R. *Analysis of the urethro-vesical region for urinary toxicity prediction after prostate radiotherapy.* Radiother. Oncol. 2018;127:432-433. (in ESTRO as poster)

- ✓ **Mylona E**, Acosta O, Lizee T, Lafond C, Crehange G, Magné N, Chiavassa S, Supiot S, Ospina JD, Campillo-Gimenez B, Castelli J, de Crevoisier R. *Bladder and urethra subregions predicting urinary toxicity after prostate cancer radiotherapy.* Radiother. Oncol. 2019;133:449. (in ESTRO as poster)

- ✓ **Mylona E**, Cicchetti A, Rancati T, Palorini F, Supiot S, Magne N, Creange G, Acosta O, de Crevoisier R. *Predicting urinary toxicity via 2D and 3D dose map analyses in prostate cancer radiotherapy.* Radiother. Oncol. 2019;133:326. (in ESTRO as oral communication)

- ✓ Acosta O, Lafond C, Barateau A, Houede B, Largent A, **Mylona E**, Perichon N, Belaby N, Haigron P, de Crevoisier R, Improved prostate cancer radiotherapy planning with decreased dose in a rectal subregion highly predictive for toxicity. Surgetica 2019.

Communications in national conferences

- ✓ Lizée T, Acosta O, **Mylona E**, Le Dain M, Lafond C, Riet FG, de Crevoisier. *Développement d'une méthode de segmentation automatique de l'urètre sur tomodensitométrie de planification permettant d'évaluer la dose urétrale en cas de radiothérapie prostatique.* Cancer/Radiothérapie 2017;21:706. doi:10.1016/J.CANRAD.2017.08.062.

- ✓ Acosta O, **Mylona E**, Lafond C, Créhange G, Supiot S, Castelli J, de Crevoisier R. *Identification de sous-régions rectale et urétrovésicales hautement prédictives de toxicité en cas d'irradiation prostatique.* Cancer/Radiothérapie 2019;23:791. doi:10.1016/J.CANRAD.2019.07.010.

International Mobilities

- ✓ Project: Spatial characterization of urinary toxicity using dose-surface maps. Responsible: Tiziana Rancati, Place: Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy. Project: Period: 9-23 September 2018. Grant: ESTRO mobility grant.

- ✓ Project: External validation of dose-volume map models. Responsible: Martin Ebert, Place: Sir Charles Gairdner Hospital, University of Western Australia, Perth, Australia. Period: December 2018-April 2019. Grant: PHC FASIC mobility grant & UBL mobility grant

Bibliography

[1] Rawla P. Epidemiology of Prostate Cancer. World J Oncol 2019;10:63–89. doi:10.14740/wjon1191.

[2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424. doi:10.3322/caac.21492.

[3] Stangelberger A, Waldert M, Djavan B. Prostate cancer in elderly men. Rev Urol 2008;10:111–9.

[4] Heidenreich A, Aus G, Bolla M, Joniau S, Matveev VB, Schmid HP, et al. EAU Guidelines on Prostate Cancer. Eur Urol 2008;53:68–80. doi:10.1016/j.eururo.2007.09.002.

[5] Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, et al. EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent—Update 2013. Eur Urol 2014;65:124–37. doi:10.1016/j.eururo.2013.09.046.

[6] VARENHORST E, BERGLUND K, LÖFMAN O, PEDERSEN K. Inter-observer Variation in Assessment of the Prostate by Digital Rectal Examination. Br J Urol 1993;72:173–6. doi:10.1111/j.1464-410X.1993.tb00682.x.

[7] Ankerst DP, Thompson IM. Sensitivity and specificity of prostate-specific antigen for prostate cancer detection with high rates of biopsy verification. Arch Ital Di Urol Androl Organo Uff [Di] Soc Ital Di Ecogr Urol e Nefrol 2006;78:125–9.

[8] Stenman U-H, Abrahamsson P-A, Aus G, Lilja H, Bangma C, Hamdy FC, et al. Prognostic value of serum markers for prostate cancer. Scand J Urol Nephrol 2005;39:64–81. doi:10.1080/03008880510030941.

[9] Sobin LH, Gospodarowicz MK (Mary K., Wittekind C (Christian), International Union against Cancer. TNM classification of malignant tumours. Wiley-Blackwell; 2009.

[10] D'Amico A V., Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al. Biochemical Outcome After Radical Prostatectomy, External Beam Radiation Therapy, or Interstitial Radiation Therapy for Clinically Localized Prostate Cancer. JAMA 1998;280:969. doi:10.1001/jama.280.11.969.

[11] Wallace TJ, Torre T, Grob M, Yu J, Avital I, Brücher B, et al. Current approaches, challenges and future directions for monitoring treatment response in prostate cancer. J Cancer 2014;5:3–24. doi:10.7150/jca.7709.

[12] Welty CJ, Cooperberg MR, Carroll PR. Meaningful end points and outcomes in men on active surveillance for early-stage prostate cancer. Curr Opin Urol 2014;24:288–92. doi:10.1097/MOU.0000000000000039.

[13] Adolfsson J. Watchful waiting and active surveillance: the current position. BJU Int 2008;102:10–4. doi:10.1111/j.1464-410X.2008.07585.x.

[14] Haar G Ter, Coussios C. High intensity focused ultrasound: physical principles and devices. Int J Hyperthermia 2007;23:89–104.

[15] Poissonnier L, Murat F-J, Belot A, Bouvier R, Rabilloud M, Rouviere O, et al. Locally recurrent prostatic adenocarcinoma after exclusive radiotherapy: results of high intensity focused ultrasound. Prog Urol 2008;18:223–9. doi:10.1016/j.purol.2008.02.006.

[16] Barker HE, Paget JTE, Khan AA, Harrington KJ. The tumour microenvironment after radiotherapy: mechanisms of resistance and recurrence. Nat Rev Cancer 2015;15:409–25. doi:10.1038/nrc3958.

[17] Romeijn HE, Ahuja RK, Dempsey JF, Kumar A. A column generation approach to radiation therapy treatment planning using aperture modulation. SIAM J Optim 2005. doi:10.1137/040606612.

[18] Bortfeld T. IMRT: a review and preview. Phys Med Biol 2006;51:R363–79. doi:10.1088/0031-9155/51/13/R21.

