Evaluation of the impact of teaching on delineation variation during a virtual Stereotactic Ablative Radiotherapy contouring workshop

Finbar Slevin[a, b, †, #], Romélie Rieu[c, †], Matthew Beasley[a], Richard Speight[a], Katharine Aitken[c, d], James Good[e], Fiona McDonald[c, d], Thomas Rackley[f], Ganesh Radhakrishna[g], Anoop Haridass[h], Louise J Murray[a, b], Alison C Tree[c, d, ‡], Ann M Henry[a, b, ‡]

[a]Leeds Cancer Centre, Leeds Teaching Hospitals NHS Trust, Beckett Street, Leeds, LS9 7TF, UK
[b]University of Leeds, Leeds, LS2 9JT, UK
[c]The Royal Marsden NHS Foundation Trust, Fulham Road, London, SW3 6JJ, UK
[d]The Institute of Cancer Research, 123 Old Brompton Road, London, SW7 3RP
[e]University Hospitals Birmingham NHS Foundation Trust, Mindelsohn Way, Edgebaston, Birmingham, B15 2GW, UK
[f]Velindre Cancer Centre, Velindre Road, Whitchurch, Cardiff, CF14 2TL, UK
[g]The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK
[h]The Clatterbridge Cancer Centre NHS Foundation Trust, 65 Pembroke Place, Liverpool, L7 8YA, UK

[†]Joint first authors
[‡]Joint senior authors
[#]Corresponding author: Finbar Slevin
Address: Leeds Cancer Centre, Leeds Teaching Hospitals NHS Trust, Beckett Street, Leeds LS9 7TF, UK
Email address: finbarslevin@nhs.net

Romélie Rieu email address: romelie.rieu@nhs.net
Matthew Beasley email address: matthew.beasley@nhs.net
Richard Speight email address: richard.speight@nhs.net
Katharine Aitken email address: katharine.aitken@rmh.nhs.uk
James Good email address: james.good@uhb.nhs.uk
Fiona McDonald email address: fiona.mcdonald@rmh.nhs.uk
Thomas Rackley email address: thomas.rackley@wales.nhs.uk
Ganesh Radhakrishna email address: g.radhakrishna@nhs.net
Anoop Haridass email address: anoopharidass1@nhs.net
Louise J Murray email address: louise.murray8@nhs.net
Alison C Tree email address: alison.tree@icr.ac.uk
Ann M Henry email address: a.henry@leeds.ac.uk

Acknowledgements: The authors would like to acknowledge Scott Kaylor and EduCase™ for their support of the workshop/analysis of participant contouring data. We thank Paul Elbourn and team at Profile Productions, as well as the representatives at the Royal College of Radiologists and UK SABR Consortium for their assistance in developing and organising the workshop.

Declarations/conflicts of interests: James Good reports honoraria from ViewRay and GenesisCare. Fiona McDonald reports consulting for Accuray and

AstraZeneca, speaker fees from AstraZeneca and Elekta and a research grant from Merck Sharp & Dohme. Alison C Tree declares research funding from Elekta, Accuray and Varian, and honoraria from Elekta and Genesis healthcare.

Sources of financial support: Finbar Slevin is a Clinical Research Fellow supported by a Cancer Research UK Centres Network Accelerator Award to the ART-NET consortium (grant number A21993). Matthew Beasley and Richard Speight are supported by a Cancer Research UK Centres Network Accelerator Award to the ART-NET consortium (grant number A21993). Louise J Murray is an associate professor supported by a Yorkshire Cancer Research funded by Yorkshire Cancer Research (award number L389LM). Alison C Tree acknowledges support from Cancer Research UK (C33589/A28284) and the CRUK ICR/RMH RadNet centre. This project represents independent research supported by the National Institute for Health research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and the Institute of Cancer Research, London. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. Ann M Henry is supported by grants from Cancer Research UK (award number 108036), National Institute for Health Research (NIHR) (award number 111218), Medical Research Council (MRC) (award number 107154) and Sir John Fisher Foundation (charity, no award number).

Abstract

Introduction

Variation in delineation of target volumes/organs at risk (OARs) is well recognised in radiotherapy and may be reduced by several methods including teaching. We evaluated the impact of teaching on contouring variation for thoracic/pelvic Stereotactic Ablative Radiotherapy (SABR) during a virtual contouring workshop.

Materials and Methods

Target volume/OAR contours produced by workshop participants for three cases were evaluated against reference contours using DICE similarity co-efficient (DSC) and line domain error (LDE) metrics. Pre and post-workshop

DSC results were compared using Wilcoxon signed ranks test to determine the impact of teaching during the workshop.

Results

Of 50 workshop participants, paired pre and post-workshop contours were available for 21 (42%), 20 (40%) and 22 (44%) participants for primary lung cancer, pelvic bone metastasis and pelvic node metastasis cases respectively. Statistically significant improvements post-workshop in median DSC and LDE results were observed for 6 (50%) and 7 (58%) of 12 structures respectively, although the magnitude of DSC/LDE improvement was modest in most cases. An increase in median DSC post-workshop ≥0.05 was only observed for GTVbone, IGTVlung and SacralPlex and reduction in median LDE >1 mm was only observed for GTVbone, CTVbone and SacralPlex. Post-workshop, median DSC values were >0.7 for 75% of structures. For 92% of structures, post-workshop contours were considered to be acceptable or within acceptable variation following review by the workshop faculty.

Conclusion

This study has demonstrated that virtual SABR contouring training is feasible and was associated with some improvements in contouring variation for multiple target volumes/OARs.

Introduction

Delineation of target volumes and organs at risk (OARs) is a key component of radiotherapy planning, but inter/intra-observer variation in contouring is well recognised and is a significant source of error within treatment workflows[1, 2]. Potential reasons for this variation may include the influence of disease site experience/expertise and skills in cross-sectional image interpretation[2-4]. The consequences of contouring variation may be profound; incorrect delineation is associated with inferior survival outcomes in clinical trials[5, 6].

Various methods exist to minimise contouring variation including delineation protocols, atlases, auto-contours, peer review and teaching[2, 3, 7, 8]. Radiotherapy is a craft specialty, necessitating the acquisition and refinement of contouring skills during clinical practice[9]. To mitigate the potential impact on training of the reduction in junior doctor working hours, smarter and more efficient methods of delivering training are required[10]. Dedicated contouring workshops may be a valuable source of experiential learning especially concerning new radiotherapy techniques[11-13].

