Physics in Medicine & Biology

New target volume delineation and PTV strategies to further personalise radiotherapy

To cite this article before publication: David Bernstein *et al* 2021 *Phys. Med. Biol.* in press https://doi.org/10.1088/1361-6560/abe029

Title: New target volume delineation and PTV strategies to further personalise radiotherapy

1 Authors

David Bernstein [a], Dr Alexandra Taylor [b], Dr Simeon Nill [c], Prof. Uwe Oelfke [c].

a) Joint Department of Physics, The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, Fulham Road, London, SW3 6JJ, United Kingdom

b) Gynaecology Unit, Royal Marsden NHS Foundation Trust, Fulham Road, London, SW3 6JJ, United Kingdom

c) Joint Department of Physics, The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London, SM2 5PT, United Kingdom

David Bernstein orchid: https://orcid.org/0000-0001-7442-0698

2 Keywords

Target Delineation. Delineation Uncertainty. PTV Margins. Geometric Uncertainties. Personalised Medicine.

3 Abstract

Target volume delineation uncertainty (DU) is arguably one of the largest geometric uncertainties in radiotherapy that are accounted for using Planning Target Volume (PTV) margins. Geometrical uncertainties are typically derived from a limited sample of patients. Consequently, the resultant margins are not tailored to individual patients. Furthermore, standard PTVs cannot account for arbitrary anisotropic extensions of the target volume originating from DU. We address these limitations by developing a method to measure DU for each patient by a single clinician. This information is then used to produce PTVs that account

for each patient's unique DU, including any required anisotropic component. We do so using a two-step uncertainty evaluation strategy that does not rely on multiple samples of data to capture the DU of a patient's Gross Tumour Volume (GTV) or clinical target volume (CTV). For simplicity, we will just refer to the GTV in the following. First, the clinician delineates two contour sets; one which bounds all voxels believed to have a probability of belonging to the GTV of 1, while the second includes all voxels with a probability greater than 0. Next, one specifies a probability density function for the true GTV boundary position within the boundaries of the two contours. Finally, a patient-specific PTV, designed to account for all systematic errors, is created using this information along with measurements of the other systematic errors. Clinical examples indicate that our margin strategy can produce significantly smaller PTVs than the van Herk margin recipe. Our new radiotherapy target delineation concept allows delineation uncertainties to be quantified by the clinician for each patient, leading to PTV margins that are tailored to each unique patient, thus paving the way to a greater personalisation of radiotherapy.

4 Introduction

In radiotherapy, dose distributions are designed for each patient with the aim of achieving acceptable probabilities of tumour control and normal-tissue toxicity. The target volumes consist of the Gross Tumour Volume (GTV) and the Clinical Target Volume (CTV), as defined by ICRU (ICRU, 2010).

The Planning Target Volume (PTV) accounts for the various geometrical uncertainties, which limit the accuracy and precision of delivering planned doses to tumour targets. The PTV is a geometrical concept, created by enlarging the CTV by a margin that is designed to ensure the CTV is covered by the intended dose over the course of treatment, with a predefined level of confidence.

Target volume delineation is a fundamental source of geometrical targeting uncertainty, often being a dominant contributor to the PTV (Njeh, 2008; Segedin and Petric, 2016; Thwaites, 2013). The magnitude of delineation uncertainty (DU) is currently estimated by measuring variations between contours produced by different observers (BIR, 2003; Tudor et al., 2020). This inter-observer variability is reported in the literature for a range of tumour sites (Chung et al., 2012; Duane et al., 2014; Feng et al., 2012; Hellebust et al., 2013; Leunens et al., 1993; Li et al., 2009; Logue et al., 1998; Meijer et al., 2003; Persson et al., 2011; Petric et al., 2013; Peulen et al., 2015; Segedin and Petric, 2016; Seravalli et al., 2015; Song et al., 2006; Weiss and Hess, 2003) and organs at risk (OARs) (Gay et al., 2012; Li et al., 2009; Sandström et al., 2016). A limitation of current practice is that DU is only measured for samples of patients, which prevents margins from accounting for DU associated with each individual patient.

The importance of margin anisotropies were demonstrated for both targets and OARs (Bell et al., 2016; Gurney-Champion et al., 2017; Meijer et al., 2003). However, a limitation in DU measurement and PTV growing tools provided by commercial Treatment Planning Systems is that margins can only account for uncertainties specified along three cardinal axes.

The common approach of estimating uncertainty via the statistical analysis of a series of measurements, such as delineations, is classified as a Type A evaluation of uncertainty (JCGM, 2012, 2008; Kuyatt, 1994). An alternative method, referred to as Type B evaluation, is based on scientific judgement using “all the available relevant information on the variability” of the quantity being measured (JCGM, 2012, 2008; Kuyatt, 1994). Both methods generate a standard deviation (SD) as an estimate of the SD for a population. Type B uncertainty estimates can be as reliable as Type A, particularly when Type A evaluations are derived from small sample sizes (JCGM, 2008; Kuyatt, 1994). Furthermore, Type A and Type B uncertainty estimates can be combined to give a combined uncertainty.

Our standard CTV-PTV margin is determined by the van Herk margin recipe (van Herk et al., 2000), henceforth referred to as MvHMR. In its simplest form the isotropic PTV-margin m, in the presence of systematic and random uncertainties, expressed by SDs Σ and σ respectively, is given by equation 1 where σ_p is a measure of the penumbra width. The parameters α and ß determine the confidence level that the CTV is covered by a specified isodose for a given fraction of the patient population.

$$m = \alpha\Sigma + \beta\sqrt{\sigma^2 + \sigma_p^2} - \beta\sigma_p \quad (1)$$

With respect to DU, the MvHMR is limited by two assumptions. First, it assumes that the patient population is sufficiently homogeneous such that Σ adequately represent the whole population even when measured in only a small sample of patients.

Second, the MvHMR assumes that all geometric uncertainties can be modelled by translations of the volume of interest (VOI). For target delineation, this implies that the clinician always delineates the target with the correct size and shape, but with errors simply being in its position. This is in conflict with the publications that show delineation error to be anisotropic, as described above. Therefore, the MvHMR is not designed to account for anisotropy which is known to exist for DU.

We propose to overcome these limitations by introducing a different strategy to account for DU based on a Type B uncertainty analysis. This concept does not rely on samples of patient populations whose targets are determined by a small number of clinicians. The key new feature is that the uncertainty is estimated by a single experienced observer, who delineates a finite boundary interval whose positions are assumed to be specified with a negligible uncertainty. The subsequent reduction of the boundary interval to a PTV is more versatile than standard margin recipes and allows an extended exploitation of personal patient images. It also accounts

for any anisotropy in the DU. We want to note that the outlined new methodology can be applied for the assessment of delineation uncertainties of GTV and/or CTV. For simplicity, we will just refer to the GTV in the following.