[19] Cho B. Intensity-modulated radiation therapy: a review with a physics perspective. Radiat Oncol J

2018;36:1–10. doi:10.3857/roj.2018.00122.
[20] Zaorsky NG, Keith SW, Shaikh T, Nguyen PL, Horwitz EM, Dicker AP, et al. Impact of Radiation Therapy Dose Escalation on Prostate Cancer Outcomes and Toxicities. Am J Clin Oncol 2018;41:409–15. doi:10.1097/COC.0000000000000285.
[21] Dearnaley DP, Jovic G, Syndikus I, Khoo V, Cowan RA, Graham JD, et al. Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: long-term results from the MRC RT01 randomised controlled trial. Lancet Oncol 2014;15:464–73. doi:10.1016/S1470-2045(14)70040-3.
[22] Cheung K. Intensity modulated radiotherapy: advantages, limitations and future developments. Biomed Imaging Interv J 2006;2:e19. doi:10.2349/biij.2.1.e19.
[23] Fischer-Valuck BW, Rao YJ, Michalski JM. Intensity-modulated radiotherapy for prostate cancer. Transl Androl Urol 2018;7:297–307. doi:10.21037/tau.2017.12.16.
[24] Schlegel W. New Technologies in 3D Conformal Radiation Therapy: Introduction and Overview. New Technol. Radiat. Oncol., Berlin/Heidelberg: Springer-Verlag; 2006, p. 1–6. doi:10.1007/3-540-29999-8_1.
[25] Sonke J-J, Aznar M, Rasch C. Adaptive Radiotherapy for Anatomical Changes. Semin Radiat Oncol 2019;29:245–57. doi:10.1016/J.SEMRADONC.2019.02.007.
[26] Alasti H, Petric MP, Catton CN, Warde PR. Portal imaging for evaluation of daily on-line setup errors and off-line organ motion during conformal irradiation of carcinoma of the prostate. Int J Radiat Oncol Biol Phys 2001. doi:10.1016/S0360-3016(00)01446-2.
[27] Moseley DJ, White EA, Wiltshire KL, Rosewall T, Sharpe MB, Siewerdsen JH, et al. Comparison of localization performance with implanted fiducial markers and cone-beam computed tomography for on-line image-guided radiotherapy of the prostate. Int J Radiat Oncol Biol Phys 2007;67:942–53. doi:10.1016/j.ijrobp.2006.10.039.
[28] Dang A, Kupelian PA, Cao M, Agazaryan N, Kishan AU. Image-guided radiotherapy for prostate cancer. Transl Androl Urol 2018;7:308–20. doi:10.21037/tau.2017.12.37.
[29] Pang EPP, Knight K, Fan Q, Tan SXF, Ang KW, Master Z, et al. Analysis of intra-fraction prostate motion and derivation of duration-dependent margins for radiotherapy using real-time 4D ultrasound. Phys Imaging Radiat Oncol 2018;5:102–7. doi:10.1016/J.PHRO.2018.03.008.
[30] Nguyen D, Lyu Q, Ruan D, O'Connor D, Low DA, Sheng K. A comprehensive formulation for volumetric modulated arc therapy planning. Med Phys 2016;43:4263–72. doi:10.1118/1.4953832.
[31] King CR, Lehmann J, Adler JR, Hai J. CyberKnife Radiotherapy for Localized Prostate Cancer: Rationale and Technical Feasibility. Technol Cancer Res Treat 2003;2:25–9. doi:10.1177/153303460300200104.
[32] Mittauer K, Paliwal B, Hill P, Bayouth JE, Geurts MW, Baschnagel AM, et al. A New Era of Image Guidance with Magnetic Resonance-guided Radiation Therapy for Abdominal and Thoracic Malignancies. Cureus 2018;10. doi:10.7759/CUREUS.2422.
[33] Moon DH, Efstathiou JA, Chen RC. What is the best way to radiate the prostate in 2016? Urol Oncol Semin Orig Investig 2016:1–10. doi:10.1016/j.urolonc.2016.06.002.
[34] 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.
[35] Michalski JM, Moughan J, Purdy J, Bosch W, Bruner DW, Bahary J-P, et al. Effect of Standard vs Dose-Escalated Radiation Therapy for Patients With Intermediate-Risk Prostate Cancer: The NRG Oncology RTOG 0126 Randomized Clinical Trial. JAMA Oncol 2018;4:e180039. doi:10.1001/jamaoncol.2018.0039.
[36] Kuban DA, Tucker SL, Dong L, Starkschall G, Huang EH, Cheung MR, et al. Long-Term Results of the M. D. Anderson Randomized Dose-Escalation Trial for Prostate Cancer. Int J Radiat Oncol Biol Phys 2008;70:67–74. doi:10.1016/j.ijrobp.2007.06.054.