Following changes to the commissioning of Stereotactic Ablative Radiotherapy (SABR) in the UK, the Royal College of Radiologists (RCR) and UK SABR Consortium organised a workshop which focused on SABR contouring for lung cancer and bone and nodal oligometastatic disease[14]. The aim of the workshop was to share expertise and experience in SABR techniques and improve participants' contouring skills. Given the Covid-19 pandemic, the workshop took place in virtual format. In this study, we evaluated the impact of teaching during the workshop on contouring variation for multiple target volumes/OARs in the thorax and pelvis.

Methods and Materials

Format of the workshop

The workshop took place on 19th and 22nd October 2020; each session lasted two hours in duration and were delivered using Adobe® Connect™ (Adobe, San Jose, CA, USA). Participants were UK-based consultants in clinical oncology and the workshop was aimed at those without prior expertise in SABR, although relative baseline experience was not assessed/recorded prior to the workshop. Participants were asked to delineate target volumes/OARs for three cases prior to the workshop using the web-based platform EduCase™ (RadOnc eLearning Centre, Inc., Fremont, CA, USA). A video tutorial was provided, which explained how to use EduCase™.

The target volumes/OARs for the three cases were:

Right upper lobe primary lung cancer

- IGTVlung (internal target volume)
- BrachialPlex
- BronchusProx
- Oesophagus
- Spinal_Canal

Left pelvic bone metastasis secondary to breast cancer

- GTVbone (gross tumour volume)
- CTVbone (clinical target volume)
- Femur_Head_Left
- Rectum

Right common iliac lymph node secondary to prostate cancer

- GTVnode
- Bowel_Large
- SacralPlex

Each case was accompanied by a clinical vignette (history, diagnosis, investigations and intended treatment) and instructions detailing which structures were to be delineated and on which axial computed tomography (CT) slices. CT axial slice thickness was 3 mm for the lung cancer case and 1 mm for the bone/node cases. Image co-registration performed in EduCase™ between CT and positron emission tomography-computed tomography (PET-CT) was available for all cases, magnetic resonance imaging (MRI) was available for the nodal and pelvic bone cases, and 4DCT was available for the primary lung case. For the lung cancer case, IGTVlung could be defined on the maximum intensity projection (MIP) scan with reference to the average intensity projection, 0% and 30% respiratory phases.

Pre-workshop contours, anonymised to clinician, were reviewed across the two workshops and teaching was provided for each case including demonstration of a reference contour produced by the workshop faculty of UK consultant clinical oncologists with a combined total of consultant experience of approximately 50 years. Relevant published contouring guidance and atlases were identified during both sessions. Teaching included clinical cases to illustrate the general principles of patient selection, planning and treatment delivery of SABR for primary lung cancer and oligometastatic disease and a dedicated session for target volume/OAR contouring.

Following each workshop, participants were invited to review/adjust their contours based on the teaching. Final attempts could be submitted up to two weeks after the second workshop session, although no further contours eligible for inclusion in the study were submitted more than 3 days after the final workshop session. The faculty provided individual written feedback to participants on their post workshop contours (this information was not provided for pre-workshop contours).

Participants were asked to provide feedback for individual speaker sessions and the overall workshop experience using a 5-point Likert scale and free text responses.

Analysis of participant contours

Each participant’s contours were compared against a reference contour, which was produced by the clinician who led each case discussion during the workshop and peer reviewed by a second faculty member. For each structure, the specific axial CT slices to be contoured was specified; these were non-contiguous and therefore a volume was not obtained. Some participants had delineated contours on slices other than those specified in the case. Therefore, to ensure a fair comparison for all participants, only contours on those pre-specified slices were considered. Participants with only one set of contours (e.g. only pre-workshop contours) were excluded. Participants with two sets of submitted contours but where no changes were made to the post-workshop contours were included.

EduCase™ provides 2-dimensional (i.e. area) Dice similarity coefficient (DSC) and line domain error (LDE) values for individual slices for participant contours compared with the reference contour. DSC is an overlap measure, which measures the intersection of two contours relative to the union and ranges from 0 (zero overlap) to 1 (perfect overlap)[1, 15, 16]. DSC can be calculated by the following formula:

$$DSC = 2 \times (Area_{reference} \cap Area_{participant})/(Area_{reference} + Area_{participant})$$

[17, 18]

where $Area_{reference} \cap Area_{participant}$ is the intersecting overlap of the two areas and $Area_{reference} + Area_{participant}$ is the union of the two areas.

LDE is a distance metric within EduCase™, which measures the average absolute Euclidean distance in millimetres between corresponding points on the reference and participant contours.

Since each structure was not a volume but instead a series of individual slices, a summary measure per structure for each participant was produced. The median value of DSC/LDE for each of these slices was calculated for each of the structures contoured by each participant. These median structure DSC/LDE values for participants with both pre and post-workshop contours were exported into IBM-SPSS Statistics for Windows version 26 (IBM Corp., Armonk, NY, USA). Each of the included contours were reviewed by two of the authors (Finbar Slevin and Romélie Rieu) to identify potential reasons for low DSC/high LDE values.

Following the workshop, the faculty reviewed participants’ post-workshop contours and provided a score (acceptable, within acceptable variation or unacceptable) and written feedback.

Statistical considerations

The median DSC/LDE and inter-quartile range (IQR) are presented as summary statistics for all the participants' median structure DSC/LDE values pre and post-workshop, since a normal distribution of data could not be assumed and also to minimise the influence of outlying values. Box and whisker plots were produced by importing data into R 3.6.1 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria) using the ggplot2 library[19]. A statistical comparison of the median DSC/LDE for each participant's structures pre and post-workshop was undertaken using the Wilcoxon signed ranks test in SPSS, since this was paired data. A P value of <0.05 was taken to indicate a statistically significant difference.

Results

Fifty participants registered for the workshop and 43 submitted at least one set of contours for each of the cases. Of these 43 participants, 21 (49%), 20 (47%) and 22 (51%) participants produced pre and post-workshop contours for the lung cancer, pelvic bone metastasis and pelvic node metastasis cases respectively. A summary of the DSC/LDE values pre and post-workshop and results of statistical comparisons are shown in **Table 1**. The spread of the median DSC/LDE values for each structure across all of the participants is illustrated in **Figure 1**.