5 Method

5.1 Type B Delineation Method

In this section, we describe our Type B uncertainty analysis for the imprecisely known 'true' GTV (GTV_T).

5.1.1 Type B Uncertainty Evaluation

A Type B uncertainty evaluation of a measurand x, requires the following be determined or estimated based on scientific judgement using all the available information:

- Containment Limits: limits on the variation of x (Castrup, 2001)
- The probability density function (PDF), φ(x), that a value x within the containment limit coincides with the unknown true value
- Containment Probability: the probability that the true value can be found within the containment limit (Castrup, 2001). We assume this to be 1, unless stated otherwise.

5.1.2 Containment limits and containment probability

In this section, we describe how the containment limits and containment probability for the imprecisely known GTV are generated.

In order to cope with the uncertainty of the GTV_T boundaries, two target structures will be drawn. First, the clinician excludes all voxels within the patient image that certainly do not belong to the GTV_T. This outer, maximal target volume, referred to as the *Outer GTV* (GTV_O),

includes all voxels g with a non-vanishing probability of belonging to GTV_T. The second new target structure, encompassed by GTV_O, is the *Inner GTV* (GTV_I) which only incudes voxels g that are considered to be part of GTV_T with certainty. GTV_I and GTV_O define the containment limits of GTV_T. They divide the imaging information into three classes of voxels g according to their probability Q of belonging to GTV_T according to equation 2. We will refer to Q also as boundary probability. Q quantifies the chance of finding the boundary of GTV_T outside a given level set of Q.

$$Q(g \in GTV_T) = \begin{cases} 1 & \text{if } g \in GTV_I \\ \in (0,1] & \text{if } g \in GTV_O \\ 0 & \text{if } g \notin GTV_O \end{cases} \qquad (2)$$

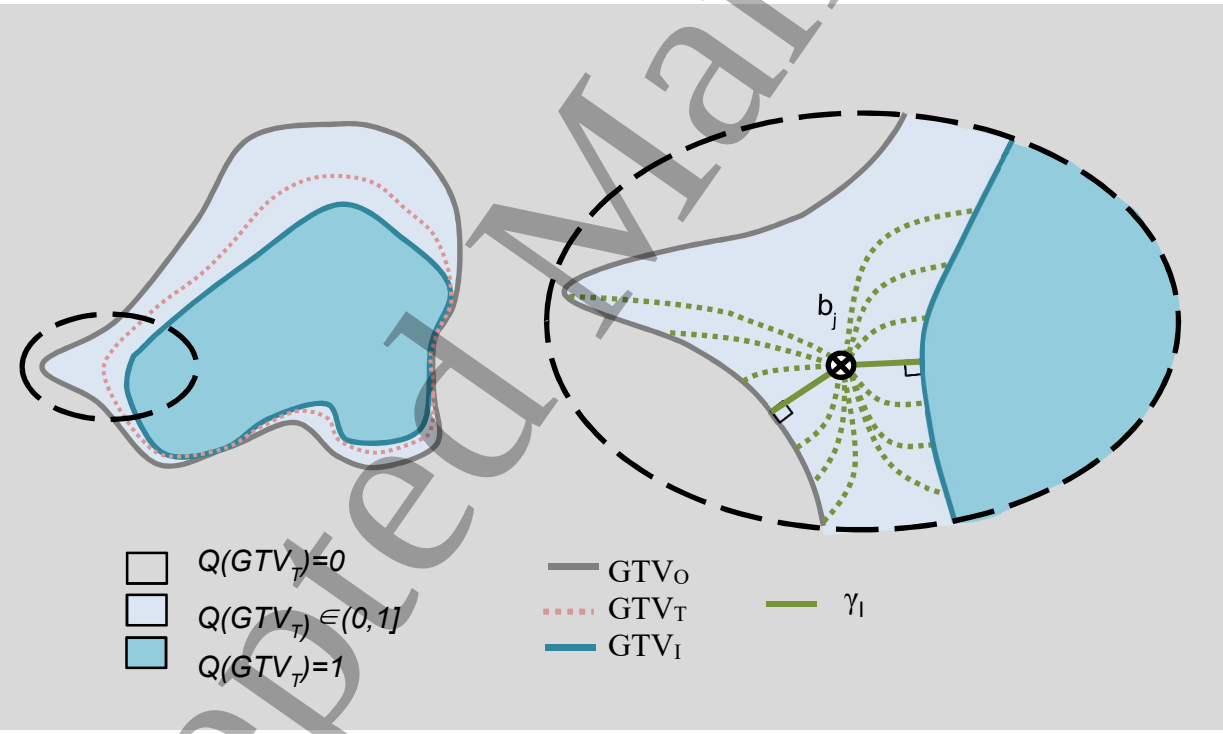

Figure 1: Left: an illustration of the containment limits and associated probabilities Q. Right, a close up illustration of the left dashed sector. Each dotted green curve illustrates a potential curve from the surface of GTV_I through the point b_j to the surface of GTV_O. The shortest curve from GTV_I through b_j to GTV_O, γ, is given by the solid green curve.

5.2 Boundary Probability Density Function (PDF)

The quantification of the boundary probability Q within the containment limits involves two essential components. First, the clinician is asked to describe the anticipated distribution of uncertainties for locating the true GTV in the boundary interval. This is facilitated by selection of a probability density function, φ, used to quantify Q(g) for each voxel g. Second, the numerical values of φ need to be derived from the surfaces of the containment limits GTV_I and GTV_O.

5.2.1 Generation of φ from the containment limits

The boundary probability function φ, related to any voxel g, is defined on the shortest trajectory γ_g connecting the surfaces of GTV_I and GTV_O and passing through g, where γ_g remains bounded by GTV_I and GTV_O along its whole length, as illustrated in figure 1. The length w_g of γ_g, referred to as the boundary width, consists of the sum of the minimal distances $d_{g,I}$ and $d_{g,O}$ of g from GTV_I and GTV_O respectively, measured whilst bounded by GTV_I and GTV_O:

$$w_g = d_{g,I} + d_{g,O} \tag{3}$$

The PDF φ depends on w_g and the length, d_g, of the trajectory measured from its starting point at GTV_I, i.e., $\varphi = \varphi(w_g, d_g)$. Its normalisation

$$\int_0^{w_g} dd_g \; \varphi(w_g, d_g) = 1 \tag{4}$$

states that for each trajectory, there will be one voxel belonging to the true boundary contour. While w_g introduces the absolute spatial scale for the specification of φ, we will work from now on with the relative distance $\rho = d_g / w_g$. The respective PDF $\varphi_r(\rho)$ is related to $\varphi(w_g, d_g)$ by

$$\varphi_r(\rho) = w_g\, \varphi\left(w_g, w_g \times \rho\right) \quad (5)$$

and satisfies the normalisation

$$\int_0^1 d\rho\ \ \varphi_r(\rho) = 1 \quad (6)$$

5.2.2 Selection of the probability density function φ_r

Now that the technical problem of defining the argument ρ of the PDF is addressed, we need to choose the detailed form of φ_r. The freedom of selecting the general form of φ_r opens the opportunity for the clinician to critically evaluate the uncertainty information contained within the boundary interval. Depending on image quality, individual patient anatomy and a reflection on the drawing process of GTV_I and GTV_O, the user can specify where they believe the true contour of GTV_T can be found. Many distributions exist that may suit this problem, however as a starting point, the following four distributions φ_r, covering a practical and plausible spectrum of functions, are offered as a choice:

1. Uniform Uncertainty

   This PDF indicates that the user assumes that any point within the boundary interval has the same chance of being part of the true contour, i.e.

   $$\varphi_r(\rho) = 1 \quad (7)$$

2. Linearly increasing from GTV_I to GTV_O

   In this case, the user believes that the true contour is located closer to GTV_O and that voxels adjacent to GTV_I are located further away from the boundary of GTV_T, i.e.

   $$\varphi_r(\rho) = 2\,\rho \quad (8)$$

3. Linearly decreasing from GTV_I to GTV_O

   The opposite bias, that GTV_T can be found closer to GTV_I, is represented by

$$\varphi_r(\rho) = 2\,(1-\rho) \tag{9}$$

The distributions of equations 8 and 9 are kept linear due to their simplicity.

4. Gaussian centred between GTV_I and GTV_O

Finally, the user can also indicate that, according their judgement, the true contour can be found within the central region of the boundary interval. This is presented by the Gaussian curve, centred around $\rho = 1/2$ and takes the form

$$\varphi_r(\rho) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(\rho-\frac{1}{2}\right)^2}{2\sigma^2}\right) \tag{10}$$

The width σ of the Gaussian determines the containment probability of the boundary interval, which is assumed to be close to 1. We chose $\sigma = 1/6$ leading to a containment probability of 0.997.

5.3 Creating the PTV

This section describes how to create a PTV using the information attained from the Type B delineation method above.

5.3.1 Boundary Probability Maps and PTV Generation

The cumulative boundary PDFs $C(\rho)$, gives the probability of finding the GTV_T boundary inside a shell of a constant level of relative distance ρ. It is given by:

$$C(\rho) = \int_0^{\rho} \varphi_r(\rho')\mathrm{d}\rho' \tag{11}$$

Whilst the probability of finding the boundary outside a shell of constant ρ is given by $q(\rho)$:

$$q(\rho) = 1 - C(\rho) \tag{12}$$

The surfaces of GTV_I/GTV_O refer to the largest/smallest shells with probabilities $q = 1$ and $q = 0$ respectively of finding the true boundary beyond these boundaries. The boundary

probability Q is defined for each voxel according to equation 13 and can be visualized as boundary probability map within each patient.

$$Q(g \in GTV_T) = \begin{cases} 1 & \text{if } g \in GTV_I \\ q(\rho) & \text{if } g \in B \\ 0 & \text{if } g \notin GTV_O \end{cases} \quad (13)$$

5.3.2 Coverage Probability Maps for Delineation Uncertainty

The accumulated pdf $C(\rho)$ in equation 11 is used to create the PTV for delineation uncertainty. If the dose level D_T is prescribed as an iso-dose to the surface defined by the condition

$$C_D(\rho) = \rho_C \quad (14)$$

then $C_D(\rho_c)$ represents the coverage probability of the target for the prescribed dose level and is referred to as PTV_D. The required value ρ_C depends on the selected coverage probability, C_D, and the explicit form of the boundary PDF $\varphi_r(\rho)$. Figure 2 illustrates the relationship between the coverage probability and fractional distance for the four specified boundary PDFs. The following equations define these relationships:

- Uniform: $C_D(\rho) = \rho$ (15)
- Linear Decreasing: $C_D(\rho) = 2\rho - \rho^2$ (16)
- Linear Increasing: $C_D(\rho) = \rho^2$ (17)
- Gaussian: $C_D(\rho) \cong \frac{1}{2}\left(1 + \text{erf}\left(\frac{n_S}{\sqrt{2}}\left(\rho - \frac{1}{2}\right)\right)\right)$ (18)

Where the value n_s is the number of standard deviations assumed to be contained within the boundary width for the Gaussian distribution. Note that equation 18 is correct to within 0.5% for $ns \geq 6$.

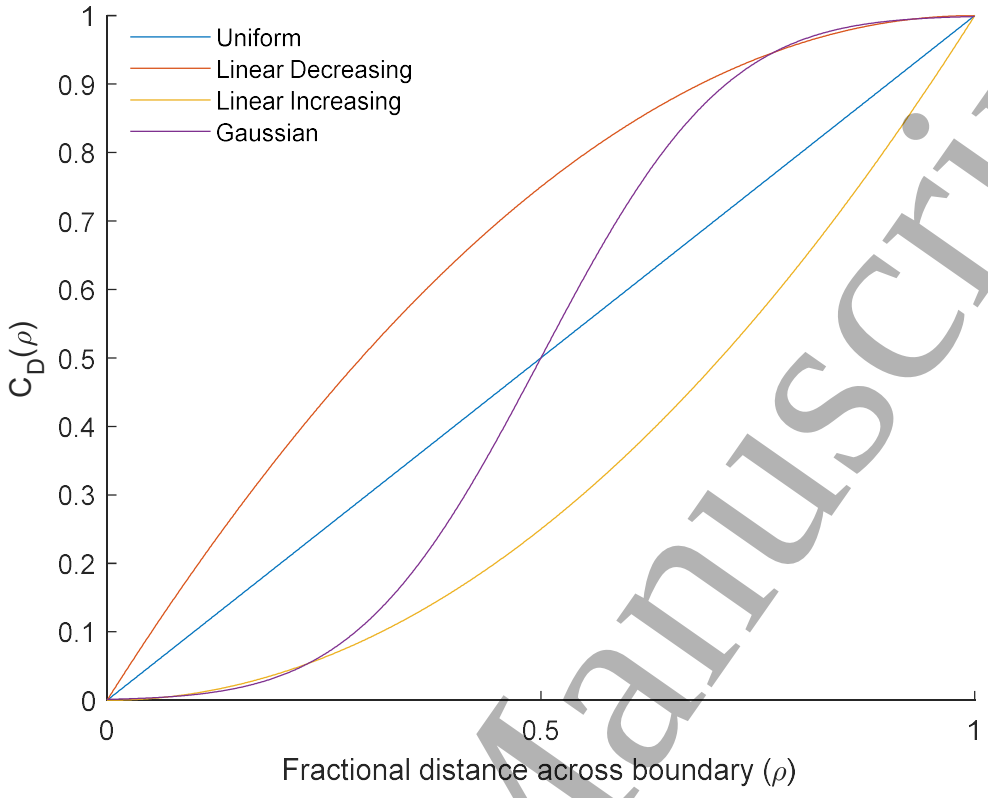

Figure 2: Coverage probability, C_D, of covering GTV_T by the prescribed isodose as a function of the position of the shell in terms of fractional distance, ρ.