[37] Peeters STH, Heemsbergen WD, Koper PCM, van Putten WLJ, Slot A, Dielwart MFH, et al. Dose-Response in Radiotherapy for Localized Prostate Cancer: Results of the Dutch Multicenter Randomized Phase III Trial Comparing 68 Gy of Radiotherapy With 78 Gy. J Clin Oncol 2006;24:1990–6. doi:10.1200/JCO.2005.05.2530.
[38] Zelefsky MJ, Fuks Z, Hunt M, Lee HJ, Lombardi D, Ling CC, et al. High dose radiation delivered by intensity modulated conformal radiotherapy improves the outcome of localized prostate cancer. J Urol 2001;166:876–81.
[39] Zietman AL, DeSilvio ML, Slater JD, Rossi CJ, Miller DW, Adams JA, et al. Comparison of Conventional-Dose vs High-Dose Conformal Radiation Therapy in Clinically Localized Adenocarcinoma of the Prostate. JAMA 2005;294:1233. doi:10.1001/jama.294.10.1233.
[40] Cheung R, Tucker SL, Lee AK, de Crevoisier R, Dong L, Kamat A, et al. Dose–response characteristics of low- and intermediate-risk prostate cancer treated with external beam radiotherapy. Int J Radiat Oncol 2005;61:993–1002. doi:10.1016/j.ijrobp.2004.07.723.
[41] Cheung R, Tucker SL, Dong L, Kuban D. Dose-response for biochemical control among high-risk prostate cancer patients after external beam radiotherapy. Int J Radiat Oncol Biol Phys 2003;56:1234–40. doi:10.1016/s0360-3016(03)00278-5.
[42] Zietman AL, Bae K, Slater JD, Shipley WU, Efstathiou JA, Coen JJ, et al. Randomized Trial Comparing Conventional-Dose With High-Dose Conformal Radiation Therapy in Early-Stage Adenocarcinoma of the Prostate: Long-Term Results From Proton Radiation Oncology Group/American College of Radiology 95-09. J Clin Oncol 2010;28:1106–11. doi:10.1200/JCO.2009.25.8475.
[43] Heemsbergen WD, Al-Mamgani A, Slot A, Dielwart MFH, Lebesque J V. Long-term results of the Dutch randomized prostate cancer trial: Impact of dose-escalation on local, biochemical, clinical failure, and survival. Radiother Oncol 2014;110:104–9. doi:10.1016/j.radonc.2013.09.026.
[44] Dulaney CR, Osula DO, Yang ES, Rais-Bahrami S. Prostate Radiotherapy in the Era of Advanced Imaging and Precision Medicine. Prostate Cancer 2016;2016:4897515. doi:10.1155/2016/4897515.
[45] Fiorino C, Valdagni R, Rancati T, Sanguineti G. Dose-volume effects for normal tissues in external radiotherapy: Pelvis. Radiother Oncol 2009;93:153–67. doi:10.1016/j.radonc.2009.08.004.
[46] C.M. G, R.V. N. Pelvic Irradiation and Its Effects on the Lower Urinary Tract: a Literature Review. Curr Bladder Dysfunct Rep 2015;10:295–302. doi:10.1007/s11884-015-0316-5.
[47] Leplège A, Hunt S. The Problem of Quality of Life in Medicine. JAMA J Am Med Assoc 1997;278:47. doi:10.1001/jama.1997.03550010061041.
[48] Berkey FJ. Managing the adverse effects of radiation therapy. Am Fam Physician 2010.
[49] Litwin MS, Lubeck DP, Henning JM, Carroll PR. Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: Results of the capsure database. J Urol 1998. doi:10.1016/S0022-5347(01)63222-1.
[50] Clark JA, Inui TS, Silliman RA, Bokhour BG, Krasnow SH, Robinson RA, et al. Patients' perceptions of quality of life after treatment for early prostate cancer. J Clin Oncol 2003. doi:10.1200/JCO.2003.02.115.
[51] Eton DT, Lepore SJ. Prostate cancer and health-related quality of life: A review of the literature. Psychooncology 2002. doi:10.1002/pon.572.
[52] Marks LB, Carroll PR, Dugan TC, Anscher MS. The response of the urinary bladder, urethra, and ureter to radiation and chemotherapy. Int J Radiat Oncol Biol Phys 1995;31:1257–80. doi:10.1016/0360-3016(94)00431-J.
[53] Haldar S, Dru C, Bhowmick NA. Mechanisms of hemorrhagic cystitis. Am J Clin Exp Urol 2014;2:199–208.
[54] Olsson CE, Jackson A, Deasy JO, Thor M. A Systematic Post-QUANTEC Review of Tolerance Doses

for Late Toxicity After Prostate Cancer Radiation Therapy. Int J Radiat Oncol 2018;102:1514–32. doi:10.1016/j.ijrobp.2018.08.015.
[55] Budäus L, Bolla M, Bossi A, Cozzarini C, Crook J, Widmark A, et al. Functional outcomes and complications following radiation therapy for prostate cancer: A critical analysis of the literature. Eur Urol 2012;61:112–27. doi:10.1016/j.eururo.2011.09.027.
[56] Zelefsky MJ, Kollmeier M, Cox B, Fidaleo A, Sperling D, Pei X, et al. Improved Clinical Outcomes With High-Dose Image Guided Radiotherapy Compared With Non-IGRT for the Treatment of Clinically Localized Prostate Cancer. Int J Radiat Oncol 2012;84:125–9. doi:10.1016/j.ijrobp.2011.11.047.
[57] 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.
[58] Ghadjar P, Zelefsky MJ, Spratt DE, Munck Af Rosenschöld P, Oh JH, Hunt M, et al. Impact of dose to the bladder trigone on long-term urinary function after high-dose intensity modulated radiation therapy for localized prostate cancer. Int J Radiat Oncol Biol Phys 2014;88:339–44. doi:10.1016/j.ijrobp.2013.10.042.
[59] Hatton JA, Greer PB, Tang C, Wright P, Capp A, Gupta S, et al. Does the planning dose-volume histogram represent treatment doses in image-guided prostate radiation therapy? Assessment with cone-beam computerised tomography scans. Radiother Oncol 2011;98:162–8. doi:10.1016/j.radonc.2011.01.006.
[60] 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.
[61] Thomas C, Keyes M, Liu M, Moravan V. Segmental Urethral Dosimetry and Urinary Toxicity in Patients With No Urinary Symptoms Before Permanent Prostate Brachytherapy. Int J Radiat Oncol 2008;72:447–55. doi:10.1016/J.IJROBP.2007.12.052.
[62] Earley JJ, Abdelbaky AM, Cunningham MJ, Chadwick E, Langley SEM, Laing RW. Correlation between prostate brachytherapy-related urethral stricture and peri-apical urethral dosimetry: A matched case-control study. Radiother Oncol 2012. doi:10.1016/j.radonc.2012.06.001.
[63] Merrick GS, Butler WM, Wallner KE, Galbreath RW, Anderson RL, Allen ZA, et al. Risk factors for the development of prostate brachytherapy related urethral strictures. J Urol 2006. doi:10.1016/S0022-5347(05)00681-6.
[64] 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.
[65] Lyman JT. Complication probability as assessed from dose-volume histograms. Radiat Res Suppl 1985.
[66] Kutcher GJ, Burman C, Brewster L, Goitein M, Mohan R. Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. Int J Radiat Oncol Biol Phys 1991. doi:10.1016/0360-3016(91)90173-2.
[67] Zelefsky MJ, Levin EJ, Hunt M, Yamada Y, Shippy AM, Jackson A, et al. Incidence of Late Rectal and Urinary Toxicities After Three-Dimensional Conformal Radiotherapy and Intensity-Modulated Radiotherapy for Localized Prostate Cancer. Int J Radiat Oncol Biol Phys 2008;70:1124–9. doi:10.1016/j.ijrobp.2007.11.044.
[68] Cahlon O, Zelefsky MJ, Shippy A, Chan H, Fuks Z, Yamada Y, et al. Ultra-High Dose (86.4 Gy) IMRT for Localized Prostate Cancer: Toxicity and Biochemical Outcomes. Int J Radiat Oncol 2008;71:330–7. doi:10.1016/j.ijrobp.2007.10.004.
[69] Karlsdóttir Á, Muren LP, Wentzel-Larsen T, Dahl O. Late Gastrointestinal Morbidity After Three-

