Radiotherapy and Oncology 188 (2023) 109868

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Review Article

Voxel-based analysis: Roadmap for clinical translation

Alan McWilliam [a,b,*,1], Giuseppe Palma [c,*,1], Azadeh Abravan [a,b], Oscar Acosta [d], Ane Appelt [e], Marianne Aznar [a,b], Serena Monti [f], Eva Onjukka [g], Vanessa Panettieri [h,i,j], Lorenzo Placidi [k], Tiziana Rancati [l], Eliana Vasquez Osorio [a,b], Marnix Witte [m], Laura Cella [f]

[a] The Division of Cancer Sciences, The University of Manchester; [b] The Christie NHS Foundation Trust, Manchester, UK; [c] Institute of Nanotechnology, National Research Council, Lecce, Italy; [d] University Rennes, CLCC Eugène Marquis, INSERM, LTSI - UMR 1099, F-35000, Rennes, France; [e] Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK; [f] Institute of Biostructures and Bioimaging, National Research Council, Naples, Italy; [g] Department of Medical Radiation Physics and Nuclear Medicine, Karolinska University Hospital, Sweden; [h] Department of Physical Sciences, Peter MacCallum Cancer Centre; [i] Central Clinical School, Monash University, Melbourne, VIC, Australia; [j] Sir Peter MacCallum Department of Oncology, The University of Melbourne, Victoria 3010, Australia; [k] Fondazione Policlinico Universitario "Agostino Gemelli" IRCCS, Rome; [l] Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy; [m] Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands

ARTICLE INFO

Article history:
Received 15 February 2023
Received in revised form 11 July 2023
Accepted 20 August 2023
Available online 6 September 2023

Keywords:
Radiotherapy
Voxel-based analysis
Outcome modelling
Validation
Clinical translation

ABSTRACT

Voxel-based analysis (VBA) allows the full, 3-dimensional, dose distribution to be considered in radiotherapy outcome analysis. This provides new insights into anatomical variability of pathophysiology and radiosensitivity by removing the need for a priori definition of organs assumed to drive the dose response associated with patient outcomes. This approach may offer powerful biological insights demonstrating the heterogeneity of the radiobiology across tissues and potential associations of the radiotherapy dose with further factors. As this methodological approach becomes established, consideration needs to be given to translating VBA results to clinical implementation for patient benefit. Here, we present a comprehensive roadmap for VBA clinical translation. Technical validation needs to demonstrate robustness to methodology, where clinical validation must show generalisability to external datasets and link to a plausible pathophysiological hypothesis. Finally, clinical utility requires demonstration of potential benefit for patients in order for successful translation to be feasible. For each step on the roadmap, key considerations are discussed and recommendations provided for best practice.

Voxel based analysis (VBA) is becoming an established analysis tool in radiotherapy (RT), allowing a deeper understanding of how the full 3-dimensional (3D) dose distribution affects patients' response to radiation. VBA includes the full, complex, dose distribution in the dose–response analysis, making it possible to move away from traditional approaches which reduce the dimensionality and typically distil the dose down to a limited number of representative values. In VBA, no prior assumption is made about anatomical dose dependence; this allows the dose to 'talk to us' directly, potentially identifying associations between local dose and patient outcomes. This approach becomes more crucial as biological insights, from in-vivo experiments, provide growing evidence that the radiosensitivity of several organs is heterogenous (e.g., due to some sub-structures being particularly radiosensitive) [1–3]. VBA can identify these heterogeneous dose sensitivities through defining 'clusters' of voxels in specific sub-regions of organs. It is therefore essential to move away from whole organ dose to VBA in order to map these insights into patient outcome modelling. It is worth noting that, although making no assumption on the homogeneity of organ response to radiation, VBA assumes the existence of a static anatomical location whose irradiation drives the outcome. This explains why the majority of the VBA literature focuses on normal tissue toxicity: Disease recurrence of anatomically varying tumours (such as lung or brain lesions in different positions) represents a dose–response relationship that cannot be studied with a classical VBA analysis due to the variability in recurrence sites.

VBA requires several technical steps with careful validation at each stage, each of which requires specific considerations. Additionally, given the complexity of the analysis, it is particularly important that such research is driven by clinically focused questions, with the ultimate aim to provide patient benefit. This requires engagement with multi-disciplinary clinical teams from the inception of any planned analysis. This engagement needs to guide the development of the study analysis plan to focus on

* Corresponding authors at: The University of Manchester c/o The Christie NHS Foundation Trust (dept 58), Wilmslow Road, Manchester M16 7QS, UK(A. McWilliam), Institute of Nanotechnology, National Research Council, Lecce, Italy(G. Palma).

E-mail addresses: alan.mcwilliam@manchester.ac.uk (A. McWilliam), giuseppe.palma@ibb.cnr.it (G. Palma).

[1] Joint first and corresponding authors.

https://doi.org/10.1016/j.radonc.2023.109868
0167-8140/

endpoints where clinical gains are needed. The increasing availability of radiotherapy planning archives, structured data collection in electronic patient notes, and linkage to registry data, makes it feasible for VBA to generate these novel results.

There have been significant methodological developments in VBA with numerous discovery studies performed. In section 2 we present an overview of the key methodological steps and summarise key papers using VBA in the thorax, pelvis, and the head and neck. The current collective of work in the VBA field makes the value of this methodology evident; and demonstrate the need to provide a roadmap to ensure results can be translated successfully into patient benefits. In this paper, we start from the position that VBA as a methodology has value and present guidance on how to fully report and validate results, as well as the steps needed to translate VBA results into the clinic for patient benefit.

Current use of VBA

Recently, there has been an increasing number of studies applying voxel-based methods in radiation oncology. Though some variation exists between the different implementations of VBA, the approach mainly consists of the two following processes.

1. spatial normalisation of individual anatomies to a common reference anatomy;
2. statistical inference on the spatial signature of dose response.

Table 1 includes a representative sample of the VBA literature which focuses on studies using an anatomically based common coordinate system (CCS) to drive the registration. Alternative approaches have been introduced which either implement dose mapping across the surface of an organ or define a structure driven coordinate system (supplementary table S1).

The technical aspects to consider have been previously described [4]. Fig. 1 reproduces the flow diagram by Palma et al. and describes the key steps required in successful VBA [4]. In this section, we highlight some of the key considerations, while recommendations for reporting standards will be introduced in the following sections.

Spatial normalisation

The first step of VBA involves the choice of an appropriate reference anatomy or common coordinate system (CCS), to which dose distributions from each individual will be mapped [5]. The definition of the optimal CCS must consider the tissue contrast of the available images (typically the planning CT) with several strategies presented in the literature:

- Using an average anatomy obtained by iterative registration and computing intensity means [6] or via group-wise registration [7]. Groupwise registration approaches have been used successfully to identify population representative CCS for paediatric studies [8].
- Identifying multiple clusters among the population dataset in order to repeat the voxel-based analysis on different templates [9,10].
- Selecting an optimal template, i.e., a typical individual anatomy, from the studied cohort or an external cohort [11–14]. The optimal patient can be selected through affinity propagation clustering [15,16] to identify the individual closest to all other anatomies. Alternatively, the CCS can be chosen by organ-at-risk mean parameters, for example bladder and urethral length [17] or lung volumes [18].
- Using a structural anatomical model as the CCS [19]. This approach may help overcome challenges with poor soft-tissue contrast and large inter-individual anatomic variability, for example in bladder and rectum [9].
- Using population atlases, such as synthetic CT phantoms [20–23]. This also acts to mask the lesion, minimising the impact of the tumour volume within the DIR.

Once the choice of the CCS has been established, inter-patient spatial normalisation needs to be performed to allow propagation of the associated dose maps onto the CCS for statistical inference. Many papers in Table 1 use deformable image registration (DIR) algorithms, with the most used transformation models being Demons and B-spline algorithms. Both algorithms have shown excellent performance in inter-patient DIR, with a robust match of anatomical structures in pelvis, head-and-neck (H&N), and thoracic regions. Both intensity-based DIR and structure driven DIR [4], or a combination, have been used in the VBA literature. Those strategies can all lead to good results with other imaging modalities than CT. For example, this approach has been successfully used in the brain to better characterise the cerebral anatomy [26].

The fundamental assumption is that the mapping to the CCS perfectly aligns tissues, organs, and sub-structures of organs across the patient population. This needs to be true for any imaging modality or multi-modality registration. This, realistically, never fully achievable; and, crucially, geometric uncertainties need to be quantified and accounted for in VBA. Section 4.2 describes approaches used in the literature, and presents recommendations.

Statistical inference

The basic strategy for statistical inference consists of testing, voxel-wise, the null hypothesis on local dose differences between groups of patients evaluated against a given outcome. The choice for the appropriate statistical model is dictated by the representation of the outcome to be analysed. Importantly, when any statistical inference is run across each voxel it is necessary to account for multiple comparisons, *i.e.*, it is likely that the null hypothesis was incorrectly rejected for some voxels (type I errors) [27]. Accounting for multiple comparisons within voxel-based dose analyse can be done in several ways, as described in 'The methodological cookbook' [4]. Focusing on RT-driven outcomes, VBA has been used to analyse binary events for survival [13] and toxicity [15,18], or continuous toxicity variables [28], by 2-sample t-tests or Spearman rank-correlation respectively. For time-to-event outcome analysis, VBA relies on voxel-wise Cox regressions (*e.g.*, survival [11], pericardial effusion [25] and lymphopenia [23]).