Table 1: Summary of median Dice similarity coefficient (DSC) and line domain error (LDE) measurements before/after teaching for each structure

Structure	***Number of participants***	***Median DSC pre (IQR)***	***Median DSC post (IQR)***	***P value from Wilcoxon signed ranks test (* indicates statistically significant result)***	***Median LDE pre (mm) (IQR)***	***Median LDE post (mm) (IQR)***	***P value from Wilcoxon signed ranks test (* indicates statistically significant result)***	***Comments***
GTVnode	21	0.74 (0.71-0.76)	0.75 (0.73-0.82)	0.003*	2.56 (2.23-2.76)	2.28 (1.85-2.61)	0.005*	
Bowel_Large	22	0.86 (0.72-0.87)	0.87 (0.82-0.88)	0.023*	3.61 (3.05-19.92)	3.31 (2.97-8.93)	0.028*	
SacralPlex	22	0 (0-0.04)	0.37 (0.21-0.68)	<0.001*	46.39 (33.11-	3.80 (2.31-	<0.001*	Some participants

					49.11)	32.56)		delineated SacralPlex based on CT atlas and others used MRI
GTVbone	20	0.77 (0.72-0.83)	0.85 (0.78-0.87)	0.002*	4.45 (3.15-4.94)	2.76 (2.39-3.70)	0.001*	
CTVbone	20	0.83 (0.78-0.87)	0.87 (0.83-0.88)	0.035*	3.73 (2.43-5.09)	2.53 (2.34-3.78)	0.037*	
Rectum	20	0.85 (0.78-0.88)	0.86 (0.81-0.89)	0.023*	2.81 (2.03-3.57)	2.40 (1.75-3.25	0.009*	
IGTVlung	20	0.71 (0.63-0.79)	0.76 (0.66-0.79)	0.311	2.07 (1.86-2.41)	1.94 (1.82-2.18)	0.029*	

BronchusProx	21	0.81 (0.72-0.83)	0.78 (0.72-0.84)	0.730	2.84 (2.37-3.80)	2.83 (2.41-3.87)	0.953	
Oesophagus	21	0.74 (0.66-0.79)	0.76 (0.66-0.81)	0.140	2.93 (2.50-3.06)	2.67 (2.20-3.02)	0.308	
Spinal_Canal	21	0.84 (0.83-0.85)	0.85 (0.83-0.86)	0.333	1.91 (1.80-2.09)	1.91 (1.78-2.10)	0.345	

LDE, line domain error; DSC, Dice similarity coefficient

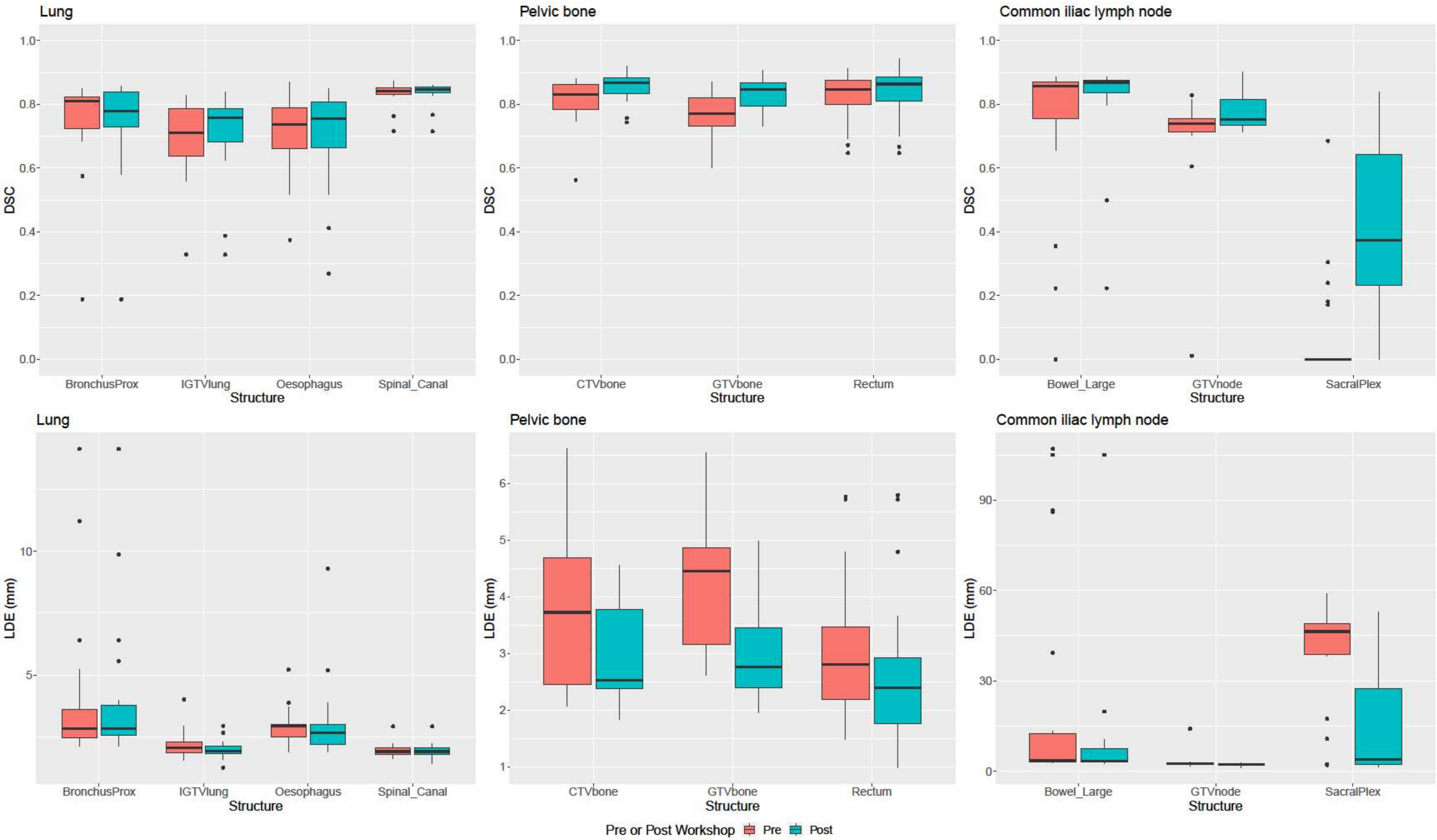
Lung
Pelvic bone
Common iliac lymph node
DSC
LDE (mm)
Structure
BronchusProx
IGTVlung
Oesophagus
Spinal_Canal
CTVbone
GTVbone
Rectum
Bowel_Large
GTVnode
SacralPlex
Pre or Post Workshop
Pre
Post

Figure 1: Box and whisker plots for the target volume/organs at risk structures for the lung cancer, pelvic bone metastasis and common iliac nodal metastasis cases. The top row represents Dice similarity coefficient (DSC) results and the bottom row represents line domain error (LDE) results. The box represents the middle 50% of the data and is bounded by the upper (Q3) and lower (Q1) quartiles and the horizontal line indicates the median value. The upper whiskers represent Q3+1.5*IQR and the lower whiskers represent Q1-1.5*IQR. Any outliers beyond these ranges are indicated as dots.