5.3.3 Incorporating the remaining systematic errors into the PTV

PTV_D presented above is designed to account for delineation error only, with a confidence level of C_D. To account for all systematic errors, we follow the method proposed by Stroom et al (Stroom et al., 1999). Firstly, the boundary probability map, Q, is convolved with a probability distribution that describes the remaining systematic errors, resulting in a coverage probability map representing all systematic errors. The PTV corresponding to the desired coverage probability is given by the voxels bound by the corresponding level set on this coverage probability map. We refer to this final PTV as PTV_B, since it is stems from Type B uncertainty evaluations and to distinguish it from the concept of the PTV.

5.3.4 Clinical Examples

Two clinical examples are presented to illustrate the differences in PTV_B arising from the different boundary PDFs, and the differences with respect to the PTV created using the MvHMR (PTV_{MvH}). PTVs were created to give a 90 % coverage probability for PTV_B, and 90 % confidence level for PTV_{MvH}. The first case is a recurrent gynaecological cancer (RGC) GTV, and the second a prostate tumour. Typical cases were chosen based from a database selected for an extended clinical study yet to be published. Delineations were performed by clinicians highly experienced in treating these cases.

In these examples, we assumed the random errors to be negligible, i.e. to be 0. We assumed systematic errors, excluding delineation, to be 1.0 mm, based on data published by McNair et al for prostate treatments using fiducial markers and an online correction strategy (McNair et al., 2008). PTV_{MvH} was grown from the GTV delineated in accordance with local clinical protocols (GTV_C) for both examples.

The prostate example was delineated on CT alone. The delineation error used in the MvHMR was assumed to be 2.0 mm, based on data published by Alasti et al for the prostate delineated on CT (Alasti et al., 2017).

The RGC example was delineated on co-registered CT and MRI. The delineation error Σ_D used in the MvHMR were those measured locally of 2.9 mm in the superior-inferior axes, 2.2 mm in the left-right axes and anterior-posterior axes. Although the MvHMR is not strictly designed to be used with varying margins, we do so to reflect common clinical practice.

6 Results

Figures 3 and 4 show the outlines and PTVs for the RGC and prostate cases respectively. The gaps between GTV_I and GTV_O show that the clinicians had uncertainty when delineating these cases. It follows that the clinicians had to use their judgement to determine where to put the clinical GTV boundary given that uncertainty. The variation in the gaps demonstrates the uncertainty to be anisotropic. For the RGC case, the position of GTV_C varied with respect to GTV_I and GTV_O. In contrast, for the prostate case, the clinician took a conservative approach to delineating GTV_C by making it the largest volume given all their uncertainties, and therefore it coincided with GTV_O.

Figures 3B-E and 4B-E show that PTV_B is significantly smaller than PTV_{MvH} at all points in the target. These also show PTV_B to be more anisotropic due to the variation in its distances from GTV_C.

Figures 3C-E and 4C-E also illustrate the different PTV_B boundary positions resulting from the different boundary PDFs. Naturally, we observe that the difference is negligible where the boundary width is small. As the boundary width increases, the uniform and linear increasing PTV_B's remain similar to each other, but become increasingly larger than the Gaussian and linear decreasing PTV_B's.