Dimensional Conformal Radiation Therapy for Prostate Cancer Fades With Time in Contrast to Genitourinary Morbidity. Int J Radiat Oncol 2008;70:1478–86. doi:10.1016/j.ijrobp.2007.08.076.
[70] Kupelian PA, Willoughby TR, Reddy CA, Klein EA, Mahadevan A. Hypofractionated Intensity-Modulated Radiotherapy (70 Gy at 2.5 Gy Per Fraction) for Localized Prostate Cancer: Cleveland Clinic Experience. Int J Radiat Oncol 2007;68:1424–30. doi:10.1016/j.ijrobp.2007.01.067.
[71] Al-Mamgani A, Heemsbergen WD, Peeters STH, Lebesque J V. Role of Intensity-Modulated Radiotherapy in Reducing Toxicity in Dose Escalation for Localized Prostate Cancer. Int J Radiat Oncol Biol Phys 2009. doi:10.1016/j.ijrobp.2008.04.063.
[72] Thor M, Olsson C, Oh JH, Petersen SE, Alsadius D, Bentzen L, et al. Urinary bladder dose-response relationships for patient-reported genitourinary morbidity domains following prostate cancer radiotherapy. Radiother Oncol 2016;119:117–22. doi:10.1016/j.radonc.2016.01.013.
[73] Harsolia A, Vargas C, Yan D, Brabbins D, Lockman D, Liang J, et al. Predictors for Chronic Urinary Toxicity After the Treatment of Prostate Cancer With Adaptive Three-Dimensional Conformal Radiotherapy: Dose-Volume Analysis of a Phase II Dose-Escalation Study. Int J Radiat Oncol Biol Phys 2007;69:1100–9. doi:10.1016/j.ijrobp.2007.04.076.
[74] Bakhshandeh M, Hashemi B, Mahdavi SRM, Nikoofar A, Vasheghani M, Kazemnejad A. Normal tissue complication probability modeling of radiation-induced hypothyroidism after head-and-neck radiation therapy. Int J Radiat Oncol Biol Phys 2013. doi:10.1016/j.ijrobp.2012.03.034.
[75] Cheung MR, Tucker SL, Dong L, de Crevoisier R, Lee AK, Frank S, et al. Investigation of bladder dose and volume factors influencing late urinary toxicity after external beam radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2007;67:1059–65. doi:10.1016/j.ijrobp.2006.10.042.
[76] Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of Normal Tissue Complication Probability Models in the Clinic. Int J Radiat Oncol Biol Phys 2010. doi:10.1016/j.ijrobp.2009.07.1754.
[77] Söhn M, Alber M, Yan D. Principal Component Analysis-Based Pattern Analysis of Dose-Volume Histograms and Influence on Rectal Toxicity. Int J Radiat Oncol Biol Phys 2007. doi:10.1016/j.ijrobp.2007.04.066.
[78] Dean JA, Wong KH, Gay H, Welsh LC, Jones AB, Schick U, et al. Functional Data Analysis Applied to Modeling of Severe Acute Mucositis and Dysphagia Resulting From Head and Neck Radiation Therapy. Int J Radiat Oncol Biol Phys 2016. doi:10.1016/j.ijrobp.2016.08.013.
[79] Gulliford SL, Webb S, Rowbottom CG, Corne DW, Dearnaley DP. Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate. Radiother Oncol 2004. doi:10.1016/j.radonc.2003.03.001.
[80] Tomatis S, Rancati T, Fiorino C, Vavassori V, Fellin G, Cagna E, et al. Late rectal bleeding after 3D-CRT for prostate cancer: Development of a neural-network-based predictive model. Phys Med Biol 2012. doi:10.1088/0031-9155/57/5/1399.
[81] Pella A, Cambria R, Riboldi M, Jereczek-Fossa BA, Fodor C, Zerini D, et al. Use of machine learning methods for prediction of acute toxicity in organs at risk following prostate radiotherapy. Med Phys 2011. doi:10.1118/1.3582947.
[82] Ospina JD, Zhu J, Chira C, Bossi A, Delobel JB, Beckendorf V, et al. Random forests to predict rectal toxicity following prostate cancer radiation therapy. Int. J. Radiat. Oncol. Biol. Phys., 2014. doi:10.1016/j.ijrobp.2014.04.027.
[83] Yahya N, Ebert M a., Bulsara M, House MJ, Kennedy A, Joseph DJ, et al. Statistical-learning strategies generate only modestly performing predictive models for urinary symptoms following external beam radiotherapy of the prostate: A comparison of conventional and machine-learning methods. Med Phys 2016;43:2040–52. doi:10.1118/1.4944738.
[84] Acosta O, de Crevoisier R. Beyond DVH: 2D/3D-Based Dose Comparison to Assess Predictors of Toxicity. Model. Radiother. Side Eff., CRC Press; 2019, p. 415–40. doi:10.1201/b21956-17.