Statistically significant improvements in DSC post-workshop were observed for each structure except for IGTVlung, Spinal_Canal, Oesophagus and BronchusProx. Only BronchusProx was associated with a worsening in median DSC post-workshop, but this difference was not statistically significant. The magnitude of increase in DSC post-workshop was often small; only GTVbone (0.08), IGTVlung (0.05) and SacralPlex (0.37) were associated with a ≥0.05 increase in median DSC. A median value of DSC >0.7 and >0.8 post-workshop was observed for nine (75%) and five (42%) of the 12 structures respectively; no median DSC value was >0.9.

Statistically significant improvements in LDE post-workshop were observed for each structure except for BronchusProx, Oesophagus and Spinal_Canal. Similar to DSC results, BronchusProx was associated with a worsening in median LDE post-workshop although this difference was not statistically significant. Again, the magnitude of improvement was often small; only GTVbone (1.7 mm), CTVbone (1.2 mm) and SacralPlex (42 mm) were associated with >1 mm reduction in median LDE post-workshop.

Some post-workshop contours were unchanged from pre-workshop: GTVnode (5 participants, 24%), Bowel_Large (10 participants, 46%), GTVbone (2 participants, 10%), CTVbone (2 participants, 10%), Rectum (8 participants, 40%), IGTVlung (8 participants, 40%), Spinal_Canal (11 participants, 52%), Oesophagus (7 participants, 33%), BronchusProx (7 participants, 33%). When the data was re-analysed without these unchanged structures, no significant differences were observed.

Regarding BrachialPlex, the case instructions did not specify that only the ipsilateral structure was to be delineated and some participants contoured bilateral structures. Similarly for Femur_Head_Left, the femoral head (i.e. excluding the femoral neck) was to be delineated but several participants delineated both the femoral head and neck and/or produced bilateral structures. Therefore, these two structures were omitted from statistical comparisons.

Regarding post-workshop contours, a summary of the feedback provided to participants is shown in **Table 2**. Ninety-two per cent of post-workshop contours were considered to be acceptable or within acceptable variation.

Eighty-four per cent of participants provided feedback on the workshop; of these, feedback regarding the overall workshop experience and each of the individual speakers was considered to be ‘good’ or ’very good’ in 82% and 99% of responses respectively. Ten per cent of feedback concerned technical issues during the workshop (e.g. sound quality)

Table 2: Summary of qualitative feedback on participants' post-workshop contours

Structure	***Number of participants***	***Number of contours acceptable (%)***	***Number of contours within acceptable variation (%)***	***Number of contours unacceptable (%)***
GTVnode	11	6 (55%)	5 (45%)	0
Bowel_Large	11	4 (36%)	6 (55%)	1 (9%)
SacralPlex	11	9 (82%)	1 (9%)	1 (9%)
GTVbone	10	2 (20%)	7 (70%)	1 (10%)
CTVbone	10	4 (40%)	5 (50%)	1 (10%)
Femur_Head_Left	10	8 (80%)	2 (20%)	0
Rectum	10	6 (60%)	2 (20%)	2 (20%)

IGTVlung	13	6 (46%)	7 (54%)	0
BrachialPlex	13	0	9 (69%)	4 (31%)
BronchusProx	13	7 (54%)	5 (38%)	1 (8%)
Oesophagus	13	4 (31%)	9 (69%)	0
Spinal_Canal	13	12 (92%)	1 (8%)	0
Total contours	**138**	**68 (49%)**	**59 (43%)**	**11 (8%)**

Discussion

This study has evaluated the impact of teaching during a SABR contouring workshop for a relatively large number of participants and multiple target volume/OARs in the thorax and pelvis. The positive feedback provided by participants about the workshop suggests that it is feasible to deliver contouring teaching in a virtual capacity. This is important since, accelerated by the changes adopted during the Covid-19 pandemic, it is likely that medical training and teaching will increasingly be delivered using virtual methods. Another important message is that it is possible to continue to provide contouring training during the Covid-19 pandemic when in person meetings are restricted. Virtual training also has the potential to reach a larger audience, without geographical restrictions, than could typically be achieved in person. We demonstrated that median DSC/LDE values for participants who completed pre and post-workshop contours for most of the target volume/OARs were similar to the reference contour, with DSC >0.7 for 75% of structures and LDE <5 mm for 83% of structures. While statistically significant improvements post-workshop in DSC and LDE were observed for 50% and 58% of structures respectively, the magnitude of improvement was small in most cases and the clinical significance of such modest improvements remains uncertain.

Although multiple studies on the effect of teaching on contouring variation have been reported, several factors make direct comparison between these and our study challenging[16]. Heterogeneity exists between studies concerning the numbers of participants, types of teaching, the structures for which contouring variation is evaluated and the types of metrics used to evaluate this variation and the use of statistical tests[1, 2, 16]. However, systematic reviews of such studies have demonstrated that an improvement in contouring variation through teaching can be achieved[2, 20]. We did not observe a large increase in DSC/reduction in LDE post-workshop, and a number of limitations of our work may explain this. While participants were asked to review their pre-workshop contours after teaching and produce a post-workshop submission, only approximately half of participants did so which reduced the number for which an analysis of teaching impact could be performed. Furthermore, even for those who did re-submit a second set of contours in some cases no changes were made. Possible reasons for this could include satisfaction with pre-workshop contours, insufficient time to re-contour every structure and a lack of hands-on time during the workshop to practise/fully compare contours with the reference contour. The latter point may be particularly relevant since it has been previously suggested that active participation is more likely to improve learning during contouring workshops[21]. Insufficient provision of practical experience was raised as a potential explanation for failure to observe improved contouring post-teaching in a previous study of a head and neck contouring programme, although there may be time/resource challenges to effectively deliver this especially for larger audiences and during the Covid-19 pandemic where face-to-face meetings are restricted[22]. Residual differences in knowledge/ability between participants despite teaching were also suggested as a possible reason why significant improvements in prostate/rectal contouring were not observed in a previous evaluation of the impact of teaching[23].