Most VBA studies listed in Table 1 only considered dose to identify spatial dose–response patterns. However, both survival and the risk of radiation induced morbidity will be influenced by other factors. Many studies considered patient-level adjustment variables in post-hoc models, most often in multivariable analyses including a dose statistic extracted from the VBA identified region [13,14,17]. This approach, although common in the literature, is limited as the definition of a 'region' based on a selected significance level is intrinsically dependent on the patient dataset size. Current best practice includes adjustment variables directly into VBA via voxel-wise Generalised Linear Models (GLMs) [16] or voxel-wise Cox proportional hazard regressions [11 23, 25]. By doing so, their effect is correctly accounted for in the estimation of the significance maps.

Dose-response for several anatomical sites has been explored using VBA techniques, with most research so far focused (although not exclusively) on the thorax, pelvis and the head and neck (HN).

Table 1
Representative sample of VBA results using an anatomical based common coordinate system with deformable image registration (DIR) into a template anatomy. The table summarises the methodologies, statistical analysis and outcomes analysed. CCS- common coordinate system, DSC – Dice Similarity Coefficient, DOO – Dose Organ Overlap, HD – Hausdorff Distance, MHD – Modified Hausdorff Distance, GLM – General Linear Model, TFCE – Threshold Free Cluster Enhancement, COM_{SD} – Standard Deviation of Centre of Mass, FWER – Family-Wise Error Rate.

Authors	Clinical domain	Outcome	Patient number	CCS	DIR algorithm	Lesion masking	Registration / Dose warping QA metrics	Dose variability assessment	VBA statistic	Correction for Multiple comparison	Filter	Adjustment variables included
Acosta et al. 2013 [15]	Prostate	rectal bleeding G1 +	105	patient CT in the cohort (found by affinity propagation clustering)	Demons	no	DSC, HD, DOO	no	t-test	no	no	no
Drean et al. 2016 [9]	Prostate	rectal bleeding G1 +	173	personalised (using each patient as template) and generic templates	Demons	no	DSC, DOO	no	Mann-Whitney test	yes (FDR - Benjamini-Hochberg)	no	no
Mylona et al. 2019 [17]	Prostate	urinary toxicity	272	patient CT in the cohort	Elastix	no	centreline distance (urethra), DSC, HD	no	Mann-Whitney test	yes (non-parametric permutation test max T)	no	no
Monti et al. 2017 [12]	Head and Neck	dysphagia G3+	42	patient CT in the cohort	Demons	no	DSC, modified HD, DOO	no	GLM	yes (non-parametric permutation test max T)	yes (TFCE)	no
Beasley et al. 2018 [28]	Head and Neck	trismus	86	rigid registration within a clip box	B-spline	no	COM_{SD}	no	Spearman's correlation	yes (non-parametric permutation test max)	no	no
Palma et al. 2016 [18]	Thorax	lung toxicity (CT changes)	98	patient CT in the cohort	Demons	no	DSC, MHD, DOO	no	GLM	yes (non-parametric permutation test max T)	yes (TFCE)	no
McWilliam et al. 2017 [13]	Thorax	mortality	1101	patient CT in the cohort	B-spline	no	Visual inspection of DIR, COM_{SD}	no	t-test	yes (non-parametric permutation test max T)	no	no
Palma et al. 2019 [21]	Thorax	Radiation pneumonitis G1+	178	synthetic CT phantom	B-spline	yes	DSC, DOO	voxel-wise dose standard deviation	GLM	yes (non-parametric permutation test max T)	yes (TFCE)	yes
Palma et al. 2019 [22]	Thorax-SBRT	RTOG lung toxicity G1+	106	synthetic CT phantom	B-spline	yes	DSC, MHD, DOO	voxel-wise dose standard deviation	GLM	yes (non-parametric permutation test max T)	yes (TFCE)	yes
Green et al. 2020 [11]	Thorax	survival	1101	patient CT in the cohort	B-spline	no	COM_{SD}	no	Cox	yes (non-parametric permutation test max β)	no	yes
Abravan et al. 2020 [14]	Thorax	lymphopenia G3+	901	patient CT in the cohort	B-spline	no	not specified	no	t-test	yes (non-parametric permutation test max T)	no	no
Bourbonne et al. 2021 [63]	Thorax	CTCAE lung toxicity G2+	167	not specified thoracic phantom	MIM maestro	no	DSC on segmented structures	no	GLM	yes (FWER Palm tool)	no	no
Cella et al. 2021 [25]	Thorax	heart toxicity (Pericardial effusion G2 +)	178	synthetic CT phantom	B-spline Elastix	yes	DSC, DOO, COM_{SD}	voxel wise dose mean and standard deviation, Pica, connectograms	GLM, Cox	yes (non-parametric permutation test max T, max Z)	yes (TFCE)	yes
Monti et al. 2022 [23]	Thorax	lymphopenia G1+	164	synthetic CT phantom	B-spline	yes	DSC, DOO, COM_{SD}	voxel-wise dose mean and standard deviation	GLM, Cox	yes (non-parametric permutation test max T, max Z)	yes (TFCE)	yes
Cho et al. 2022 [29]	Thorax	lymphopenia G3+	66	not specified	B-spline	yes	not specified	no	GLM	yes (non-parametric permutation test max T)	yes (TFCE)	yes
Monti et al. 2022 [24]	Thorax	esophagitis G2+	173	synthetic CT phantom	B-spline	yes	DSC, DOO, COM_{SD}	voxel-wise dose mean and standard deviation	GLM, Cox	yes (non-parametric permutation test max T, max Z)	yes (TFCE)	yes

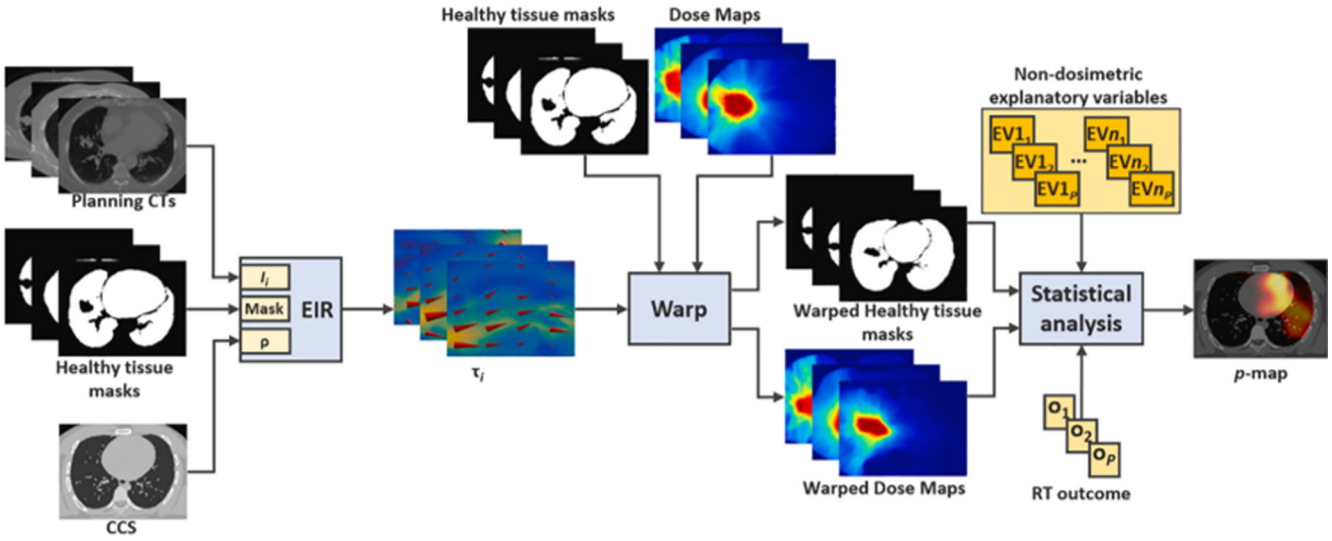

Fig. 1. A flow chart of a VBA workflow. Images, masks of segmentations are registered using non-rigid registration (elastic image registration) to a CCS. The non-rigid transformations are used to warp the dose maps into the frame-of-reference of the CCS allowing statistical analysis to be performed. The output of the statistically analysis is displayed on the CCS allowing localisation to the underlying anatomy. Figure reproduced from Palma et al. [4].

Each region presents specific considerations and opportunities; we present a brief summary for each below.

Voxel based analysis in thoracic regions

There has been considerable focus on VBA analysis of thoracic cohorts for survival and toxicity end-points. In analysing survival, the base of the heart was identified in a large group of non-small cell lung cancer (NSCLC) patients as a particularly dose-sensitive region, where higher doses are associated with poorer survival [11,13]. Such analyses have provided insight into the anatomical drivers of toxicities. When exploring radiation pneumonitis (RP) in NSCLC patients treated with either passive scattering proton therapy or intensity modulated radiation therapy (IMRT), VBA identified regions in the lower lungs (mainly the right lung) and in the heart where the dose was highly correlated with RP, supporting previously hypothesised biophysical relationships between dose to the heart and RP [21]. In the same cohort of patients, a Cox-VBA methodology was employed to analyse pericardial effusion, highlighting regions both in the heart and lungs [25]. Furthermore, correlations between local dose patterns and radiation induced lung damage (RILD) were investigated for a cohort of lung cancer patients treated with stereotactic body radiation therapy. The discovered relevance of dose to the heart and lower lungs in the development of RILD supports the previously reported results on lung toxicity [22]. VBA analysis of radiation induced oesophagitis has highlighted spatial dose patterns encompassing the thoracic oesophagus [24].

