Spatial descriptions of radiotherapy dose: normal tissue complication models and statistical associations

Martin Andrew Ebert, Sarah Gulliford, Oscar Acosta, Renaud de Crevoisier, Todd Mcnutt, Wilma D Heemsbergen, Marnix G Witte, Giuseppe Palma, Tiziana Rancati, Claudio Fiorino

PHYSICS IN MEDICINE AND BIOLOGY

TOPICAL REVIEW

Spatial descriptions of radiotherapy dose: Normal tissue complication models and statistical associations

- Martin A Ebert[1,2,3]
- Sarah Gulliford[4,5]
- Oscar Acosta[6]
- Renaud de Crevoisier[6a]
- Todd McNutt[7]
- Wilma D Heemsbergen[8]
- Marnix Witte[9]
- Giuseppe Palma[10]
- Tiziana Rancati[11]
- Claudio Fiorino[12]

[1] School of Physics, Mathematics and Computing, University of Western Australia, Crawley, Western Australia

[2] Department of Radiation Oncology, Sir Charles Gairdner Hospital, Nedlands, Western Australia

[3] 5D Clinics, Claremont, Western Australia

[4] Department of Radiotherapy Physics, University College Hospitals London, United Kingdom

[5] Department of Medical Physics and Bioengineering, University College London, United Kingdom

[6] Univ Rennes, CLCC Eugène Marquis, INSERM, LTSI - UMR 1099, F-35000 Rennes, France

[7] Johns Hopkins University, Baltimore, Maryland, USA

[8] Department of Radiotherapy, Erasmus MC Cancer Institute, the Netherlands

[9] The Netherlands Cancer Institute, Amsterdam, The Netherlands

[10] Institute of Biostructures and Bioimaging, National Research Council, Napoli, Italy

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

[12] Medical Physics, San Raffaele Scientific Institute, Milano, Italy

Corresponding author:

Dr Martin A Ebert

Radiation Oncology

Sir Charles Gairdner Hospital

Hospital Ave

Nedlands, Western Australia 6009

Australia

Tel: +61 423 976 746

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

Abstract

For decades, dose-volume information for segmented anatomy has provided the essential data for correlating radiotherapy dosimetry with treatment-induced complications. Dose-volume information has formed the basis for modelling those associations via normal tissue complication (NTCP) models and for driving treatment planning. Limitations to this approach have been identified. Many studies have emerged demonstrating that the incorporation of information describing the spatial nature of the dose distribution, and potentially its correlation with anatomy, can provide more robust associations with toxicity and seed more general NTCP models. Such approaches are culminating in the application of computationally intensive processes such as machine learning and the application of neural networks. The opportunities these approaches have for individualising treatment, predicting toxicity and expanding the solution space for radiation therapy are substantial and have clearly widespread and disruptive potential. Impediments to reaching that potential include issues associated with data collection, model generalisation and validation.

This review examines the role of spatial models of complication and summarises relevant published studies. Sources of data for these studies, appropriate statistical methodology frameworks for processing spatial dose information and extracting relevant features are described. Spatial complication modelling is consolidated as a pathway to guiding future developments towards effective, complication-free radiotherapy treatment.

Keywords: radiotherapy, complications, modelling, dosimetry

Word count: ~15,500 (~22,000 with bibliography)

1 Introduction

In radiotherapy, the risk of treatment-induced toxicity is the limiting factor for dose escalation in pursuit of an increase in local control. The prediction of radio-induced side-effects guides the physician and the patient between treatment alternatives and enables treatment optimisation by integrating predictive models within computerised planning.

Radio-induced toxicity is classically linked to the dose-volume relationship, patient clinical parameters (such as medical history and adjuvant treatments) and intrinsic radiosensitivity. With steady increases in computational capabilities and increased efforts to gather and analyse relevant data (Deasy *et al.*, 2010), exploiting information from more available data with integrative approaches is now feasible.

The dose-volume toxicity relationship has been widely investigated. In 2010, the Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) review summarized the three-dimensional dose/volume/outcome data to update and refine the related normal tissue tolerance guidelines (Marks *et al.*, 2010), initially provided by Emami *et al.* (1991). Dose-volume histogram (DVH) based 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). Different approaches have been historically developed to model NTCP, with the Lyman–Kutcher–Burman (LKB) model being one of the first and most commonly employed (Lyman, 1985). Even if prediction of toxicity and treatment plan evaluation with the NTCP-based models is still common practice, these kinds of models present limitations reducing their prediction capability. DVHs reduce the 3D (or even 4D) dose distribution within an organ to a unidimensional and discrete representation of the dose-volume relationship, inhibiting the ability of models to account for the actual underlying complexity.

Spatial NTCP models have sought to geometrically represent the 3D dose distribution. This allows information on the pattern as well as the amount of dose to be characterised. Recent spatial NTCP models have sought to geometrically represent the 3D dose distribution in a single coordinate system via a spatial normalisation for a joint analysis of dose at the lowest sampling scale (pixel and voxel levels, referred to from here as “pixel-wise” in 2D and “voxel-wise” in 3D) (e.g., (Marcello *et al.*, 2020a; Mylona *et al.*, 2020b; Palma *et al.*, 2020a; Jiang *et al.*, 2019; Palma *et al.*, 2019b)). These low spatial-scale methodologies have allowed the unravelling of the local dose-effect relationship across a population at each single voxel in a common coordinate system in different organs. Models

can also be created by defining and analysing spatial features of the 3D dose distribution (e.g. Buettner *et al.* (2012b)) or abstractions such as the dose surface map (DSM) (e.g. Heemsbergen *et al.* (2020)). These spatial methods, and others described below, have been pursued to improve prediction and classification. Such models may also facilitate identification of the underlying aetiology of radio-induced injury and be used to improve patient-specific treatment planning. They are likely to reduce toxicity (Drean *et al.*, 2016b; Lafond *et al.*, 2020), and may one day inform or help validate *in silico* models of treatment toxicity (e.g. (Cicchetti *et al.*, 2020)).

The goal of this review is to describe these recent spatial dose-effect investigations and NTCP models and provide some guidance around their development.

2 Strategies for characterising dose distributions

2.1 Dose-volume approaches

The concept of the dose-volume relationship of a defined region of interest became commonplace when both 3D dose computation and 3D segmentation ("contouring") of regions became practical. The cumulative DVH synthesises the dose vs volume relationship as a function representing the percentage of volume that receives at least a certain dose.

2.1.1 The advantages of dose-volume approaches

The primary advantage of the dose-volume approach is linked to the wealth of knowledge obtained through prior studies of radiation and the resounding clinical success of such approaches. Today's radiation therapy is driven by dose-volume constraints based on the results of published studies and meta-analyses. So much so, that today's dose distributions contain little information outside the bounds of these dose volume parameters, as they are controlled for in clinical practice.

Dose-volume metrics are easily understood and are based on the natural compartmentalisation of the body into organs. Reporting of them can be reduced to a table of numeric entries representing the quality of the complex 3D treatment plan. To even further simplify their presentations, software applications have reduced them to colour codes to indicate alerts when a plan may violate one of the treatment goals. In busy clinics, this facilitates rapid evaluation. Dose-volume metrics are also convenient when defining the goals for optimization in inverse treatment planning.

Radiobiological models have been developed to bridge the gap between the physical dose-based objectives to drive treatment planning and the clinical dose goals reflecting the toxicity risks. Those commonly studied, such as NTCP, tumour control probability (TCP) and the complication-free

tumour control probability (P+) (Källman *et al.*, 1992), have typically been designed to operate on DVH information.

2.1.2 The disadvantages of dose-volume approaches

Fundamentally the DVH assumes that every sub-volume of the region is of equal importance to the function of that tissue and is equally sensitive to radiation dose. Realistically, the segmented regions in radiotherapy are typically bulk anatomy and do not reflect the microstructure of anatomy that may be impacted by radiation. Therefore, the DVH may be too course of a feature to adequately model the impact radiation may have on the anatomy.

The assumption that each element of tissue is equally important to the NTCP function and equally sensitive to radiation dose is simply not true for many anatomical regions typically segmented in treatment planning. For example, the parotid glands consist of acinar cells producing saliva and a ductal region that carries the saliva to the oral cavity. Similarly, a kidney is made up of several cell types and structures. In other cases, such as the oesophagus and rectum, organ structure consists of a mucosal layer surrounded by muscle tissue. These structures may have different risks when the dose is high to the entire circumference of the structure versus when it has the same volume of dose oriented longitudinally along the structure. Understanding the true causal relationships between radiation dose and normal tissue dysfunction is limited with dose volume metrics that are naive to the detailed components of the anatomy.

The spinal cord has a complex spatial arrangement of functional sub-units (FSUs - compartments that accomplish part of an organ's function), and subsequently a complex inter-relationship with overall organ function. Precise pre-clinical experiments performed by Bijl *et al.* (2003) identified large variations in dose-volume based predictors of paralysis in rats when the spatial patterns of irradiation were changed. Conventionally, simple maximum cord dose has been used to predict subsequent complications (Kirkpatrick *et al.*, 2010). However, the inhomogeneity in irradiation now afforded with stereotactic spinal radiosurgery exceeds the predictive capability of dose-volume analysis (Medin and Boike, 2011). Similarly, models of lung complication had focused on their parallel-like nature and mean lung dose (MLD) had conventionally been used as a principal predictor. However, evidence for more localised dose-response in humans emerged nearly two decades ago (Seppenwoolde *et al.*, 2004), following extensive animal experiments (as well reviewed by Voshart *et al.* (2021)).

Further, analysis has traditionally been limited to dose-volume metrics for single organs. Many human functions involve multiple components of anatomy. Swallowing, for example involves many

muscles in the head and neck region. These muscles may be able to compensate for one another and the impact of a dose pattern across the set of muscles and its impact on swallowing can be quite complicated. In many cases, a significant portion of the anatomy is not contoured at all, and the dose-volume metrics can only be computed for contoured regions. Contouring with high spatial detail in routine workflow remains burdensome. Similarly, in a shift of spatial focus for dose-toxicity association, the impact on lung toxicity from cardiac irradiation has been identified (Tucker *et al.*, 2014; van Luijk *et al.*, 2005).

Multiple spatial dose distributions (an essentially infinite number) will yield the same or similar DVH. Dependence on a dose-volume approach requires an assumption that all those distributions will lead to the same toxicity – the problem of *degeneracy*. Conversely, dose-volume derived NTCP models from studies involving specific irradiation techniques will have been derived with minimal variation in DVH between patients. Extrapolation of DVH and NTCP metrics beyond the specific context in which they were derived is known to be dangerous. Due to this limitation, as well as many other sources of variations between cohorts, DVH-based complication models derived for one treatment approach tend not to be applicable to alternative irradiation strategies in the same sites (Troeller *et al.*, 2015).

2.2 Approaches that preserve spatial information

To overcome the limitation of whole organ DVHs, recent approaches have investigated the existence of spatial signatures of dose distributions across dimensionalities and at diverse spatial scales. Here we describe the processing of treatment planning data (Figure 1) required to achieve extraction of features describing spatial distributions at the various spatial scales and development of subsequent toxicity models (Figure 2). Practical applications of these features and models are described in Section 5.

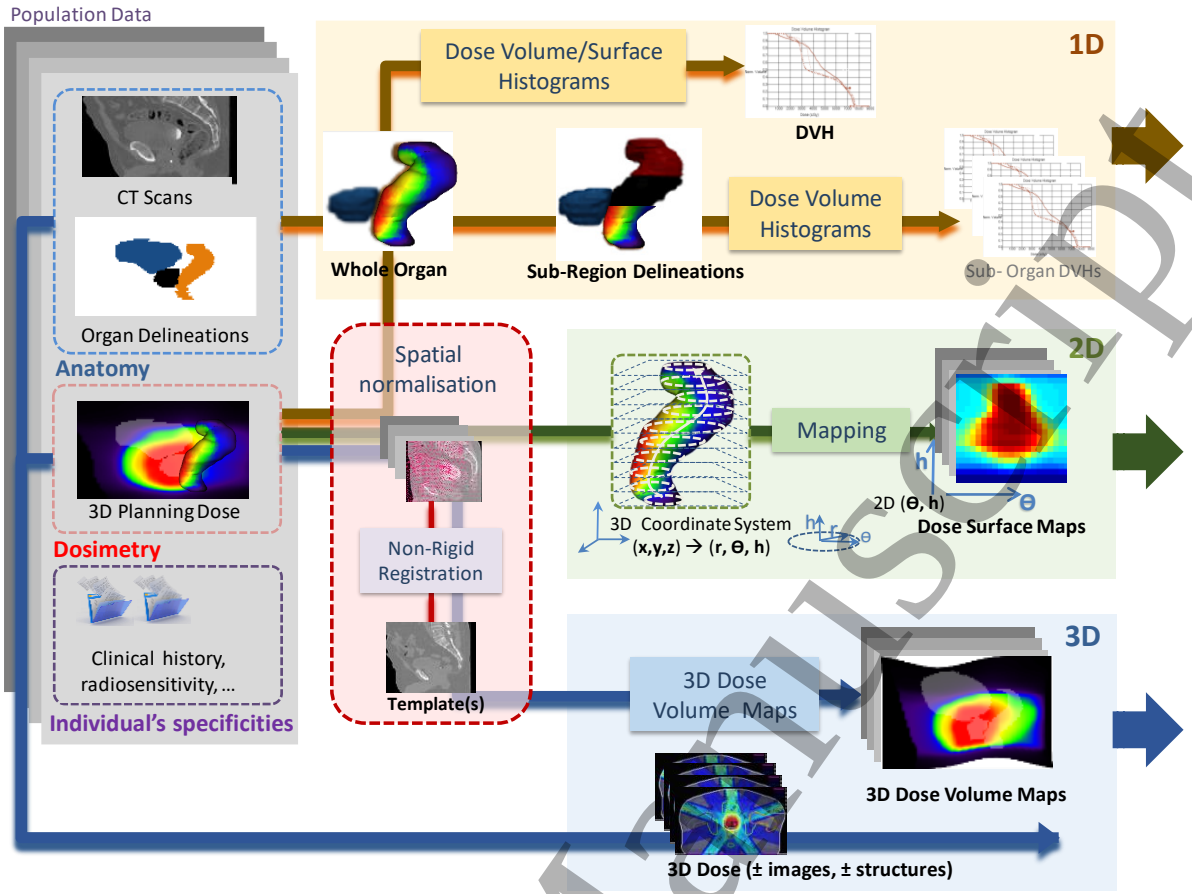


Figure 1. Processing workflows for preparing data for toxicity modelling across dimensionalities. Orange, path for histogram development (1D data); green, path for 2D dose surface maps (2D data); and blue, path for 3D dose volume maps (3D data). Some data sources and processes may not be used in all approaches, and these are indicated with dashed borders.

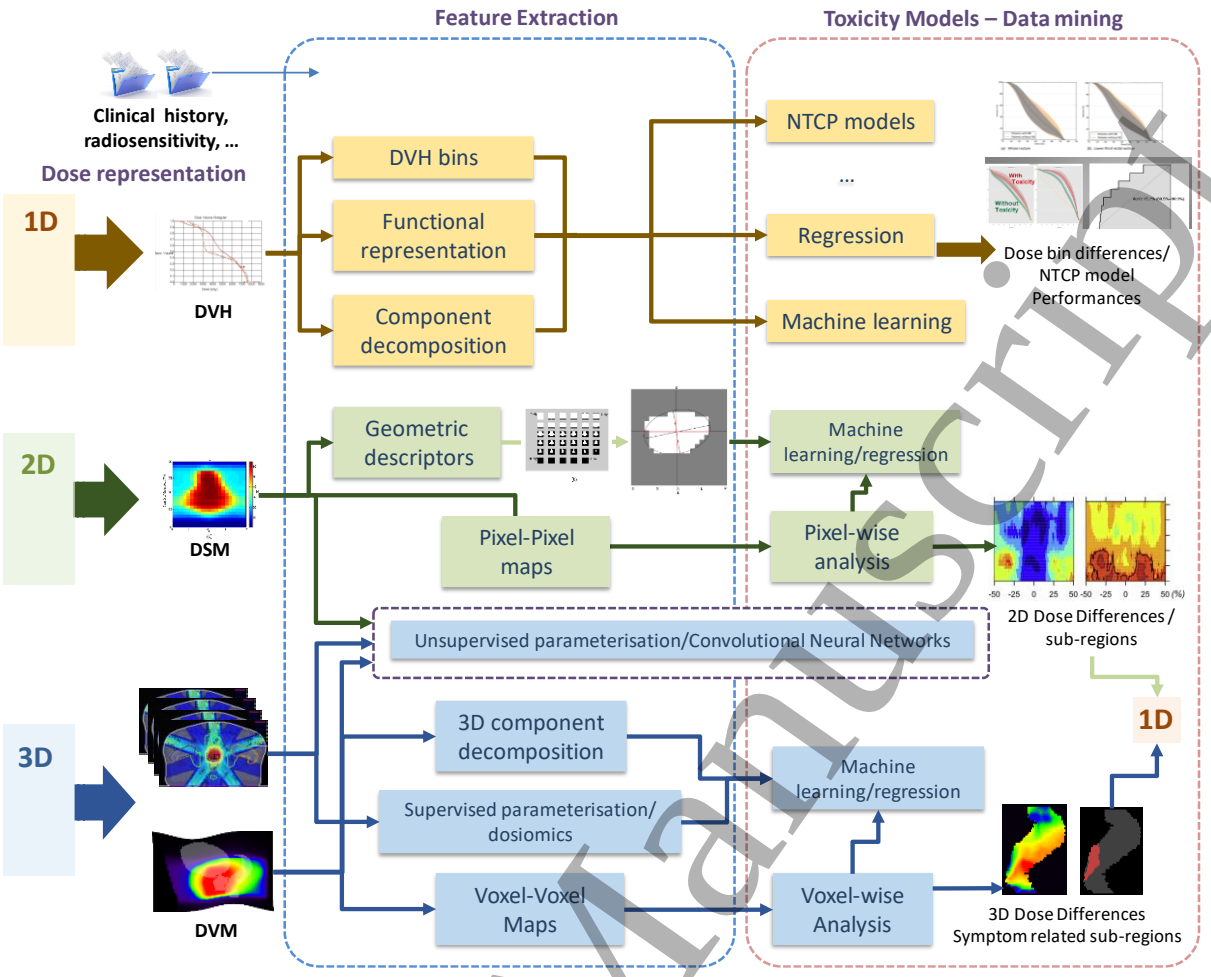


Figure 2. Data flow in the extraction of dosimetric features and construction of toxicity models. Features extracted from 1D, 2D or 3D data are exploited following different strategies, leading to different kinds of predictive models (NTCP, machine learning or general regression). 2D DSM and 3D DVM models may require the entire population dose to be mapped to a single coordinate system before being analysed.

2.2.1 1D precision dose-volume approaches

The simplest approach is to identify a more precise sub-region of the organ where dosimetry and DVH metrics are most correlated with outcome. Improvements in NTCP models, and evidence of correlations between local dose and side-effects, have been provided by undertaking DVH analysis (or analysis with related histogram information) at spatial scales below the organ level. Partitioning the organs for computing sub-region DVHs for example has demonstrated a sub-anatomical dependence for specific toxicities (Ebert *et al.*, 2015b; Heemsbergen *et al.*, 2005; Peeters *et al.*, 2006b; Stenmark *et al.*, 2014). The question that may arise is whether the organ partitions are anatomically-equivalent across individuals allowing DVH comparisons. If sub-region partitions between patients are generated following the same geometrical criteria, then they can refer to the same anatomo-physiological regions. The identification of correlative regions can be derived

manually (e.g. Gulliford *et al.* (2017)), or by identifying clusters of correlated pixels and voxels in 2D and 3D representations (e.g. Drean *et al.* (2016b)). DVH-based features of those sub-regions can be used to validate their association with complications.