Low DSC/high LDE values for certain structures in our study could be related to interpretation of the case instructions, especially for BrachialPlex and Femur_Head_Left. The latter structure was also only to be delineated on a single axial CT slice at the very inferior aspect of the femoral head. Different methods for contouring BrachialPlex exist, and there remains variation in practice[24-26]. Given the high dose per fraction used with SABR and variable reliance on MRI across different treatment centres, the UK SABR Consortium Guidance recommends contouring the subclavian/axillary vessels as a surrogate for BrachialPlex[26]. National consensus is needed, and future iterations of the recently published OAR harmonisation guidance will support this[25]. For SacralPlex, some participants delineated the visible nerve using the MRI while others delineated a larger surrogate structure using the CT. Both of these may be legitimate approaches, although contouring as per the Yi et al guidance does not rely on expert MRI interpretation of nerve position and may therefore be simpler for those learning[27]. However, unfamiliarity with the contouring of certain OARs might have contributed to low DSC/high LDE results. Although not statistically significant, the median DSC/LDE for BrochusProx appeared slightly worse post-workshop. The reason for this was not clearly apparent and the magnitude of difference was small, but it could possibly be related to delineation uncertainties regarding the distal extent of the lobar bronchi. A visual guide to delineation of BrachialPlex, BronchusProx and SacralPlex is illustrated in **Figure 2** while recommended contouring guidance/atlases are collated in **Table 3**[25-30].

Table 3: A summary of resources to support target volume/organ at risk delineation

Structure	***References***
Standardised nomenclature guidance	AAPM TG-263 [28]
GTVnode	UK SABR Consortium guidance version 6.1, 2019 [26]
GTVbone/CTVbone	De la Pinta, 2020 [29]
Lung primary	UK SABR Consortium guidance version 6.1, 2019 [26]
OAR contouring summary resources	UK SABR Consortium guidance version 6.1, 2019 [26] Mir, 2019 [25] Wright, 2019 [30]
SacralPlex	Yi, 2012 [27]

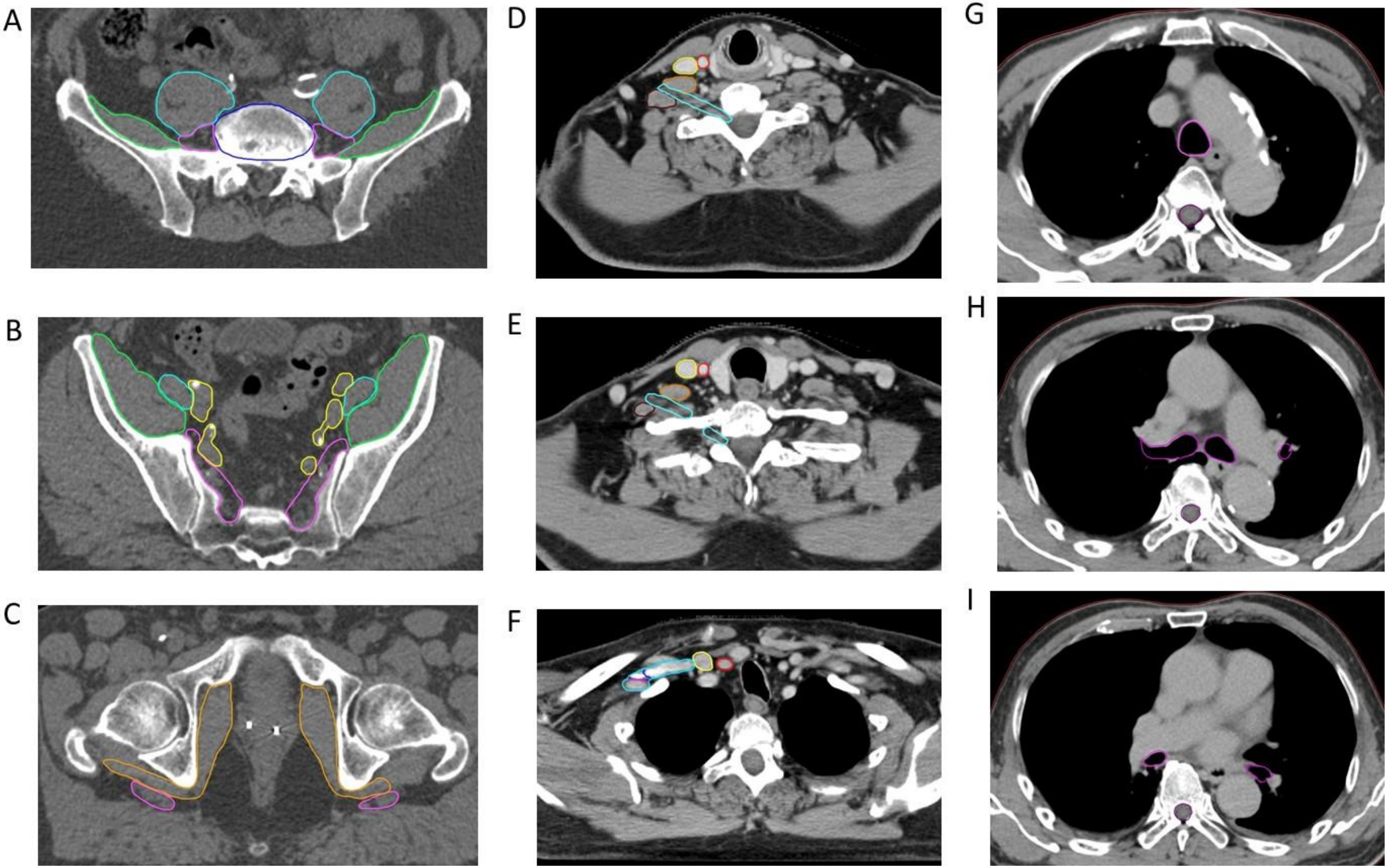
A
B
C
D
E
F
G
H
I

Figure 2: Visual guide to delineation of SacralPlex, BrachialPlex and BronchusProx

In Figure 2A-C, SacralPlex; iliacus muscle (green), L5 vertebral body (dark blue), obturator internus muscle (orange), psoas muscle (light blue), SacralPlex (purple), vessels (yellow) are shown. SacralPlex is contoured using a 5 mm diameter roller ball. In 2A, superior border of SacralPlex is shown at L4/5 vertebral interspace; SacralPlex is shown bordered by (ilio)psoas muscle anteriorly and vertebral body posteriorly. In 2B, at the sacro-iliac foramen, SacralPlex is shown bordered by vessels anteriorly, iliacus muscle laterally and sacral ala posteriorly. In 2C, inferior border of SacralPlex is shown at the level of the superior femoral neck bordered by obturator internus muscle anteriorly, gluteus maximus muscle posteriorly.