[85] Shelley LEA, Scaife JE, Romanchikova M, Harrison K, Forman JR, Bates AM, et al. Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy. Radiother Oncol 2017;123:466. doi:10.1016/J.RADONC.2017.04.008.
[86] 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.
[87] Stenmark MH, Conlon ASC, Johnson S, Daignault S, Litzenberg D, Marsh R, et al. Dose to the inferior rectum is strongly associated with patient reported bowel quality of life after radiation therapy for prostate cancer. Radiother Oncol 2014. doi:10.1016/j.radonc.2014.01.007.
[88] Ebert MA, Foo K, Haworth A, Gulliford SL, Kennedy A, Joseph DJ, et al. Gastrointestinal dose-histogram effects in the context of dose-volume-constrained prostate radiation therapy: Analysis of data from the radar prostate radiation therapy trial. Int J Radiat Oncol Biol Phys 2015. doi:10.1016/j.ijrobp.2014.11.015.
[89] Peeters STH, Lebesque J V., Heemsbergen WD, Van Putten WLJ, Slot A, Dielwart MFH, et al. Localized volume effects for late rectal and anal toxicity after radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2006. doi:10.1016/j.ijrobp.2005.10.002.
[90] Heemsbergen WD, Hoogeman MS, Hart GAM, Lebesque J V., Koper PCM. Gastrointestinal toxicity and its relation to dose distributions in the anorectal region of prostate cancer patients treated with radiotherapy. Int J Radiat Oncol Biol Phys 2005. doi:10.1016/j.ijrobp.2004.07.724.
[91] Hoogeman MS, Van Herk M, De Bois J, Muller-Timmermans P, Koper PCM, Lebesque J V. Quantification of local rectal wall displacements by virtual rectum unfolding. Radiother Oncol 2004. doi:10.1016/j.radonc.2003.11.015.
[92] Witte MG, Heemsbergen WD, Bohoslavsky R, Pos FJ, Al-Mamgani A, Lebesque J V., et al. Relating Dose Outside the Prostate With Freedom From Failure in the Dutch Trial 68 Gy vs. 78 Gy. Int J Radiat Oncol Biol Phys 2010. doi:10.1016/j.ijrobp.2009.04.040.
[93] 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.
[94] 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 IMRT. Radiother Oncol 2016;119:388–97. doi:10.1016/j.radonc.2016.04.023.
[95] 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.
[96] 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.
[97] 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.
[98] Munbodh R, Jackson A, Bauer J, Schmidtlein CR, Zelefsky MJ. Dosimetric and anatomic indicators of late rectal toxicity after high-dose intensity modulated radiation therapy for prostate cancer. Med Phys 2008;35:2137–50. doi:10.1118/1.2907707.
[99] Tucker SL, Zhang M, Dong L, Mohan R, Kuban D, Thames HD. Cluster model analysis of late rectal bleeding after IMRT of prostate cancer: A case-control study. Int J Radiat Oncol Biol Phys 2006;64:1255–64. doi:10.1016/j.ijrobp.2005.10.029.
[100] 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.
[101] Buettner F, Gulliford SL, Webb S, Partridge M. Modeling late rectal toxicities based on a parameterized representation of the 3D dose distribution. Phys Med Biol 2011;56:2103–18. doi:10.1088/0031-9155/56/7/013.
[102] Buettner F, Gulliford SL, Webb S, Sydes MR, Dearnaley DP, Partridge M. The dose-response of the anal sphincter region - An analysis of data from the MRC RT01 trial. Radiother Oncol 2012. doi:10.1016/j.radonc.2012.03.002.
[103] Moulton CR, House MJ, Lye V, Tang CI, Krawiec M, Joseph DJ, et al. Spatial features of dose-surface maps from deformably-registered plans correlate with late gastrointestinal complications. Phys Med Biol 2017. doi:10.1088/1361-6560/aa663d.
[104] Wortel RC, Witte MG, van der Heide UA, Pos FJ, Lebesque J V., van Herk M, et al. Dose–surface maps identifying local dose–effects for acute gastrointestinal toxicity after radiotherapy for prostate cancer. Radiother Oncol 2015;117:515–20. doi:10.1016/J.RADONC.2015.10.020.
[105] 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;32:1681–9. doi:10.1016/J.EJMP.2016.08.013.
[106] 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.
[107] Zhen X, Chen J, Zhong Z, Hrycushko B, Zhou L, Jiang S, et al. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study. Phys Med Biol 2017. doi:10.1088/1361-6560/aa8d09.
[108] Monti S, Pacelli R, Cella L, Palma G. Inter-patient image registration algorithms to disentangle regional dose bioeffects. Sci Rep 2018;8:4915. doi:10.1038/s41598-018-23327-0.
[109] Dréan G, Acosta O, Lafond C, Simon A, de Crevoisier R, Haigron P. Interindividual registration and dose mapping for voxelwise population analysis of rectal toxicity in prostate cancer radiotherapy. Med Phys 2016;43:2721–30. doi:10.1118/1.4948501.
[110] Ashburner J, Friston KJ. Voxel-Based Morphometry—The Methods. Neuroimage 2000;11:805–21. doi:10.1006/nimg.2000.0582.
[111] Palma G, Monti S, Avino VD, Conson M, Liuzzi R, Pressello MC, et al. A Voxel-Based Approach to Explore Local Dose Differences Associated With Radiation-Induced Lung Damage. Radiat Oncol Biol 2016;96:127–33. doi:10.1016/j.ijrobp.2016.04.033.
[112] Acosta O, De Crevoisier R. Beyond DVH: 2D/3D based dose comparison to assess predictors of toxicity n.d.
[113] Castadot P, Lee JA, Parraga A, Geets X, Macq B, Grégoire V. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. Radiother Oncol 2008. doi:10.1016/j.radonc.2008.04.010.
[114] Salguero FJ, Saleh-Sayah NK, Yan C, Siebers J V. Estimation of three-dimensional intrinsic dosimetric uncertainties resulting from using deformable image registration for dose mapping. Med Phys 2011. doi:10.1118/1.3528201.
[115] Rohlfing T. Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. IEEE Trans Med Imaging 2012. doi:10.1109/TMI.2011.2163944.
[116] Fiorino C, Palorini F. SP-0215: Advanced methods for 2D/3D dose map correlation in modelling toxicity. Radiother Oncol 2015;115:S108–9. doi:10.1016/S0167-8140(15)40213-0.
[117] Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage 2009;44:62–70. doi:10.1016/J.NEUROIMAGE.2008.05.021.
[118] SMITH S, NICHOLS T. Threshold-free cluster enhancement: Addressing problems of smoothing,

threshold dependence and localisation in cluster inference. Neuroimage 2009;44:83–98. doi:10.1016/j.neuroimage.2008.03.061.
[119] 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.
[120] McWilliam A, Faivre-Finn C, Kennedy J, Kershaw L, van Herk MB. Data Mining Identifies the Base of the Heart as a Dose-Sensitive Region Affecting Survival in Lung Cancer Patients. Int J Radiat Oncol 2016. doi:10.1016/j.ijrobp.2016.06.128.
[121] Holmes AP, Blair RC, Watson JDG, Ford I. Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. J Cereb Blood Flow Metab 1996;16:7–22. doi:10.1097/00004647-199601000-00002.
[122] Pederson AW, Fricano J, Correa D, Pelizzari CA, Liauw SL. Late toxicity after intensity-modulated radiation therapy for localized prostate cancer: An exploration of dose-volume histogram parameters to limit genitourinary and gastrointestinal toxicity. Int J Radiat Oncol Biol Phys 2012;82:235–41. doi:10.1016/j.ijrobp.2010.09.058.
[123] 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.
[124] Boersma LJ, Van Den Brink M, Bruce AM, Shouman T, Gras L, Te Velde A, 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;41:83–92. doi:10.1016/S0360-3016(98)00037-6.
[125] Storey MR, Pollack A, Zagars G, Smith L, Antolak J, Rosen I. Complications from radiotherapy dose escalation in prostate cancer: preliminary results of a randomized trial. Int J Radiat Oncol 2000;48:635–42. doi:http://dx.doi.org/10.1016/S0360-3016(00)00700-8.
[126] Rosewall T, Catton C, Currie G, Bayley A, Chung P, Wheat J, et al. The relationship between external beam radiotherapy dose and chronic urinary dysfunction - A methodological critique. Radiother Oncol 2010;97:40–7. doi:10.1016/j.radonc.2010.08.002.
[127] 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.
[128] Pearson D, Gill SK, Campbell N, Reddy K. Dosimetric and volumetric changes in the rectum and bladder in patients receiving CBCT-guided prostate IMRT: analysis based on daily CBCT dose calculation. J Appl Clin Med Phys 2016;17:107–17. doi:10.1120/jacmp.v17i6.6207.
[129] 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.
[130] 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.
[131] 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: Ad interim results of a cohort study. Eur Urol Suppl 2014;13:114–5.
[132] 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.