2.2.2 2D surface mapping

Spatial considerations on the distribution of dose to an organ surface can be achieved with dose surface mapping (DSM). DSMs present a virtual unfolded planar representation of the dose distribution across an organ wall. Such mapping has been implemented following different strategies (Sanchez-Nieto *et al.*, 2001; Hoogeman *et al.*, 2004; Munbodh *et al.*, 2008; Tucker *et al.*, 2006b; Witztum *et al.*, 2016). A 2D image is constructed via parametric mapping from the 3D coordinate system of the organ wall. The general idea is depicted with a rectal DSM in Figure 3, where a direct relationship exists between the 3D cylindrical coordinates and the 2D (Θ,h) space. Thus, each pixel in 2D corresponds to a portion of the organ wall where the dose is mapped and propagated by interpolation. Once constructed, dose surface maps can be used to undertake "pixel-wise" analysis (Yahya *et al.*, 2017), or parameterised using geometric descriptors such as lateral and longitudinal extent (Buettner *et al.*, 2009b) or texture features (Chen *et al.*, 2018).

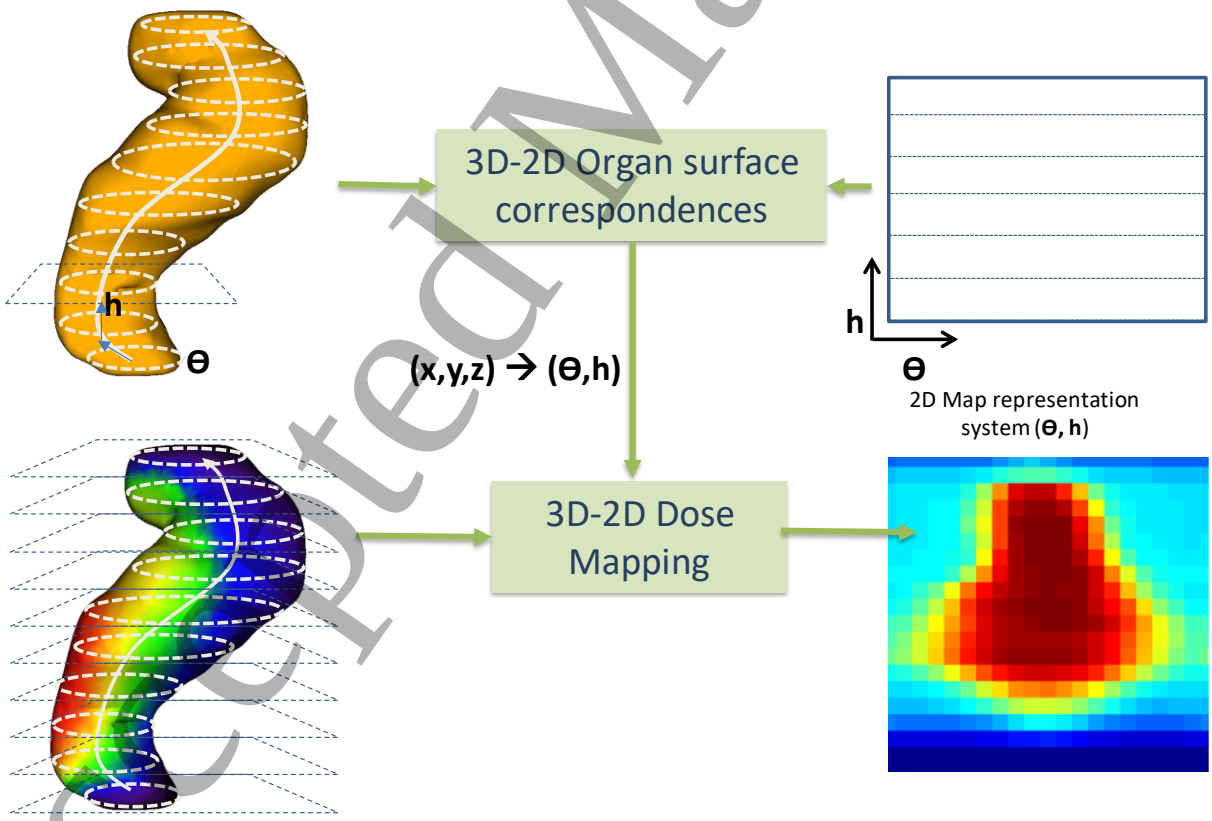


Figure 3: Construction of a dose surface map (DSM) for the rectum by establishment of a direct relationship between the 3D coordinate system and a planar O(Θ,h) space.

Crucial aspects in this construction are the definition of the origin (i.e. 0,0) and the resolution and size of the 2D images. If the rectum was the organ to be studied (e.g. (Buettner *et al.*, 2009a; Moulton *et al.*, 2017)) a cylindrical coordinate system for building the DSM has been used. In Buettner *et al.* (2009b) the contour was thus cut at the posterior-most position on each CT-slice and unwrapped to a map of 21x21 pixels. Witztum *et al.* (2016) raised some of the issues concerning tortuous structures. They developed a raytracing approach to create dose surface maps for the duodenum accounting for the bend in the structure, following an inner path.

In other hollow organs such as the bladder a similar slice-based methodology has been applied. In works from Palorini *et al.* (2016a) and Yahya *et al.* (2017), 1 mm-resolution DSMs were generated (cranial-caudal direction), by virtually cutting bladder contours at the points intersecting the sagittal plane passing through its centre-of-mass. Because of the large inter-individual bladder variability some issues arise when having large and small bladders to map together for population analysis or where some parts of the bladder are not equally mapped. In Mylona *et al.* (2020a) this was addressed with an anisotropic vertical interpolation to the smallest bladder, aligned at the bladder base.

2.2.3 3D feature extraction

It is feasible to reduce the complex 3D voxel-level dose information to a smaller number of features via an appropriate spatial parameterisation. One such approach is to describe the spatial distribution within an organ via 3D moments (Buettner *et al.*, 2012b; Dean *et al.*, 2016). Alternatively, borrowing from the world of imaging analytics, supervised descriptions can be obtained via spatial texture features ("dosiomics" (Liang *et al.*, 2019; Rossi *et al.*, 2018), "dosomics" (Placidi *et al.*, 2020) or "radiomorphology" (Jiang *et al.*, 2019)), or unsupervised learning can be employed via neural networks.

2.2.4 3D volume mapping

At a fine scale, dose-outcome correlations can be investigated at the voxel level. For voxel-wise comparisons to be meaningful, anatomical correspondence across the individuals must be ensured. This pre-processing step is referred to as "spatial normalisation", whose goal is to define geometrical transformations aimed at registering and resampling inter-individual anatomies and doses into a common coordinate system as depicted in Figure 1 (e.g. (Monti *et al.*, 2020; Acosta *et al.*, 2013; Rigaud *et al.*, 2019; Acosta and De Crevoisier, 2019)). This 3D-3D dose mapping to a common coordinate system to create a dose-volume map (DVM) remains challenging. Such mapping may be obtained via a parametric representation of the anatomy in a spherical or cylindrical coordinate

system as in Chen *et al.* (2013). It may be more precisely computed through existing non-rigid (deformable) registration methods (McWilliam *et al.*, 2017; Monti *et al.*, 2018; Marcello *et al.*, 2020a), or tailored to a particular anatomy as proposed for the rectum in Drean *et al.* (2016a), or for the bladder in Mylona *et al.* (2019) using spatial descriptors. Depending on the investigated anatomical site, organ-driven registration methods may be more precise than the ones based on intensity levels. This is the case in Acosta *et al.* (2013), Drean *et al.* (2016a) and Mylona *et al.* (2019) where anatomical mapping based on 3D structural models of the considered organs were proposed. These approaches require, nevertheless, a precedent segmentation of some of the considered structures such as the urethra (Acosta *et al.*, 2017). However, when inter-individual registration is to new patients without identified structures or is to be structure-agnostic, image information alone must be used (McWilliam *et al.*, 2017; Monti *et al.*, 2018; Abravan *et al.*, 2020).

The 3D spatial normalisation approach can also be used to align anatomy for derivation of DSMs, especially in the case of pixel-wise analysis, or for the purpose of sub-region identification.

3 Practical Considerations

The development of spatial response models places specific demands on the nature of technical data collected for their construction. When interpreting, utilising or publishing a spatial complication model, factors impacting the underlying technical data should be considered and appropriately reported and accommodated (see section 4.2.3). The relevance and quality of patient outcome data is of similar or even greater importance for the derivation of useful models. Additional data types can constitute modifying and stratifying co-variates, such as patient demographics and co-morbidities, disease staging, treatment characteristics (techniques, timing, adjuvant treatments), pathologic and genetic information.

3.1 Required technical data

Due to the computational nature of spatial models, it is assumed that required data will be available in digital form which could be arbitrary in-house, native or proprietary formats, or more generally in prescribed formats such as Digital Imaging and Communications in Medicine (DICOM) (NEMA), Neuroimaging Informatics Technology Initiative (NIfTI) (NIfTI, 2020) or Nearly Raw Raster Data (NRRD) (SourceForge, 2020). The three principal technical data ingredients for model development are briefly described below.

3.1.1 Imaging

Anatomical imaging typically provides the reference space for derived models, guides definition of segmented structures and facilitates intra- and inter-individual registration. The robustness of spatial models can depend significantly on the sensitivity and specificity of imaging, particularly through influence on the definition of structures (e.g. (Roach *et al.*, 2019)).

3.1.2 Structures

Many of the processes for characterising spatial dose distributions presented in Sections 2.2 and 5 operate on information related to anatomical and functional structures. The definition of such structures can be made manually by observers at the time of patient treatment planning or manually through retrospective review of collated data. Alternatively, autosegmentation routines utilising anatomical atlases (Kennedy *et al.*, 2019) or artificial intelligence approaches (Fu *et al.*, 2020) can be used. Structure segmentation can represent a significant source of uncertainty in the derivation and application of models, with multiple contributing factors:

- Geometric variability: The location and extent of structures will depend on multiple factors relating to image quality, image sensitivity and specificity, inter-observer variability (e.g. (Roach *et al.*, 2019)), organ deformation and motion (e.g. (Palorini *et al.*, 2016a)), errors and limitations in image registration, bias propagated from atlas definitions or neural network learning environments or selection of a patient template (see Section 2.2) (Acosta *et al.*, 2010).
- Structure definition: A common source of undesired variability, particularly when pooling data sources or during validation, is variable definition of anatomical structures (e.g. (Nitsche *et al.*, 2017)). Models need to operate on like-definitions. Variability and ambiguity can be reduced through the use of consensus definitions, reviews of definitions such as within the QUANTEC reports (Bentzen *et al.*, 2010), or published standards (Wright *et al.*, 2019).
- Structure naming: Structure naming can often be problematic for scripting model development, particularly when data comes from multiple institutions. This can be ameliorated through use of naming conventions (e.g. (Mayo *et al.*, 2018; Santanam *et al.*, 2012)) or ontologies (Phillips *et al.*, 2020). Note that spatial models may utilise or give rise to non-standard structures (such as predictive clusters identified in DSMs and DVMs).

3.1.3 Dose

As indicated in Figure 1 and Figure 2, access to multi-dimensional descriptions of dose distributions, or features derived from them, represents a common minimum level of required technical data. When deriving and applying spatial models, several aspects of these data should be considered:

- Accuracy: Although dose distributions are frequently available based on planned or intended treatments, correct models will be based on dose distributions which have been verified or accumulated as delivered (e.g. (Shelley *et al.*, 2017; Jaffray *et al.*, 2010)). Accuracy should ideally have been assessed independently, such as via participation in credentialing exercises (e.g., (Ebert *et al.*, 2011; Molineu *et al.*, 2013; Weber *et al.*, 2014)). Deformations of dose, due perhaps to the intra-individual accumulation process (Tilly *et al.*, 2013) or inter-individual co-registration (see Section 2.2.3) will impact on the accuracy of dose representation.
- Precision: Spatial resolution in the description of dose can impact the ability to precisely represent the underlying response effects. The resolution of dose calculation has previously been shown to impact even dose-volume based models (Ebert *et al.*, 2010; Kim *et al.*, 2018). Variation in resolution can have a moderate impact on dosimetric texture features (Placidi *et al.*, 2020). With an increasing need to develop models for precision stereotactic treatments, precise descriptions of steep dose gradients across spatially-limited structures are required (e.g. (Ryu *et al.*, 2007; Hrycushko *et al.*, 2019; Gale *et al.*, 2017; Kim *et al.*, 2014)).
- Completeness: Dose calculations are often limited in extent relative to potentially-involved anatomy, such as when based on cone beam CT data obtained with accelerator-mounted imaging systems. This can inhibit spatial models, particularly those relating to low-doses over extensive regions of anatomy.
- Temporal features: Dose fractionation, inter- and intra-fraction dose temporal patterns can impact complication incidence (Dörr, 2015). Changes in response due to variable dose-per-fraction, either between voxels or due to variable treatment phases, may need to be incorporated into the model. Such variations may also be accounted for using equieffective dose estimates (Bentzen *et al.*, 2012), noting that this leads to spatial discontinuities where parameters vary between tissues. The complexity of temporal dose effects increases significantly when intra-treatment variations due to organ motion or the pharmacokinetics of radionuclide deliveries are considered.

3.1.4 Treatment description

Treatment factors, such as patient set-up at imaging and treatment, patient preparation, the use of immobilisation and fixation devices, may be co-variates of importance to the specificity of a model. This information is often not captured in DICOM fields or through oncology information systems.

3.2 Outcome Data

Outcome information, providing the known output for a model (the "endpoint" or "event incidence"), comes in diverse forms. For complication outcome, we are typically concerned with organ-specific symptoms of radiation injury which may manifest over months or years. These can be graded at discrete (ordinal) levels using standardised clinician- or patient-reported instruments such as provided by the Common Terminology Criteria for Adverse Events (Trotti *et al.*, 2003) developed by the United States (US) National Cancer Institute (NCI), instruments developed in-house or by various international collaboratives. The trend is towards the use of patient-reported complications for outcome. This is because the severity of symptoms are often under-reported by clinicians (Xiao *et al.*, 2013), and follows recognition of the importance of focussing on symptoms with the most impact on patients' quality of life. Although definitions can vary, complications are typically graded according to indicative symptoms and required interventions (GX – Grade X):

- G0 – symptoms are absent
- G1 – the complication is mild and no interventions are required
- G2 – the complication is moderate and some local intervention might be required
- G3 – the complication is severe and intervention is required, though is not life-threatening
- G4 – the complication is life-threatening and major intervention is required
- G5 – the complication has caused death

Whilst some models can utilise continuous outcomes, for NTCP models it is common to convert measures to a binary endpoint classification. These may be either determined at fixed time-points following treatment, as incidence at any time during follow-up, or the time-to-event incidence if temporal features can be incorporated in the model. The definition, interpretation, collection and application of complication outcome measures are notorious sources of uncertainty in outcome modelling. Multiple factors should be kept in mind related to model accuracy and generalisability:

- Specificity of the included patient cohort.
- The relevance of an outcome to patient quality-of-life.
- Variations in scoring mechanisms and criteria.
- Variations in follow-up time or time between measurements.

- The identification and influence of comorbidities, concurrent treatments or pre-existing morbidities.
- The influence of social and/or technical factors on measures.
- The nature of the data source, as discussed below.

3.3 Data sources

When considering sources of data for spatial complication models, we can consider the ability of those sources to meet specific criteria for development of generalizable, robust and powerful models. A source should provide large volumes of high quality, well-curated data for patients with diverse characteristics and treated with diverse techniques (noting that data diversity can lead to unexpectedly biased results, as discussed in Section 6.1). The sub-optimal performance of many radiotherapy outcomes models can largely be blamed on the paucity and lack of diversity of available data (Luo *et al.*, 2020).

Table 1 lists specific criteria, provides some examples of sources and attempts to describe, via generalisations, how likely each source is to meet the criteria. In Table 1, *quality* infers the completeness, accuracy and consistency of technical and outcome data. *Diversity* relates to the variability in studied populations, radiotherapy technique and overall patient treatment, including trial vs non-trial contexts (Chen *et al.*, 2016; Krauss, 2018). Diversity also pertains to inter-individual variations in spatially-localised dose (note also the implications of diversity for model generalizability, as discussed in Section 6.2).

Some points to note in relation to Table 1:

- **Single-institution studies** enable ready access to appropriate high-quality data though with minimal variability and typically only small patient numbers. Collated data is rarely made available outside the institution.
- **Multi-centre clinical trials** often employ rigorous data collation. However, such trials will rarely be statistically powered specifically for the purpose of spatial response modelling and so the sample size may be insufficient. Software systems developed over the last couple of decades, both in-house and commercially, have facilitated quality assessment of technical data by multicentre trials groups (e.g. (Ebert *et al.*, 2010; Deasy *et al.*, 2003; La Macchia *et al.*, 2012; Meroni *et al.*, 2019; Roelofs *et al.*, 2014; Deasy and Adita, 2013; Purdy, 2008; Purdy *et al.*, 1998)). Although the quality of clinical trial data can be advantageous, variations from trial conditions in the clinic, including participant selection, can bias model predictions relative to non-trial practice (Ohri *et al.*, 2013).

- **Data pooling and publication.** International policies are trending towards data availability and interoperability (e.g. (Hayman *et al.*, 2019; Taichman *et al.*, 2017)). In Table 1 we distinguish "public" pooling and publication, such as provided by the Cancer Imaging Archive (www.cancerimagingarchive.net, (Clark *et al.*, 2013)), from "private" pooling, such as might be achieved via manufacturer-led knowledge base collaboratives and user-communities. Both public and private data pools have the potential for development of large cohorts with data variability, though data quality may be ambiguous if not well documented.
- **Federated data access** can enable accessing large patient cohorts spanning multiple repositories, including clinical systems at individual treatment centres. Ethical and socio-political issues can be minimised if model parameters can be estimated for data at each site, before being combined centrally (Deist *et al.*, 2017). Although no published evidence was found that spatial complication models have been derived through this approach, the potential for validation of developed models is significant.

Table 1: Potential sources of data for spatial models and their ability to meet desirable criteria for forming statistically-powerful, generalisable models that meet current standards for validation and translation.

Source	Technical data quality	Outcome data quality	Variability/ diversity	Sample size	FAIR	Facilitates validation
Single institution studies	High	High	Low	Low	Low	Low
Multicentre clinical trials	High	High	Medium	Medium	High	Medium
Public data pooling and publication	Variable	Variable	Medium	Medium	High	High
Private data pooling	Variable	Low	Medium	High	Medium	Variable
Federated data access	Low	Low	High	High	Variable	High

[a] Including dosimetric accuracy.

[b] A data source will have a high ability to satisfy this criterion if it meets the FAIR principles (Wilkinson *et al.*, 2016) defined by the FORCE 11 (Future of Research Communications and e-Scholarship) community, of data being findable, accessible, interoperable and re-usable.

[c] e.g., Treatment planning system manufacturer-facilitated knowledge base consortia.

4 Statistical and Modelling Considerations

A central aim of using spatial dose descriptors to model dose-complication is to reduce the impact of degeneracy relative to dose-volume approaches. It is important that the process utilised maintains the principles associated with robust, unambiguous statistical analysis and interpretation. Here we summarise such relevant considerations.