In Figure 2D-F, BrachialPlex contoured as suggested by UK SABR Consortium Guidelines[26]; anterior scalene muscle (orange), BrachialPlex (light blue), common carotid artery (red), internal jugular vein (posterior scalene muscle (brown), subclavian artery (pink) and subclavian vein (dark blue) are shown. Intravenous contrast is helpful, and BrachialPlex is contoured using a 5 mm diameter roller ball. 2D shows a proximal slice: the superior border of BrachialPlex is at the bifurcation of the brachiocephalic trunk into the jugular/subclavian veins (or carotid/subclavian arteries). In 2E, a middle section of BrachialPlex is shown; the plexus sits between anterior and middle scalene muscles. In 2F, the neurovascular complex including the subclavian and axillary vessels are contoured as a surrogate for the brachial plexus, ending after the neurovascular structures cross the second rib.

In Figure 2G-I, BronchusProx; BronchusProx (purple) is shown. In 2G, superior border of BronchusProx is the distal 2 cm of trachea including carina. In 2.2H, the mid-section of BronchusProx is shown and includes right/left upper lobe bronchi, bronchus intermedius, right middle lobe bronchus, lingular bronchus and right/left lower lobe bronchi. In 2I, contouring of lobar bronchi stops immediately at the site of a segmental bifurcation.

The metric thresholds that correlate to a minimum expected standard of contouring are uncertain but it has been previously suggested that DSC >0.7 indicates a good level of agreement[2]. However, previous studies have demonstrated discrepancies between contours considered to be acceptable based on expert review and the results of overlap measure comparisons[31]. In this study, 92% of the post-workshop contours were considered to be acceptable/within acceptable variation while 75% of structures had a DSC >0.7. A range of comparison metrics exists and each provides different information about the relationship between two contours and each has its limitations[16]. A summary of commonly used metrics for contour comparison is shown in **Supplementary Table**; it is unclear which is the optimum metric to use[1, 2, 16, 18, 32-37]. For this reason, it has previously been recommended that multiple metrics ideally be reported including measures of volume, overlap and distance[1, 16]. In this study, we only reported DSC and LDE since we did not have volumetric contouring data. It should be emphasised that DSC may provide less reliable results when applied to very small contours and it may lack discrimination for very large volumes[18]. However, it does provide some insight into both the volumetric and spatial relationship between two contours and it is frequently reported in contouring studies[1, 11].

Quantitative concordance in target volume/OAR delineation does not necessarily equate to a clinically acceptable contour; incorrect delineation of even a small proportion of a target volume or an OAR could have profound clinical consequences, especially for SABR where tight margins, steep dose gradients and ablative doses are used[2, 38, 39]. This risk means that quantitative metrics should ideally be accompanied by visual review of contours and provision of qualitative feedback, analogous to the peer review process used in clinical practice and recommended by the RCR[7]. This approach is used in clinical trials for pre-trial approval for participation or on-trial individual case evaluation. Qualitative feedback can be provided detailing acceptable/unacceptable variation from the protocol and a similar process was used in this study for feedback on post-workshop contours[3, 40-42]. However, this approach may be time consuming and an efficient/reliable method of assessment which can identify clinically relevant discrepancies is needed[1, 16, 31].

The practice of clinical oncology takes place against an increasingly complex backdrop of developments in imaging and novel methods of treatment delivery. Alongside ever increasing pressures in healthcare services, considerable challenges exist for training and continuous professional development of trainees and consultants respectively[9]. Formal training initiatives have been established to deliver the acquisition, and maintenance, of contouring competences in an attempt to improve target volume/OAR delineation beyond what could be achieved by a single workshop in isolation. The Fellowship in Anatomic deLineation and CONtouring (FALCON) programme is a European Society of Radiation Oncology (ESTRO) initiative that provides access to e-learning contouring resources in addition to its use within dedicated workshops[18, 21]. The RCR ARENA and Clinical Oncology Planning Project (COPP) are some example of initiatives to increase access to expert/peer-led structured outlining training to promote consistency in target volume and OAR outlining, and facilitate robust assessment of outlining practice for all grades of Clinical Oncologists[43].

This study has a number of additional limitations. The workshop was limited in its time/level of interactivity because of restrictions imposed during the pandemic and this could have impacted on the educational experience/DSC and LDE results that we observed (although participant feedback for the workshop remained positive). The same cases were used for both pre and post-workshop contouring; while this enabled the analysis of paired data, it meant that post-workshop contour performance could have been influenced by familiarity with the case and thus extrapolation of similar levels of performance to other cases would not necessarily be guaranteed. We did not stratify by prior experience when undertaking our analysis; this was because this information was not available to the authors but it could have influenced the results that were obtained. The workshop was aimed at those without prior experience in SABR but experience with OAR delineation would have varied depending on disease site expertise. We also did not evaluate longer-term maintenance of contouring competences by provision of further cases for contouring as part of this workshop, although response rates for such interventions for a single workshop in isolation may be limited[20]. It might be expected that initial educational gains could progressively negate over time, meaning that ongoing evaluation of performance during training programmes/as part of consultant continuous professional development will be required. Finally, feedback on post-workshop contours was only available for approximately half of participants included in our analyses; this affected the conclusions that can be drawn regarding the qualitative feedback but does reflect the challenge of providing such information in a timely manner.