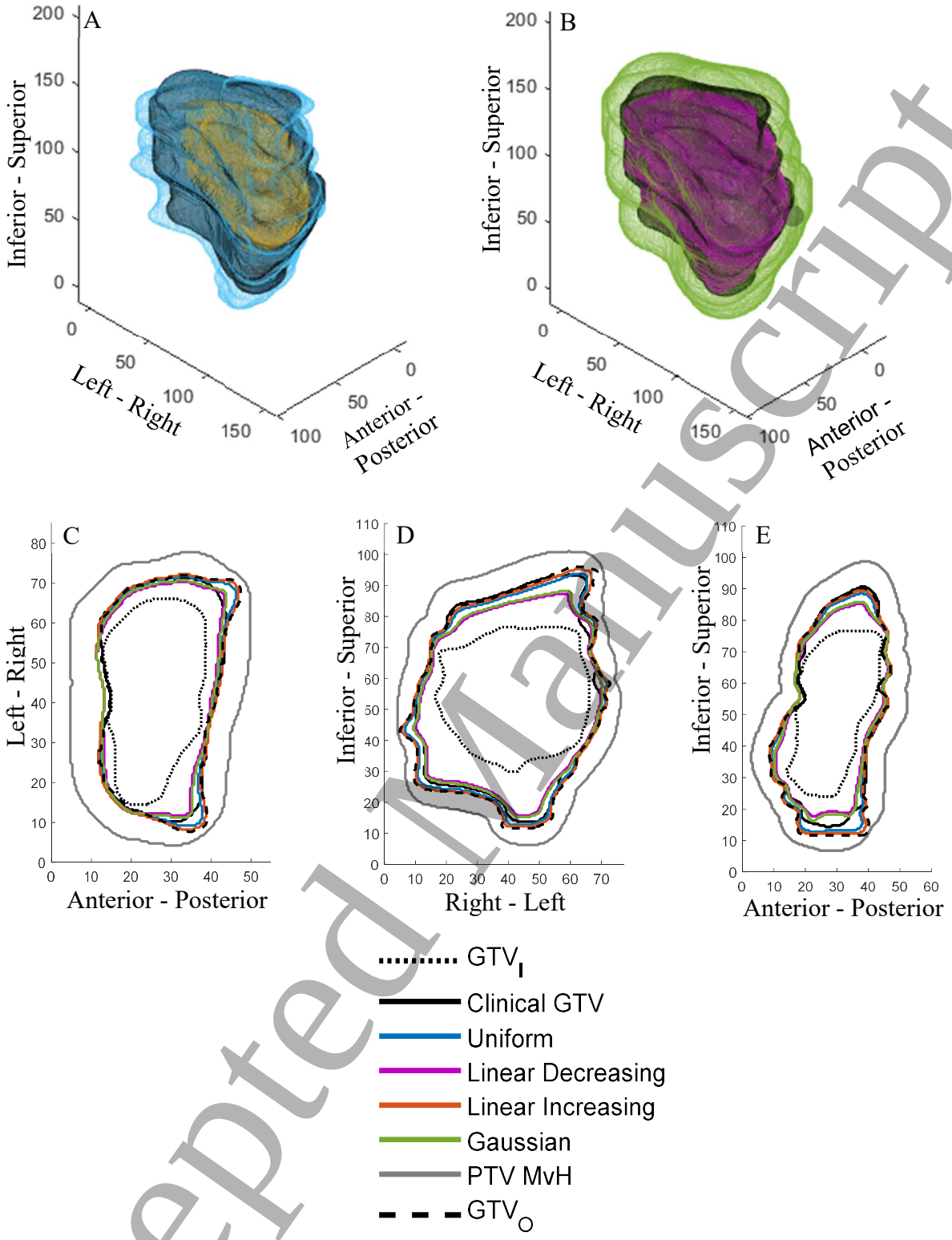

Figure 3: Clinical gynaecological example. A: 3D rendering of GTV_I (yellow), the clinical GTV_C (black) and GTV_O (light blue). B: 3D rendering of PTV_B (pink), using the Gaussian PDF, the clinical GTV_C (black) and PTV_{MvH} (green). C, D and E: GTV_I and GTV_O contours, PTV_B and PTV_{MvH} boundaries through central axial, coronal and sagittal slices respectively. Axes are in mm.

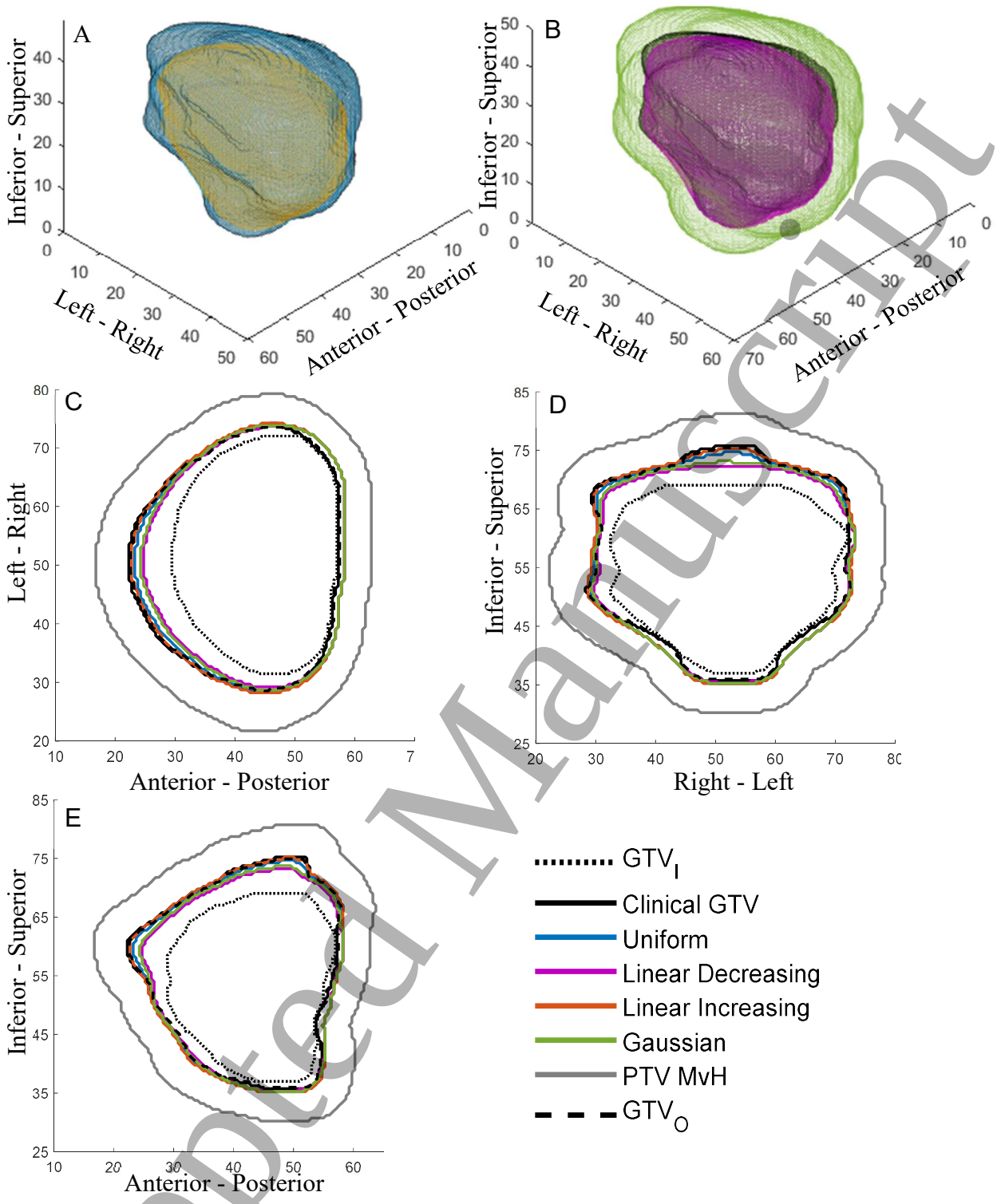

Figure 4: Clinical prostate example. A: 3D rendering of GTV_I (yellow), the clinical GTV_C (black) and GTV_O (light blue). Note that GTV_C and GTV_O coincide over the majority of the surface. B: 3D rendering of PTV_B (pink), using the Gaussian PDF, the clinical GTV_C (black) and PTV_{MvH} (green). C, D and E: GTV_I and GTV_O contours, PTV_B and PTV_{MvH} boundaries through central axial, coronal and sagittal slices respectively. Axes are in mm.

7 Discussion

In current practice, geometric uncertainties in radiotherapy are mitigated by adding PTV margins. These are based on estimates of the magnitude of each source of uncertainty, which are combined according to a margin recipe. The quality of each uncertainty estimate affects the appropriateness of the determined margin. The primary aim of this paper was to create PTVs that are based on the delineation uncertainty associated with each individual patient, as opposed to a sample of patients from a population, secondly to produce PTVs that reflect the anisotropy in the uncertainty, whilst thirdly, still accounting for all sources of systematic geometric uncertainties.

Population delineation error is currently estimated using a Type A approach, in which multiple clinicians delineate several cases. In this paper, we present a method for measuring delineation error based on Type B uncertainty evaluation methods. In this approach, the clinician uses their knowledge and experience to delineate the containment limits for the GTV (i.e. GTV_I and GTV_O) and specify the PDF for the unknown 'true' GTV (i.e. GTV_T) boundary based on where they believe the most likely true boundary position to be.

