Technical Innovations & Patient Support in Radiation Oncology 3–4 (2017) 41–47

Contents lists available at ScienceDirect

Technical Innovations & Patient Support in Radiation Oncology

journal homepage: www.elsevier.com/locate/tipsro

Research article

A dose based approach for evaluation of inter-observer variations in target delineation

Ingrid Kristensen [a,b,*], Kristina Nilsson [c], Måns Agrup [d], Karin Belfrage [b], Anna Embring [e], Hedda Haugen [f], Anna-Maja Svärd [g], Tommy Knöös [b,h], Per Nilsson [b,h]

[a] Department of Oncology, Clinical Sciences, Lund University, Lund, Sweden
[b] Department of Haematology, Oncology and Radiation Physics, Skåne University Hospital, Lund, Sweden
[c] Department of Immunology, Genetics and Pathology, Experimental and Clinical Oncology, Clinical Oncology, Uppsala University Hospital, Uppsala, Sweden
[d] Department of Oncology, Linköping University Hospital, Linköping, Sweden
[e] Department of Oncology, Karolinska University Hospital, Stockholm, Sweden
[f] Department of Oncology, Sahlgrenska University Hospital, Gothenburg, Sweden
[g] Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden
[h] Department of Medical Radiation Physics, Clinical Sciences, Lund University, Lund, Sweden

ARTICLE INFO

Article history:
Received 24 August 2017
Received in revised form 6 October 2017
Accepted 9 October 2017
Available online 4 November 2017

Keywords:
Inter-observer variation
Pediatric radiotherapy
Target delineation
Treatment planning

ABSTRACT

Background and purpose: Substantial inter-observer variations in target delineation have been presented previously. Target delineation for paediatric cases is difficult due to the small number of children, the variation in paediatric targets, the number of study protocols, and the individual patient's specific needs and demands. Uncertainties in target delineation might lead to under-dosage or over-dosage. The aim of this work is to apply the concept of a consensus volume and good quality treatment plans to visualise and quantify inter-observer target delineation variations in dosimetric terms in addition to conventional geometrically based volume concordance indices.

Material and methods: Two paediatric cases were used to demonstrate the potential of adding dose metrics when evaluating target delineation diversity; Hodgkin's disease (case 1) and rhabdomyosarcoma of the parotid gland (case 2). The variability in target delineation (PTV delineations) between six centres was quantified using the generalised conformity index, *CIgen*, generated for volume overlap. The STAPLE algorithm, as implemented in CERR, was used for both cases to derive a consensus volumes. STAPLE is a probabilistic estimate of the true volume generated from all observers. Dose distributions created by each centre for the original target volumes were then applied to this consensus volume.

Results: A considerable variation in target segmentation was seen in both cases. For case 1 the variation was 374–960 cm^3 (average 669 cm^3) and for case 2; 65–126 cm^3 (average 109 cm^3). *CIgen* were 0.53 and 0.70, respectively. The DVHs in absolute volume displayed for the delineated target volume as well as for the consensus volume adds information on both "compliant" target volumes as well as outliers which are hidden with just the use of concordance indices.

Conclusions: The DVHs in absolute volume add valuable and easily understood information to various indices for evaluating uniformity in target delineation.

Introduction

Substantial inter-observer variations in target delineation have been presented in a number of previous studies [1–9]. The variations can be due to differences in interpretation of the diagnostic material, ambiguities in treatment protocols, lack of guidelines and/or inadequate, differences in local policies, the availability and use of multi-modality imaging, the subjective assessment of disease dissemination and/or the individual training and experience of the radiation oncologists. In a recent review, Vinod et al. [10] concluded that guidelines and atlases or atlas-based delineation tools would improve delineation [11,12], as well as training and the use of multi-modality imaging. Studies have also shown that delineation workshops [13] and peer reviews [14] can

* Corresponding author at: Department of Haematology, Oncology and Radiation Physics, Skåne University Hospital, Lund, Sweden.
E-mail address: ingrid.kristensen@skane.se (I. Kristensen).