When planning a contouring workshop, the following considerations may be relevant based on prior recommendations/the authors' experience[3, 16, 20, 31, 44]:

- Workshop format; incorporation of time to practise contouring/re-contouring is recommended in addition to didactic teaching (the duration of the workshop should be considered in relation to this)
- Clarity of instructions for cases to be contoured including detailed delineation guidance and specification of laterality, where relevant
- Timely access to relevant target volume/OAR guidance/atlases
- Provision of co-registered imaging
- Target audience; disease sites, numbers of target volume/OARs, number/complexity of cases
- Choice of assessment; quantitative metrics (such as volume, distance and overlap metrics) should ideally be used in conjunction with qualitative feedback. Be realistic about how much qualitative feedback can be provided in a timely manner to each participant
- Post workshop, provision of expert contour (where available) for participant comparison
- Where a reference contour is used; discussion regarding variation that may occur between even 'expert' outliners. One approach could be to use three expert contours and demonstrate the union and overlap as the maximum and minimum acceptable contours
- Identification of common errors/sources of variation for particular target volume/OARs
- Highlight available e-learning resources for self-directed learning
- Design of workshop feedback to evaluate participant confidence in contouring before/after the workshop
- Audiovisual/technological considerations; including undertaking a 'trial run' of the equipment prior to the workshop, including a method of quality assurance for displayed imaging, provision for participants with disabilities and recommendations that participants ensure they have a stable internet connection and adequate audiovisual equipment to fully participate

Conclusion

This study has demonstrated that virtual contouring training is feasible and that teaching during a virtual SABR contouring workshop for multiple target volumes/OARs was associated with some improvements in contouring variation.

Virtual contouring workshops could play an important role in aiding the acquisition of contouring competences alongside formal training initiatives.

References

1. Jameson, M.G., L.C. Holloway, P.J. Vial, S.K. Vinod, and P.E. Metcalfe, *A review of methods of analysis in contouring studies for radiation oncology.* Journal of Medical Imaging and Radiation Oncology, 2010. **54**(5): p. 401-410.

2. Vinod, S.K., M. Min, M.G. Jameson, and L.C. Holloway, *A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology.* Journal of Medical Imaging and Radiation Oncology, 2016. **60**(3): p. 393-406.

3. Gwynne, S., D. Gilson, J. Dickson, S. McAleer, and G. Radhakrishna, *Evaluating Target Volume Delineation in the Era of Precision Radiotherapy: FRCR, Revalidation and Beyond.* Clinical Oncology, 2017. **29**(7): p. 436-438.

4. Njeh, C.F., *Tumor delineation: The weakest link in the search for accuracy in radiotherapy.* J Med Phys, 2008. **33**(4): p. 136-40.

5. Peters, L.J., B. O'Sullivan, J. Giralt, T.J. Fitzgerald, A. Trotti, J. Bernier, et al., *Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02.* J Clin Oncol, 2010. **28**(18): p. 2996-3001.

6. Weber, D.C., M. Tomsej, C. Melidis, and C.W. Hurkmans, *QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy.* Radiotherapy and Oncology, 2012. **105**(1): p. 4-8.

7. Royal College of Radiologists. *Radiotherapy target volume definition and peer review – RCR guidance*. 2017. [Accessed 20th January 2021]; Available from: https://www.rcr.ac.uk/publication/radiotherapy-target-volume-definition-and-peer-review.

8. Weiss, E. and C.F. Hess, *The Impact of Gross Tumor Volume (GTV) and Clinical Target Volume (CTV) Definition on the Total Accuracy in Radiotherapy.* Strahlentherapie und Onkologie, 2003. **179**(1): p. 21-30.

9. Walls, G.M., G.G. Hanna, and J.J. McAleer, *Learning radiotherapy: the state of the art.* BMC Medical Education, 2020. **20**(1): p. 150.

10. Datta, S.T. and S.J. Davies, *Training for the future NHS: training junior doctors in the United Kingdom within the 48-hour European working time directive.* BMC medical education, 2014. **14 Suppl 1**(Suppl 1): p. S12-S12.

11. De Bari, B., M. Dahele, M. Palmu, S. Kaylor, L. Schiappacasse, and M. Guckenberger, *Short interactive workshops reduce variability in contouring treatment*

volumes for spine stereotactic body radiation therapy: Experience with the ESTRO FALCON programme and EduCase™ training tool. Radiotherapy and Oncology, 2018. **127**(1): p. 150-153.

12. Dewas, S., J.-E. Bibault, P. Blanchard, C. Vautravers-Dewas, Y. Pointreau, F. Denis, et al., *Delineation in thoracic oncology: a prospective study of the effect of training on contour variability and dosimetric consequences.* Radiation oncology (London, England), 2011. **6**: p. 118-118.

13. Khoo, E.L.H., K. Schick, A.W. Plank, M. Poulsen, W.W.G. Wong, M. Middleton, et al., *Prostate Contouring Variation: Can It Be Fixed?* International Journal of Radiation Oncology, Biology, Physics, 2012. **82**(5): p. 1923-1929.

14. NHS England. *Stereotactic ablative radiotherapy (SABR) for patients with metachronous extracranial oligometastatic cancer (all ages)*. 2020. [Accessed 20th January 2021]; Available from: https://www.england.nhs.uk/publication/stereotactic-ablative-radiotherapy-sabr-for-patients-with-metachronous-extracranial-oligometastatic-cancer-all-ages/.

15. Vaassen, F., C. Hazelaar, A. Vaniqui, M. Gooding, B. van der Heyden, R. Canters, et al., *Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy.* Physics and Imaging in Radiation Oncology, 2020. **13**: p. 1-6.

16. Vinod, S.K., M.G. Jameson, M. Min, and L.C. Holloway, *Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies.* Radiotherapy and Oncology, 2016. **121**(2): p. 169-179.

17. Dice, L.R., *Measures of the Amount of Ecologic Association Between Species.* 1945. **26**(3): p. 297-302.

18. Rivin del Campo, E., S. Rivera, M. Martínez-Paredes, P. Hupé, A. Slocker Escarpa, I. Borget, et al., *Assessment of the novel online delineation workshop dummy run approach using FALCON within a European multicentre trial in cervical cancer (RAIDs).* Radiotherapy and Oncology, 2017. **124**(1): p. 130-138.

19. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. 2016. [Accessed 26th January 2021]; Available from: https://ggplot2.tidyverse.org.

20. Cacicedo, J., A. Navarro-Martin, S. Gonzalez-Larragan, B. De Bari, A. Salem, and M. Dahele, *Systematic review of educational interventions to improve contouring in radiotherapy.* Radiotherapy and Oncology, 2020. **144**: p. 86-92.

21. Eriksen, J.G., C. Salembier, S. Rivera, B. De Bari, D. Berger, G. Mantello, et al., *Four years with FALCON – An ESTRO educational project: Achievements and perspectives.* Radiotherapy and Oncology, 2014. **112**(1): p. 145-149.