4.1 Feature Selection

4.1.1 Candidate dosimetric features and collinearity

An important step in developing phenomenological NTCP models (van der Schaaf *et al.*, 2015) is to start off with a list of potential prognostic factors based on the literature and underlying radiobiological assumptions (e.g. assumed α/β ratio). This can reduce the number of false positive findings and guide the feature reduction process (Palma *et al.*, 2020a; Heinze *et al.*, 2018). The inhomogeneous physical dose distribution can be aggregated into dose features (Figure 2) that represent the biological dose received and are predictive for the toxicity endpoint of interest. The result may be just a small number of features as derived from a spatial parameterisation. However, hundreds to thousands of dose features can be retrieved from a spatial voxel-wise 3D dose distribution, even though the sample size may be quite limited, and collinearity is likely. Candidate prognostic factors selected from a group of correlated variables are typically those that have the highest predictive power at univariate analysis compared to the correlated variables that are a priori excluded. A general rule of thumb is that correlation between candidate variables for a multivariable model should be below ≈ 0.7 (El Naqa *et al.*, 2009; Schaake *et al.*, 2016).

4.1.2 Feature reduction

The generally accepted rule of thumb is that regression models should be used with a minimum of 10 “events per variable” EPV (Peduzzi *et al.*, 1996). This rule has been criticized as being too strict - Vittinghoff and McCulloch (2007) instead recommend a minimum of 7 EPV. After pre-processing the dataset to a candidate list of features considered for modelling, a variable selection algorithm must be chosen (Heinze *et al.*, 2018; Steyerberg and Vergouwe, 2014). Valid approaches to reduce the number of features (and clinical co-variates) to the most predictive in a multivariate model are: 1) select variables for the final multivariable model based on their univariate model estimates, using a p value threshold; 2) backward and forward selection tools like Wald, Likelihood Ratio and conditional regression methods; and 3) the LASSO method (least absolute shrinkage and selection operator) which is a logistic regression analysis with a penalty for the magnitude of the regression coefficients to prevent overfitting (Tibshirani, 1996; Buettner *et al.*, 2011; Gabryś *et al.*, 2018). Consideration can be given to reduction of features through use of their principal components

(e.g.(Chen *et al.*, 2011)). Additionally, feature selection can be combined with the method to determine association with outcome through algorithms such as random forest, and through the stability of features in associative models derived from sampled sub-sets of the full data (i.e. "bootstrapping"). Adequate feature reduction is vital to ensuring the ability for a model to be generalised. An excellent overview of techniques is provided in Guyon and Elisseeff (2003).

4.1.3 Co-variates

The inclusion of clinical factors in NTCP models may improve the predictive power of the model considerably (Defraene *et al.*, 2012; Morimoto *et al.*, 2019; Rancati *et al.*, 2011; Dean *et al.*, 2017; Palma *et al.*, 2020b). A preselection of all treatment- patient- and tumour- related factors by an educated guess is needed to avoid false positive results. For this purpose, a literature search is recommended to define candidate clinical factors to be considered subsequently in model building (Steyerberg and Vergouwe, 2014).

4.1.4 Models and Algorithms

To parametrize the dose-dependence of an organ at risk, typically a sigmoid-shaped function is fitted, like the LKB model, the Relative Seriality (RS) model, and the general logistic regression model (Trott *et al.*, 2012). It has been demonstrated that the general applied logistic regression technique produces very similar dose-effect curves as the LKB and RS model (Defraene *et al.*, 2012). A prerequisite is that the type and pattern of toxicity (i.e. the dependent variable) has to be translated and captured in a 'present (1)/not present (0)'' for logistic regression modelling.

As an alternative in the current information age, data mining and machine learning approaches for toxicity prediction research are increasingly applied (Robertson *et al.*, 2015; Beasley *et al.*, 2018; Gabryś *et al.*, 2018; Luo *et al.*, 2020; Dean *et al.*, 2018; Palma *et al.*, 2019a). Commonalities and differences between the more conventional methods of model-based statistical inference and the rapidly progressing field of data driven machine learning have given rise to an active debate (c.f. the field of imaging in neuroscience (Bzdok, 2017)). It has been shown that machine learning approaches do not, by default, result in better predictions (Yahya *et al.*, 2016; Dean *et al.*, 2018). Unsupervised machine learning approaches aim to produce accurate predictions for unseen data based on a large body of training data, and do not depend on tractable relations between variables, which can limit sensible extrapolation of the associated models. Conventional regression, on the other hand, may reveal the specific dependence of a given variable on several independent variables within a data set. From this comes the opportunity to extrapolate beyond the initial model fitting, beyond the initial conditions under which data were acquired, by adaptation.

Selection of the appropriate statistical test(s) depends on the nature of the predicted outcome. If time to event is considered important, parameters of a proportional hazards model may be inspected (provided proportionality of the hazard is valid), or e.g. accelerated failure time models may be employed (Bradburn *et al.*, 2003). On the other hand, when fixed time point differences or incidence over multiple time points are considered sufficiently descriptive, parametric t-tests or nonparametric signed-rank tests can be performed (Lumley *et al.*, 2002). Rather than to seek rejection of a null-hypothesis, Bayesian analysis may provide a more informative description of observed differences (Kruschke, 2013).

4.1.5 Voxel-wise models

Although conventional statistics can be applied at a pixel-wise or voxel-wise level, a comparison of the aggregated data dichotomised by endpoint is a commonly used approach. Detailed descriptions and formalisms of the process for voxel-wise analysis for toxicity studies are provided by Acosta and De Crevoisier (2019) and Palma *et al.* (2020a). The idea of identifying local dose-response patterns by voxel-wise analysis based on two-sample tests was derived from neuroimaging studies where the aim is to discover voxel-wise changes due to a specific disease (Ashburner and Friston, 2000; Whitwell, 2009). When comparing DSMs/DVMs, the null-hypothesis is that there is no difference between the dose distributions of the patients with and without toxicity, which can be tested either using parametric (e.g. Student's T-test) or nonparametric tests (e.g. the Mann-Whitney U test or the Wilcoxon rank-sum test). In both cases, a map of p-values can be filled in voxel by voxel, pinpointing where are the significant differences between the groups of patients.

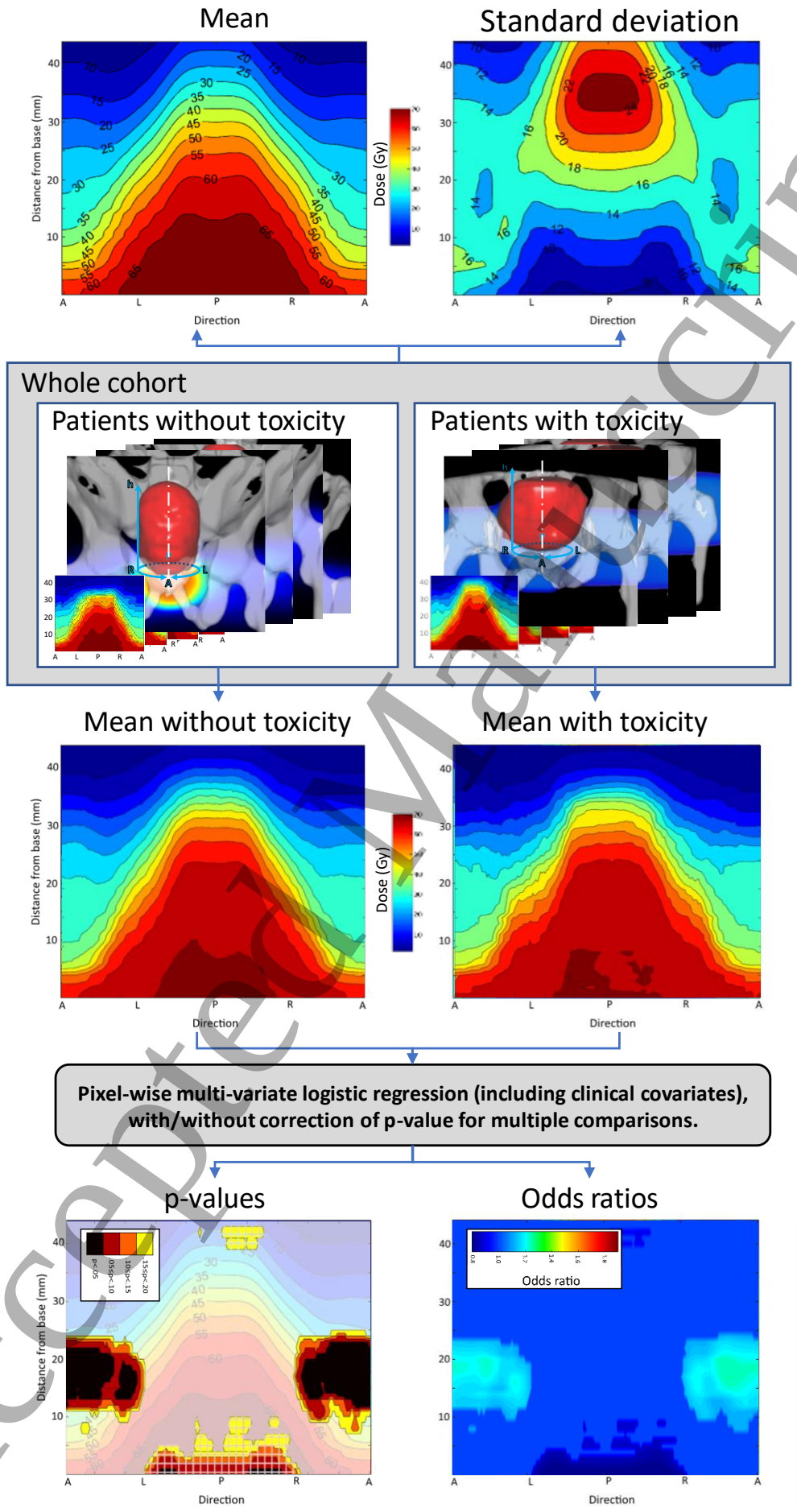


Figure 4: Example of voxel (pixel)-wise assessment of a dose-complication relationship for a change in international prostate symptom score (IPSS) of ≥ 10 in a prostate radiotherapy trial cohort (Yahya *et al.*, 2017). Bladder DSMs have been derived by cylindrical mapping using 200 equally-spaced radial samples at each interpolated 1 mm slice vertically from the bladder base (up to 45 mm for all patients). Pixel-wise logistic regression incorporating clinical factors leads to the (uncorrected) p-value and odds ratio maps shown, including demonstration of a confounding protective effect of dose at the bladder base.

Logistic regression, LKB modelling or Logit dose-response modelling are possible alternative approaches to studying local dose-response effects at the voxel-level (as illustrated via an example for DSMs in Figure 4). For each voxel in the DSM/DVM, the relationship between the dose and the incidence of a selected toxicity endpoint is calculated. When the actuarial incidence of the side-effect is considered, Cox regression constitutes a suitable choice (Marcello *et al.*, 2020a; Marcello *et al.*, 2020b). This analysis produces a map of best-fit parameters, constant and b-coefficient for dose for the logistic regression, TD50 (uniform dose corresponding to 50% complication probability) and slope at TD50 for LKB and Logit models, H0(t) and β-coefficient for dose when Cox is used. This kind of analysis allows identifying regions where the dose-response is steeper vs areas where it is shallow, thus providing a hypothesis for treatment optimization on selected sub-regions.

Clinical risk factors can be included, with the inclusion of multiple b-coefficients/ β-coefficients in logistic and Cox regressions and with the addition of dose modifying factors in LKB and Logit models (Peeters *et al.*, 2006a). Of note, in this case, a map of effect sizes for the clinical risk factors is produced, with a variation of effect sizes at the voxel level. Discussion is still open on the meaning of these variations, with the possibility of a clinical factor to be a protective factor in some voxels and a risk factor in others. A possible alternative way to include clinical risk factors is to use local dose-based modelling to determine areas with different dose-response curves and apply an adjustment for clinical risk factors at a sub-region level or a patient level.

4.1.6 Significance

From a modelling perspective a large variation over the population provides the best opportunity to derive a high-quality dose-effect model (see Table 1). Techniques that result in high rates of toxicity do not necessarily exhibit a large variation over patients. When deriving statistics at the voxel-level, dose deposited by external beams gives rise to correlations between dose variables. Establishing significance based on per-voxel null hypothesis testing (see Section 5.2) severely suffers from multiple testing issues – the likelihood of incorrect rejection of that hypothesis. Methods based on estimated false discovery rate (FDR) have been proposed, which have been shown to hold under positive dependencies (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Storey, 2002). Permutation methods can be used to establish significance based on test statistics aggregated over

the individual voxels (Manly, 1997; Chen *et al.*, 2013; Palma *et al.*, 2020a): a pertinent global threshold of the single-voxel test statistic is derived, leading to the selection of voxels that exceed that value. Reporting the adjusted map beyond arbitrary thresholds for significance (such as the commonly-used p = 0.05) might be suggested, allowing readers to make a more informed conclusion by also considering the trends and spatial patterns of association, rather than focusing on specific highly significant voxels (Palorini *et al.*, 2016b).

4.2 Performance, validity and reporting

Three main purposes of statistical models can be identified: 1) predictive/prognostic models, focussing on event prediction; 2) explanatory models explaining difference in outcome via explanatory variables, focussing on (causal) relationships and the magnitude of effects; and 3) descriptive models with the main purpose to capture accurately the association between the dependent variable and the independent variables, which may focus on both elements of prediction, relationships and magnitude of effects (Shmueli, 2010).

4.2.1 Model performance

NTCP models are descriptive models, describing the relationship between biological dose, clinical cofactors, and toxicity risks. To evaluate discriminative (predictive) power, the performance of the model is commonly reported through the area under the receiver operating characteristic curve (AUC) which is a measure that combines the specificity and sensitivity in one number (Dean *et al.*, 2018; Men *et al.*, 2019). In case of a large imbalance in the data, the F-score based on precision-recall could additionally be considered (Saito and Rehmsmeier, 2015).

4.2.2 Model validity

The internal validity of a prediction model concerns the reproducibility of the underlying data. To avoid overfitting and unstable models, preferred methods for internal validation are cross-validation and bootstrap resampling techniques (Heinze *et al.*, 2018; Steyerberg and Vergouwe, 2014; Xu *et al.*, 2012). For the external validation of the model, concerning the generalizability of the results to other similar patient populations outside the database and outside the institution, independent validation datasets are required (Bentzen *et al.*, 2010). A relevant example is provided by Mylona *et al.* (2020b), where dosimetry for sub-regions in the bladder was found to be more predictive of complications than that for the whole organ, as validated in an external cohort.

4.2.3 Model reporting

It is recommended to report at least the following characteristics of a developed phenomenological (data-driven) NTCP model (Jackson *et al.*, 2010; Collins *et al.*, 2015): study population, received treatment, definition and measurement of predicted outcome, dose-volume information of full organs and relevant sub-volumes, basic statistical data on incidence of toxicity including number of subjects and number of events, complication rates associated with constraints, available follow-up time, statistical motivation of sample size, handling of missing data, numerical range and median of the dosimetric variables of interest, model parameter estimates and their standard errors, applied feature selection method (model building algorithm), candidate variable list, applied validation methods, goodness-of-fit and discriminative power of the final model. For spatial models, it is also recommended to report dose-grid resolution and dose calculation algorithm (Placidi *et al.*, 2020), and a definition for the algorithms used in extraction of features (e.g. (Zwanenburg *et al.*, 2020)). A checklist for transparent reporting is available through the TRIPOD initiative (Collins *et al.*, 2015).

5 Review of methods – spatial dose associations with complications and applications to NTCP calculation

Section 2 defined, in general terms, approaches that may be used, in various dimensions, to represent dose information in ways that retain spatial information from which features may be extracted. Section 3 detailed where the data may be obtained from to inform those processes, and for describing the complication outcomes with which the features will be correlated, using the statistical processes described in Section 4. We can now review publications which attempt to combine these to derive NTCP models and for examining associations of spatial dose information with complication incidence.

Evidence of improved predictive capabilities with models which are inclusive of spatial information have been emerging from analysis of isolated data sets over the last 10 – 15 years. Table 2 provides a summary of some previously published analyses where a comparison has been made between histogram-based toxicity models and those incorporating various forms of spatial dose information.

560

561 **Table 2: A selection of published studies comparing histogram-based toxicity modelling to models incorporating spatial information. Note that some studies incorporated multiple**
562 **approaches to spatial feature extraction.**

Reference	Tumour site	Evaluated region	Toxicity endpoint	Spatial method	Comparison	Impact on complication prediction
Heemsbergen *et al.* (2005)	Prostate	Rectum	Various	Pixel-wise, DSM sub-regions	Total rectum DSH vs sub-regions alone	Several specific toxicities only associated with spatially-localised dose
Peeters *et al.* (2005)	Prostate	Rectum	Acute rectal ≥ G2	Dose-length parameters	Total rectum DVH and DSH vs addition of spatial features	Most significant DVH and dose-length parameter both improved final model
Peeters *et al.* (2006b)	Prostate	Anorectum	Various	DVHs for sub-regions (rectum, anus)	Total anorectum DVH vs sub-regions	Specific toxicities better predicted by sub-region dosimetry
Acosta *et al.* (2013)	Prostate	Rectum	Rectal bleeding	Voxel-wise dose, DVM	Rectum DVH vs voxel-wise	Rectal bleeding only correlates with identified local dose levels, not with total rectum DVH.
Drean *et al.* (2016b)	Prostate	Rectum	Rectal bleeding	Voxel-wise and manually identified sub-region	Rectum DVH vs different sub-regions	DVH-based inferior-anterior hemi rectum (voxel-wise identified sub-region) performed best.
Casares-Magaz *et al.* (2017)	Prostate	Rectum	Various	Pixel-wise DSM	Rectal DVH and DSH vs pixel-wise	For all endpoints DSM-based parameters showed better AUCs (mean 0.64) than the best DSH/DSH-based parameters (mean 0.61)
Rossi *et al.* (2018)	Prostate	Rectum	Rectal bleeding Faecal leakage	3D texture features	Rectal DVH vs addition of spatial features	Bleeding - AUC increased 0.68 to 0.72; leakage - AUC increased from 0.68 to 0.75
Buettner *et al.* (2009a)	Prostate	Rectum	Rectal bleeding	CNN on DSM	Rectal DSH vs addition of spatial features	AUC increased from 0.59 to 0.64
Buettner *et al.* (2011)	Prostate	Rectum	Various	Parameterised DSM	NTCP based on rectal vs addition of spatial features	AUC increased from 0.59 to 0.63 – 0.67
Zhen *et al.* (2017)	Cervix	Rectum	General toxicity	DSM texture features and CNN (with transfer learning)	Peak dose-indices vs texture features vs CNN	AUC 0.47-0.58 (dose-indices), 0.70 (texture features), 0.89 (CNN)
Wilkins *et al.* (2020)	Prostate	Anorectum	Various	Parameterised DSM; manual sub-regions	Rectal DVH vs sub-region DVH vs DSM	DSM-based parameters did not improve prediction compared to DVH-based parameters;