For patients with lung cancer there has also been much interest in investigating the development of lymphopenia. In a cohort of 901 NSCLC patients, significant correlations were identified between the risk of severe lymphopenia ($\geq$G3) and dose to the thoracic vertebrae, total heart and lungs [14]. In a smaller group of patients, treated with concurrent chemoradiation, a correlation was found between lymphocyte depletion at the end of treatment and the dose to the large vessels, in particular the aortic arch [29]. In addition, a Cox-VBA, incorporating an analysis of the kinetics or radiation induced lymphocyte loss, identified the risks of irradiation of lymphoid organs and organs with abundant blood pools [23].

Voxel based analysis in the pelvic region

In the pelvis, much of the work has centred on rectal and urinary toxicities for patients treated for prostate cancer [15,30]. The relationship between rectal bleeding (RB) and the spatial dose information was investigated for a cohort of 105 prostate patients treated with IMRT; a significant correlation was highlighted in the anterior part of the rectum [15]. This analysis was then methodologically refined and extended to a larger cohort of patients to analyse risk of Grade $\geq$ 1 and Grade $\geq$ 2 RB, and demonstrated that the irradiation of the inferior-anterior hemi anorectum and upper part of the anal canal drives the RB development [9]. An analysis of urinary toxicities relating to the dose to the urethra and bladder has identified heterogeneous correlations between local dose and five Grade $\geq$ 1 symptoms: high correlations were found in the prostatic urethra for acute incontinence, the bladder trigone for acute and late urinary retention, the posterior bladder for dysuria, and the superior bladder for late haematuria [17].

Voxel based analysis in the head and neck region

Due to the complexity of toxicity profiles and the potential for multi-organ interplay, VBA has been proposed to provide better understanding of the spatial signature of radiation sensitivity in composite regions like in the head and neck (H&N) [12,28]. In a small cohort of patients treated at a single institution with a combination of 3D-conformal radiotherapy (3D-CRT) and IMRT, the cricopharyngeal muscle and cervical oesophagus received significantly higher doses in patients developing radiation induced acute dysphagia [12]. VBA has also been used to investigate radiation induced trismus, and to identify the most significant correlations in the ipsilateral masseter [28].

Overview of the validation process

As with all forms of model development, models developed using VBA techniques need to show that they are repeatable, reproducible, and generalisable. In the VBA domain, consideration needs to be given to technical (methodological) validation, clinical validation, and translation into clinical use (Fig. 2). Here we define the following:

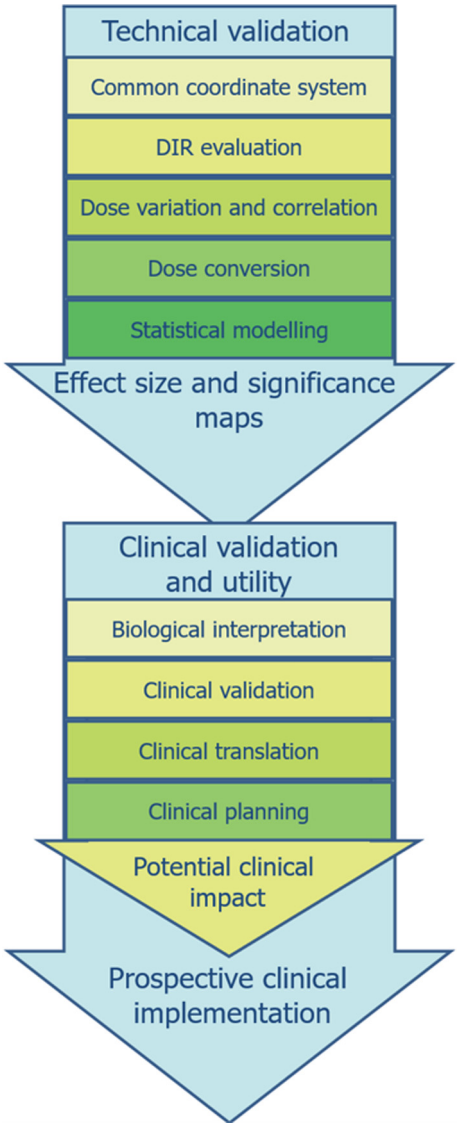

Fig. 2. Illustration of the roadmap through the technical validation, clinical validation and clinical utility leading to prospective clinical implementation.

- Technical validation (section 4) needs to demonstrate that results are repeatable to reasonable changes in the analysis pipeline, associated, for example, with the choice of CCS or DIR algorithm. In this context, accurate reporting and standardisation of the analysis procedure is crucial.
- Clinical validation (section 5) needs to demonstrate that findings are reproducible and generalisable to external cohorts, indicating findings are suitable for clinical translation. Reproducibility refers to the ability of VBA to generate similar results on slightly different input data. This could mean a second population of patients treated at the same centre or a population from an external RT centre, but receiving the same form of RT; i.e., technique, dose prescription and goals on target coverage, fractionation, constraints to OARs, and the same definition of the study outcome. Generalisability refers to the ability of VBA to generate similar results on independent populations, which could differ more extensively from the original population. This could include patients treated with different RT techniques, doses/fractionation and constraints to OARs.
- Clinical utility needs to demonstrate that it is feasible to change clinical practice to incorporate VBA findings, providing an estimate of the benefit to inform appropriate prospective implementation strategies.

Typically, VBA results in a map of effect sizes and their statistical significance (p-map). Concordance metrics can be used to assess repeatability, reproducibility, or generalisability of the VBA results during the technical validation steps described. These concordance metrics are included and described in Appendix A.

Technical validation

In addition to the methodological variability associated with the definition of the VBA statistical model (false discovery rate vs family-wise error rate control, threshold-free cluster-enhancement filter, etc.), the major consideration for technical validation remains the accuracy of the spatial normalisation, involving the assumption that the dose delivered to corresponding locations is correctly mapped. This section will first focus on the choice of the CCS and on the evaluation of the DIR before discussing methods to assess dose variability; recommendations for best practice are provided for each step.

Choice of CCS

The choice of CSS needs to be robust to large anatomical and imaging variability among patients. Small misalignments across the population are expected and need to be assessed. Various options for CCS used in the literature are described in Section 2.1. However, studies comparing the choice of CCS have not been widely performed. Methodologies selecting the CCS using clustering strategies have been proposed to avoid selecting the CCS arbitrarily, identifying the exemplar patient by competition as the most representative shape within each cluster according to an inter-individual distance based on geometric features of organ shapes [9,15–17]. Repeatability of a VBA study should be assessed by selecting multiple CCS and comparing results of the VBA from each. An example is included as Fig. 3, this VBA study used multiple CCS in an analysis of dysphagia in head and neck cancer treatments [31]. Three CCS were used allowing identified regions to be directly compared between template anatomies. The authors used the information from different CCS to define the most robust region, extracting the mean dose from the region on each template anatomy. A further paper explored the potential of selecting multiple CCS for handling paediatric populations, selecting based on age categories to maximise the accuracy of the registration step [32]. Further methodologies for the comparison of the full VBA maps can be performed using the methodologies described in Appendix A.

DIR accuracy evaluation

Evaluating the accuracy of inter-individual mapping is complex. Besides the lack of ground truth, anatomical differences between patients limit the usability of traditional intra-patient DIR performance metrics, such as those recommended in the report of American Association for Medical Physics (AAPM) task group 132[33]. In the VBA literature, geometric and image-based metrics are complementarily used to estimate inter-individual anatomical mapping accuracy. Dice similarity coefficient (DSC) and Modified Hausdorff Distance are among the more commonly used evaluation metrics. These metrics reflect geometric overlap between transformed structures. Intensity-based scores, on the other hand, allow estimating image similarity between the CCS and the warped images. As intensity-based metrics do not rely on segmentations,