The effectiveness of the proposed method is dependent on the quality of the Type B uncertainty evaluation. Therefore, it is important that clinicians have the necessary knowledge and experience to delineate the containment limits. Even with this condition met, observer variability is likely to affect GTV_I and GTV_O delineations since delineating them is ultimately a subjective process. Prior to using the approach presented here, it is important that measures are implemented to; ensure clinicians have the necessary knowledge and experience, minimise any observer variability, and maximise consistency between clinicians. We hypothesise that measures used to reduce observer variability when standard delineation protocols are used, will

also be effective in reducing any observer variability in the delineated containment limits and ultimately improve the quality of the Type B uncertainty evaluation. For example, interventions recommended by Chang et al (Chang et al., 2017) and Vinod et al (Vinod et al., 2016b) could be developed and applied, these include; the use of atlases and guidelines, teaching for example through workshops, and peer review of outlines.

This Type B approach is proposed to address several limitations associated with using Type A methods for measuring delineation error in the clinic. The first limitation is its resource intensive nature; this arises from the need for multiple clinicians to delineate each case used for the uncertainty estimate. Performing studies in this way can also be logistically challenging due to the limited availability of clinicians. Limited resources restricts us to estimating delineation error on only a sample of patients. This prevents the creation of margins tailored to each individual patient. By using a Type B approach, a single clinician can estimate the delineation error for each patient, without relying on multiple clinicians. This approach would be less resource intensive than the Type A approach and so it may be feasible to do this for each and every patient in a clinical setting.

The second limitation the Type B delineation method aims to address is that DU is typically only measured along the cardinal axes, therefore, general information on the anisotropic nature of the uncertainty is lost. There are several studies in which anisotropic delineation error has been recorded (Bell et al., 2016; Deurloo et al., 2005; Meijer et al., 2003; Peulen et al., 2015; Remeijer et al., 1999). However, these all rely on multiple observers making it unfeasible to do so routinely for each patient.

Some studies assume adequate spatial correlation in delineation error between patients in order to produce anisotropic PTVs (Bell et al., 2016; Nijkamp et al., 2012). This assumption is unlikely to be valid for all tumour sites, in such cases anisotropic population-based margins

would be inappropriate. Xu et al (Xu et al., 2015) present an alternative approach to estimating delineation error on a patient-by-patient basis for prostate cancer. To do this they proposed using the contours produced by a single observer, along with the contrast in the CT image, to model the delineation uncertainty. The authors then used coverage probability techniques to produce PTVs, which resulted in plans with improved target and/or OAR doses when compared with PTVs created using the MvHMR. One of the key limitations of their method is the reliance on a model to estimate delineation uncertainty, as opposed to using clinical data and clinician knowledge as done in this paper.

It is also not clear how delineation uncertainty would be modelled when multi-modality imaging is used for delineation, as is often the case in modern radiotherapy, or how well the model would translate to other tumour sites. These limitations will affect the produced PTVs. By measuring the uncertainty on a patient-by-patient basis, for example using the Type B method presented here, the delineation uncertainty can be measured anisotropically for each individual VOI by a single clinician regardless of anatomical site or imaging modalities used. The methods presented here could then be used to account for those uncertainties without also needing to make assumptions about the spatial correlations in Σ_D between patients. Furthermore, unlike the methods in this paper, Xu et al do not show how to account for other sources of systematic geometric uncertainties.

The third limitation the Type B delineation method aims to address is associated with the sample sizes used to measure delineation error. Inter-observer variability studies, which are used to measure delineation error, are generally limited by having a small number of observers. As described in the introduction, the uncertainty associated with a population standard deviation estimate depends on the sample size used, with the uncertainty decreasing with increasing sample size. For example, the authors of a review of 131 DU publications, showed

them to have a median of 9 participants for GTV delineation error assessments (Vinod et al., 2016a). The Chi-Square distribution shows that for a study with nine observers, the 95% confidence interval for the SD is 0.68 to 1.92 times the measured SD, which translates into an arguable significant uncertainty in the calculated PTV margin. The consequence of small sample sizes, and the reliance on sample data, means that Type B uncertainty evaluations can be as reliable as Type A evaluations (JCGM, 2008; Kuyatt, 1994).

The approach of delineating the inner and outer limits of a target has been previously presented (Waschek et al., 1997). In that paper, the authors used fuzzy logic to derive a PTV based on estimated impact of including different voxels within the PTV on tumour control and normal tissue complication probabilities. Using fuzzy logic to determine the PTV is conceptually very different to the geometric uncertainty based methods widely used, such as MvHMR, which is perhaps why such methods are rarely considered in clinical practice. The methods presented in this paper have the advantage that they are consistent with our standard geometric uncertainty based methods. Another advantage of the methods in this paper is that they account for all sources of systematic geometric uncertainty, unlike the fuzzy logic method.

Clinical examples presented to illustrate the methods and concepts, show that the clinician can establish regions of uncertainty using the method presented in this paper. They show that, unlike MvHMR, PTV_B mirrors the anisotropy in the delineation uncertainty. They show that, regardless of the boundary PDF used, PTV_B seems to be consistently smaller than PTV_{MvH}. This reduction is a result of addressing the assumptions made in the MvHMR that are not valid for DU. The reduction presents the potential for dose escalation to the tumour and/or a reduction in toxicities by reducing the dose to surrounding tissue. However, as with any new technique, these methods should be assessed through a clinical study to ensure there are no unintended consequences from any reduction in the margin.

We have assumed that the selected PDF is appropriate for the whole VOI. This is a pragmatic assumption as defining multiple PDFs for one VOI would not be practical in a clinical setting, and may not result in a significant benefit. Four PDF options were presented to illustrate the methodology, however, there may be alternative appropriate distributions. The clinical examples show that where the boundary widths are narrow, the differences between the PTVs resulting from the different boundary PDFs are negligible. As the boundary width increases, the PTVs separate into two groups, with the uniform and linear increasing PDFs resulting in larger PTVs than the Gaussian and linear decreasing. These differences show that selecting an appropriate PDF can be important for larger boundary widths, as it can affect the PTV boundary position.

The method used to create PTV_D uses the concept of shells in a similar manner to Shusharina et al (Shusharina et al., 2018), who used shells in place of a CTV and applied this concept for treatment optimization. Unfortunately, the authors do not show how these shells are derived in detail.

Where an expansion is required from the GTV to account for macroscopic spread, giving a CTV, any uncertainty in the expansion required will add to the overall uncertainty in the delineated target. In our approach, we do not aim to address the problem of uncertainty in GTV to CTV expansion. Instead, we rely on the clinician to consider it when delineating the containment limits for the CTV, i.e. CTV_I and CTV_O. The CTV containment limits may be delineated directly, or by expanding GTV_I and GTV_O by the required GTV to CTV margin and modifying them according to anatomical boundaries if required. This approach is consistent with that taken in the MvHMR, in which uncertainty in the final CTV outline is considered, rather than any uncertainty in the GTV to CTV expansion. A possible future extension of the method present here would be to incorporate uncertainty in the GTV to CTV expansion and/or

information on the distribution of microscopic disease around the GTV, for example as done by Stroom et al (Stroom et al., 2014) who developed a GTV to PTV margin.