				(rectum, anal-canal)	features	sub-region dosimetry not identified as more predictive
Heemsbergen *et al.* (2010)	Prostate	Bladder	Urinary obstruction	DVM, voxel-wise (specific local dose points)	DSH-based total bladder vs addition of local point dose trigone	Both DVH point and local trigone dose point added to final model
Improta *et al.* (2016)	Prostate	Bladder	IPSS toxicity score	Pixel-wise DSM	Bladder DSH vs addition of spatial features	AUC increased from 0.58-0.71 to 0.66-0.77
Palorini *et al.* (2016b)	Prostate	Bladder	Acute urinary symptoms	Pixel-wise and parameterised DSM	Bladder DSH vs parameterised DSM	For all endpoints DSM-based parameters showed better AUCs than the best DSH-based parameters
Rossi *et al.* (2018)	Prostate	Bladder	Nocturia Incontinence	3D texture features	Bladder DVH vs addition of texture features	Nocturia - AUC increased from 0.63 to 0.67; Incontinence - AUC increased from 0.71 to 0.73
Mylona *et al.* (2019)	Prostate	Bladder, urethra	Acute and late urinary symptoms	Sub-regions derived from voxel-wise DVM analysis	Bladder DVH vs sub-regions DVHs	AUC improvements in both acute and late toxicity in several sub-regions including the urethra (AUCs ≥ 0.62)
Beasley *et al.* (2018)	H&N	Head region	Trismus	Voxel-wise DVM	Organ vs sub-region DVH	Identified voxel cluster most significant
Buettner *et al.* (2012b)	H&N	Salivary glands	Xerostomia	Parameterised 3D organ dose distribution	Mean dose vs 3D moments	AUC increased from ~0.7 to > 0.8
Gabryś *et al.* (2018)	H&N	Parotid glands	Xerostomia	Parameterised 3D dose	Mean dose and parotid DVH vs addition of multiple spatial features	AUC increased from < 0.6 to 0.68 – 0.78 for dose-gradient features
Men *et al.* (2019)	H&N	Glands	Xerostomia	3D dose CNN and CT images	Combinations of basic dose-volume metrics, clinical parameters, and CNN based on images, structures and dose	AUC increased from 0.56 for dose metrics alone to 0.84 with all CNN information
Monti *et al.* (2017)	H&N	Neck region	Dysphagia	Voxel-wise DVM	Sub-region mean dose and multi-organ DVH (Alterio *et al.*, 2017)	AUC confirmed between multi-organ vs voxel-wise analysis (~0.8)
Dean *et al.* (2018)	H&N	Pharyngeal mucosa	Dysphagia	3D spatial parameterisation	Organ DVH vs spatial features	DVH features as predictive as spatial features (AUC ~ 0.71-0.82) and maintained on external

						validation
Dean *et al.* (2016)	H&N	Approximated oral mucosa	Acute mucositis	3D moments	Organ DVH vs addition of spatial features	No improvement
Dean *et al.* (2017)	H&N	Oral cavity	Acute mucositis	Sub-region definition (mucosal surface)	Organ DVH vs sub-region	No improvement
Palma *et al.* (2016)	Thorax	Lung	Lung fibrosis	Voxel-wise DVM identified sub-regions	Whole-lung mean dose vs sub-region-based mean dose	AUC increased from 0.60 to 0.75
Palma *et al.* (2019a)	Thorax	Lung	Lung fibrosis	Voxel-wise DVM	LKB vs 3D model (PACE)	AUC increased from 0.66 to 0.85
Lee *et al.* (2020)	Lung	Oesophagus	Weight loss	3D texture features	Combinations of DVH and radiomic/dosiomic features	Substantial increases in AUC though addition of spatial features
Liang *et al.* (2020)	Lung	Lung	Pneumonitis	3D dose texture features and CNN	DVH vs NTCP vs dosiomics vs 3D CNN	AUC increased from 0.676 (DVH) to 0.744 (NTCP) to 0.782 (dosiomics) to 0.842 (CNN)

Figure 5 illustrates the progression from dose-volume to spatial models of varying complexity, using the relationship of pelvic radiotherapy dose to gastrointestinal complications as an example. References describing studies are provided, grouped according to the complexity of anatomical information used and by the spatial dose features used in the investigations. Many of the cited studies are discussed in more detail below.

Anatomical information

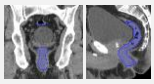
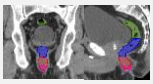
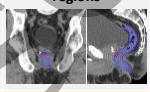
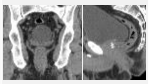
Primary feature definitions		2D	3D				
		2D rectum BEV	3D rectum	Manual sub-regions	Broader pelvic anatomy	Statistical sub-regions	Structure agnostic
							
1D (Histograms)	Estimated DVH	Emami et al 1991					
	Calculated DVH		Fiorino et al 2002; Gulliford et al 2004; Rancati et al 2004; Sohn et al 2007; Michalski et al 2010; Tomatis et al 2012; Ospina et al 2014; Fargeas et al 2018	Peeters et al 2006b; Buettner et al 2012; Stenmark et al 2014; Ebert et al 2015b; Gulliford et al 2017; Wilkins et al 2020	Smeenk et al 2012; Shaake et al 2016	Drean et al 2016b; Mylona et al 2020b	
	DSH/DWH/zDVH		Cheng and Das 1999; Meijer et al 1999; Fiorino et al 2003	Buettner et al 2012; Kim et al 2014; Ebert et al 2015b; Wilkins et al 2020			
2D (Dose-surface maps)	2D pixel-wise		Wortel et al 2015; Onjukka et al 2019				
	Parameterised		Heemsbergen et al 2005; Munbodh et al 2008; Buettner et al 2009b; Buettner et al 2011; Moulton et al 2017; Shelley et al 2017; Vanneste et al 2018; Henderson et al 2018; Casares-Magaz et al 2017, 2019	Buettner et al 2012; Wilkins et al 2020			
	Cluster model		Tucker et al 2006b				
	Supervised		Casares-Magaz et al 2017; Chen et al 2018				
	Unsupervised		Buettner et al 2009a;				
3D (Dose-volume maps)	3D voxel-wise		Acosta et al 2013; Fargeas et al 2013;			Drean et al 2013	Marcello 2020; Ospina et al 2013;
	Parameterised						Chen et al 2011;
	Supervised		Rossi et al 2018				
	Unsupervised		Zhen et al 2017				Cologner et al 2015

Figure 5: Illustration of the variety and evolution of methods for incorporation of dosimetric features into dose-complication association studies and NTCP models in the context of gastro-intestinal toxicity. References are provided as examples for studies involving various combinations of anatomical information and dosimetric feature extraction and are not exhaustive. Studies can be further broken down according to the model used for association with complication (see e.g., Acosta and De Crevoisier (2019)). (BEV – beam's eye view).

5.1 Use of histogram-based features

5.1.1 Description

The degeneracy of the spatial dose distribution into the associated DVH of a structure may be moderated if the dose distribution can be correlated with more specific descriptions of the underlying functional structures themselves. This can be achieved, for example, by breaking a given structure down spatially into more precise or component sub-structures according to some anatomical or statistical criterion (as described in Section 2.2.1). The DVH characteristics of each sub-structure can be considered independently. Analysis of more specific structures can also reveal that dose to the originally-hypothesised structure of interest may be less correlated with complication than alternative adjacent structures. It is also possible to utilise additional spatial information regarding the structure (such as medical imaging scans) to modify the basic DVH information being used as input to a dose-volume based NTCP model.

5.1.2 Examples

A first class of models is based on the assumption that the organs can be thought of as organized in functional sub-units (FSUs). If the density of FSUs $f(\vec{r})$ is not homogeneous throughout the considered structure Ω, a more informative version of the DVH would be weighted by the corresponding $f(\vec{r})$ yielding $fDVH(D_0)$, defined as:

$$fDVH(D_0) = \frac{\int_\Omega f(\vec{r}) H[D(\vec{r}) - D_0] d\vec{r}}{\int_\Omega f(\vec{r}) d\vec{r}}$$

where $H(\cdot)$ is the Heaviside step function (Lu *et al.*, 1997). Though DVH-based NTCP models would be better recast on fDVH, it has been recognized that the derivation of the detailed underlying arrangement of FSUs in most anatomical sites still requires dedicated studies from techniques such as functional imaging (e.g. (Arslan *et al.*, 2018; Lee and Park, 2020)).

For lung, a low-cost variation on the fDVH concept is represented by the dose-mass histogram (DMH), in which the mass density (easily estimated from the planning CT) is considered as a surrogate of FSU density. As expected, the DMH results to be independent of breathing phase (Nioutsikou *et al.*, 2005; Cella *et al.*, 2015). Interestingly, however, a study on the risk of postoperative pulmonary complications among oesophageal cancer patients found no evidence of significant benefits from the substitution of DVHs with DMHs within the NTCP model (Tucker *et al.*, 2006a).

Similarly, for hollow organs such as the rectum, the absence of FSUs within the wall content led to the development of the dose-wall histogram (DWH). DWHs represent the DVH of the organ wall only based on the segmented outer organ contour (Meijer *et al.*, 1999). The dose-surface histogram

(DSH) lies, instead, on the histogram of the dose delivered to a representative surface of the organ. Two main approaches have been proposed for the DSH computation: one based on the interpolation of the dose on the organ surface (Lu *et al.*, 1995), and one normalizing the DVH of the organ wall by the shell depth in the limit of vanishing thickness (Palma and Cella, 2019). There is often a strong correlation between the various histogram types (Fiorino *et al.*, 2003; Carillo *et al.*, 2012; Hoogeman *et al.*, 2005). An exception is when the irradiation technique delivers a dose gradient that is steep relative to the organ size, such as found by Kim *et al* for prostate cancer patients treated with stereotactic radiotherapy (Kim *et al.*, 2014).

A first hybrid approach for including a notion of spatial dose distribution within a histogram framework is the zDVH (Cheng and Das, 1999), which expresses the volume receiving a given dose at a given cranio-caudal position in the form of a 2D histogram.

An effective approach based on pathophysiological knowledge of the toxicity aetiology consists in splitting a heterogeneous district into component substructures to achieve better DVH-response predictions. This approach has been made for the anorectum (Peeters *et al.*, 2006b; Ebert *et al.*, 2015a) and the bladder trigone (Ghadjar *et al.*, 2014; Henderson *et al.*, 2018). Outcome associations have also been undertaken over broader spatial ranges of anatomy than conventionally hypothesised. For rectal toxicity in pelvic radiotherapy for example, although the gastrointestinal tract is usually targeted for derivation of associations, alternative structures can provide stronger associations with specific toxicities. Smeenk *et al* showed that incontinence was more strongly associated to dose to the pelvic floor muscles (Smeenk *et al.*, 2012), whilst Gulliford *et al* discovered the importance of dose to the peri-rectal fat space for control-like symptoms (Gulliford *et al.*, 2017).

The emergence of voxel-wise toxicity analyses in radiation oncology has fostered a data-driven evolution of this approach. This is aimed at defining, on a statistical basis, the relevant anatomical substructures involved in the development of radiation induced morbidity and from which histogram-based features can be extracted. This approach is described in Sections 5.2 and 5.4.

5.2 Voxel-wise assessment

5.2.1 Description

In contrast to analyses based on known or hypothesised FSUs as in Section 5.1, the use of voxel-wise methods points to an “agnostic”/bottom-up approach. Once the DSMs/DVMs in a cohort are spatially registered to a common coordinate system (see Section 2.2 for relevant details) in a way that they can be compared voxel-wise, the regions which are significantly associated to the

particular (toxicity) outcome are identified by statistical inference. Different approaches can be used, as described in Section 4.1.5. In general, the final goal of voxel-wise analysis is to identify regions driving the clinical manifestation of radio-induced side effects, i.e. to find clusters of voxels where the dose is significantly different in patients with/without toxicity (see also Section 5.4). The resulting organ sub-regions do not consider any prior anatomical or functional division. They can provide information to make inferences on the differential radio-sensibility of some organs or the simultaneous implication of different structures on some radio-induced toxicities.

Voxel-wise assessment does not by default generate an NTCP model. DVHs in the regions that were highlighted as statistically associated with the selected outcome should be considered to derive NTCP models following a classical dose-response analysis. Alternatively, a total complication risk can be formed from aggregation of risks determined at the voxel level.

5.2.2 Examples

2D dose-surface outcome mapping

2D DSMs (Section 2.2.2) are usually generated from an anatomical structure and restricted to the surface of this structure. This choice produces results which can be easily translated into organ sub-regions to be spared. Historically, the first analyses of DSMs in the radiotherapy field were related to hollow organs whose geometry could be easily associated with a cylinder, such as the oesophagus (Chen *et al.*, 2013; Dankers *et al.*, 2017) and the rectum (Casares-Magaz *et al.*, 2019; Munbodh *et al.*, 2008; Onjukka *et al.*, 2019; Sanchez-Nieto *et al.*, 2001; Tucker *et al.*, 2006b; Wortel *et al.*, 2015). Although pixel-wise assessment can be made to derive patterns of response, significant progress has been made by parameterising the DSMs, reducing the number of features and providing parameters for NTCP models, as discussed in Section 5.3.

Pixel-wise studies have related DSMs for the bladder with a number of early and late urinary endpoints (Palorini *et al.*, 2016b; Mylona *et al.*, 2020a; Palorini *et al.*, 2016a; Yahya *et al.*, 2017; Improta *et al.*, 2016). Recently a method for the calculation of DSM for the heart was implemented by using a modified cylindrical coordinate system (McWilliam *et al.*, 2020). DSMs of the heart were analysed to infer possible local dose effect for survival after lung cancer radiotherapy. The rationale for considering heart DSMs rather than DVMs resides in the location on the surface of some clinically relevant sub-regions, such as the coronary arteries, the electrical conduction system and the myocardium.

3D voxel-wise outcome mapping

DVMs (Section 2.2.4) can be generated either starting from an anatomical structure and restricted to its volume or independently from any structure. The second choice has the power to embrace a totally agnostic approach regarding which organs/tissues are involved in radio-induced toxicity and entails the possibility of highlighting the interaction between different organs and FSUs. Notably, special care should be taken in order to counteract the possibility of finding significant areas which offer no feasible anatomical explanation and which could lead to inappropriate organ-sparing objectives in treatment planning.

Organ-based DVMs were considered in the literature for the analysis of both rectal (Acosta *et al.*, 2013; Drean *et al.*, 2016b; Mylona *et al.*, 2020b; Shelley *et al.*, 2017; Marcello *et al.*, 2020b) and urinary (Mylona *et al.*, 2019; Mylona *et al.*, 2020a; Mylona *et al.*, 2020b; Marcello *et al.*, 2020a) toxicity. These kinds of analysis heavily build upon robust co-registration methods, which become even more critical when organs highly prone to organ motion and variable filling are considered.

The first published example of the use of quasi-organ-agnostic DVMs was from Heemsbergen *et al.* (2010) investigating urinary toxicity. In this case, the 3D reconstruction started with the definition of the outer surface of the prostate and with the identification of the spatial coordinates of the prostate centre of mass. After that, for every patient, a spherical surface was considered, extending 6 cm from the prostate. Every voxel inside this region was identified through polar coordinates (distance from the prostate centre of mass and two angles identifying the vector connecting the single voxel to the prostate centre of mass) and the absorbed dose in each voxel was calculated by trilinear interpolation of the nearest dose points of the individual dose grid.

Regression coefficients associated to each voxel specifically in the salivary glands have been used to shed light on the regional radio-sensitivity of the glands (Jiang *et al.*, 2019). Other studies considered DVMs without any restriction to specific contoured organs for investigation of local dose effects in the thoracic/head and neck region (Beasley *et al.*, 2018; McWilliam *et al.*, 2017; Monti *et al.*, 2017; Palma *et al.*, 2016; Palma *et al.*, 2019d; Palma *et al.*, 2019c; Green *et al.*, 2020), with interest in the association of dose pattern with lung toxicity, acute dysphagia, trismus and survival.

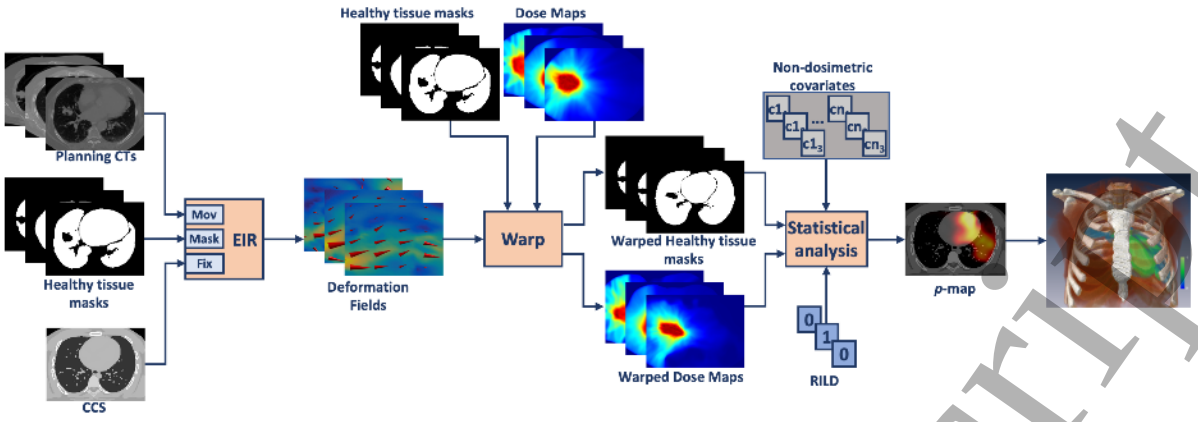


Figure 6: Flowchart for 3D voxel-wise analysis of dosimetric association with lung toxicity following stereotactic body lung radiotherapy, reproduced from Palma *et al.* (2019c). Structures and dose distributions are spatially normalised to a common coordinate system (CCS) via course alignment of structures and CT elastic image registration (EIR). Voxel-wise significance maps are then derived by calculating a test statistic at the voxel level and adjusting for multiple comparisons via a permutation algorithm.

NTCP from voxel-wise methods

As already pointed out, although voxel-wise analysis can identify important organ sub-regions, it does not provide an NTCP.

NTCP can be derived by classical dose-response modelling on the specific sub-regions, either including the whole DVH as calculated in the specific identified areas or choosing some particular DVH cut-points. Examples of this kind of approach can be found in (Buettner *et al.*, 2009b; Heemsbergen *et al.*, 2010; Mylona *et al.*, 2019; Onjukka *et al.*, 2019; Palma *et al.*, 2016; Casares-Magaz *et al.*, 2017). In Drean *et al.* (2016b), parameters for the LKB model were computed within the voxel-wise derived sub-region.

More sophisticated and more global approaches were also developed, taking information from voxel-wise analysis directly into account. Vinogradskiy *et al.* (2012) proposed a modified LKB model where the lung dose in each voxel was weighted using a user-defined spatial weighting matrix which could be derived by a previous voxel-wise analysis. Jiang *et al.* (2019) demonstrated prediction of xerostomia induced by the irradiation of the salivary glands in head-and-neck cancer patients using a ridge logistic regression model directly dealing with the local dose delivered to each voxel of the organ at risk. The framework was naturally able to include non-dosimetric predictors in the NTCP model.

Palma *et al.* (2019a) established a new formalism, called PACE (Probabilistic Atlas for normal tissue Complication Estimation in radiation therapy), which incorporates regional dose information coming

from regression performed at the voxel-wise with clinical risk factors. PACE builds upon the LKB model and substitutes the generalized equivalent uniform dose (EUD) with a weighted combination of risks as calculated by regression at the voxel level, using confidence intervals for predicted risks as weights (thus giving more weight to more certain predictions).