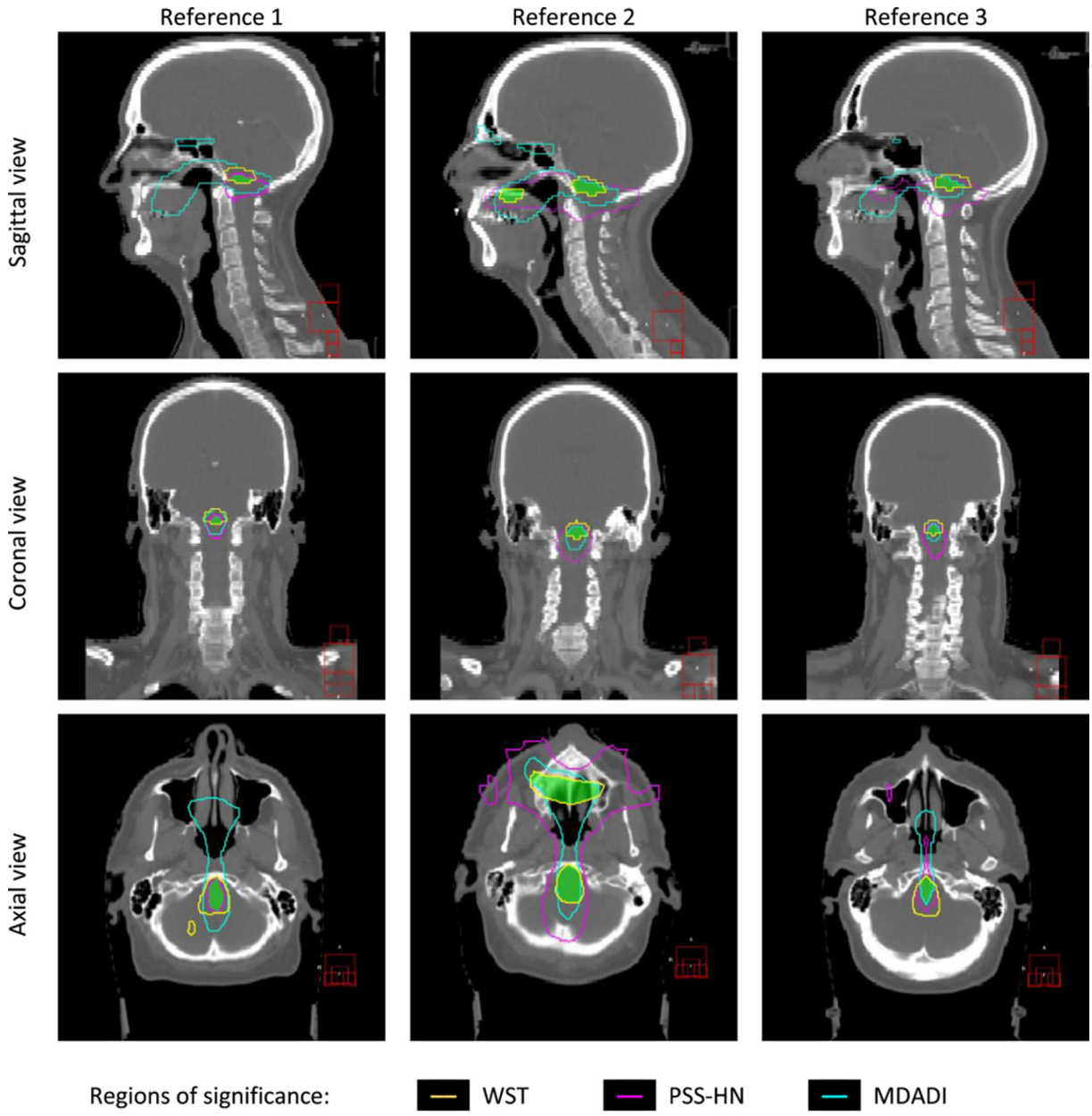

Fig. 3. Three CCS are shown in the three cardinal directions for the analysis of dysphagia in patients treated for head and neck cancer. Regions of significance ($p < 0.005$) for three different dysphagia metrics. MDADI = MD Anderson Dysphagia Inventory; PSS-HN = performance status scale for normalcy of diet; WST = water swallowing test. In each CCS the green region shows the common region across all three measures. These regions were used to extract the mean dose received for each patient plan on each CCS allowing a measure of uncertainty to be determined. Figure reproduced from Vasquez Osorio et al. [31].

they are not impacted by inaccuracies in their definition (inter-observer variability). However, good performance estimated by these metrics does not necessarily imply good point-to-point mapping.

In the context of dose VBA, the Dose Organ Overlap (DOO) can be an insightful metric, especially where high dose gradients are present, common within OARs. In this situation, registration errors may result in large local differences in dose on the reference template. This metric measures the ratio between dose distributions on the intersection and union of the considered region, and is similar to a DSC weighted by the dose [19]. The most important consideration in using this metric is the assumption that the dose distributions will be similar across the patient population.

Once assessed, the Euclidean uncertainty in the DIR can be accounted for, prior to statistical VBA, by blurring the patients' dose distributions by a Gaussian kernel with the corresponding full width half maximum (FWHM). This approach relies on the assumption that the registration errors are spatially uniform and randomly distributed between patients, and therefore impact the dose mapping in a random fashion.

Assessment of the dose variation and correlation

Spatial characterisation of the dose should form part of the technical validation for every VBA study in RT [34]. In radiation oncology, VBA can only identify dose patterns that are within the range of the heterogeneity of the dose distributions included in the dataset. Therefore, it is essential to acknowledge the intrinsic features of the dose that can limit the robustness of the VBA results and impact the radiobiological detail that can be identified.

The homogeneity of the statistical power in the analysed anatomical region strongly depends on the uniformity of voxel-

wise dose cumulants. Homogeneous mean (μ) and standard deviation (σ) maps rule out the hypothesis that relevant radiosensitive regions were dampened in the VBA results due to the non-uniform variability of the dose maps. Consequently, μ and σ maps of normalised dose distributions should be computed voxel-wise over the patients and their uniformity can be quantitatively evaluated by the Michelson contrast [34] as described in Appendix B. In addition, the spatial resolution of the VBA results (maps of the statistical model coefficients and their significance) may be affected by the correlation between the doses delivered to different anatomical substructures [25]. The spatial independence of the doses delivered in the CCS to each substructure can be investigated by means of probabilistic independent component analysis (PICA) [35] and connectogram analysis [36].

The PICA blindly infers the model order of the analysed dataset, which corresponds to the number of statistically significant independent spatial maps that generate the input. When applied to a dose distribution dataset (in the form of n [patient] $\times$ M [voxel]), a PICA model order comparable to the cohort size n signals a significant mutual independence between patients' dose distributions.

The connectograms, instead, highlight the most relevant associations between each pair of substructures according to the pairwise significance of the Spearman correlation between the mean doses related to the substructures. Indeed, the VBA will have more chance to discriminate the statistical properties arising from each substructure where there is a weaker correlation between substructure pairs. In the connectogram rings, outwards from the centre, the average and standard deviation of the mean doses across the patient population is presented for each substructure. This approach is a powerful visual representation to aid understanding of the correlations of dose and identifying potential limitations within the VBA.

Dose accuracy and dose conversion

A further potential source of uncertainty comes from the use of data from older versions of treatment planning systems where there exists uncertainty in the dose calculation. For example, where a lung plan was calculated using a pencil beam algorithm compared to more modern calculations using collapsed cone or Monte Carlo techniques. Care is needed to describe the dose calculation used within the manuscript. Where results from the VBA are presented, consideration and discussion of whether these are likely to be affected by regions of high uncertainty in the dose calculation, i.e., where scatter conditions are not fully modelled.

Dose distributions should be corrected for fraction size effects, by rescaling to an equieffective dose (e.g., a Biologically Effective Dose, BED), commonly done using the linear quadratic (LQ) model [37]. The choice of the α/β value applied in the LQ model needs to be driven by evidence, where available, for the study end-point analysed (i.e., tumour control versus normal tissue acute or late toxicity). Finally, where there is uncertainty regarding the correct choice of α/β value, a sensitivity analysis can be performed, changing the value across a reasonable accepted range and evaluating any potential impact on the reported results.

Clinical validation and utility

Clinical validation and demonstration of clinical utility are essential steps in the translation of VBA results into decision making and thus ultimately improvement of patient outcomes. Clinical validation needs to demonstrate that the VBA results are generalisable and robustly associated to a meaningful outcome, across multiple datasets and patient populations. Clinical utility refers to the ability to interpret and apply the results from the VBA in the clinical pathway to achieve an overall benefit for patients. Here, we follow the workflow described in Fig. 2.

Pre-clinical and clinical radiobiological interpretation

Clinical plausibility needs to be considered for any VBA spatial map. As a minimum, the VBA results need to be evaluated against current understanding of the pathophysiological processes associated with the anatomy in question. In other words, from which potential biological mechanisms could the dose–effect heterogeneity stem? These discussions should incorporate expert domain knowledge through input from relevant medical specialists. The degree of biological plausibility may guide the level of evidence needed for further validation. These discussions will also inform endpoints for future prospective studies, and generate hypotheses to direct future avenues of research, both pre-clinical and translational.

A deeper understanding of VBA findings can arise from a back-translation process, using the population level findings to formulate testable pre-clinical hypotheses. Potential mechanisms identified through discussions with medical specialists can guide the design of experiments to understand the potential underlying pathophysiology. This may include experiments in animal models, for example selective irradiation of the identified anatomical regions under controlled conditions. In an example of this process, VBA results identifying the most radiosensitive regions in the heart of lung cancer patients [13] were back-translated into controlled mice experiments, validating the findings and identifying potential mechanistic explanations [1,38].

Clinical validation

Validation of VBA goes beyond standard model validation in that two interlinking aspects need consideration and separate validation:

1. spatial dose–response maps;
2. predictive models which include the knowledge obtained by VBA.

The efforts required to validate the VBA model are proportionate to the patient risk associated with its implementation. The risk assessment should take into consideration the complexity of the hypothesis, the trade-offs involved in its implementation and the expected gain [39].

Extensive validation and reporting guidelines for predictive models are available, presenting a best practice approach that can be adapted for the VBA paradigm. In particular, predictive models which include knowledge obtained by VBA should be reported using the TRIPOD consortium checklist [40]. As discussed in section 3, validation should involve both assessment of reproducibility (the ability of VBA to generate similar results on slightly different input data) and generalisability (the ability of VBA to generate similar results on independent populations). These two steps may involve both internal and external validation cohorts [41]. An example of external validation of VBA on multiple external and independent cohorts is included as Fig. 4.

For validation of dose response maps, the following approaches should be considered depending on availability of datasets:

- Internal resampling 1) bootstrapping: repeatedly resample the population to define a spatial dose response map defining a consensus result with confidence intervals. 2) cross-validation: validation of the model prediction in the resampled test portion of the dataset [44].