8 Conclusions

A new concept for radiotherapy target delineation and how to design a corresponding PTV were presented to address several shortcomings of currently used margin recipes. The key innovative feature is that delineation uncertainties are quantified by the clinician for each patient, leading to PTV margins that are tailored to the unique patient and the set of images representing the radiotherapy relevant anatomy, thus paving the way to a greater personalisation of radiotherapy. The two clinical examples considered seem to indicate that conventional margin strategies are less flexible and may be too conservative in ensuring dose coverage of the radiation target.

9 Acknowledgements

This publication presents independent research part funded by the National Institute for Health Research (NIHR, HCS DRF-2014-05-005). This research was also part funded by the NIHR Biomedical Research Centre at The Royal Marsden and The Institute of Cancer Research. Research at The Institute of Cancer Research is also supported by Cancer Research UK under Programme C33589/A19727. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

10 References

Alasti, H., Cho, Y.-B., Catton, C., Berlin, A., Chung, P., Bayley, A., Vandermeer, A., Kong, V., Jaffray, D., 2017. Evaluation of high dose volumetric CT to reduce inter-observer delineation variability and PTV margins for prostate cancer radiotherapy. Radiother Oncol 125, 118–123. https://doi.org/10.1016/j.radonc.2017.08.012

Bell, L.R., Pogson, E.M., Metcalfe, P.E., Holloway, L.C., 2016. Defining and assessing an anisotropic delineation margin for modern radiotherapy. Med Phys 43, 6644. https://doi.org/10.1118/1.4967942

BIR, 2003. Geometric Uncertainties in Radiotherapy—Defining the Planning Target Volume. The British Institute of Radiology.

Castrup, H., 2001. Distributions for Uncertainty Analysis. Proceedings of the International Dimensional Workshop.

Chang, A.T.Y., Tan, L.T., Duke, S., Ng, W.-T., 2017. Challenges for Quality Assurance of Target Volume Delineation in Clinical Trials. Frontiers in Oncology 7, 221. https://doi.org/10.3389/fonc.2017.00221

Chung, E., Stenmark, M.H., Evans, C., Narayana, V., McLaughlin, P.W., 2012. Expansion/de-expansion tool to quantify the accuracy of prostate contours. Int. J. Radiat. Oncol. Biol. Phys. 83, 33–7. https://doi.org/10.1016/j.ijrobp.2011.05.040

Deurloo, K.E.I., Steenbakkers, R.J.H.M., Zijp, L.J., de Bois, J.A., Nowak, P.J.C.M., Rasch, C.R.N., van Herk, M., 2005. Quantification of shape variation of prostate and seminal vesicles during external beam radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 61, 228–238. https://doi.org/10.1016/j.ijrobp.2004.09.023

Duane, F.K., Langan, B., Gillham, C., Walsh, L., Rangaswamy, G., Lyons, C., Dunne, M., Walker, C., McArdle, O., 2014. Impact of delineation uncertainties on dose to organs at risk in CT-guided intracavitary brachytherapy. Brachytherapy 13, 210–8. https://doi.org/10.1016/j.brachy.2013.08.010

Feng, M., Demiroz, C., Vineberg, K.A., Eisbruch, A., Balter, J.M., 2012. Normal tissue anatomy for oropharyngeal cancer: contouring variability and its impact on optimization. Int. J. Radiat. Oncol. Biol. Phys. 84, e245-9. https://doi.org/10.1016/j.ijrobp.2012.03.031

Gay, H.A., Barthold, H.J., O'Meara, E., Bosch, W.R., El Naqa, I., Al-Lozi, R., Rosenthal, S.A., Lawton, C., Lee, W.R., Sandler, H., Zietman, A., Myerson, R., Dawson, L.A., Willett, C., Kachnic, L.A., Jhingran, A., Portelance, L., Ryu, J., Small Jr, W., Gaffney, D., Viswanathan, A.N., Michalski, J.M., 2012. Pelvic Normal Tissue Contouring Guidelines for Radiation Therapy: A Radiation Therapy Oncology Group Consensus Panel Atlas. Int J Radiat Oncol Biol Phys 83, e353–e362. https://doi.org/10.1016/j.ijrobp.2012.01.023

Gurney-Champion, O.J., Versteijne, E., van der Horst, A., Lens, E., Rütten, H., Heerkens, H.D., Paardekooper, G.M.R.M., Berbee, M., Rasch, C.R.N., Stoker, J., Engelbrecht, M.R.W., van Herk, M., Nederveen, A.J., Klaassen, R., van Laarhoven, H.W.M., van Tienhoven, G., Bel, A., 2017. Addition of MRI for CT-based pancreatic tumor delineation: a feasibility study. Acta Oncol 56, 923–930. https://doi.org/10.1080/0284186X.2017.1304654

Hellebust, T.P., Tanderup, K., Lervag, C., Fidarova, E., Berger, D., Malinen, E., Potter, R., Petric, P., 2013. Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy. Radiother Oncol 107, 13–9. https://doi.org/10.1016/j.radonc.2012.12.017

ICRU, 2010. Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT): Contents. Journal of the ICRU 10, NP. https://doi.org/10.1093/jicru/ndq002

JCGM, 2012. 200:2012 International Vocabulary of Metrology - Basic and General Concepts and Assocaited Terms (VIM). 3rd Edition. Joint Committee for Guides in Metrology.

JCGM, 2008. 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement JCGM 100:2008 (GUM 1995 with minor corrections). BIPM Joint Committee for Guides in Metrology Paris.

Kuyatt, B.N.T. and C.E., 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. NIST Technical Note 1297, Gaithersburg, MD.

Leunens, G., Menten, J., Weltens, C., Verstraete, J., van der Schueren, E., 1993. Quality assessment of medical decision making in radiation oncology: variability in target volume delineation for brain tumours. Radiother Oncol 29, 169–75.

Li, X.A., Tai, A., Arthur, D.W., Buchholz, T.A., Macdonald, S., Marks, L.B., Moran, J.M., Pierce, L.J., Rabinovitch, R., Taghian, A., Vicini, F., Woodward, W., White, J.R., Radiation Therapy Oncology Group, M.-I., Multiobserver, S., 2009. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study. Int. J. Radiat. Oncol. Biol. Phys. 73, 944–51. https://doi.org/10.1016/j.ijrobp.2008.10.034

Logue, J.P., Sharrock, C.L., Cowan, R.A., Read, G., Marrs, J., Mott, D., 1998. Clinical variability of target volume description in conformal radiotherapy planning. Int J Radiat Oncol Biol Phys 41, 929–31.