https://doi.org/10.1016/j.tipsro.2017.10.002
2405-6324/

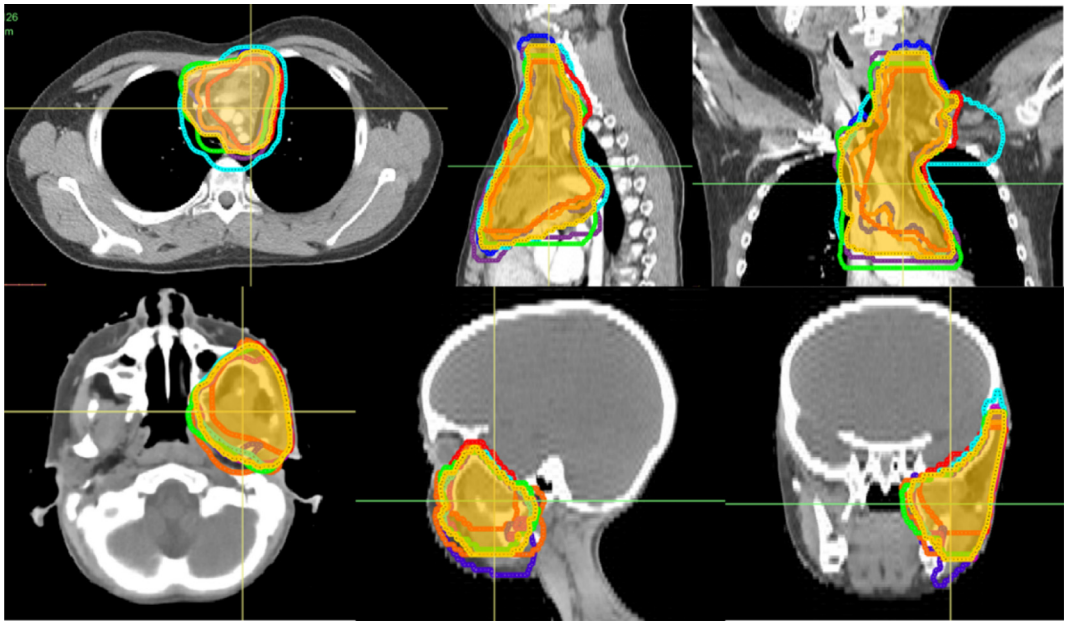

Fig. 1. Volume delineations fromall six centres for case 1 (top panel) and case 2 (bottom panel) as well as the consensus volume in transparent yellow.

Table 1
Volume related metrics for delineated target volumes.

	Case 1	Case 2
Volume (cm^3) average (range)	669 (374–960)	109 (65–126)
Intersection volume (cm^3)	293	53
Union volume (cm^3)	1189	131
CI_{gen}	0.53	0.70

Table 2
Volume related metrics for the STAPLE derived consensus volume.

	Case 1	Case 2
Volume (cm^3)	706	92
Agreement sensitivity (mean ± SD)	0.78 ± 0.20	0.88 ± 0.13
Agreement specificity (mean ± SD)	0.96 ± 0.03	0.98 ± 0.01
K	0.63	0.78

Table 3
Dose-volume metrics for each centre's target volume.

	Case 1 Average dose, (range)	Case 2 Average dose, (range)
$V_{95\%}$ (%)	91% (76–98)	95% (87–99)
$D_{98\%}$ (Gy)	18.3 (16.7–19.1)	38.9 (37.5–39.9)
$D_{50\%}$ (Gy)	19.9 (19.7–20.3)	41.5 (41.5–41.7)
$D_{2\%}$ (Gy)	20.6 (20.3–21.1)	42.8 (42.7–43.1)
HI	0.12 (0.09–0.19)	0.09 (0.07–0.13)
RCI	1.00 (0.90–1.16)	0.90 (0.68–1.09)

improve target delineation concordance and reduce inter-observer variability. Target delineation for paediatric cases is even more difficult due to the small number of children at most centres, the large variation in paediatric targets, the large number of study protocols, and the individual patient's specific needs and demands [15–17]. Uncertainties in target delineation might lead to under-dosage or over-dosage, causing a decrease in tumour control probability (TCP) or an increase in normal tissue complications (NTCP).

The evaluation of differences in segmented volumes in inter-observer studies can be done in several ways [18]. There is, however, no consensus among researchers on the methodologies to be applied and which metrics to report; e.g. differences in volume sizes, centre of mass variations, concordance indices, etc., making comparison between studies difficult to interpret. Valentini et al. describes a methodology for auto-segmentation which also could be used for studies on inter-observer variations [19]. Applying concordance indices is the most common method. It converts the variation in positions and sizes of delineated structures in relation to each other into a numerical value. The numerical value of different concordance indices are, however, dependent on the size of the structure studied and it is hence difficult to judge the resulting index value or when e.g. an improvement has occurred and to which degree. There is also an uncertainty in target delineation studies regarding which volume should be considered the "golden standard" or reference volume [20]. This volume is chosen in dissimilar ways in different studies. It could be segmented by an "expert" or a group of "experts". Another more objective method is to derive a "consensus volume" by applying an algorithm that computes a probabilistic estimate of the "true" segmentation based on the delineated volumes, e.g. STAPLE (Simultaneous Truth And Performance Level Estimation) [21]. This method has previously been introduced for radiotherapy [22] and used in target delineations studies [23–34].

Dose metrics are, however, not routinely reported in delineation studies, even though it might be helpful making the consequences of target delineation variations easier to interpret [20]. If treatment plans are created as a part of the target delineation process and these plans are clinically acceptable it would be an attractive complement to evaluate the quality of the resulting dose distribution on a consensus volume rather than only the volume metrics *per se*.

At an internal target delineation workshop, performed by The Swedish Workgroup for Paediatric Radiotherapy, these concepts were discussed. The group has previously performed and reported on an inter-observer study evaluated with conventional volume metrics [35].

The aim of this paper is to apply the concept of a consensus volume and good quality treatment plans for two paediatric cases to visualise target delineation variation in dosimetric terms in addition to conventional geometrically based volume concordance indices.