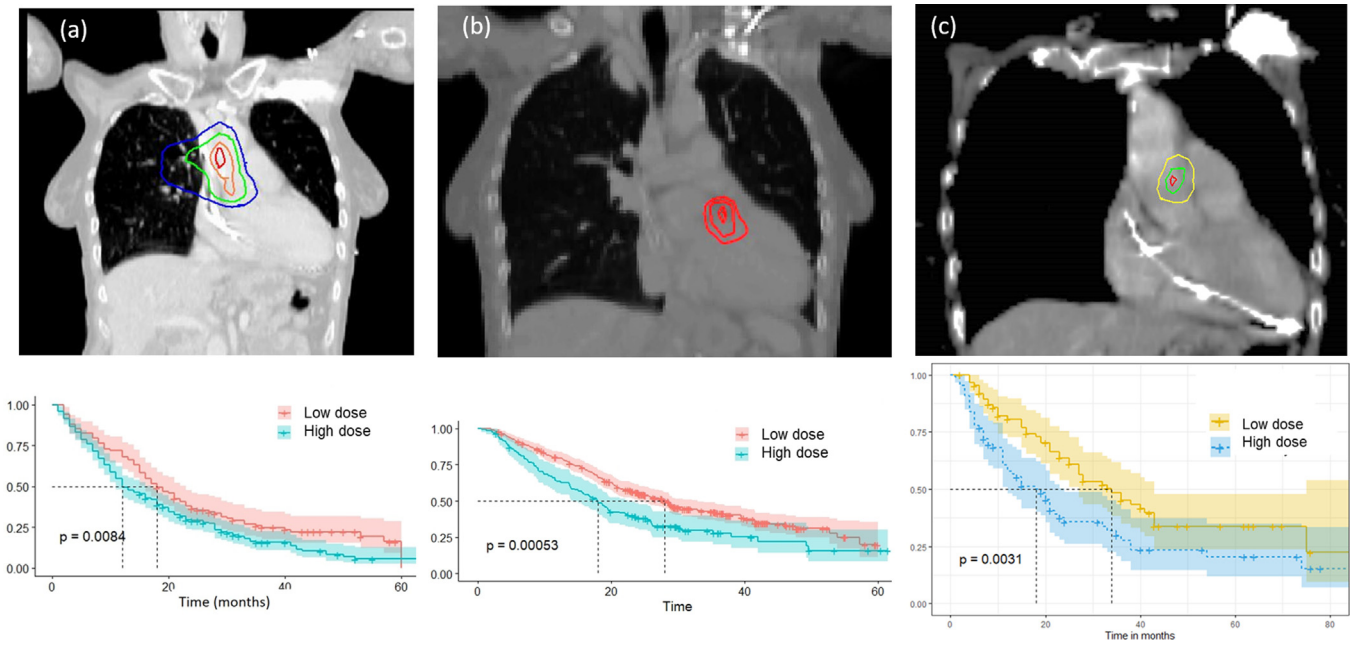

Fig. 4. Example of external validation in VBA in multiple cohorts. The VBA associated higher dose to the base of the heart with excess mortality after radiotherapy for patients with lung cancer. The same CCS is used in each case allowing regions defined by the VBA to be compared across cohorts. The lines in each analysis represent the 95th, 90th, and 85th significance level. In (a), 1101 patients from a single centre were analysed, all treated with 55 Gy in 20 fractions [13], in (b) 490 patients from the RTOG 0617 trial are analysed [42], and in (c) 205 patients from the PET-plan were analysed [43].".

- Mapping the identified anatomical structure directly to an internal/external validation cohort, extracting dose to this region and validating the derived models [19,41].
- Generation of the spatial dose response map in the internal/external validation cohort for direct comparison against the training result. The form of suitable comparison will follow from the VBA model applied as described in Section 4 and appendix A.

Importantly, validation of spatial dose response maps may fail both due to inherent issues with the developed VBA model (lack of reproducibility), and due to differences in treatment techniques or patient data between the development and validation cohorts (lack of generalisability). As in any clinical prediction model development, consideration needs to be given to patient level factors, and how this might impact the generalisability of spatial dose map findings. Geographical population factors will interact with dose and impact VBA results. For example, higher incidence of smoking or multimorbidity burdens will likely interact with the dose–response against the end-points of interest for VBA – overall survival, recurrence, and toxicity. Recommendations are available to aid in understanding and quantifying these impacts [40].

Clinical utility and implementation

This section outlines a framework and recommendations required to perform the in-silico comparative studies whose results will guide clinical studies to show feasibility and will provide the initial estimate of expected clinical benefit for patients. Finally, approaches for prospective clinical implementation studies will be suggested to measure real patient outcomes.

Clinical planning

The potential use of knowledge from VBA in treatment planning (TP) optimisation is numerous, with increasingly complex implementation requirements. This section will focus primarily on current standard of practice approaches with a view towards future optimisation strategies.

1. Identification of an anatomical structure as outcome driver, for which conventional dose metrics (DVH/EUD/max dose) can be extracted and constrained in TP.
2. Implementation of multi-criteria TP optimisation process of the 3D dose map based on different VBA-based NTCP models of treatment outcomes.

Due to the nature of the methodology, results from VBA will not inherently align with standard anatomical structures. For plan optimisation, this presents some difficulties: current standard planning practices require a defined region with which to apply a dosimetric constraint for plan optimisation. Section 5.1 discussed the need for biological plausibility in the underlying anatomy. This process should be extended so that segmentation of the most dose-sensitive regions can be protocolised based on an anatomical definition. The definition of this region should involve clinical experts to ensure inclusion of structures where irradiation may impact organ function.

Most importantly, as discussed throughout the paper, the segmentation of dose-sensitive regions, as opposed to dose-insensitive regions, represents a simplification to work with capabilities of commercial treatment planning systems and does not fully utilise the potential of VBA. Instead, the future direction of VBA is towards optimising plans directly on the dose response maps. Such dose response maps lend themselves to dose painting approaches; where optimisation constraints can be applied in a per-voxel level (directly from dose response maps for individual patient outcomes) in combination with optimisation constraints based on segmentations. Methodologies from studies optimising plans using functional imaging information can be readily adopted [45–47]. Where multiple dose response maps are being used within the same plan then multicriterial optimisation and prioritisation need to be considered.

Next, the ability to optimise deliverable plans has to be demonstrated through planning and delivery verification studies. Recommendations are published for completing and reporting planning studies [48], including studies comparing photons with proton plans [49]. Clinical acceptability, including the ability to meet existing protocoled constraints, needs to be shown alongside implementation of any further dose constraints based on the VBA result. Furthermore, visualisation of differences in the dose distribution and in organ DVHs between 'traditional' and VBA-based plans will provide confidence that no unacceptable detriment in dose distributions is present.

Where new optimisation constraints impact target coverage or existing OAR dose levels then prioritisation among planning goals is needed. This requires clinical consensus, discussions with domain experts as part of an iterative planning study to define what is achievable. This process can develop planning approaches and identify patient specific metrics to determine which patients may require a higher level of input and expertise to define their optimal plan. Flagging such cases early in the process as complex allows additional time or resources to be allocated.

Estimating potential clinical impact

Retrospective planning studies will provide an estimation of the ability to clinically implement the VBA insights. Combined with understanding of other toxicity risk factors, this will enable initial estimates of the potential benefit for patients [9,50], as well as identification of sub-groups where absolute benefit may be highest / lowest. While the impact of the new approach can be expected to vary greatly, depending on the extent to which existing planning strategies already incidentally suppress the dose to the VBA-defined region, VBA insights may generally result in more deliberate planning strategies. Such work will provide guidance on the number of patients where a dose reduction is possible. There may be a role for use of separate, validated NTCP models to further quantify the expected patient benefit. These in-silico experiments will inform prospective implementation, allowing an estimate of the number of patients needed to demonstrate the effect. Based on this, the most appropriate form of prospective implementation can be chosen, with possibilities discussed in the following section.

Prospective clinical implementation

Ultimately, the dose distribution across the patient's anatomy will be altered, even if dose statistics are acceptable and the DVH curves comparable to existing practice. Any implementation in the clinic needs appropriate monitoring for any unexpected increase in patient toxicities.

Robust comparisons between treatments optimised using novel VBA dose limits could be made using randomised controlled trials (RCT). However, the practicalities of implementing RCTs, and the challenges of using them to evaluate changes in rapidly evolving radiation treatment delivery techniques, mean they are often not suitable for this type of innovation [51]. Alternatives for implementation of technical changes in RT optimisation include:

1. Continuous learning cycle [52,53] where potential impacts to patient outcomes, due to changes in treatment, are monitored using routinely collected data. A proof-of-principle clinical study using this framework for implementing a VBA result is underway, as described by Price et al. [54].
2. The R-IDEAL framework [55] uses a five-stage development process, aiming for quick clinical implementation without exposing patients to additional risk. The steps described above fall under R-IDEAL stage 0, with stage 1 the first time used in patients, through to long term outcomes (stage 4).

Recommendations / best practice guidelines

The following section summarises recommendations for best practice to perform VBA experiments and report within scientific papers.

VBA reporting

1. The choice of CCS(s) and the rationale behind it should be fully described.

- Consider making the chosen CCS(s) available open source alongside study publications, to allow for reproducibility experiments by independent groups. This can be facilitated by using open-source patient images or digital phantoms as the CCS.