5.3 Spatial parameterisation of dose distributions

5.3.1 Description

Attempts to reduce the number of features, reduce collinearity and generalise models from voxel-wise analyses can be made by parameterising the dose distribution. For analyses restricted to specific organs, this will typically involve functional parameterisation of DSMs and organ-constrained DVMs, with or without registration to a template geometry. The resulting parameters can then become co-variates in regression models or supervised machine learning models. More widespread dose distributions can be parameterised using techniques borrowed from imaging analytics – namely, the supervised derivation of specific feature classes ("dosiomics" (Liang *et al.*, 2019)). Unsupervised classification of outcome based on the dose distribution can also be attempted with convolutional neural networks, with or without the inclusion of anatomical and functional imaging information.

5.3.2 Examples

Parameterisation of 2D dose

The process of derivation of parameters for geometric descriptors from a DSM is illustrated in Figure 2, particularly in the context of investigating rectal complications due to prostate radiotherapy. Concentric isodoses on the rectal wall from prostate radiotherapy can be thresholded systematically at different doses and fitted with an ellipse. Parameterised geometrical features can then be calculated (Buettner *et al.*, 2009b).

Previous studies on rectal toxicity following prostate radiotherapy indicate that spatial dose metrics such as lateral extent of dose around the circumference of the rectum, longitudinal extent and eccentricity derived from rectal dose surface maps (DSM), are related to toxicities including rectal bleeding and loose stools (Buettner *et al.*, 2009b; Moulton *et al.*, 2017). Interestingly, a recent test of this approach failed to demonstrate any improvement over DVH-based prediction of rectal toxicity (Wilkins *et al.*, 2020). This result may be confounded by the differences between planned and delivered dose distributions (see Section 3.1.3), with Shelley *et al.* (2017) finding parameters derived

from DSMs for accumulated dose being more predictive than those from planned dose, as also found by Casares-Magaz *et al.* (2019) at the pixel-level.

Casares-Magaz *et al.* (2017) demonstrated that parameterised DSMs performed slightly better than DSHs when predicting rectal toxicity and produced results for more endpoints by quantifying the dose when a DSM was subdivided to a 3x3 matrix. Vanneste *et al.* (2018) used DSMs to evaluate the effect of hydrogel rectal spacers on dose to the rectum for prostate radiotherapy.

Although most published data relating to parameterised rectal DSMs is from patients who received prostate radiotherapy, Chen *et al.* (2018) detail the use of DSM to relate the dose from both external beam and brachytherapy for a cohort of cervix patients. The two dose distributions were non-rigidly registered, and a rectal DSM created from the summed dose distribution. Both volumetric and texture metrics were calculated, and principal component analysis used to provide inputs to a support vector machine-based model. Area and texture parameters were found to be important and to have an improved AUC compared to the standard Groupe Européen de Curiethérapie/European Society for Radiotherapy (GEC-ESTRO) model.

Parameterisation of 3D dose

For a solid structure such as the parotid it is possible to define metrics to quantify the relative 3D spatial distribution of dose to the whole organ. Buettner *et al.* (2012b) used 3D spatial invariant moments to characterise the morphology of the dose distribution to the parotid in terms of centre of mass, spread and skewness. They showed that minimising the dose to cranial and lateral regions of the parotid gland would decrease the incidence of xerostomia. The model containing spatial metrics had a significantly-improved performance compared to the standard predictive of model of mean dose. 3D moments were also used in a comparison of the conventional oral mucosa outline and a novel segmentation to predict acute mucositis (Dean *et al.*, 2016). Dose distributions to the two organ-as-risk (OAR) structures were calculated and used as inputs to both penalised logistic regression and random forest models. In this example, using the novel segmentation and spatial metrics did not improve model performance compared to a model built on fractional dose-volume data for the conventional structure. (Dean *et al.*, 2018) studied acute dysphagia using moments and dose–volume-length and -circumference data for the pharyngeal mucosa. They demonstrated that although the length and circumference that received over 1 Gy per fraction were shown to be important, a penalised logistic regression NTCP model trained purely on dose-volume data performed equally well on internal validation and was superior when applied to an external validation cohort.

Supervised broad spatial descriptors

The papers described so far have used a variety of bespoke methods to parameterise the spatial distribution of dose. However, synergy with the field of radiomics allows spatial dose distributions to be characterised by a vast array of standardised metrics (Zwanenburg *et al.*, 2020). Here the 3D radiotherapy dose distribution can be characterised in full with or without reference to structure segmentation. Radiomic features from patient images can be integrated to derive models predictive of complication (Talamonti *et al.*, 2019).

One study which assessed this concept and compared predictions to previous work is described by Gabryś *et al.* (2018). This study compared models to predict xerostomia starting with a standard model based on mean dose and parotid-specific spatial metrics described above (Buettner *et al.*, 2012b). Spatial descriptors were extended for the parotid gland to consider entropy along with dosiomic descriptors of DVH shape and general dosiomic features describing the gradient of the entire 3D dose distribution. The manuscript describes comparisons of many models including conventional statistical and machine learning approaches. Additionally, feature selection and class balance approaches were compared. Overall the strongest features identified were parotid gland volume eccentricity and the spread of the contralateral parotid dose distribution. The contralateral dose gradient of the 3D dose distribution (right to left) was also identified on univariate analysis but did not feature strongly in the final multivariate analysis. Similarly, Lee *et al.* (2020) informed machine learning algorithms with combinations of dose-volume, radiomics and dosiomics features, together with clinical co-variates. Resulting predictive models of weight loss in lung cancer radiotherapy with greater accuracy than models based on dose-volume and clinical co-variates alone, though still with a relatively modest AUC of 0.71.

Liang *et al.* (2019) compared conventional dosimetric models with models derived from dosiomic features to predict radiation pneumonitis. It was demonstrated that a multivariate NTCP model including the Grey Level Co-occurence Matrix (GLCM) contrast and Grey Level Run Length Matrix (GLRLM) (which has similarities to lateral and longitudinal extent described previously) outperformed models based on dose-volume parameters and conventional NTCP model parameters. Rossi *et al.* (2018) included texture analysis features in a study to predict genitourinary and gastrointestinal toxicity following prostate radiotherapy. The 3D texture analysis features for the rectum and bladder were derived from standard radiomics and used alongside non-treatment related features (such as age, staging and comorbidities) and DVH-based metrics to build multivariate logistic regression NTCP models. It was demonstrated that for gastrointestinal endpoints inclusion of texture features improved the AUC compared to models containing only

clinical and DVH-based features. Results for genitourinary toxicity were generally not improved by any dosimetric features.

Unsupervised broad spatial descriptors

An alternative to utilising crafted dosiomic descriptors of broad distributions is to apply neural networks. For example, Buettner *et al.* (2009a) used DSMs derived by the rectum unfolding as input for a rectal bleeding model based on locally-connected neural networks able to account for the topology of the dose distribution. The higher performance achieved by such models, compared to the more traditional fully-connected conventional neural nets based on DSHs, suggested that morphological aspects of the dose distributions play a relevant role in the development of radiation induced morbidity. Zhen *et al.* (2017) utilised a convolutional neural network (CNN) to distinguish rectal DSMs indicative of toxicity, incorporating transfer learning to compensate limited patient data.

CNNs can be used to extract unspecified higher-level features of 3D dose distributions which can directly classify the distributions as likely to lead to complications, and studies have begun to emerge demonstrating this with varying combinations of ancillary information. Ibragimov et al (Ibragimov *et al.*, 2018; Ibragimov *et al.*, 2019) utilised CNNs incorporating 3D dose information, supplemented with transfer learning from previous abdominal imaging, for hepatobiliary toxicity prediction following stereotactic liver radiotherapy. Incorporating the CNNs with more conventional features including dose-volume parameters, dose-fractionation and clinical co-variates increased the model predictions (increase in AUC from 0.79 to 0.85). In a strategy which preferences identifying likely toxicity (i.e., minimising false negatives), the CNN approach halved the number of false positive predictions relative to DVH-based prediction. Ibragimov et al were able to extend this approach to a structure-agnostic spatial assessment to map anatomical regions where dose variations associate with toxicity. This revealed regions associated with the hepatobiliary tract and liver as specific focus regions to guide dose planning (Ibragimov *et al.*, 2020).

In a progression from the dosiomics approach, Liang *et al.* (2020) utilised CNNs incorporating the 3D dose distribution for predicting pneumonitis following volumetric-modulated radiotherapy. A superior prediction (AUC 0.842) was achieved relative to regression models incorporating dosimetric, NTCP and dosiomics features (AUC < 0.782). Class activation maps were used to identify lung regions associated with increased or reduced high-grade toxicity.

In head and neck cancer radiotherapy, Men *et al.* (2019) used CNNs which incorporated one or more of the planning CT images, planned 3D dose and segmented anatomy, for prediction of grade ≥ 2

xerostomia, and compared prediction against regression models incorporating dose with or without clinical co-variates. The CNNs provided greater accuracy (AUC < 0.84), compared to the regression models, for all combinations of 3D information except for when 3D dose was removed.

5.4 Spatial clustering

5.4.1 Description

Thames *et al.* (2004) proposed that hot spots distributed as small areas throughout an organ at risk are likely to cause a different response than if the highest dose covers one contiguous region. This difference would translate through to a difference in toxicity prediction using NTCP models which describe the clustering of damage to FSUs. This concept of spatial dose clusters forms bridges between voxel-wise assessment, definition of sub-regions and spatial analysis based on spatial parameterisation. The cluster models highlight the relevance of including both the number and the spatial location of radiation-sterilized FSUs in a comprehensive NTCP model (Thames *et al.*, 2004). In a general sense, these models suggest that a volume receiving at least a given dose value is more likely associated with a radiation-induced toxicity if it corresponds to a connected spatial cluster rather than if spatially scattered (Deasy and El Naqa, 2008).

5.4.2 Examples

Tucker et al demonstrated a practical application of the method described by Thames using rectal DSMs (Tucker *et al.*, 2006b). Nine case-control pairs with very similar absolute DSH but with and without grade 2 rectal bleeding were used to fit a local-effect cluster model. The logistic function describing the probability of damage for each voxel in each DSM had 2 unknown parameters. The model was fitted to maximise the relationship between maximum cluster size (considering 2-connectivity) between the cases and controls. Although the cohort was very small, the authors were able to find parameter values which separated cases from controls and inferred that dose distributions in the region of 30 Gy were important for the prediction of rectal bleeding.

Chao et al (Chao *et al.*, 2020; Chao *et al.*, 2018) also developed spatial cluster metrics based on the method proposed by Thames et al. They demonstrated that maximum cluster size for the superior 5 cm of the oesophagus was not related to conventional dose-volume and NTCP metrics and inferred that spatial distributions were not represented by conventional dose metrics. They applied a cluster-based approach to model xerostomia (Chao *et al.*, 2019). The metrics included mean cluster size and largest cluster size normalised to the volume of the gland which were incorporated into LKB models. Although no conventional (DVH-based) LKB model was derived from the cohort a comparison was

made with published models utilising mean dose (n=1). TD50 was higher for the thresholded cluster model at just below 40 Gy compared to 26 Gy from the literature.

The concept of spatial cluster models can be expanded using percolation theory, which has origins in statistical physics and considers how clusters are connected. Originally proposed at a similar time to Thames' (Thames *et al.*, 2004) work on cluster models, Myers and Niemierko (2004) presented the use of percolation theory for predicting NTCP from clusters. Gale *et al.* (2017) describe how the concept can be applied to geometric arrangements of FSUs to predict toxicity for both parallel and serial organs.

Several studies considered the clusters of organ voxel L_p^- whose dose-toxicity association exceeded some statistical significance threshold p. They showed that the mean dose in such clusters could be a more powerful predictor of toxicity than traditional metrics associated to the organ considered as a whole structure. Hence, an NTCP model can be proficiently trained as a logistic regression of the patients' outcomes as a function of simple dose metrics in the cluster L_p^- propagated from the common coordinate system of the voxel-wise analysis (see Section 5.2) to each individual native space. In this way, sub-regions have been identified in different locations such as the lungs (Palma *et al.*, 2016), the heart (McWilliam *et al.*, 2017), head and neck (Monti *et al.*, 2017), the rectum (Acosta *et al.*, 2013; Drean *et al.*, 2016b) and the bladder(Mylona *et al.*, 2019).

6 Ongoing Endeavours

6.1 Model development and validation

As for other approaches to radiotherapy complication modelling, a major issue is represented by the quantity and quality of data available to researchers. Relative to DVH-based models, spatial methods require more comprehensive data (see Section 3). Despite the abundance of relevant data generated continuously around the world and the technical capability to collect it, and despite decades of pleas (e.g. (Deasy *et al.*, 2010)), remarkably little data has become available to progress this type of analysis. Based on legislative constraints (i.e., ownership, privacy and patient consent needs) it is likely, at least in the next few years, that data will prevalently come from clinical trials where their recovery, storage and access are already planned.

The implementation of innovative trials including large cohorts of clinical data (Baumann *et al.*, 2016) could rapidly change the landscape. Such trials could multiply the opportunities for developing models, provide opportunities for validating models, and enable the merging of different large cohorts to increase feature diversity. A specific issue may concern the possibility of introducing

unpredictable biases if pooling together cohorts of patients treated, for instance, at different dose levels with largely different spatial locations of the high-dose volumes. Uncontrolled voxel-wise comparisons could lead to "false" spatial effects due, for instance, to the higher incidence or prevalence of side effects in cohorts delivering systematically higher doses and/or treating larger volumes. Ideally, the availability of large cohorts should be accompanied by a proper grouping of patients to make the different patient groups comparable.

6.2 Model generalisation and extension

Apart from the critical issues related to generalizability of NTCP models such as technical, temporal or geographical variabilities (van der Schaaf *et al.*, 2015), a few specific points deserve discussion.

The interplay between the spatial patterns of a certain modality/technique and the inter-individual variability is a challenging issue: well driven studies may help in quantifying the real impact of a modality with respect to another. The generalizability of models across different modalities need high-quality studies and extensive validation. One confounding problem is that the patterns of dose delivered today already reflect the existing knowledge based on dose volume metrics. As these models mature, there is the potential for radiobiological predictions that consider the spatial pattern of dose that can drive the optimization of treatment plans towards more favourable dose patterns beyond that of the traditional dose-volume metrics.

Another important field of investigation regarding model generalisation is represented by the challenge of combined therapies. Data from studies testing radiotherapy-only vs combined therapy (for instance chemotherapy, immunotherapy) could help in assessing spatial dosimetry correlations specifically linked to the action, for instance, of a drug and making possible local dose corrections incorporating its effect. Similarly, highly non-conventional dose and dose-rate distributions, such as from ultra-fast irradiation (Esplen *et al.*, 2020) or molecular radiotherapy (Stokke *et al.*, 2017) will offer new data sources with which to generalise derived models.

6.3 Including intra and inter-fraction changes

In many situations, both intra and inter-fraction anatomical and geometrical changes may have a significant impact in modifying the delivered dose with respect to the planned one. In particular, the prevalence of systematic over random changes may potentially blur (or even hide) the correlation with toxicities; consequently, investigations quantifying these effects are needed. As an example, the impact of variable bladder filling on bladder DSM can be assessed from daily cone-beam CT imaging: one recent study showed a relatively small impact of variable filling on bladder DSM during image-guided radiotherapy of prostate cancer (Palorini *et al.*, 2016a). A statistical approach based on

Gaussian-like variations of local doses likely works in several situations but is expected to fail in others, such as when the phenomenon itself is prevalently non-Gaussian. Shelley *et al.* (2017) demonstrated superiority in toxicity prediction from rectal DSMs formed from estimated delivered rather than planned dose.

When the toxicity rate is small (say, <10%), those few patients with large systematic changes resulting in a relevant increase of dose to proximal OARs may jeopardise results. Greater efforts may especially be expected in trying to incorporate individually-assessed anatomical modifications in stereotactic body radiation therapy (SBRT) (Magallon-Baro *et al.*, 2019), looking to the 3D dose-of-the-day and/or to the accumulated dose instead of the planned dose distribution. SBRT is also prone to be associated with larger effects due to both the reduced margins and the high dose per fraction, dealing with an enhanced impact on critical regions even with small anatomical/geometrical changes. Relevant effects due to systematic deviations between the planned and the delivered dose may occur even in unexpected situations and the availability of in-room imaging information is of paramount importance to identify them. The recently reported correlation between shift toward the heart of field isocentre during delivery and poorer survival in lung cancer patients treated with SBRT is a highly paradigmatic example (Johnson-Hart *et al.*, 2018).

Similarly, intra-fraction changes are known to significantly affect the delivered dose in specific sites. Breathing-induced motion can be highly anisotropic and variable between patients in the different thoracic and abdominal areas. Although, to our knowledge, no studies have reported on the impact of intra-fraction motion on spatial models, more relevant research in this area is needed.

6.4 Potential applications of artificial intelligence

The rise of deep learning approaches for image segmentation, pattern recognition and patient classification adds many opportunities to extend this field (El Naqa and Das, 2020). Ready access to advanced deep learning tools is making this kind of analysis more popular (with examples given in Section 5.3). A merit of these methods is the opportunity to consider features mostly “hidden” to the human eye and to find complex correlations in a multi-layer approach. On the other hand, this same merit may also constitute a disadvantage from the point of view of interpretability of the results and consequent confidence in clinically applying them; in fact, any attempt to maintain some causality to explain any correlation is largely lost. A major issue regarding artificial intelligence models is their intrinsically higher difficulty in being validated. Valdes and Interian (2018) provide a timely summary of the potential for mis-interpretation in such complex approaches. Keeping the models as simple and interpretable as possible should be considered valuable: the benefit of the

addition of deep learning based spatial signatures should always be demonstrated and quantified in validation cohorts.

6.5 Understanding pathophysiology

An intriguing and relevant field of investigation related to NTCP models based on 3D/2D similarity comparisons concerns the meaning of the resulting regions whose dose differences are predictive of toxicity. As already underlined, the information resulting from these analyses cannot be automatically associated to a specific cause, being intrinsically a phenomenological finding (i.e.: simply reflecting some statistical correlation). Moreover, the assessment of specific volumes/surfaces within the body/OARs apparently more "sensitive" to radiation can be biased by unknown factors or just due to geometrical or technical issues. Any hypothetical causality has to be considered as a strength of such models, in case the found results are consistent with known physiological processes/functionalities. As examples, identification of the bladder trigone as a structure likely to be highly sensitive (Rancati *et al.*, 2017; Henderson *et al.*, 2018; Yahya *et al.*, 2017) is consistent with the involvement of the trigone in the physiology of urination, and the physiological connection between the heart and lungs (Ghobadi *et al.*, 2012) adds validity to correlation of heart dose with lung toxicity (Palma *et al.*, 2019c; Palma *et al.*, 2019d). Any hypothesis generated by such models would deserve to be tested in pre-clinical and clinical studies. Animal models may be well used to verify the existence of spatial effects. Conversely, pre-clinical research may first explain specific patterns of toxicity that may be confirmed later by studies dealing with dose similarity comparison. An interesting example is the evidence of spatial dosimetry effects within parotids impacting xerostomia, due to the sparing (or not) of stem cells contained in the ductal region. Such observations have been reported in animal experiments (van Luijk *et al.*, 2015) and confirmed by a 3D dose comparison investigation on data from a large patient cohort treated for head-neck cancer (Jiang *et al.*, 2019).

6.6 Model application

Although examples of practical applications of NTCP models incorporating spatial dosimetric features are rare, it is likely that a few of the most robust results will increasingly influence planning optimization. When a causal relationship between a spatial effect and the pattern of the corresponding side effects is apparent, changes may be easily implemented in clinical practice. Two examples are the previously-cited cases of the bladder trigone for prostate cancer and the ductal region of the parotid glands. The latter, cited above as originating in pre-clinical studies, is being assessed within a clinical trial (van Luijk *et al.*, 2015), which is probably the first example of a trial

specifically looking to the possibility to exploit information regarding the spatial dose distribution within an OAR to reduce toxicity.