22. D'Souza, L., J. Jaswal, F. Chan, M. Johnson, K.Y. Tay, K. Fung, et al., *Evaluating the impact of an integrated multidisciplinary head & neck competency-based anatomy & radiology teaching approach in radiation oncology: a prospective cohort study.* BMC medical education, 2014. **14**: p. 124-124.

23. Szumacher, E., N. Harnett, S. Warner, V. Kelly, C. Danjoux, R. Barker, et al., *Effectiveness of Educational Intervention on the Congruence of Prostate and Rectal Contouring as Compared With a Gold Standard in Three-Dimensional Radiotherapy for Prostate.* International Journal of Radiation Oncology*Biology*Physics, 2010. **76**(2): p. 379-385.

24. Hall, W.H., M. Guiou, N.Y. Lee, A. Dublin, S. Narayan, S. Vijayakumar, et al., *Development and Validation of a Standardized Method for Contouring the Brachial Plexus: Preliminary Dosimetric Analysis Among Patients Treated With IMRT for Head-and-Neck Cancer.* International Journal of Radiation Oncology*Biology*Physics, 2008. **72**(5): p. 1362-1367.

25. Mir, R., S.M. Kelly, Y. Xiao, A. Moore, C.H. Clark, E. Clementel, et al., *Organ at risk delineation for radiation therapy clinical trials: Global Harmonization Group consensus guidelines.* Radiotherapy and Oncology, 2020. **150**: p. 30-39.

26. UK SABR Consortium. *Stereotactic Ablative Radiotherapy (SABR): A Resource (version 6.1)*. 2019. [Accessed 8th February 2021]; Available from: https://www.sabr.org.uk/wp-content/uploads/2019/04/SABRconsortium-guidelines-2019-v6.1.0.pdf.

27. Yi, S.K., W. Mak, C.C. Yang, T. Liu, J. Cui, A.M. Chen, et al., *Development of a Standardized Method for Contouring the Lumbosacral Plexus: A Preliminary Dosimetric Analysis of this Organ at Risk Among 15 Patients Treated With Intensity-Modulated Radiotherapy for Lower Gastrointestinal Cancers and the Incidence of Radiation-Induced Lumbosacral Plexopathy.* International Journal of Radiation Oncology*Biology*Physics, 2012. **84**(2): p. 376-382.

28. American Association of Physicists in Medicine. *Report No. 263 - Standardizing Nomenclatures in Radiation Oncology*. 2018. [Accessed 8th February 2021]; Available from: https://www.aapm.org/pubs/reports/detail.asp?docid=171.

29. De la Pinta, C., *SBRT in non-spine bone metastases: a literature review.* Med Oncol, 2020. **37**(12): p. 119.

30. Wright, J.L., S.S. Yom, M.J. Awan, S. Dawes, B. Fischer-Valuck, R. Kudner, et al., *Standardizing Normal Tissue Contouring for Radiation Therapy Treatment Planning: An ASTRO Consensus Paper.* Practical Radiation Oncology, 2019. **9**(2): p. 65-72.

31. Duke, S.L., L.-T. Tan, N.B.K. Jensen, T. Rumpold, A.A.C. De Leeuw, C. Kirisits, et al., *Implementing an online radiotherapy quality assurance programme with supporting continuous medical education – report from the EMBRACE-II evaluation of cervix cancer IMRT contouring.* Radiotherapy and Oncology, 2020. **147**: p. 22-29.

32. Ackerly, T., J. Andrews, D. Ball, M. Guerrieri, B. Healy, and I. Williams, *Discrepancies in volume calculations between different radiotherapy treatment planning systems.* Australas Phys Eng Sci Med, 2003. **26**(2): p. 91-3.

33. Brock, K.K., L.A. Dawson, M.B. Sharpe, D.J. Moseley, and D.A. Jaffray, *Feasibility of a novel deformable image registration technique to facilitate classification, targeting, and monitoring of tumor and normal tissue.* International Journal of Radiation Oncology, Biology, Physics, 2006. **64**(4): p. 1245-1254.

34. Fotina, I., C. Lütgendorf-Caucig, M. Stock, R. Pötter, and D. Georg, *Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy.* Strahlentherapie und Onkologie, 2012. **188**(2): p. 160-167.

35. Hanna, G.G., A.R. Hounsell, and J.M. O'Sullivan, *Geometrical Analysis of Radiotherapy Target Volume Delineation: a Systematic Review of Reported Comparison Methods.* Clinical Oncology, 2010. **22**(7): p. 515-525.

36. Maurer, C.R., Q. Rensheng, and V. Raghavan, *A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. **25**(2): p. 265-270.

37. Taha, A.A. and A. Hanbury, *Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool.* BMC medical imaging, 2015. **15**: p. 29-29.

38. Murray, P., K. Franks, and G.G. Hanna, *A systematic review of outcomes following stereotactic ablative radiotherapy in the treatment of early-stage primary lung cancer.* Br J Radiol, 2017. **90**(1071): p. 20160732.

39. Trakul, N., A.C. Koong, and D.T. Chang, *Stereotactic body radiotherapy in the treatment of pancreatic cancer.* Semin Radiat Oncol, 2014. **24**(2): p. 140-7.

40. Cox, S., A. Cleves, E. Clementel, E. Miles, J. Staffurth, and S. Gwynne, *Impact of deviations in target volume delineation – Time for a new RTQA approach?* Radiotherapy and Oncology, 2019. **137**: p. 1-8.

41. Gwynne, S., E. Spezi, D. Sebag-Montefiore, S. Mukherjee, E. Miles, J. Conibear, et al., *Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case.* The British journal of radiology, 2013. **86**(1024): p. 20120398-20120398.

42. Gwynne, S., G. Jones, R. Maggs, D. Eaton, E. Miles, J. Staffurth, et al., *Prospective review of radiotherapy trials through implementation of standardized multicentre workflow and IT infrastructure.* Br J Radiol, 2016. **89**(1064): p. 20160020.

43. Evans, E., G. Radhakrishna, D. Gilson, P. Hoskin, E. Miles, F. Yuille, et al., *Target Volume Delineation Training for Clinical Oncology Trainees: The Role of ARENA and COPP.* Clin Oncol (R Coll Radiol), 2019. **31**(6): p. 341-343.

44. Roques, T.W., *Patient Selection and Radiotherapy Volume Definition — Can We Improve the Weakest Links in the Treatment Chain?* Clinical Oncology, 2014. **26**(6): p. 353-355.