2. The DIR algorithm and its tuning parameters should be included for reproducibility.

- If any registration has been fine-tuned (e.g., for maximising the numbers in small datasets), the number of instances and deviating workflow should be reported.
- The number of patients with a failed spatial normalisation of the dose should be reported and what were the encountered issues (e.g., atypical anatomical differences, excessive image artefacts).

3. DIR accuracy should be described, including the metric(s) used for evaluation.
4. If dose was converted, details of the model should be reported (e.g., α/β, BED vs EQDx).
5. The results of VBA (statistical effect size and significance) should be presented as thoroughly as possible:

- The effects and significance maps should be fully presented, no thresholding should be applied to demonstrate levels of significance;
- Snapshots of multi-planar reconstruction of the maps should be shown along the main three anatomical planes;
- Consider including a binary file of the full 3-dimensional VBA maps as supplementary material or in a public git-hub repository (e.g., zenodo, git-hub, academic website), linked to the spatial coordinates of the CCS.

VBA technical validation

1. VBA results from multiple CCS should be compared to assess the robustness to the choice of the CSS according to the validation metrics.
2. An assessment of the dose variability and autocorrelation in the population should be performed by computing:

- Voxel wise mean and standard deviation of the dose distributions;
- Consider: Present dosimetric PICA and connectograms if relevant structures are available.

3. If a model of biologically effective dose has been used, consider including a sensitivity analysis of potential impact of changing key parameters (α/β) as supplementary material.

VBA clinical validation

1. TRIPOD reporting guidelines should be followed, with reproducibility and generalisability reported.
2. Differences in patient demographics, treatment characteristics and dose variability between cohorts (assuming prior publication includes information) should be acknowledged.
3. Recommendations for technical validation should be followed in new cohorts.

Considerations for translation

For translation, guidelines exist and are discussed above. The key considerations are:

1. Definition of the anatomical region for planning with clear guidelines for reproducing. Or, how the VBA results are to be incorporated into planning optimisation (e.g., dose painting).
   - Design of the expert consensus process.
2. Definition of the optimisation constraints for the VBA region and prioritisation of goals in RT treatment planning.
3. Implementation of any VBA results within a prospective study is essential.

Discussion

VBA is providing new insights into anatomical variability of pathophysiology and radiosensitivity by removing the need for prior definition of organs assumed to drive the dose response associated with patient outcomes. In the last decade, the methodology has been developed and toolkits expanded to handle a variety of clinical questions [4,56]. Exemplar VBA results are now moving towards clinical translation with the aim to improve patient outcomes after RT [55]. Translating novel research into clinical practice for patient benefit is a difficult task in RT. To ensure that VBA results do not fall into the (so-called) translational gaps [57], here we presented a roadmap paper for bringing VBA results from conception, through technical and clinical validation, to demonstration of clinical utility and implementation.

Current state-of-the-art VBA includes adjustment variables directly in the analysis rather than in a post-VBA step. Two multi-variable approaches are gaining ground with voxel-based Cox-Proportional hazard models [11] and generalised linear models [18]. The interpretation of these analyses becomes increasingly difficult as they produce one effect map per variable. These maps show how the impact of each variable changes across the CCS with the dose variability in the patient population. With each additional variable included in the analysis, the more complex the output becomes. Better methods to display and interrogate this multidimensional output are needed to fully utilise these approaches. It is important to note that the maps of each included covariate are analogous to multiple estimates from a single (multivariable) model. Analysing these in isolation can introduce biases (the 'table 2 fallacy' [58,59]). Interpretation of such estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. It is therefore recommended to focus on interpreting the maps of the variable under scrutiny (in this context, the dose) to avoid over-interpretation and biases. Further, as in any statistical analysis, proper variability of the explanatory variables (e.g., the dose) and the dependent variable (e.g., the RT outcome) is required to power the analysis. Performing a prospective study power calculation remains a difficult and unsolved problem in VBA.

VBA will continue to be developed and additional tools developed to facilitate better understanding. With a view to future developments, there is a need to move to a causal interpretation of VBA results. As above, interpreting the output of VBA accounting for pre-selected adjustment variables may inflate the risk of table 2 fallacy [58]. Moving to causal inference techniques [60] will reduce this risk and would open possibilities to ask 'what if?' questions in understanding the best approach to mitigate the observed risk.

Current approaches to translate VBA require the definition of a VBA-based dose/volume metric or equivalent uniform dose (EUD) objective for inclusion in treatment plan optimisation. This is reflected in the roadmap presented above. Future approaches should use the radiosensitivity maps within a dose painting optimisation approach. These maps will break the link to predefined anatomical structures and will account for patient factors. Where multiple radiosensitivity maps exist (for example, where multiple toxicities have been modelled), then these can be used within a multi-criteria optimisation. The validation of the full VBA sensitivity map will need a different approach to that presented and needs to be fully developed by the community.

In conjunction with the analysis of the full dose information, future studies may consider accounting for patient genetics. There is increasing understanding of genetic markers for individual radiosensitivity [61]. These markers are likely to interact with the 3-dimensional dose distributions as a modulating factor for risk of toxicity [21]. Depending on the patient population and the outcome of interest, the optimal analysis may account for both, marrying individual genomic markers with individual dose.

This paper presents a roadmap to report VBA findings and steps needed for technical and clinical validation to move VBA findings into clinical practice. Crucially, for ease of multi-centre validation, we need standardised, open access tools [62] and the creation of an international VBA community. This VBA community will work together to develop and champion VBA to ensure results define the optimal RT for patients.

CRediT authorship contribution statement

Alan McWilliam: Conceptualization, Writing – original draft, Writing – review & editing. **Giuseppe Palma:** Conceptualization, Writing – original draft, Writing – review & editing. **Azadeh Abravan:** Conceptualization, Writing – original draft, Writing – review & editing. **Oscar Acosta:** Conceptualization, Writing – original draft, Writing – review & editing. **Ane Appelt:** Conceptualization, Writing – original draft, Writing – review & editing. **Marianne Aznar:** Conceptualization, Writing – original draft, Writing – review & editing. **Serena Monti:** Conceptualization, Writing – original draft, Writing – review & editing. **Eva Onjukka:** Conceptualization, Writing – original draft, Writing – review & editing. **Vanessa Panettieri:** Conceptualization, Writing – original draft, Writing – review & editing. **Lorenzo Placidi:** Conceptualization, Writing – original draft, Writing – review & editing. **Tiziana Rancati:** Conceptualization, Writing – original draft, Writing – review & editing. **Eliana Vasquez Osorio:** Conceptualization, Writing – original draft, Writing – review & editing. **Marnix Witte:** Conceptualization, Writing – original draft, Writing – review & editing. **Laura Cella:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Dr Alan McWilliam, Dr Giuseppe Palma and Dr Laura Cella would like to thank ESTRO for the opportunity to chair the physics workshop theme "Mining the radiotherapy dose, exploring dose response patterns in radiation therapy". This paper is a direct output from the discussions during this workshop.

Dr Alan McWilliam and Dr Marianne Aznar are supported by NIHR Manchester Biomedical Research Centre and by Cancer Research UK via funding to the Cancer Research Manchester Centre [C147/A25254]. Dr Alan McWilliam, Dr Marianne Aznar, Dr Eliana Vasquez Osorio and Dr Azadeh Abravan are also supported by the

Cancer Research UK RadNet Manchester [C1994/A28701]. Dr Marianne Aznar also acknowledges the support of the Engineering and Physical Research Council (Grant number EP/T028017/1). Dr Ane Appelt is supported by Cancer Research UK RadNet [C19942/A28832] and Yorkshire Cancer Research [L389AA]. Dr Serena Monti is supported by PROGETTO CIR01_00023 -IMPARA -IMAGING DALLE MOLECOLE ALLA PRECLINICA -RAFFORZAMENTO DEL CAPITALE UMANO. Dr Eva Onjukka is supported by a fellowship from the Swedish Cancer Society [22 0531]. Dr Oscar Acosta acknowledge fundings from the French National Research Agency (ANR) in the framework of the "Investing for the Future" program, CominLabs Excellence Laboratory (ANR-10-LABX-07-01), via the European Frame ERA Permed (PerPlanRT ERAPERMED2020-110) and Fondation Rennes 1 via the program Semesters for innovation. Dr Marnix Witte is supported by Research institutional grants awarded to the Netherlands Cancer Institute from the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport.

Appendix A. Validation metrics

A possible concordance metric for the β maps obtained in repeatability, reproducibility or generalizability experiments is the Minkowski distance. Given two maps A and B, it is defined as:

$$d_a(A,B) = \left(\frac{1}{\lambda(\Omega)}\int_\Omega |A-B|^a d\lambda\right)^{\frac{1}{a}}$$

where $\lambda(\Omega)$ is the measure of the analysed region-of-interest. This distance, parametrized by the real parameter $a \geq 1$, can be summarised by the global metrics [61]:

$$\underline{d}(A,B) = \int_0^1 d_{\frac{1}{n}}(A,B)dn$$

which ranges from 0 (perfect match) to 1 (worst match) [64].

The same metric can be applied for the repeatability check of the p maps (in this case, it ranges from 0 (perfect match) to 1 (worst match)), since a repeatable significance map would require an absolute matching of the p values. In this case, the DSC Index over p (DI_p) has been suggested as an alternative concordance measure [4]. Given two significance maps, A and B, and defining SP as the sublevel set of a significance map at a given p value, the DI_p is computed, for each $(A,B) < P \leq 1$, as:

$$DI_p[A,B](P) = DI(S_P[A], S_P[B])$$

where DI is the DSC Index over two sets. The average value of the DI_p function can be used to summarise the match and it ranges from 0 (worst match) to 1 (perfect match).