A likely progression will be the incorporation of spatial models into tools to evaluate the planned 3D dose distribution and for generating NTCP and risk estimates. This could be accomplished, for instance, within clinical trials or as an additional tool for plan quality assurance, in complement with conventional DVH-based EUD/NTCP estimates. The propagation of identified sensitive sub-regions to an individual would facilitate toxicity-minimised planning, without the need to modify current optimisation methods (Acosta and De Crevoisier, 2019). This has been demonstrated by Lafond *et al.* (2020). A subsequent natural extension would be the possibility to implement these models directly into the optimization engine. However, the general adoption of spatial models is greatly inhibited by the prior evolution of the planning process and optimisation engines in the context of dose-volume approaches. For spatial models that cannot be formulated via dose-volume terminology, research planning systems are required to enable inclusion of the relevant predicted complication models in optimisation constraints and objectives or via scripting capabilities of commercial planning systems (e.g. (Voutilainen, 2016)). With the growth of artificial intelligence based planning systems, there is considerable scope for building automated planning algorithms that directly incorporate spatial models to augment or replace dose-volume based optimisation (Wang *et al.*, 2019).

Intriguingly, for models which are agnostic to segmented structures, plan optimisation could in principle be feasible without the incorporation of dose-volume data for OARs. This would permit a segmentation-free plan optimization. In the same direction, this kind of approach could also find applications in overall treatment optimization, directly considering patient outcome as the goal and incorporating possible “systemic” effects due to the irradiation of multiple organs and to the interaction with the immune system (for instance through the implicit consideration of the incidental irradiation of nodes and of the vascular system). Similarly, one could hypothesize applications in combined treatments to include the effect of modifying agents at the voxel-level, and to “virtual human” simulation in the optimisation of patient-specific treatments.

7 Conclusion

The field reviewed in in this paper is still in its infancy. However, models which consider the spatial characteristics of radiotherapy dose will permit the expansion, or at least fine-graining, of the solution space for radiotherapy treatment planning and improving the prediction of treatment complications. The potential for large-scale relevant applications in treatment personalization, plan

optimization and evaluation are more than promising. Rapid developments and extensive applications are expected in the coming years.

8 Acknowledgements

ME acknowledges funding support from the National Health and Medical Research Council (NHMRC grant 1077788). TR was partially supported by the Fondazione Italo Monzino. OA and RD acknowledge partial funding from a French government grant (through the CominLabs excellence laboratory and managed by the National Research Agency in the ‘‘Investing for the Future” program, under reference ANR-10-LABX-07-01). SG is supported by a Cancer Research UK Centres Network Accelerator Award Grant (A21993) to the ART-NET Consortium.

References

Abravan A, Faivre-Finn C, Kennedy J, McWilliam A and van Herk M 2020 Radiotherapy-Related Lymphopenia Affects Overall Survival in Patients With Lung Cancer *Journal of Thoracic Oncology* **15** 1624-35

Acosta O and De Crevoisier R 2019 *Modelling Radiotherapy Side Effects - Practical Applications for Planning Optimisation,* ed T Rancati and C Fiorino (Boca Raton: CRC Press) pp 415-40

Acosta O, Dowling J, Cazoulat G, Simon A, Salvado O, de Crevoisier R and Haigron P 2010 *Prostate Cancer Imaging. Computer-Aided Diagnosis, Prognosis, and Intervention: International Workshop, Held in Conjunction with MICCAI 2010, Beijing,China, September 24, 2010. Proceedings,* ed A Madabhushi*, et al.* (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 42-51

Acosta O, Drean G, Ospina J D, Simon A, Haigron P, Lafond C and de Crevoisier R 2013 Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy *Phys Med Biol* **58** 2581-95

Acosta O, Mylona E, Le Dain M, Voisin C, Lizee T, Rigaud B, Lafond C, Gnep K and de Crevoisier R 2017 Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy *Radioth Oncol* **125** 492-9

Alterio D, Gerardi M A, Cella L, Spoto R, Zurlo V, Sabbatini A, Fodor C, D'Avino V, Conson M, Valoriani F, Ciardo D, Pacelli R, Ferrari A, Maisonneuve P, Preda L, Bruschini R, Cossu Rocca M, Rondi E, Colangione S, Palma G, Dicuonzo S, Orecchia R, Sanguineti G and Jereczek-Fossa B A 2017 Radiation-induced acute dysphagia : Prospective observational study on 42 head and neck cancer patients *Strahlenther Onkol* **193** 971-81

Arslan S, Ktena S I, Makropoulos A, Robinson E C, Rueckert D and Parisot S 2018 Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex *NeuroImage* **170** 5-30

Ashburner J and Friston K J 2000 Voxel-Based Morphometry—The Methods *NeuroImage* **11** 805-21

Baumann M, Krause M, Overgaard J, Debus J, Bentzen S M, Daartz J, Richter C, Zips D and Bortfeld T 2016 Radiation oncology in the era of precision medicine *Nat Rev Cancer* **16** 234-49

Beasley W, Thor M, McWilliam A, Green A, Mackay R, Slevin N, Olsson C, Pettersson N, Finizia C, Estilo C, Riaz N, Lee N Y, Deasy J O and van Herk M 2018 Image-based Data Mining to Probe Dosimetric Correlates of Radiation-induced Trismus *Int J Radiat Oncol Biol Phys* **102** 1330-8

Benjamini Y and Hochberg Y 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing *J Royal Stat Soc Ser B* **57** 289-300

Benjamini Y and Yekutieli D 2001 The Control of the False Discovery Rate in Multiple Testing under Dependency *Annal Stat* **29** 1165-88

Bentzen S M, Constine L S, Deasy J O, Eisbruch A, Jackson A, Marks L B, Ten Haken R K and Yorke E D 2010 Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues *Int J Radiat Oncol Biol Phys* **76** S3-S9

Bentzen S M, Dorr W, Gahbauer R, Howell R W, Joiner M C, Jones B, Jones D T, van der Kogel A J, Wambersie A and Whitmore G 2012 Bioeffect modeling and equieffective dose concepts in radiation oncology--terminology, quantities and units *Radioth Oncol* **105** 266-8

Bijl H P, van Luijk P, Coppes R P, Schippers J M, Konings A W and van der Kogel A J 2003 Unexpected changes of rat cervical spinal cord tolerance caused by inhomogeneous dose distributions *Int J Radiat Oncol Biol Phys* **57** 274-81

Bradburn M J, Clark T G, Love S B and Altman D G 2003 Survival analysis part II: multivariate data analysis--an introduction to concepts and methods *Br J Cancer* **89** 431-6

Buettner F, Gulliford S L, Webb S and Partridge M 2009a Using dose-surface maps to predict radiation-induced rectal bleeding: a neural network approach *Phys Med Biol* **54** 5139

Buettner F, Gulliford S L, Webb S and Partridge M 2011 Modeling late rectal toxicities based on a parameterized representation of the 3D dose distribution *Phys Med Biol* **56** 2103-18

Buettner F, Gulliford S L, Webb S, Sydes M R, Dearnaley D P and Partridge M 2009b Assessing correlations between the spatial distribution of the dose to the rectal wall and late rectal toxicity after prostate radiotherapy: an analysis of data from the MRC RT01 trial (ISRCTN 47772397) *Phys Med Biol* **54** 6535-48

Buettner F, Gulliford S L, Webb S, Sydes M R, Dearnaley D P and Partridge M 2012a The dose-response of the anal sphincter region--an analysis of data from the MRC RT01 trial *Radioth Oncol* **103** 347-52

Buettner F, Miah A B, Gulliford S L, Hall E, Harrington K J, Webb S, Partridge M and Nutting C M 2012b Novel approaches to improve the therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised phase III trial *Radioth Oncol* **103** 82-7

Bzdok D 2017 Classical Statistics and Statistical Learning in Imaging Neuroscience *Front Neurosci* **11** 543

Carillo V, Cozzarini C, Chietera A, Perna L, Gianolini S, Maggio A, Botti A, Rancati T, Valdagni R and Fiorino C 2012 Correlation between surrogates of bladder dosimetry and dose-volume histograms of the bladder wall defined on MRI in prostate cancer radiotherapy *Radioth Oncol* **105** 180-3

Casares-Magaz O, Bülow S, Pettersson N J, Moiseenko V, Pedersen J, Thor M, Einck J, Hopper A, Knopp R and Muren L P 2019 High accumulated doses to the inferior rectum are associated with late gastro-intestinal toxicity in a case-control study of prostate cancer patients treated with radiotherapy *Acta oncologica* **58** 1543-6

Casares-Magaz O, Muren L P, Moiseenko V, Petersen S E, Pettersson N J, Høyer M, Deasy J O and Thor M 2017 Spatial rectal dose/volume metrics predict patient-reported gastro-intestinal symptoms after radiotherapy for prostate cancer *Acta oncologica* **56** 1507-13

Cella L, D'Avino V, Palma G, Conson M, Liuzzi R, Picardi M, Pressello M C, Boboc G I, Battistini R, Donato V and Pacelli R 2015 Modeling the risk of radiation-induced lung fibrosis: Irradiated heart tissue is as important as irradiated lung *Radioth Oncol* **117** 36-43

Chao M, Wei J, Lo Y-C and Peñagarícano J A 2020 Dose cluster model parameterization of the parotid gland in irradiation of head and neck cancer *Phys Eng Sci Med* **43** 143-53

Chao M, Wei J, Lo Y C and Penagaricano J A 2019 Percolation Based Cluster Models Fully Incorporating Spatial Dose Distribution in Assessment of Parotid Gland Radiation Induced Complication in Head and Neck Cancer Treatment *Int J Radiat Oncol Biol Phys* **105** S169-S70

Chao M, Wei J, Narayanasamy G, Yuan Y, Lo Y-C and Peñagarícano J A 2018 Three-dimensional cluster formation and structure in heterogeneous dose distribution of intensity modulated radiation therapy *Radioth Oncol* **127** 197-205

Chen B, Acosta O, Kachenoura A, Ospina Arango J, Dréan G, Simon A, Bellanger J-J, Haigron P and Crevoisier R 2011 Spatial Characterization and Classification of Rectal Bleeding in Prostate Cancer Radiotherapy with a Voxel-Based Principal Components Analysis Model for 3D Dose Distribution. In: *14th Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions - International Workshop, Held in Conjunction with MICCAI · MICCAI 2011,* ed A Madabhushi*, et al.* pp 60-9

Chen C, Witte M, Heemsbergen W and Herk M v 2013 Multiple comparisons permutation test for image based data mining in radiotherapy *Radiat Oncol* **8** 293-

Chen J, Chen H, Zhong Z, Wang Z, Hrycushko B, Zhou L, Jiang S, Albuquerque K, Gu X and Zhen X 2018 Investigating rectal toxicity associated dosimetric features with deformable accumulated rectal surface dose maps for cervical cancer radiotherapy *Radiat Oncol* **13** 125-

Chen Y W, Mahal B A, Muralidhar V, Nezolosky M, Beard C J, Den R B, Feng F Y, Hoffman K E, Martin N E, Orio P F and Nguyen P L 2016 Association Between Treatment at a High-Volume Facility and Improved Survival for Radiation-Treated Men With High-Risk Prostate Cancer *Int J Radiat Oncol Biol Phys* **94** 683-90

Cheng C W and Das I J 1999 Treatment plan evaluation using dose-volume histogram (DVH) and spatial dose-volume histogram (zDVH) *Int J Radiat Oncol Biol Phys* **43** 1143-50

Cicchetti A, Laurino F, Possenti L, Rancati T and Zunino P 2020 In silico model of the early effects of radiation therapy on the microcirculation and the surrounding tissues *Phys Med* **73** 125-34

Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L and Prior F 2013 The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository *J Digit Imaging* **26** 1045-57

Collins G S, Reitsma J B, Altman D G and Moons K G M 2015 Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement *BMC Medicine* **13** 1

Coloigner J, Fargeas A, Kachenoura A, Wang L, Dréan G, Lafond C, Senhadji L, Crevoisier R d, Acosta O and Albera L 2015 A Novel Classification Method for Prediction of Rectal Bleeding in Prostate Cancer Radiotherapy Based on a Semi-Nonnegative ICA of 3D Planned Dose Distributions *IEEE J Biomed Health Inform* **19** 1168-77

Dankers F, Wijsman R, Troost E G, Monshouwer R, Bussink J and Hoffmann A L 2017 Esophageal wall dose-surface maps do not improve the predictive performance of a multivariable NTCP model for acute esophageal toxicity in advanced stage NSCLC patients treated with intensity-modulated (chemo-)radiotherapy *Phys Med Biol* **62** 3668-81

Dean J, Wong K, Gay H, Welsh L, Jones A-B, Schick U, Oh J H, Apte A, Newbold K, Bhide S, Harrington K, Deasy J, Nutting C and Gulliford S 2018 Incorporating spatial dose metrics in machine learning-based normal tissue complication probability (NTCP) models of severe acute dysphagia resulting from head and neck radiotherapy *Clin Transl Radiat Oncol* **8** 27-39

Dean J A, Welsh L C, Wong K H, Aleksic A, Dunne E, Islam M R, Patel A, Patel P, Petkar I, Phillips I, Sham J, Schick U, Newbold K L, Bhide S A, Harrington K J, Nutting C M and Gulliford S L 2017 Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk *Clin Oncol (R Coll Radiol)* **29** 263-73

Dean J A, Wong K H, Welsh L C, Jones A B, Schick U, Newbold K L, Bhide S A, Harrington K J, Nutting C M and Gulliford S L 2016 Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy *Radioth Oncol* **120** 21-7

Deasy J and Adita A 2013 *Informatics in Radiation Oncology,* ed G Starkschall and R A Siochi (Bosa Roca: CRC Press Inc)

Deasy J O, Bentzen S M, Jackson A, Ten Haken R K, Yorke E D, Constine L S, Sharma A and Marks L B 2010 Improving normal tissue complication probability models: the need to adopt a "data-pooling" culture *Int J Radiat Oncol, Biol, Phys* **76** S151-S4

Deasy J O, Blanco A I and Clark V H 2003 CERR: a computational environment for radiotherapy research *Med Phys* **30** 979-85

Deasy J O and El Naqa I 2008 Image-based modeling of normal tissue complication probability for radiation therapy *Cancer Treat Res* **139** 215-56

Defraene G, Van den Bergh L, Al-Mamgani A, Haustermans K, Heemsbergen W, Van den Heuvel F and Lebesque J V 2012 The Benefits of Including Clinical Factors in Rectal Normal Tissue Complication Probability Modeling After Radiotherapy for Prostate Cancer *Int J Radiat Oncol, Biol, Phys* **82** 1233-42

Deist T M, Jochems A, van Soest J, Nalbantov G, Oberije C, Walsh S, Eble M, Bulens P, Coucke P, Dries W, Dekker A and Lambin P 2017 Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT *Clin Transl Radiat Oncol* **4** 24-31

Dörr W 2015 Radiobiology of tissue reactions *Annals of the ICRP* **44** 58-68

Drean G, Acosta O, Lafond C, Simon A, de Crevoisier R and Haigron P 2016a Interindividual registration and dose mapping for voxelwise population analysis of rectal toxicity in prostate cancer radiotherapy *Med Phys* **43** 2721-30

Drean G, Acosta O, Ospina J D, Fargeas A, Lafond C, Correge G, Lagrange J L, Crehange G, Simon A, Haigron P and de Crevoisier R 2016b Identification of a rectal subregion highly predictive of rectal bleeding in prostate cancer IMRT *Radioth Oncol* **119** 388-97

Dréan G, Acosta O, Ospina J D, Voisin C, Rigaud B, Simon A, Haigron P and de Crevoisier R 2013 How to identify rectal sub-regions likely involved in rectal bleeding in prostate cancer radiotherapy. In: *IX International Seminar on Medical Information Processing and Analysis,* ed SPIE (Mexico DF: SPIE) p 9

Ebert M A, Bulsara M, Haworth A, Kearvell R, Richardson S, Kennedy A, Spry N A, Bydder S A, Joseph D J and Denham J W 2015a 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* **59** 99-108

Ebert M A, Foo K, Haworth A, Gulliford S L, Kennedy A, Joseph D J and Denham J W 2015b 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* **91** 595-603

Ebert M A, Harrison K M, Howlett S J, Cornes D, Bulsara M, Hamilton C S, Kron T, Joseph D J and Denham J W 2011 Dosimetric intercomparison for multicenter clinical trials using a patient-based anatomic pelvic phantom *Med Phys* **38** 5167-75

Ebert M A, Haworth A, Kearvell R, Hooton B, Hug B, Spry N A, Bydder S A and Joseph D J 2010 Comparison of DVH data from multiple radiotherapy treatment planning systems *Phys Med Biol* **55** N337-N46

El Naqa I, Bradley J D, Lindsay P E, Hope A J and Deasy J O 2009 Predicting radiotherapy outcomes using statistical learning techniques *Phys Med Biol* **54** S9-s30

El Naqa I and Das S 2020 The role of machine and deep learning in modern medical physics *Med Phys* **47** e125-e6

Emami B, Lyman J, Brown A, Coia L, Goitein M, Munzenrider J E, Shank B, Solin L J and Wesson M 1991 Tolerance of normal tissue to therapeutic irradiation *Int J Radiat Oncol, Biol, Phys* **21** 109-22

Esplen N M, Mendonca M S and Bazalova-Carter M 2020 Physics and biology of ultrahigh dose-rate (FLASH) radiotherapy: a topical review *Phys Med Biol*

Fargeas A, Acosta O, Ospina Arrango J D, Ferhat A, Costet N, Albera L, Azria D, Fenoglietto P, Créhange G, Beckendorf V, Hatt M, Kachenoura A and de Crevoisier R 2018 Independent component analysis for rectal bleeding prediction following prostate cancer radiotherapy *Radioth Oncol* **126** 263-9

Fargeas A, Kachenoura A, Acosta O, Albera L, Drean G and De Crevoisier R 2013 Feature extraction and classification for rectal bleeding in prostate cancer radiotherapy: A PCA based method *IRBM* **34** 296-9

Fiorino C, Gianolini S and Nahum A E 2003 A cylindrical model of the rectum: comparing dose–volume, dose–surface and dose–wall histograms in the radiotherapy of prostate cancer *Phys Med Biol* **48** 2603-16

Fiorino C, Vavassori V, Sanguineti G, Bianchi C, Cattaneo G M, Piazzolla A and Cozzarini C 2002 Rectum contouring variability in patients treated for prostate cancer: impact on rectum dose-volume histograms and normal tissue complication probability *Radioth Oncol* **63** 249-55

Fu Y, Lei Y, Wang T, Curran W J, Liu T and Yang X 2020 Deep learning in medical image registration: a review *Phys Med Biol* **65** 20TR01

Gabryś H S, Buettner F, Sterzing F, Hauswald H and Bangert M 2018 Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia *Front Oncol* **8** 35

Gale N, House M and Ebert M A 2017 Using percolation networks to incorporate spatial-dose information for assessment of complication probability in radiotherapy *Australas Phys Eng Sci Med* **40** 869-80

Ghadjar P, Zelefsky M J, Spratt D E, Munck af Rosenschöld P, Oh J H, Hunt M, Kollmeier M, Happersett L, Yorke E, Deasy J O and Jackson A 2014 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* **88** 339-44