[133] 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.
[134] 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.
[135] Barnett GC, Coles CE, Elliott RM, Baynes C, Luccarini C, Conroy D, et al. Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study. Lancet Oncol 2012;13:65–77. doi:10.1016/S1470-2045(11)70302-3.
[136] Smith JJ, Wasserman I, Milgrom SA, Chow OS, Chen C-T, Patil S, et al. Single Nucleotide Polymorphism TGFβ1 R25P Correlates with Acute Toxicity during Neoadjuvant Chemoradiotherapy in Rectal Cancer Patients. Int J Radiat Oncol Biol Phys 2017;97:924–30. doi:10.1016/j.ijrobp.2016.12.015.
[137] Someya M, Hori M, Gocho T, Nakata K, Tsuchiya T, Kitagawa M, et al. Prediction of acute gastrointestinal and genitourinary radiation toxicity in prostate cancer patients using lymphocyte microRNA. Jpn J Clin Oncol 2018;48:167–74. doi:10.1093/jjco/hyx181.
[138] Bucci J, Spadinger I, Hilts M, Sidhu S, Smith C, Keyes M, et al. Urethral and periurethral dosimetry in prostate brachytherapy: Is there a convenient surrogate? Int J Radiat Oncol Biol Phys 2002. doi:10.1016/S0360-3016(02)03054-7.
[139] Waterman FM, Dicker AP. Determination of the urethral dose in prostate brachytherapy when the urethra cannot be visualized in the postimplant CT scan. Med Phys 2000. doi:10.1118/1.598912.
[140] Commowick O, Grégoire V, Malandain G. Atlas-based delineation of lymph node levels in head and neck computed tomography images. Radiother Oncol 2008. doi:10.1016/j.radonc.2008.01.018.
[141] Martin S, Daanen V, Troccaz J. Atlas-based prostate segmentation using an hybrid registration. Int J Comput Assist Radiol Surg 2008. doi:10.1007/s11548-008-0247-0.
[142] Aljabar P, Heckemann RA, Hammers A, Hajnal J V., Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage 2009. doi:10.1016/j.neuroimage.2009.02.018.
[143] Langerak TR, Van Der Heide UA, Kotte ANTJ, Viergever MA, Van Vulpen M, Pluim JPW. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans Med Imaging 2010. doi:10.1109/TMI.2010.2057442.
[144] Martin Ś, Troccaz J, Daanen V. Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model. Med Phys 2010. doi:10.1118/1.3315367.
[145] Ramus L, Thariat J, Marcy P-Y, Pointreau Y, Bera G, Commowick O, et al. Automatic segmentation using atlases in head and neck cancers: Methodology. Cancer/Radiotherapie 2010. doi:10.1016/j.canrad.2010.01.005.
[146] Van Rikxoort EM, Prokop M, De Hoop B, Viergever MA, Pluim JPW, Van Ginneken B. Automatic segmentation of pulmonary lobes robust against incomplete fissures. IEEE Trans Med Imaging 2010. doi:10.1109/TMI.2010.2044799.
[147] Rohlfing T, Brandt R, Menzel R, Maurer CR. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 2004. doi:10.1016/j.neuroimage.2003.11.010.
[148] Acosta O, Dowling J, Drean G, Simon A, De Crevoisier R, Haigron P. Multi-atlas-based segmentation of pelvic structures from ct scans for planning in prostate cancer radiotherapy.

Abdomen Thorac. Imaging An Eng. Clin. Perspect., 2014. doi:10.1007/978-1-4614-8498-1_24.
[149] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004. doi:10.1109/TMI.2004.828354.
[150] Mccormick M, Liu X, Jomier J, Marion C, Ibanez L. Itk: Enabling reproducible research and open science. Front Neuroinform 2014. doi:10.3389/fninf.2014.00013.
[151] Acosta O, Simon A, Monge F, Commandeur F, Bassirou C, Cazoulat G, et al. Evaluation of multi-atlas-based segmentation of CT scans in prostate cancer radiotherapy. Proc. - Int. Symp. Biomed. Imaging, 2011. doi:10.1109/ISBI.2011.5872795.
[152] Išgum I, Staring M, Rutten A, Prokop M, Viergever MA, Van Ginneken B. Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. IEEE Trans Med Imaging 2009. doi:10.1109/TMI.2008.2011480.
[153] 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.
[154] Kataria PT, Medicity M-T, Medicity M-T. Simple diagrammatic method to delineate male urethra in prostate cancer radiotherapy: An MRI based approach. 2016.
[155] Lee HK, D'Souza WD, Yamal JMJ, Pollack A, Lee AK, Palmer MB, et al. Dosimetric consequences of using a surrogate urethra to estimate urethral dose after brachytherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2003. doi:10.1016/S0360-3016(03)00583-2.
[156] 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.
[157] 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.
[158] 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.
[159] 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.
[160] Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B 1996;58:267–88. doi:10.2307/2346178.
[161] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385–95.
[162] Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1–22.
[163] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics n.d.
[164] 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.
[165] Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Trans Med Imaging 2010;29:196–205. doi:10.1109/TMI.2009.2035616.
[166] Dubuisson M-P, Jain AK. A modified Hausdorff distance for object matching. Proc. 12th Int. Conf. Pattern Recognit., vol. 1, IEEE Comput. Soc. Press; n.d., p. 566–8. doi:10.1109/ICPR.1994.576361.
[167] Yahya N, Ebert MA, Bulsara M, House MJ, Kennedy A, Joseph DJ, et al. Urinary symptoms following external beam radiotherapy of the prostate: Dose-symptom correlates with multiple-