On the other hand, the comparison of significance maps for reproducibility or generalizability studies requires a measure not focusing on absolute matching of the p values. A valuable alternative is provided by the DSC Index over Volume (DI_V). Denoting as $f_X[P]$ the relative volume of $S_P[X]$, for each $0 < V \leq 1$ the DI_V is defined as:

$$DI_V[A,B](V) = DI\left(S_{f_A^{-1}(V)}[A], S_{f_B^{-1}(V)}[B]\right)$$

DI_V measures the match for an equal volume of the significance maps and quantifies the spatial trends of significance, looking at the overlap of equally ranked voxels. It has been shown that its AUC ranges between 1-log 2 (worst match) and 1 (perfect match) [4].

Appendix B. Uniformity of dose cumulants

For a function I, the Michelson contrast can be computed as:

$$C_M = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

where I_{max} and I_{min} are the highest and lowest values of the function. For a positive-valued function I (such as μ and σ maps), $0 \leq C_M \leq 1$. For a given fraction, $0 < f \leq 1$ of the volume V of the support (i.e., the analysed organ or anatomical apparatus) of I, $C_M[I](f)$ can be defined as the minimum C_M assumed by the restrictions of I over the subsets of V $\{\|S_i\| = f\|V\|\}$. Since $C_M[I](f)$ is a monotonically increasing function, a summary description of the uniformity of I is provided by its area under curve (AUC). The AUC value of $C_M[I](f)$ is 0 for the constant maps and tends to 1 for highly inhomogeneous maps.

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2023.109868.

References

[1] Ghita M, Gill EK, Walls GM, Edgar KS, McMahon SJ, Osorio EV, et al. Cardiac sub-volume targeting demonstrates regional radiosensitivity in the mouse heart. Radiother Oncol 2020;152:216–21. https://doi.org/10.1016/j.radonc.2020.07.016.
[2] Konings AW, Cotteleer F, Faber H, van Luijk P, Meertens H, Coppes RP. Volume effects and region-dependent radiosensitivity of the parotid gland. Int J Radiat Oncol Biol Phys 2005;62:1090–5. https://doi.org/10.1016/j.ijrobp.2004.12.035.
[3] Spampinato S, Fokdal L, Marinovskij E, Axelsen S, Pedersen EM, Pötter R, et al. Assessment of dose to functional sub-structures in the lower urinary tract in locally advanced cervical cancer radiotherapy. Phys Med 2019;59:127–32. https://doi.org/10.1016/j.ejmp.2019.01.0174.
[4] Palma G, Monti S, Cella L. Voxel-based analysis in radiation oncology: A methodological cookbook. Phys Med 2020;69:192–204. https://doi.org/10.1016/j.ejmp.2019.12.013.
[5] Rigaud B, Simon A, Castelli J, Lafond C, Acosta O, Haigron P, et al. Deformable image registration for radiation therapy: principle, methods, applications and evaluation. Acta Oncol 2019;58:1225–37. https://doi.org/10.1080/0284186X.2019.1620331.
[6] Rohlfing T, Brandt R, Menzel R, Maurer Jr CR. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 2004;21:1428–42. https://doi.org/10.1016/j.neuroimage.2003.11.010.
[7] Yigitsoy M, Wachinger C, Navab N. Temporal groupwise registration for motion modelling. Inf Process Med Imaging. 2011;22:648-659. doi:10.1007/978-3-642-22092-0_53.
[8] Veiga C, Lim P, Anaya VM, Chandy E, Ahmad R, D'Souza D, et al. Atlas construction and spatial normalisation to facilitate radiation-induced late effects research in childhood cancer. Phys Med Biol 2021;66:. https://doi.org/10.1088/1361-6560/abf010105005.
[9] Dréan G, Acosta O, Ospina JD, Fargeas A, Lafond C, Corrégé G, et al. Identification of a rectal subregion highly predictive of rectal bleeding in prostate cancer IMRT. Radiother Oncol 2016;119:388–97. https://doi.org/10.1016/j.radonc.2016.04.023.
[10] Rohlfing T, Sullivan EV, Pfefferbaum A. Subject-matched templates for spatial normalization. Med Image Comput Comput Assist Interv 2009;12:224–31. https://doi.org/10.1007/978-3-642-04271-3_28.
[11] Green A, Vasquez Osorio E, Aznar MC, McWilliam A, van Herk M. Image based data mining using Per-voxel Cox regression. Front Oncol 2020;10:1178. https://doi.org/10.3389/fonc.2020.01178.
[12] Monti S, Palma G, D'Avino V, Gerardi M, Marvaso G, Ciardo D, et al. Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients. Sci Rep 2017;7:7220. https://doi.org/10.1038/s41598-017-07586-x.
[13] McWilliam A, Kennedy J, Hodgson C, Vasquez Osorio E, Faivre-Finn C, van Herk M. Radiation dose to heart base linked with poorer survival in lung cancer patients. Eur J Cancer 2017;85:106–13. https://doi.org/10.1016/j.ejca.2017.07.053.
[14] Abravan A, Faivre-Finn C, Kennedy J, McWilliam A, van Herk M. Radiotherapy-related lymphopenia affects overall survival in patients with lung cancer. J Thorac Oncol 2020;15:1624–35. https://doi.org/10.1016/j.jtho.2020.06.008.
[15] Acosta O, Drean G, Ospina JD, Simon A, Haigron P, Lafond C, et al. Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. Phys Med Biol 2013;58:2581–95. https://doi.org/10.1088/0031-9155/58/8/2581.
[16] Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007;315:972–6. https://doi.org/10.1126/science.1136800.
[17] Mylona E, Acosta O, Lizee T, Lafond C, Crehange G, Magné N, et al. Voxel-based analysis for identification of urethrovesical subregions predicting urinary