Ghobadi G, van der Veen S, Bartelds B, de Boer R A, Dickinson M G, de Jong J R, Faber H, Niemantsverdriet M, Brandenburg S, Berger R M, Langendijk J A, Coppes R P and van Luijk P 2012 Physiological interaction of heart and lung in thoracic irradiation *Int J Radiat Oncol Biol Phys* **84** e639-46

Green A, Vasquez Osorio E, Aznar M C, McWilliam A and van Herk M 2020 Image Based Data Mining Using Per-voxel Cox Regression *Front Oncol* **10** 1178

Gulliford S L, Ghose S, Ebert M A, Kennedy A, Dowling J, Mitra J, Joseph D J and Denham J W 2017 Radiotherapy dose-distribution to the perirectal fat space (PRS) is related to gastrointestinal control-related complications *Clin Transl Radiat Oncol* **7** 62-70

Gulliford S L, Webb S, Rowbottom C G, Corne D W and Dearnaley D P 2004 Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate *Radioth Oncol* **71** 3-12

Guyon I and Elisseeff A 2003 An introduction to variable and feature selection *J. Mach. Learn. Res.* **3** 1157–82

Hayman J A, Dekker A, Feng M, Keole S R, McNutt T R, Machtay M, Martin N E, Mayo C S, Pawlicki T, Smith B D, Kudner R, Dawes S and Yu J B 2019 Minimum Data Elements for Radiation Oncology: An American Society for Radiation Oncology Consensus Paper *Pract Radiat Oncol* **9** 395-401

Heemsbergen W D, Al-Mamgani A, Witte M G, van Herk M, Pos F J and Lebesque J V 2010 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* **78** 19-25

Heemsbergen W D, Hoogeman M S, Hart G A M, Lebesque J V and Koper P C M 2005 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* **61** 1011-8

Heemsbergen W D, Incrocci L, Pos F J, Heijmen B J M and Witte M G 2020 Local Dose Effects for Late Gastrointestinal Toxicity After Hypofractionated and Conventionally Fractionated Modern Radiotherapy for Prostate Cancer in the HYPRO Trial *Front Oncol* **10**

Heinze G, Wallisch C and Dunkler D 2018 Variable selection - A review and recommendations for the practicing statistician *Biom J* **60** 431-49

Henderson D R, Murray J R, Gulliford S L, Tree A C, Harrington K J and Van As N J 2018 An Investigation of Dosimetric Correlates of Acute Toxicity in Prostate Stereotactic Body Radiotherapy: Dose to Urinary Trigone is Associated with Acute Urinary Toxicity *Clin Oncol (R Coll Radiol)* **30** 539-47

Hoogeman M S, Peeters S T, de Bois J and Lebesque J V 2005 Absolute and relative dose-surface and dose-volume histograms of the bladder: which one is the most representative for the actual treatment? *Phys Med Biol* **50** 3589-97

Hoogeman M S, van Herk M, de Bois J, Muller-Timmermans P, Koper P C M and Lebesque J V 2004 Quantification of local rectal wall displacements by virtual rectum unfolding *Radioth Oncol* **70** 21-30

Hrycushko B, van der Kogel A J, Phillips L, Folkert M R, Sayre J W, Vernino S, Hassan-Rezaeian N, Foster R D, Yamada Y, Timmerman R and Medin P M 2019 Spinal Nerve Tolerance to Single-Session Stereotactic Ablative Radiation Therapy *Int J Radiat Oncol Biol Phys* **104** 845-51

Ibragimov B, Toesca D, Chang D, Yuan Y, Koong A and Xing L 2018 Development of deep neural network for individualized hepatobiliary toxicity prediction after liver SBRT *Med Phys* **45** 4763-74

Ibragimov B, Toesca D A S, Chang D T, Yuan Y, Koong A C, Xing L and Vogelius I R 2020 Deep learning for identification of critical regions associated with toxicities after liver stereotactic body radiation therapy *Med Phys* **47** 3721-31

Ibragimov B, Toesca D A S, Yuan Y, Koong A C, Chang D T and Xing L 2019 Neural Networks for Deep Radiotherapy Dose Analysis and Prediction of Liver SBRT Outcomes *IEEE J Biomed Health Inform* **23** 1821-33

Improta I, Palorini F, Cozzarini C, Rancati T, Avuzzi B, Franco P, Degli Esposti C, Del Mastro E, Girelli G, Iotti C, Vavassori V, Valdagni R and Fiorino C 2016 Bladder spatial-dose descriptors correlate with acute urinary toxicity after radiation therapy for prostate cancer *Phys Med* **32** 1681-9

Jackson A, Marks L B, Bentzen S M, Eisbruch A, Yorke E D, Ten Haken R K, Constine L S and Deasy J O 2010 The lessons of QUANTEC: recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome *Int J Radiat Oncol, Biol, Phys* **76** S155-S60

Jaffray D A, Lindsay P E, Brock K K, Deasy J O and Tomé W A 2010 Accurate accumulation of dose for improved understanding of radiation effects in normal tissue *Int J Radiat Oncol, Biol, Phys* **76** S135-S9

Jiang W, Lakshminarayanan P, Hui X, Han P, Cheng Z, Bowers M, Shpitser I, Siddiqui S, Taylor R H, Quon H and McNutt T 2019 Machine Learning Methods Uncover Radiomorphologic Dose Patterns in Salivary Glands that Predict Xerostomia in Patients with Head and Neck Cancer *Adv Radiat Oncol* **4** 401-12

Johnson-Hart C N, Price G J, Faivre-Finn C, Aznar M C and van Herk M 2018 Residual Setup Errors Towards the Heart After Image Guidance Linked With Poorer Survival in Lung Cancer Patients: Do We Need Stricter IGRT Protocols? *Int J Radiat Oncol Biol Phys* **102** 434-42

Källman P, Lind B K and Brahme A 1992 An algorithm for maximizing the probability of complication-free tumour control in radiation therapy *Phys Med Biol* **37** 871-90

Kennedy A, Dowling J, Greer P B, Holloway L, Jameson M G, Roach D, Ghose S, Rivest-Henault D, Marcello M and Ebert M A 2019 Similarity clustering-based atlas selection for pelvic CT image segmentation *Med Phys* **46** 2243-50

Kim D W, Cho L C, Straka C, Christie A, Lotan Y, Pistenmaa D, Kavanagh B D, Nanda A, Kueplian P, Brindle J, Cooley S, Perkins A, Raben D, Xie X J and Timmerman R D 2014 Predictors of rectal tolerance observed in a dose-escalated phase 1-2 trial of stereotactic body radiation therapy for prostate cancer *Int J Radiat Oncol Biol Phys* **89** 509-17

Kim K-H, Chung J-B, Suh T S, Kang S-W, Kang S-H, Eom K-Y, Song C, Kim I-A and Kim J-S 2018 Dosimetric and radiobiological comparison in different dose calculation grid sizes between Acuros XB and anisotropic analytical algorithm for prostate VMAT *PLOS ONE* **13** e0207232

Kirkpatrick J P, van der Kogel A J and Schultheiss T E 2010 Radiation Dose–Volume Effects in the Spinal Cord *Int J Radiat Oncol, Biol, Phys* **76** S42-S9

Krauss A 2018 Why all randomised controlled trials produce biased results *Annals of Medicine* **50** 312-22

Kruschke J K 2013 Bayesian estimation supersedes the t test *J Exp Psychol Gen* **142** 573-603

La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, Lomax A J and Widesott L 2012 Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer *Radiat Oncol* **7** 160-

Lafond C, Barateau A, N'Guessan J, Perichon N, Delaby N, Simon A, Haigron P, Mylona E, Acosta O and de Crevoisier R 2020 Planning With Patient-Specific Rectal Sub-Region Constraints Decreases Probability of Toxicity in Prostate Cancer Radiotherapy *Front Oncol* **10** 1597

Lee S H, Han P, Hales R, Voong K R, Noro K, Sugiyama S, Haller J W, McNutt T and Lee J 2020 Multi-view radiomics and dosiomics analysis with machine learning for predicting acute-phase weight loss in lung cancer patients treated with radiotherapy *Phys Med Biol*

Lee S J and Park H J 2020 Single photon emission computed tomography (SPECT) or positron emission tomography (PET) imaging for radiotherapy planning in patients with lung cancer: a meta-analysis *Scientific Reports* **10** 14864

Liang B, Tian Y, Chen X, Yan H, Yan L, Zhang T, Zhou Z, Wang L and Dai J 2020 Prediction of Radiation Pneumonitis With Dose Distribution: A Convolutional Neural Network (CNN) Based Model *Front Oncol* **9**

Liang B, Yan H, Tian Y, Chen X, Yan L, Zhang T, Zhou Z, Wang L and Dai J 2019 Dosiomics: Extracting 3D Spatial Features From Dose Distribution to Predict Incidence of Radiation Pneumonitis *Front Oncol* **9**

Lu Y, Li S, Spelbring D, Song P, Vijayakumar S, Pelizzari C and Chen G T Y 1995 Dose-surface histograms as treatment planning tool for prostate conformal therapy *Med Phys* **22** 279-84

Lu Y, Spelbring D R and Chen G T Y 1997 Functional dose - volume histograms for functionally heterogeneous normal organs *Phys Med Biol* **42** 345-56

Lumley T, Diehr P, Emerson S and Chen L 2002 The importance of the normality assumption in large public health data sets *Annu Rev Public Health* **23** 151-69

Luo Y, Chen S and Valdes G 2020 Machine learning for radiation outcome modeling and prediction *Med Phys* **47** e178-e84

Lyman J T 1985 Complication Probability as Assessed from Dose-Volume Histograms *Rad Res* **8** 13-9

Magallon-Baro A, Loi M, Milder M T W, Granton P V, Zolnay A G, Nuyttens J J and Hoogeman M S 2019 Modeling daily changes in organ-at-risk anatomy in a cohort of pancreatic cancer patients *Radioth Oncol* **134** 127-34

Manly B F J 1997 *Randomization, Bootstrap and Monte Carlo Methods in Biology* (London: Chapman and Hall)

Marcello M, Denham J W, Kennedy A, Haworth A, Steigler A, Greer P, Holloway L, Dowling J, Jameson M, Roach D, Joseph D J, Gulliford S L, Dearnaley D P, Sydes M R, Hall E and Ebert M A 2020a Increased dose to organs in urinary tract associates with measures of genitourinary toxicity in pooled voxel-based analysis of 3 randomized Phase III trials *Front Oncol* **10**

Marcello M, Denham J W, Kennedy A, Haworth A, Steigler A, Greer P B, Holloway L C, Dowling J A, Jameson M G, Roach D, Joseph D J, Gulliford S L, Dearnaley D P, Sydes M R, Hall E and Ebert M A 2020b Relationships between rectal and perirectal doses and rectal bleeding or tenesmus in pooled voxel-based analysis of 3 randomised phase III trials *Radiotherapy and Oncology* **150** 281-92

Marks L B, Yorke E D, Jackson A, Ten Haken R K, Constine L S, Eisbruch A, Bentzen S M, Nam J and Deasy J O 2010 Use of normal tissue complication probability models in the clinic *Int J Radiat Oncol, Biol, Phys* **76** S10-S9

Mayo C S, Moran J M, Bosch W, Xiao Y, McNutt T, Popple R, Michalski J, Feng M, Marks L B, Fuller C D, Yorke E, Palta J, Gabriel P E, Molineu A, Matuszak M M, Covington E, Masi K, Richardson S L, Ritter T, Morgas T, Flampouri S, Santanam L, Moore J A, Purdie T G, Miller R C, Hurkmans C, Adams J, Jackie Wu Q R, Fox C J, Siochi R A, Brown N L, Verbakel W, Archambault Y, Chmura S J, Dekker A L, Eagle D G, Fitzgerald T J, Hong T, Kapoor R, Lansing B, Jolly S, Napolitano M E, Percy J, Rose M S, Siddiqui S, Schadt C, Simon W E, Straube W L, St James S T, Ulin K, Yom S S and Yock T I 2018 American Association of Physicists in Medicine Task Group 263: Standardizing Nomenclatures in Radiation Oncology *Int J Radiat Oncol Biol Phys* **100** 1057-66

McWilliam A, Dootson C, Graham L, Banfill K, Abravan A and van Herk M 2020 Dose surface maps of the heart can identify regions associated with worse survival for lung cancer patients treated with radiotherapy *Phys Imag Radiat Oncol* **15** 46-51

McWilliam A, Kennedy J, Hodgson C, Vasquez Osorio E, Faivre-Finn C and van Herk M 2017 Radiation dose to heart base linked with poorer survival in lung cancer patients *Eur J Can* **85** 106-13

Medin P M and Boike T P 2011 Spinal cord tolerance in the age of spinal radiosurgery: lessons from preclinical studies *Int J Radiat Oncol, Biol, Phys* **79** 1302-9

Meijer G J, van den Brink M, Hoogeman M S, Meinders J and Lebesque J V 1999 Dose-wall histograms and normalized dose-surface histograms for the rectum: a new method to analyze the dose distribution over the rectum in conformal radiotherapy *Int J Radiat Oncol Biol Phys* **45** 1073-80

Men K, Geng H, Zhong H, Fan Y, Lin A and Xiao Y 2019 A Deep Learning Model for Predicting Xerostomia Due to Radiation Therapy for Head and Neck Squamous Cell Carcinoma in the RTOG 0522 Clinical Trial *Int J Radiat Oncol Biol Phys* **105** 440-7

Meroni S, Cavatorta C, Barra S, Cavagnetto F, Scarzello G, Scaggion A, Pecori E, Diletto B, Alessandro O, Massimino M, Gianolini S, Pignoli E and Gandola L 2019 A dedicated cloud system for real-time upfront quality assurance in pediatric radiation therapy *Strahlenther Onkol* **195** 843-50

Michalski J M, Gay H, Jackson A, Tucker S L and Deasy J O 2010 Radiation dose-volume effects in radiation-induced rectal injury *Int J Radiat Oncol, Biol, Phys* **76** S123-S9

Molineu A, Hernandez N, Nguyen T, Ibbott G and Followill D 2013 Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom *Med Phys* **40** 022101

Monti S, Pacelli R, Cella L and Palma G 2018 Inter-patient image registration algorithms to disentangle regional dose bioeffects *Scientific Reports* **8** 4915

Monti S, Paganelli C, Buizza G, Preda L, Valvo F, Baroni G, Palma G and Cella L 2020 A novel framework for spatial normalization of dose distributions in voxel-based analyses of brain irradiation outcomes *Phys Med* **69** 164-9

Monti S, Palma G, D'Avino V, Gerardi M, Marvaso G, Ciardo D, Pacelli R, Jereczek-Fossa B A, Alterio D and Cella L 2017 Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients *Scientific Reports* **7** 7220

Morimoto M, Bijl H P, A V D S, Xu C J, Steenbakkers R, Chouvalova O, Yoshioka Y, Teshima T and Langendijk J A 2019 Development of Normal Tissue Complication Probability Model for Trismus in Head and Neck Cancer Patients Treated With Radiotherapy: The Role of Dosimetric and Clinical Factors *Anticancer Res* **39** 6787-98

Moulton C R, House M J, Lye V, Tang C I, Krawiec M, Joseph D J, Denham J W and Ebert M A 2017 Spatial features of dose–surface maps from deformably-registered plans correlate with late gastrointestinal complications *Phys Med Biol* **62** 4118

Munbodh R, Jackson A, Bauer J, Schmidtlein C R and Zelefsky M J 2008 Dosimetric and anatomic indicators of late rectal toxicity after high-dose intensity modulated radiation therapy for prostate cancer *Med Phys* **35** 2137-50

Murphy K, Ginneken B v, Reinhardt J M, Kabus S, Ding K, Deng X, Cao K, Du K, Christensen G E, Garcia V, Vercauteren T, Ayache N, Commowick O, Malandain G, Glocker B, Paragios N, Navab N, Gorbunova V, Sporring J, Bruijne M d, Han X, Heinrich M P, Schnabel J A, Jenkinson M, Lorenz C, Modat M, McClelland J R, Ourselin S, Muenzing S E A, Viergever M A, Nigris D D, Collins D L, Arbel T, Peroni M, Li R, Sharp G C, Schmidt-Richberg A, Ehrhardt J, Werner R, Smeets D, Loeckx D, Song G, Tustison N, Avants B, Gee J C, Staring M, Klein S, Stoel B C, Urschler M, Werlberger M, Vandemeulebroucke J, Rit S, Sarrut D and Pluim J P W 2011 Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge *IEEE Trans Med Imaging* **30** 1901-20

Myers C and Niemierko A 2004 Percolation-based cluster models of dose-volume effects *Int J Radiat Oncol Biol Phys* **60** S157

Mylona E, Acosta O, Lizee T, Lafond C, Crehange G, Magné N, Chiavassa S, Supiot S, Arango Ospina J D, Campillo-Gimenez B, Castelli J and de Crevoisier R 2019 Voxel-based analysis for identification of urethro-vesical subregions predicting urinary toxicity after prostate cancer radiotherapy *Int J Radiat Oncol Biol Phys* **104** 343-54

Mylona E, Cicchetti A, Rancati T, Palorini F, Fiorino C, Supiot S, Magne N, Crehange G, Valdagni R, Acosta O and de Crevoisier R 2020a Local dose analysis to predict acute and late urinary toxicities after prostate cancer radiotherapy: Assessment of cohort and method effects *Radioth Oncol* **147** 40-9

Mylona E, Ebert M, Kennedy A, Joseph D, Denham J, Steigler A, Supiot S, Acosta O and de Crevoisier R 2020b Rectal and Urethro-vesical Subregions for Toxicity Prediction After Prostate Cancer Radiotherapy: validation of voxel-based models in an independent population *Int J Radiat Oncol Biol Phys*

NEMA Digital Imaging and Communications in Medicine (DICOM) Standard. (Rosslyn, VA, USA: National Electrical Manufacturers Association)

NIfTI 2020 Neuroimaging Informatics Technology Initiative. Neuroimaging Informatics Technology Initiative)

Nioutsikou E, Webb S, Panakis N, Bortfeld T and Oelfke U 2005 Reconsidering the definition of a dose-volume histogram *Phys Med Biol* **50** L17-9

Nitsche M, Brannath W, Brückner M, Wagner D, Kaltenborn A, Temme N and Hermann R M 2017 Comparison of different contouring definitions of the rectum as organ at risk (OAR) and dose-volume parameters predicting rectal inflammation in radiotherapy of prostate cancer: which definition to use? *The British journal of radiology* **90** 20160370-

Ohri N, Shen X, Dicker A P, Doyle L A, Harrison A S and Showalter T N 2013 Radiotherapy protocol deviations and clinical outcomes: A meta-analysis of cooperative group clinical trials *J Natl Cancer Inst* **105** 387-93

Onjukka E, Fiorino C, Cicchetti A, Palorini F, Improta I, Gagliardi G, Cozzarini C, Degli Esposti C, Gabriele P, Valdagni R and Rancati T 2019 Patterns in ano-rectal dose maps and the risk of late toxicity after prostate IMRT *Acta oncologica* **58** 1757-64

Ospina J D, Zhu J, Chira C, Bossi A, Delobel J B, Beckendorf V, Dubray B, Lagrange J L, Correa J C, Simon A, Acosta O and de Crevoisier R 2014 Random forests to predict rectal toxicity following prostate cancer radiation therapy *Int J Radiat Oncol Biol Phys* **89** 1024-31

Palma G and Cella L 2019 A new formalism of Dose Surface Histograms for robust modeling of skin toxicity in radiation therapy *Phys Med* **59** 75-8