event and event-count models. Radiother Oncol 2015;117:277–82. doi:10.1016/j.radonc.2015.10.003.
[168] Nitti VW. The prevalence of urinary incontinence. Rev Urol 2001;3 Suppl 1:S2-6.
[169] Delobel J-B, Gnep K, Ospina JD, Beckendorf V, Chira C, Zhu J, et al. Nomogram to predict rectal toxicity following prostate cancer radiotherapy. PLoS One 2017;12:e0179845. doi:10.1371/journal.pone.0179845.
[170] Denham JW, Steigler A, Joseph D, Lamb DS, Spry NA, Duchesne G, et al. Radiation dose escalation or longer androgen suppression for locally advanced prostate cancer? Data from the TROG 03.04 RADAR trial. Radiother Oncol 2015. doi:10.1016/j.radonc.2015.05.016.
[171] Yahya N, Ebert MA, Bulsara M, Kennedy A, Joseph DJ, Denham JW. Independent external validation of predictive models for urinary dysfunction following external beam radiotherapy of the prostate: Issues in model development and reporting. Radiother Oncol 2016;120:339–45. doi:10.1016/j.radonc.2016.05.010.
[172] Denham JW, Wilcox C, Joseph D, Spry NA, Lamb DS, Tai KH, et al. Quality of life in men with locally advanced prostate cancer treated with leuprorelin and radiotherapy with or without zoledronic acid (TROG 03.04 RADAR): Secondary endpoints from a randomised phase 3 factorial trial. Lancet Oncol 2012. doi:10.1016/S1470-2045(12)70423-0.
[173] Haworth A, Kearvell R, Greer PB, Hooton B, Denham JW, Lamb D, et al. Assuring high quality treatment delivery in clinical trials - Results from the Trans-Tasman Radiation Oncology Group (TROG) study 03.04 “RADAR” set-up accuracy study. Radiother Oncol 2009. doi:10.1016/j.radonc.2008.10.011.
[174] Ebert MA, Haworth A, Kearvell R, Hooton B, Coleman R, Spry N, et al. Detailed review and analysis of complex radiotherapy clinical trial planning data: Evaluation and initial experience with the SWAN software system. Radiother Oncol 2008. doi:10.1016/j.radonc.2007.11.013.
[175] Denham JW, Wilcox C, Lamb DS, Spry NA, Duchesne G, Atkinson C, et al. Rectal and urinary dysfunction in the TROG 03.04 RADAR trial for locally advanced prostate cancer. Radiother Oncol 2012. doi:10.1016/j.radonc.2012.09.018.
[176] Ebert MA, Bulsara M, Haworth A, Kearvell R, Richardson S, Kennedy A, et al. Technical quality assurance during the TROG 03.04 RADAR prostate radiotherapy trial: Are the results reflected in observed toxicity rates? J Med Imaging Radiat Oncol 2015. doi:10.1111/1754-9485.12212.
[177] Kearvell R, Haworth A, Ebert MA, Murray J, Hooton B, Richardson S, et al. Quality improvements in prostate radiotherapy: Outcomes and impact of comprehensive quality assurance during the TROG 03.04 “RADAR” trial. J Med Imaging Radiat Oncol 2013. doi:10.1111/1754-9485.12025.
[178] Denham JW, Joseph D, Lamb DS, Spry NA, Duchesne G, Matthews J, et al. Short-term androgen suppression and radiotherapy versus intermediate-term androgen suppression and radiotherapy, with or without zoledronic acid, in men with locally advanced prostate cancer (TROG 03.04 RADAR): 10-year results from a randomised, phase 3, . Lancet Oncol 2019. doi:10.1016/S1470-2045(18)30757-5.
[179] Denham JW, Nowitz M, Joseph D, Duchesne G, Spry NA, Lamb DS, et al. Impact of androgen suppression and zoledronic acid on bone mineral density and fractures in the Trans-Tasman Radiation Oncology Group (TROG) 03.04 Randomised Androgen Deprivation and Radiotherapy (RADAR) randomized controlled trial for locally advanced pr. BJU Int 2014. doi:10.1111/bju.12497.
[180] Ebert MA, Haworth A, Kearvell R, Hooton B, Hug B, Spry NA, et al. Comparison of DVH data from multiple radiotherapy treatment planning systems. Phys Med Biol 2010. doi:10.1088/0031-9155/55/11/N04.
[181] Krawczyk B. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 2016;5:221–32. doi:10.1007/s13748-016-0094-0.
[182] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (With

discussion and a rejoinder by the authors). Ann Stat 2000;28:337–407. doi:10.1214/aos/1016218223.
[183] Efron B. Bayes' Theorem in the 21st Century. Science (80-) 2013;340:1177–8. doi:10.1126/science.1236536.
[184] PANG S, GONG J. C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks. Syst Eng - Theory Pract 2009;29:94–104. doi:10.1016/S1874-8651(10)60092-0.
[185] Fordellone M, Bellincontro A, Mencarelli F. Partial least squares discriminant analysis: A dimensionality reduction method to classify hyperspectral data 2018.
[186] Friedman JH. Regularized Discriminant Analysis. J Am Stat Assoc 1989;84:165. doi:10.2307/2289860.
[187] Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002.
[188] Han H, Wang W-Y, Mao B-H. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning, Springer, Berlin, Heidelberg; 2005, p. 878–87. doi:10.1007/11538059_91.
[189] Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explor Newsl 2004;6:20. doi:10.1145/1007730.1007735.
[190] Wilson DL. Asymptotic Properties of Nearest Neighbor Rules Using Edited Data. IEEE Trans Syst Man Cybern 1972;SMC-2:408–21. doi:10.1109/TSMC.1972.4309137.
[191] Haibo He, Yang Bai, Garcia EA, Shutao Li. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE Int. Jt. Conf. Neural Networks (IEEE World Congr. Comput. Intell., IEEE; 2008, p. 1322–8. doi:10.1109/IJCNN.2008.4633969.
[192] Dumancas G, Bello G. Comparison of Machine-Learning Techniques for Handling Multicollinearity in Big Data Analytics and High Performance Data Mining, 2015. doi:10.13140/RG.2.1.1579.4641.
[193] López V, Fernández A, García S, Palade V, Herrera F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Inf Sci (Ny) 2013;250:113–41. doi:10.1016/J.INS.2013.07.007.
[194] More A. Survey of resampling techniques for improving classification performance in unbalanced datasets 2016.
[195] Pozzolo AD, Caelen O, Bontempi G. When is undersampling effective in unbalanced classification tasks? Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2015. doi:10.1007/978-3-319-23528-8_13.
[196] Ferri C, Hernández-Orallo J, Modroiu R. An experimental comparison of performance measures for classification. Pattern Recognit Lett 2009. doi:10.1016/j.patrec.2008.08.010.
[197] Ling CX, Huang J, Zhang H. AUC: A statistically consistent and more discriminating measure than accuracy. IJCAI Int. Jt. Conf. Artif. Intell., 2003.
[198] Hand DJ. Measuring classifier performance: A coherent alternative to the area under the ROC curve. Mach Learn 2009. doi:10.1007/s10994-009-5119-5.
[199] 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.
[200] Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: Relationships with clinical outcomes. J Natl Cancer Inst 2009. doi:10.1093/jnci/djp386.
[201] Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. Lancet Oncol 2006. doi:10.1016/S1470-2045(06)70910-X.
[202] Van Der Poel HG, Tillier C, De Blok WM, Acar C, Van Muilekom EHAM, Van Den Bergh RCN.