McNair, H.A., Hansen, V.N., Parker, C.C., Evans, P.M., Norman, A., Miles, E., Harris, E.J., Del-Acroix, L., Smith, E., Keane, R., Khoo, V.S., Thompson, A.C., Dearnaley, D.P., 2008. A Comparison of the Use of Bony Anatomy and Internal Markers for Offline Verification and an Evaluation of the Potential Benefit of Online and Offline Verification Protocols for Prostate Radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 71, 41–50. https://doi.org/10.1016/j.ijrobp.2007.09.002

Meijer, G.J., Rasch, C., Remeijer, P., Lebesque, J.V., 2003. Three-dimensional analysis of delineation errors, setup errors, and organ motion during radiotherapy of bladder cancer. Int. J. Radiat. Oncol. Biol. Phys. 55, 1277–1287. https://doi.org/10.1016/s0360-3016(02)04162-7

Nijkamp, J., Swellengrebel, M., Hollmann, B., de Jong, R., Marijnen, C., van Vliet-Vroegindeweij, C., van Triest, B., van Herk, M., Sonke, J.-J., 2012. Repeat CT assessed CTV variation and PTV margins for short- and long-course pre-operative RT of rectal cancer. Radiat Oncol 102, 399–405. https://doi.org/10.1016/j.radonc.2011.11.011

Njeh, C.F., 2008. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. J Med Phys 33, 136–40. https://doi.org/10.4103/0971-6203.44472

Persson, G.F., Nygaard, D.E., Munck Af Rosenschold, P., Richter Vogelius, I., Josipovic, M., Specht, L., Korreman, S.S., 2011. Artifacts in conventional computed tomography (CT) and free breathing four-dimensional CT induce uncertainty in gross tumor volume determination. Int. J. Radiat. Oncol. Biol. Phys. 80, 1573–80. https://doi.org/10.1016/j.ijrobp.2010.10.036

Petric, P., Hudej, R., Rogelj, P., Blas, M., Tanderup, K., Fidarova, E., Kirisits, C., Berger, D., Dimopoulos, J.C., Potter, R., Hellebust, T.P., 2013. Uncertainties of target volume delineation in MRI guided adaptive brachytherapy of cervix cancer: a multi-institutional study. Radiother Oncol 107, 6–12. https://doi.org/10.1016/j.radonc.2013.01.014

Peulen, H., Belderbos, J., Guckenberger, M., Hope, A., Grills, I., van Herk, M., Sonke, J.J., 2015. Target delineation variability and corresponding margins of peripheral early stage NSCLC treated with stereotactic body radiotherapy. Radiother Oncol 114, 361–6. https://doi.org/10.1016/j.radonc.2015.02.011

Remeijer, P., Rasch, C., Lebesque, J.V., van Herk, M., 1999. A general methodology for three-dimensional analysis of variation in target volume delineation. Med Phys 26, 931–40. https://doi.org/10.1118/1.598485

Sandström, H., Chung, C., Jokura, H., Torrens, M., Jaffray, D., Toma-Dasu, I., 2016. Assessment of organs-at-risk contouring practices in radiosurgery institutions around

the world - The first initiative of the OAR Standardization Working Group. Radiat Oncol. https://doi.org/10.1016/j.radonc.2016.10.014

Segedin, B., Petric, P., 2016. Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them? Radiol Oncol 50, 254–62. https://doi.org/10.1515/raon-2016-0023

Seravalli, E., van Haaren, P.M., van der Toorn, P.P., Hurkmans, C.W., 2015. A comprehensive evaluation of treatment accuracy, including end-to-end tests and clinical data, applied to intracranial stereotactic radiotherapy. Radiother Oncol 116, 131–8. https://doi.org/10.1016/j.radonc.2015.06.004

Shusharina, N., Craft, D., Chen, Y.-L., Shih, H., Bortfeld, T., 2018. The clinical target distribution: a probabilistic alternative to the clinical target volume. Phys Med Biol 63, 155001.

Song, W.Y., Chiu, B., Bauman, G.S., Lock, M., Rodrigues, G., Ash, R., Lewis, C., Fenster, A., Battista, J.J., Van Dyk, J., 2006. Prostate contouring uncertainty in megavoltage computed tomography images acquired with a helical tomotherapy unit during image-guided radiation therapy. Int J Radiat Oncol Biol Phys 65, 595–607. https://doi.org/10.1016/j.ijrobp.2006.01.049

Stroom, J., Gilhuijs, K., Vieira, S., Chen, W., Salguero, J., Moser, E., Sonke, J.J., 2014. Combined recipe for clinical target volume and planning target volume margins. Int. J. Radiat. Oncol. Biol. Phys. 88, 708–14. https://doi.org/10.1016/j.ijrobp.2013.08.028

Stroom, J.C., de Boer, H.C., Huizenga, H., Visser, A.G., 1999. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. Int. J. Radiat. Oncol. Biol. Phys. 43, 905–19.

Thwaites, D., 2013. Accuracy required and achievable in radiotherapy dosimetry: have modern technology and techniques changed our views? J Phys Conf Ser 444, 012006.

Tudor, G., Bernstein, D., Riley, S., Rimmer, Y., Thomas, S., van Herk, M., Webster, A., 2020. Geometric Uncertainties in Daily Online IGRT: Refining the CTV–PTV Margin for Contemporary Photon Radiotherapy. British Institute of Radiology. https://doi.org/10.1259/geo-unc-igrt

van Herk, M., Remeijer, P., Rasch, C., Lebesque, J.V., 2000. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol Biol Phys 47, 1121–35.

Vinod, S.K., Jameson, M.G., Min, M., Holloway, L.C., 2016a. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. Radiother Oncol. https://doi.org/10.1016/j.radonc.2016.09.009

Vinod, S.K., Min, M., Jameson, M.G., Holloway, L.C., 2016b. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol 60, 393–406. https://doi.org/10.1111/1754-9485.12462

Waschek, T., Levegrün, S., Kampen, M. van, Glesner, M., Engenhart-Cabillic, R., Schlegel, W., 1997. Determination of target volumes for three-dimensional radiotherapy of cancer patients with a fuzzy system. Fuzzy Sets Syst 89, 361–370. https://doi.org/10.1016/S0165-0114(97)00026-2

Weiss, E., Hess, C.F., 2003. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. Strahlenther Onkol 179, 21–30. https://doi.org/10.1007/s00066-003-0976-5

Xu, H., Gordon, J.J., Siebers, J.V., 2015. Coverage-based treatment planning to accommodate delineation uncertainties in prostate cancer treatment. Med Phys 42, 5435–43. https://doi.org/10.1118/1.4928490