Palma G, Monti S, Buonanno A, Pacelli R and Cella L 2019a PACE: A Probabilistic Atlas for Normal Tissue Complication Estimation in Radiation Oncology *Front Oncol* **9** 130-

Palma G, Monti S and Cella L 2020a Voxel-based analysis in radiation oncology: A methodological cookbook *Phys Med* **69** 192-204

Palma G, Monti S, Conson M, Pacelli R and Cella L 2019b Normal tissue complication probability (NTCP) models for modern radiation therapy *Semin Oncol* **46** 210-8

Palma G, Monti S, D'Avino V, Conson M, Liuzzi R, Pressello M C, Donato V, Deasy J O, Quarantelli M, Pacelli R and Cella L 2016 A Voxel-Based Approach to Explore Local Dose Differences Associated With Radiation-Induced Lung Damage *Int J Radiat Oncol Biol Phys* **96** 127-33

Palma G, Monti S, Thor M, Rimner A, Deasy J O and Cella L 2019c Spatial signature of dose patterns associated with acute radiation-induced lung damage in lung cancer patients treated with stereotactic body radiation therapy *Phys Med Biol* **64** 155006

Palma G, Monti S, Xu T, Scifoni E, Yang P, Hahn S M, Durante M, Mohan R, Liao Z and Cella L 2019d Spatial dose patterns associated with radiation pneumonitis in a randomized trial comparing intensity-modulated photon therapy with passive scattering proton therapy for locally advanced non-small cell lung cancer *Int J Radiat Oncol Biol Phys* **104** 1124-32

Palma G, Taffelli A, Fellin F, D'Avino V, Scartoni D, Tommasino F, Scifoni E, Durante M, Amichetti M, Schwarz M, Amelio D and Cella L 2020b Modelling the risk of radiation induced alopecia in brain tumor patients treated with scanned proton beams *Radioth Oncol* **144** 127-34

Palorini F, Botti A, Carillo V, Gianolini S, Improta I, Iotti C, Rancati T, Cozzarini C and Fiorino C 2016a Bladder dose–surface maps and urinary toxicity: Robustness with respect to motion in assessing local dose effects *Phys Med* **32** 506-11

Palorini F, Cozzarini C, Gianolini S, Botti A, Carillo V, Iotti C, Rancati T, Valdagni R and Fiorino C 2016b First application of a pixel-wise analysis on bladder dose surface maps in prostate cancer radiotherapy *Radioth Oncol* **119** 123-8

Peduzzi P, Concato J, Kemper E, Holford T R and Feinstein A R 1996 A simulation study of the number of events per variable in logistic regression analysis *J Clin Epidem* **49** 1373-9

Peeters S T, Hoogeman M S, Heemsbergen W D, Hart A A, Koper P C and Lebesque J V 2006a Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling *Int J Radiat Oncol Biol Phys* **66** 11-9

Peeters S T, Hoogeman M S, Heemsbergen W D, Slot A, Tabak H, Koper P C and Lebesque J V 2005 Volume and hormonal effects for acute side effects of rectum and bladder during conformal radiotherapy for prostate cancer *Int J Radiat Oncol Biol Phys* **63** 1142-52

Peeters S T H, Lebesque J V, Heemsbergen W D, van Putten W L J, Slot A, Dielwart M F H and Koper P C M 2006b Localized volume effects for late rectal and anal toxicity after radiotherapy for prostate cancer *Int. J. Radiat. Oncol. Biol. Phys.* **64** 1151-61

Phillips M H, Serra L M, Dekker A, Ghosh P, Luk S M H, Kalet A and Mayo C 2020 Ontologies in radiation oncology *Phys Med* **72** 103-13

Placidi L, Lenkowicz J, Cusumano D, Boldrini L, Dinapoli N and Valentini V 2020 Stability of dosomics features extraction on grid resolution and algorithm for radiotherapy dose calculation *Phys Med* **77** 30-5

Purdy J A 2008 Quality assurance issues in conducting multi-institutional advanced technology clinical trials *Int J Radiat Oncol, Biol, Phys* **71** S66-70

Purdy J A, Harms W B, Michalski J and Bosch W R 1998 Initial experience with quality assurance of multi-institutional 3D radiotherapy clinical trials. A brief report *Strahlentherapie und Onkologie* **174 Suppl 2** 40-2

Rancati T, Fiorino C, Fellin G, Vavassori V, Cagna E, Casanova Borca V, Girelli G, Menegotti L, Monti A F, Tortoreto F, Delle Canne S and Valdagni R 2011 Inclusion of clinical risk factors into NTCP modelling of late rectal toxicity after high dose radiotherapy for prostate cancer *Radioth Oncol* **100** 124-30

Rancati T, Fiorino C, Gagliardi G, Cattaneo G M, Sanguineti G, Borca V C, Cozzarini C, Fellin G, Foppiano F, Girelli G, Menegotti L, Piazzolla A, Vavassori V and Valdagni R 2004 Fitting late rectal bleeding data using different NTCP models: results from an Italian multi-centric study (AIROPROS0101) *Radioth Oncol* **73** 21-32

Rancati T, Palorini F, Cozzarini C, Fiorino C and Valdagni R 2017 Understanding urinary toxicity after radiotherapy for prostate cancer: first steps forward *Tumori* **103** 395-404

Rigaud B, Simon A, Castelli J, Lafond C, Acosta O, Haigron P, Cazoulat G and de Crevoisier R 2019 Deformable image registration for radiation therapy: principle, methods, applications and evaluation *Acta oncologica* **58** 1225-37

Roach D, Holloway L C, Jameson M G, Dowling J A, Kennedy A, Greer P B, Krawiec M, Rai R, Denham J, De Leon J, Lim K, Berry M E, White R T, Bydder S A, Tan H T, Croker J D, McGrath A, Matthews J, Smeenk R J and Ebert M A 2019 Multi-observer contouring of male pelvic anatomy: Highly variable agreement across conventional and emerging structures of interest *J Med Imaging Radiat Oncol* **63** 264-71

Robertson S P, Quon H, Kiess A P, Moore J A, Yang W, Cheng Z, Afonso S, Allen M, Richardson M, Choflet A, Sharabi A and McNutt T R 2015 A data-mining framework for large scale analysis of dose-outcome relationships in a database of irradiated head and neck cancer patients *Med Phys* **42** 4329-37

Roelofs E, Dekker A, Meldolesi E, van Stiphout R G P M, Valentini V and Lambin P 2014 International data-sharing for radiotherapy research: an open-source based infrastructure for multicentric clinical data mining *Radioth Oncol* **110** 370-4

Rossi L, Bijman R, Schillemans W, Aluwini S, Cavedon C, Witte M, Incrocci L and Heijmen B 2018 Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy *Radioth Oncol* **129** 548-53

Ryu S, Jin J Y, Jin R, Rock J, Ajlouni M, Movsas B, Rosenblum M and Kim J H 2007 Partial volume tolerance of the spinal cord and complications of single-dose radiosurgery *Cancer* **109** 628-36

Saito T and Rehmsmeier M 2015 The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets *PLoS One* **10** e0118432

Sanchez-Nieto B, Fenwick J, Nahum A and Dearnaley D P 2001 Biological dose surface maps: evaluation of 3D dose data for tubular organs *Radioth Oncol* **61** S52

Santanam L, Hurkmans C, Mutic S, van Vliet-Vroegindeweij C, Brame S, Straube W, Galvin J, Tripuraneni P, Michalski J and Bosch W 2012 Standardizing naming conventions in radiation oncology *Int J Radiat Oncol Biol Phys* **83** 1344-9

Schaake W, van der Schaaf A, van Dijk L V, Bongaerts A H, van den Bergh A C and Langendijk J A 2016 Normal tissue complication probability (NTCP) models for late rectal bleeding, stool frequency and fecal incontinence after radiotherapy in prostate cancer patients *Radioth Oncol* **119** 381-7

Seppenwoolde Y, De Jaeger K, Boersma L J, Belderbos J S A and Lebesque J V 2004 Regional differences in lung radiosensitivity after radiotherapy for non–small-cell lung cancer *Int J Radiat Oncol Biol Phys* **60** 748-58

Shelley L E A, Scaife J E, Romanchikova M, Harrison K, Forman J R, Bates A M, Noble D J, Jena R, Parker M A, Sutcliffe M P F, Thomas S J and Burnet N G 2017 Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy *Radioth Oncol* **123** 466-71

Shmueli G 2010 To Explain or to Predict? *Statist. Sci.* **25** 289-310

Smeenk R J, Hoffmann A L, Hopman W P, van Lin E N and Kaanders J H 2012 Dose-effect relationships for individual pelvic floor muscles and anorectal complaints after prostate radiotherapy *Int J Radiat Oncol Biol Phys* **83** 636-44

Söhn M, Alber M and Yan D 2007 Principal Component Analysis-Based Pattern Analysis of Dose-Volume Histograms and Influence on Rectal Toxicity *Int J Radiat Oncol Biol Phys* **69** 230-9

SourceForge 2020 NRRD. SourceForge)

Stenmark M H, Conlon A S C, Johnson S, Daignault S, Litzenberg D, Marsh R, Ritter T, Vance S, Kazzi N, Feng F Y, Sandler H, Sanda M G and Hamstra D A 2014 Dose to the inferior rectum is strongly associated with patient reported bowel quality of life after radiation therapy for prostate cancer *Radioth Oncol* **110** 291-7

Steyerberg E W and Vergouwe Y 2014 Towards better clinical prediction models: seven steps for development and an ABCD for validation *Eur Heart J* **35** 1925-31

Stokke C, Gabiña P M, Solný P, Cicone F, Sandström M, Gleisner K S, Chiesa C, Spezi E, Paphiti M, Konijnenberg M, Aldridge M, Tipping J, Wissmeyer M, Brans B, Bacher K, Kobe C and Flux G 2017 Dosimetry-based treatment planning for molecular radiotherapy: a summary of the 2017 report from the Internal Dosimetry Task Force *EJNMMI Physics* **4** 27

Storey J D 2002 A Direct Approach to False Discovery Rates *J Royal Stat Soc B* **64** 479-98

Taichman D B, Sahni P, Pinborg A, Peiperl L, Laine C, James A, Hong S-T, Haileamlak A, Gollogly L, Godlee F, Frizelle F A, Florenzano F, Drazen J M, Bauchner H, Baethge C and Backus J 2017 Data Sharing Statements for Clinical Trials — A Requirement of the International Committee of Medical Journal Editors *NEJM* **376** 2277-9

Talamonti C, Piffer S, Greto D, Mangoni M, Ciccarone A, Dicarolo P, Fantacci M E, Fusi F, Oliva P, Palumbo L, Favre C, Livi L, Pallotta S and Retico A 2019 Radiomic and Dosiomic Profiling of Paediatric Medulloblastoma Tumours Treated with Intensity Modulated Radiation Therapy. (Cham: Springer International Publishing) pp 56-64

Thames H D, Zhang M, Tucker S L, Liu H H, Dong L and Mohan R 2004 Cluster models of dose-volume effects *Int J Radiat Oncol Biol Phys* **59** 1491-504

Tibshirani R 1996 Regression Shrinkage and Selection via the Lasso *J Royal Stat Soc Ser B* **58** 267-88

Tilly D, Tilly N and Ahnesjö A 2013 Dose mapping sensitivity to deformable registration uncertainties in fractionated radiotherapy - applied to prostate proton treatments *BMC Med Phys* **13** 2-

Tomatis S, Rancati T, Fiorino C, Vavassori V, Fellin G, Cagna E, Mauro F, Girelli G, Monti A and Baccolini M 2012 Late rectal bleeding after 3D-CRT for prostate cancer: development of a neural-network-based predictive model *Phys Med Biol* **57** 1399

Troeller A, Yan D, Marina O, Schulze D, Alber M, Parodi K, Belka C and Söhn M 2015 Comparison and limitations of DVH-based NTCP models derived from 3D-CRT and IMRT data for prediction of gastrointestinal toxicities in prostate cancer patients by using propensity score matched pair analysis *Int J Radiat Oncol Biol Phys* **91** 435-43

Trott K-R, Doerr W, Facoetti A, Hopewell J, Langendijk J, van Luijk P, Ottolenghi A and Smyth V 2012 Biological mechanisms of normal tissue damage: Importance for the design of NTCP models *Radioth Oncol* **105** 79-85

Trotti A, Colevas A D, Setser A, Rusch V, Jaques D, Budach V, Langer C, Murphy B, Cumberlin R, Coleman C N and Rubin P 2003 CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment *Semin Radiat Oncol* **13** 176-81

Tucker S L, Liao Z, Dinh J, Bian S X, Mohan R, Martel M K and Grosshans D R 2014 Is there an impact of heart exposure on the incidence of radiation pneumonitis? Analysis of data from a large clinical cohort *Acta oncologica* **53** 590-6

Tucker S L, Liu H H, Wang S, Wei X, Liao Z, Komaki R, Cox J D and Mohan R 2006a Dose-volume modeling of the risk of postoperative pulmonary complications among esophageal cancer patients treated with concurrent chemoradiotherapy followed by surgery *Int J Radiat Oncol Biol Phys* **66** 754-61

Tucker S L, Zhang M, Dong L, Mohan R, Kuban D and Thames H D 2006b Cluster model analysis of late rectal bleeding after IMRT of prostate cancer: A case-control study *Int J Radiat Oncol Biol Phys* **64** 1255-64

Valdes G and Interian Y 2018 Comment on 'Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study' *Phys Med Biol* **63** 068001

van der Schaaf A, Langendijk J A, Fiorino C and Rancati T 2015 Embracing Phenomenological Approaches to Normal Tissue Complication Probability Modeling: A Question of Method *Int J Radiat Oncol, Biol, Phys* **91** 468-71

van Luijk P, Novakova-Jiresova A, Faber H, Schippers J M, Kampinga H H, Meertens H and Coppes R P 2005 Radiation damage to the heart enhances early radiation-induced lung function loss *Cancer Res* **65** 6509-11

van Luijk P, Pringle S, Deasy J O, Moiseenko V V, Faber H, Hovan A, Baanstra M, van der Laan H P, Kierkels R G J, van der Schaaf A, Witjes M J, Schippers J M, Brandenburg S, Langendijk J A, Wu J and Coppes R P 2015 Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer *Sci Transl Med* **7** 305ra147-305ra147

Vanneste B G L, Buettner F, Pinkawa M, Lambin P and Hoffmann A L 2018 Ano-rectal wall dose-surface maps localize the dosimetric benefit of hydrogel rectum spacers in prostate cancer radiotherapy *Clin Transl Radiat Oncol* **14** 17-24

Vinogradskiy Y, Tucker S L, Liao Z and Martel M K 2012 A novel method to incorporate the spatial location of the lung dose distribution into predictive radiation pneumonitis modeling *Int J Radiat Oncol Biol Phys* **82** 1549-55

Vittinghoff E and McCulloch C E 2007 Relaxing the rule of ten events per variable in logistic and Cox regression *Am J Epidemiol* **165** 710-8

Voshart D C, Wiedemann J, van Luijk P and Barazzuol L 2021 Regional Responses in Radiation-Induced Normal Tissue Damage *Cancers (Basel)* **13** 367

Voutilainen A 2016 Spatial Objectives in Radiation Therapy Treatment Planning. In: *School of Science*: Aalto University)

Wang C, Zhu X, Hong J C and Zheng D 2019 Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future *Technology in Cancer Research & Treatment* **18** 1533033819873922

Weber D C, Vallet V, Molineu A, Melidis C, Teglas V, Naudy S, Moeckli R, Followill D S and Hurkmans C W 2014 IMRT credentialing for prospective trials using institutional virtual phantoms: results of a joint European Organization for the Research and Treatment of Cancer and Radiological Physics Center project *Radiat Oncol* **9** 123

Whitwell J L 2009 Voxel-based morphometry: an automated technique for assessing structural changes in the brain *J Neurosci* **29** 9661-4

Wilkins A, Naismith O, Brand D, Fernandez K, Hall E, Dearnaley D and Gulliford S 2020 Derivation of Dose/Volume Constraints for the Anorectum from Clinician- and Patient-Reported Outcomes in the CHHiP Trial of Radiation Therapy Fractionation *Int J Radiat Oncol, Biol, Phys* **106** 928-38

Wilkinson M D, Dumontier M, Aalbersberg I J, Appleton G, Axton M, Baak A, Blomberg N, Boiten J W, da Silva Santos L B, Bourne P E, Bouwman J, Brookes A J, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C T, Finkers R, Gonzalez-Beltran A, Gray A J, Groth P, Goble C, Grethe J S, Heringa J, t Hoen P A, Hooft R, Kuhn T, Kok R, Kok J, Lusher S J, Martone M E, Mons A, Packer A L, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M A, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J and Mons B 2016 The FAIR Guiding Principles for scientific data management and stewardship *Sci Data* **3** 160018

Witztum A, George B, Warren S, Partridge M and Hawkins M A 2016 Unwrapping 3D complex hollow organs for spatial dose surface analysis *Med Phys* **43** 6009

Wortel R C, Witte M G, van der Heide U A, Pos F J, Lebesque J V, van Herk M, Incrocci L and Heemsbergen W D 2015 Dose-surface maps identifying local dose-effects for acute gastrointestinal toxicity after radiotherapy for prostate cancer *Radioth Oncol* **117** 515-20

Wright J L, Yom S S, Awan M J, Dawes S, Fischer-Valuck B, Kudner R, Mailhot Vega R and Rodrigues G 2019 Standardizing Normal Tissue Contouring for Radiation Therapy Treatment Planning: An ASTRO Consensus Paper *Pract Radiat Oncol* **9** 65-72

Xiao C, Polomano R and Bruner D W 2013 Comparison between patient-reported and clinician-observed symptoms in oncology *Cancer Nurs* **36** E1-e16

Xu C J, van der Schaaf A, Van't Veld A A, Langendijk J A and Schilstra C 2012 Statistical validation of normal tissue complication probability models *Int J Radiat Oncol Biol Phys* **84** e123-9

Yahya N, Ebert M A, Bulsara M, House M J, Kennedy A, Joseph D J and Denham J W 2016 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* **43** 2040-52

Yahya N, Ebert M A, House M J, Kennedy A, Matthews J, Joseph D J and Denham J W 2017 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 Biol Phys* **97** 420-6

Zhen X, Chen J, Zhong Z, Hrycushko B, Zhou L, Jiang S, Albuquerque K and Gu X 2017 Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study *Phys Med Biol* **62** 8246-63

Zwanenburg A, Vallières M, Abdalah M A, Aerts H J W L, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga R J, Boellaard R, Bogowicz M, Boldrini L, Buvat I, Cook G J R, Davatzikos C, Depeursinge A, Desseroit M-C, Dinapoli N, Dinh C V, Echegaray S, Naqa I E, Fedorov A Y,

Gatta R, Gillies R J, Goh V, Götz M, Guckenberger M, Ha S M, Hatt M, Isensee F, Lambin P, Leger S, Leijenaar R T H, Lenkowicz J, Lippert F, Losnegård A, Maier-Hein K H, Morin O, Müller H, Napel S, Nioche C, Orlhac F, Pati S, Pfaehler E A G, Rahmim A, Rao A U K, Scherer J, Siddique M M, Sijtsema N M, Fernandez J S, Spezi E, Steenbakkers R J H M, Tanadini-Lang S, Thorwarth D, Troost E G C, Upadhaya T, Valentini V, Dijk L V v, Griethuysen J v, Velden F H P v, Whybra P, Richter C and Löck S 2020 The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping *Radiology* **295** 328-38