Interview-based versus questionnaire-based quality of life outcomes before and after prostatectomy. J Endourol 2013. doi:10.1089/end.2012.0735.
[203] Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: Agreement between patient and physician reporting in three randomized trials. J Clin Oncol 2015. doi:10.1200/JCO.2014.57.9334.
[204] Andersen ES, Muren LP, Sørensen TS, Noe K, Thor M, Petersen JB, et al. Bladder dose accumulation based on a biomechanical deformable image registration algorithm in volumetric modulated arc therapy for prostate cancer. Phys Med Biol 2012. doi:10.1088/0031-9155/57/21/7089.
[205] Thor M, Bentzen L, Hysing LB, Ekanger C, Helle SI, Karlsdóttir Á, et al. Prediction of rectum and bladder morbidity following radiotherapy of prostate cancer based on motion-inclusive dose distributions. Radiother Oncol 2013;107:147–52. doi:10.1016/j.radonc.2013.03.029.
[206] Lee S, Kerns S, Ostrer H, Rosenstein B, Deasy JO, Oh JH. Machine Learning on a Genome-wide Association Study to Predict Late Genitourinary Toxicity After Prostate Radiation Therapy. Int J Radiat Oncol Biol Phys 2018;101:128–35. doi:10.1016/j.ijrobp.2018.01.054.
[207] Kerns SL, Dorling L, Fachal L, Bentzen S, Pharoah PDP, Barnes DR, et al. Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. EBioMedicine 2016;10:150–63. doi:10.1016/j.ebiom.2016.07.022.
[208] Rios R, De Crevoisier R, Ospina JD, Commandeur F, Lafond C, Simon A, et al. Population model of bladder motion and deformation based on dominant eigenmodes and mixed-effects models in prostate cancer radiotherapy. Med Image Anal 2017;38:133–49. doi:10.1016/J.MEDIA.2017.03.001.
[209] Rios R, Ospina JD, Lafond C, Acosta O, Espinosa J, de Crevoisier R. Characterization of Bladder Motion and Deformation in Prostate Cancer Radiotherapy. IRBM 2016. doi:10.1016/j.irbm.2016.09.001.

Titre: Amélioration de la prédiction de la toxicité urinaire après radiothérapie du cancer de la prostate à partir de modèles spatiaux multi-échelle de la dose: depuis les organes à risque aux sous-régions

Mots clés : radiothérapie; toxicité urinaire; modèles prédictifs; planification;

Résumé : La radiothérapie externe est un traitement locorégional du cancer. L'objectif de la radiothérapie impose un compromis entre la délivrance d'une dose maximale dans la tumeur afin d'augmenter le contrôle local et la curabilité, et d'une dose minimale aux organes sains afin de limiter la toxicité. Les symptômes urinaires peuvent être liés à l'irradiation de régions spécifiques de la vessie ou de l'urètre. Dans ce cas, la dose reçue par l'ensemble de la vessie peut ne pas suffire à expliquer la toxicité urinaire.
Dans le contexte du traitement du cancer de la prostate par radiothérapie, ce travail de thèse vise à analyser les corrélations spatiales entre la dose et les effets secondaires, cette problématique étant abordée dans un cadre d'analyse de population.
Pour évaluer la contribution de l'urètre à la toxicité urinaire, nous proposons une méthode de segmentation basée sur plusieurs atlas pour identifier avec précision cette structure sur les images CT. Nous utilisons ensuite deux méthodes pour analyser la distribution de dose spatiale. L'une basée sur la construction de cartes 2D dose-surface (DSM) couplée à des comparaisons pixel par pixel et l'autre basée sur des cartes 3D dose-volume (DVM) combinées à des comparaisons par voxel. Les sous-régions identifiées ont été validées dans des populations externes, ouvrant la perspective d'une planification de traitement spécifique du patient. Nous étudions également le potentiel d'une amélioration complémentaire de la prédiction en exploitant de méthodes d'apprentissage automatique.

Title: From global to local spatial models for improving prediction of urinary toxicity following prostate cancer radiotherapy

Keywords: radiotherapy; urinary toxicity; predictive models; treatment planning; dose calculation;

Abstract: External beam radiotherapy (EBRT) is a clinical standard for treating prostate cancer. The objective of EBRT is to deliver a high radiation dose to the tumor to maximize the probability of local control while sparing the neighboring organs (mainly the rectum and the bladder) in order to minimize the risk of complications. Developing reliable predictive models of genitourinary (GU) toxicity is of paramount importance to prevent radiation-induced side-effects, and improve treatment reliability. Urinary symptoms may be linked to the irradiation of specific regions of the bladder or the urethra, in which case the dose received by the entire bladder may not be sufficient to explain GU toxicity.
Going beyond the global, whole-organ-based models towards more local, sub-organ approaches, this thesis aimed to improve our understanding of radiation-induced urinary side-effects and ameliorate the prediction of urinary toxicity following prostate cancer radiotherapy.
With the objective to assess the contribution of urethra damage to urinary toxicity, we proposed a multi-atlas-based segmentation method to accurately identify this structure on CT images. The second objective was to identify specific symptom-related subregions in the bladder and the urethra predictive of different urinary symptoms. For this purpose, we proposed two methodologies for analyzing the spatial dose distribution; one based on the construction of 2D dose-surface maps (DSM) coupled with pixel-wise comparisons and another based on 3D dose-volume maps (DVMs) combined with voxel-wise comparisons. Identified subregions were validated in external populations, opening the perspective for patient specific treatment planning. We also implemented and compared different machine learning strategies and data augmentation techniques, paving the way to further improve urinary toxicity prediction.