toxicity after prostate cancer radiation therapy. Int J Radiat Oncol Biol Phys 2019;104:343–54. https://doi.org/10.1016/j.ijrobp.2019.01.088.
[18] Palma G, Monti S, D'Avino V, Conson M, Liuzzi R, Pressello MC, et al. A Voxel-based approach to explore local dose differences associated with radiation-induced lung damage. Int J Radiat Oncol Biol Phys 2016;96:127–33. https://doi.org/10.1016/j.ijrobp.2016.04.033.
[19] Dréan G, Acosta O, Lafond C, Simon A, de Crevoisier R, Haigron P. Interindividual registration and dose mapping for voxel wise population analysis of rectal toxicity in prostate cancer radiotherapy. Med Phys 2016;43:2721–30. https://doi.org/10.1118/1.4948501.
[20] Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. Med Phys 2010;37:4902–15. https://doi.org/10.1118/1.3480985.
[21] Palma G, Monti S, Xu T, Scifoni E, Yang P, Hahn SM, et al. 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 2019;104:1124–32. https://doi.org/10.1016/j.ijrobp.2019.02.039.
[22] Palma G, Monti S, Thor M, Rimner A, Deasy JO, Cella L. 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 2019;64:. https://doi.org/10.1088/1361-6560/ab2e16155006.
[23] Monti S, Xu T, Liao Z, Mohan R, Cella L, Palma G. On the interplay between dosiomics and genomics in radiation-induced lymphopenia of lung cancer patients. Radiother Oncol 2022;167:219–25. https://doi.org/10.1016/j.radonc.2021.12.038.
[24] Monti S, Xu T, Mohan R, Liao Z, Palma G, Cella L. Radiation-induced esophagitis in non-small-cell lung cancer patients: Voxel-based analysis and NTCP modelling. Cancers (Basel) 2022;14:1833. https://doi.org/10.3390/cancers14071833.
[25] Cella L, Monti S, Xu T, Liuzzi R, Stanzione A, Durante M, et al. Probing thoracic dose patterns associated to pericardial effusion and mortality in patients treated with photons and protons for locally advanced non-small-cell lung cancer. Radiother Oncol 2021;160:148–58. https://doi.org/10.1016/j.radonc.2021.04.025.
[26] Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL; Brain Development Cooperative Group. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage. 2011;54:313-27. doi: 10.1016/j.neuroimage.2010.07.033.
[27] Chen C, Witte M, Heemsbergen W, van Herk M. Multiple comparisons permutation test for image based data mining in radiotherapy. Radiat Oncol 2013;8:293. https://doi.org/10.1186/1748-717X-8-293.
[28] Beasley W, Thor M, McWilliam A, Green A, Mackay R, Slevin N, et al. Image-based data mining to probe dosimetric correlates of radiation-induced trismus. Int J Radiat Oncol Biol Phys 2018;102:1330–8. https://doi.org/10.1016/j.ijrobp.2018.05.054.
[29] Cho Y, Kim Y, Chamseddine I, Lee WH, Kim HR, Lee IJ, et al. Lymphocyte dynamics during and after chemo-radiation correlate to dose and outcome in stage III NSCLC patients undergoing maintenance immunotherapy. Radiother Oncol 2022;168:1–7. https://doi.org/10.1016/j.radonc.2022.01.007.
[30] Mylona E, Cicchetti A, Rancati T, Palorini F, Fiorino C, Supiot S, et al. Local dose analysis to predict acute and late urinary toxicities after prostate cancer radiotherapy: Assessment of cohort and method effects. Radiother Oncol 2020;147:40–9. https://doi.org/10.1016/j.radonc.2020.02.028.
[31] Vasquez Osorio E, Abravan A, Green A, van Herk M, Wai Lee L, Ganderton D, et al. Dysphagia at 1 year is associated with mean dose to the inferior section of the brain stem. Int J Radiat Oncol Biol Phys 2023. https://doi.org/10.1016/j.ijrobp.2023.06.004.
[32] Wilson LJ, Bryce-Atkinson A, Green A, Li Y, Merchant TE, van Herk M, et al. Image-based data mining applies to data collected from children. Phys Med 2022;99:31–43. https://doi.org/10.1016/j.ejmp.2022.05.003.
[33] Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. Med Phys. 2017;44:e43-e76. doi: 10.1002/mp.12256.
[34] Palma G, Monti S, Pacelli R, Liao Z, Deasy JO, Mohan R, et al. Radiation pneumonitis in thoracic cancer patients: Multi-center voxel-based analysis. Cancers (Basel) 2021;13:3553. https://doi.org/10.3390/cancers13143553.
[35] Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 2004;23:137–52. https://doi.org/10.1109/TMI.2003.822821.
[36] Irimia A, Chambers MC, Torgerson CM, Van Horn JD. Circular representation of human cortical networks for subject and population-level connectomic visualization. Neuroimage 2012;60:1340–51. https://doi.org/10.1016/j.neuroimage.2012.01.107.
[37] Bentzen SM, Dörr W, Gahbauer R, Howell RW, Joiner MC, Jones B, et al. Bioeffect modeling and equieffective dose concepts in radiation oncology–terminology, quantities and units. Radiother Oncol 2012;105:266–8. https://doi.org/10.1016/j.radonc.2012.10.006.
[38] Walls GM, Ghita M, Queen R, Edgar KS, Gill EK, Kuburas R, et al. Spatial gene expression changes in the mouse heart after base-targeted irradiation. Int J Radiat Oncol Biol Phys 2023;115:453–63. https://doi.org/10.1016/j.ijrobp.2022.08.031.
[39] Bodner J, Kaul, V. A Framework for In Silico Clinical Trials for Medical Devices Using Concepts From Model Verification, Validation, and Uncertainty Quantification (VVUQ)." Proceedings of the ASME 2021 Verification and Validation Symposium. ASME 2021 Verification and Validation Symposium. 2021. Doi: 10.1115/VVS2021-65094.
[40] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:. https://doi.org/10.1136/bmj.g7594g7594.
[41] Mylona E, Ebert M, Kennedy A, Joseph D, Denham J, Steigler A, et al. Rectal and urethro-vesical subregions for toxicity prediction after prostate cancer radiation therapy: validation of voxel-based models in an independent population. Int J Radiat Oncol Biol Phys 2020;108:1189–95. https://doi.org/10.1016/j.ijrobp.2020.07.019.
[42] McWilliam A, Abravan A, Banfill K, Faivre-Finn C, van Herk M. Demystifying the results of RTOG 0617: Identification of dose sensitive cardiac subregions associated with overall survival. J Thorac Oncol 2023;18:599–607. https://doi.org/10.1016/j.jtho.2023.01.085.
[43] Craddock M, Nestle U, Koenig J, Schimek-Jasch T, Kremp S, Lenz S, et al. Cardiac function modifies the impact of heart base dose on survival: A Voxel-wise analysis of patients with lung cancer from the PET-plan trial. J Thorac Oncol 2023;18:57–66. https://doi.org/10.1016/j.jtho.2022.09.004.
[44] Xu CJ, van der Schaaf A, Van't Veld AA, Langendijk JA, Schilstra C. Statistical validation of normal tissue complication probability models. Int J Radiat Oncol Biol Phys 2012;84:e123–9. https://doi.org/10.1016/j.ijrobp.2012.02.022.
[45] Petit SF, Breedveld S, Unkelbach J, den Hertog D, Balvert M. Robust dose-painting-by-numbers vs. nonselective dose escalation for non-small cell lung cancer patients. Med Phys 2021;48:3096–108. https://doi.org/10.1002/mp.14840.
[46] Yan D, Chen S, Krauss DJ, Chen PY, Chinnaiyan P, Wilson GD. Tumor Voxel dose-response matrix and dose prescription function derived using 18F-FDG PET/CT images for adaptive dose painting by number. Int J Radiat Oncol Biol Phys 2019;104:207–18. https://doi.org/10.1016/j.ijrobp.2019.01.077.
[47] Jiménez-Ortega E, Ureba A, Vargas A, Baeza JA, Wals-Zurita A, García-Gómez FJ, et al. Dose painting by means of Monte Carlo treatment planning at the voxel level. Phys Med 2017;42:339–44. https://doi.org/10.1016/j.ejmp.2017.04.005.
[48] Hansen CR, Crijns W, Hussein M, Rossi L, Gallego P, Verbakel W, et al. Radiotherapy treatment planning study guidelines (RATING): A framework for setting up and reporting on scientific treatment planning studies. Radiother Oncol 2020;153:67–78. https://doi.org/10.1016/j.radonc.2020.09.033.
[49] Lowe M, Gosling A, Nicholas O, Underwood T, Miles E, Chang YC, et al. Comparing proton to photon radiotherapy plans: UK consensus guidance for reporting under uncertainty for clinical trials. Clin Oncol (R Coll Radiol) 2020;32:459–66. https://doi.org/10.1016/j.clon.2020.03.014.
[50] McWilliam A, Khalifa J, Vasquez Osorio E, Banfill K, Abravan A, Faivre-Finn C, et al. Novel methodology to investigate the effect of radiation dose to heart substructures on overall survival. Int J Radiat Oncol Biol Phys 2020;108:1073–81. https://doi.org/10.1016/j.ijrobp.2020.06.031.
[51] van Loon J, Grutters J, Macbeth F. Evaluation of novel radiotherapy technologies: what evidence is needed to assess their clinical and cost effectiveness, and how should we get it? Lancet Oncol 2012;13:e169–77. https://doi.org/10.1016/S1470-2045(11)70379-5.
[52] van der Schaaf A, Langendijk JA, Fiorino C, Rancati T. Embracing phenomenological approaches to normal tissue complication probability modelling: a question of method. Int J Radiat Oncol Biol Phys 2015;91:468–71. https://doi.org/10.1016/j.ijrobp.2014.10.017.
[53] Price G, Mackay R, Aznar M, McWilliam A, Johnson-Hart C, van Herk M, et al. Learning healthcare systems and rapid learning in radiation oncology: Where are we and where are we going? Radiother Oncol 2021;164:183–95. https://doi.org/10.1016/j.radonc.2021.09.030.
[54] Price G, Devaney S, French DP, Holley R, Holm S, Kontopantelis E, et al. Can real-world data and rapid learning drive improvements in lung cancer survival? The RAPID-RT Study Clin Oncol (R Coll Radiol) 2022;34:407–10. https://doi.org/10.1016/j.clon.2021.12.017.
[55] Verkooijen HM, Kerkmeijer LGW, Fuller CD, Huddart R, Faivre-Finn C, Verheij M, et al. R-IDEAL: A framework for systematic clinical evaluation of technical innovations in radiation oncology. Front Oncol 2017;7:59. https://doi.org/10.3389/fonc.2017.00059.
[56] Ebert MA, Gulliford S, Acosta O, de Crevoisier R, McNutt T, Heemsbergen WD, et al. Spatial descriptions of radiotherapy dose: normal tissue complication models and statistical associations. Phys Med Biol 2021;66. https://doi.org/10.1088/1361-6560/ac0681.
[57] O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14:169–86. https://doi.org/10.1038/nrclinonc.2016.162.
[58] Westreich D, Greenland S. The Table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. Am J Epidemiol 2013;177:292–8. https://doi.org/10.1093/aje/kws412.
[59] Bandoli G, Palmsten K, Chambers CD, Jelliffe-Pawlowski LL, Baer RJ, Thompson CA. Revisiting the Table 2 fallacy: A motivating example examining preeclampsia and preterm birth. Paediatr Perinat Epidemiol 2018;32:390–7. https://doi.org/10.1111/ppe.12474.
[60] Pearl J. An introduction to causal inference. Int J Biostat 2010;6:Article 7. https://doi.org/10.2202/1557-4679.1203.
[61] Scott JG, Sedor G, Scarborough JA, Kattan MW, Peacock J, Grass GD, et al. Personalizing radiotherapy prescription dose using genomic markers of

radiosensitivity and normal tissue toxicity in NSCLC. J Thorac Oncol 2021;16:428–38. https://doi.org/10.1016/j.jtho.2020.11.008.
[62] Palma G, Cella L, Monti S. MAMBA – multi-pAradigM voxel-based analysis: A computational cookbook. Med Phys 2023. https://doi.org/10.1002/mp.16260.
[63] Bourbonne V, Lucia F, Jaouen V, Bert J, Rehn M, Pradier O, et al. Development and prospective validation of a spatial dose pattern based model predicting acute pulmonary toxicity in patients treated with volumetric arc-therapy for locally advanced lung cancer. Radiother Oncol 2021;164:43–9. https://doi.org/10.1016/j.radonc.2021.09.008.
[64] Monti S, Pacelli R, Cella L, Palma G. Inter-patient image registration algorithms to disentangle regional dose bioeffects. Sci Rep 2018;8:4915. https://doi.org/10.1038/s41598-018-23327-0.