BJR https://doi.org/10.1259/bjr.20210356

Received: 17 March 2021 | Revised: 05 July 2021 | Accepted: 19 July 2021

Cite this article as:
Piazzese C, Evans E, Thomas B, Staffurth J, Gwynne S, Spezi E. FIELD^{RT}: an open-source platform for the assessment of target volume delineation in radiation therapy. *Br J Radiol* 2021; **94**: 20210356.

FULL PAPER

FIELD^{RT}: an open-source platform for the assessment of target volume delineation in radiation therapy

[1,2,3]CONCETTA PIAZZESE, PhD, [3]ELIN EVANS, MD, [4]BETSAN THOMAS, MD, [3]JOHN STAFFURTH, MD, [4]SARAH GWYNNE, MD and [2,3]EMILIANO SPEZI, PhD

[1]University of Huddersfield, School of Computing & Engineering, Huddersfield, UK
[2]Cardiff University, School of Engineering, Cardiff, UK
[3]Clinical Oncology, Velindre Cancer Centre, Cardiff, UK
[4]Clinical Oncology, South West Wales Cancer Centre, Swansea, UK

Address correspondence to: Dr Concetta Piazzese
E-mail: *c.piazzese@hud.ac.uk*

Sarah Gwynne and Emiliano Spezi have contributed equally to this study and should be considered as senior authors.

Objectives: Target volume delineation (TVD) has been identified as a weakness in the accuracy of radiotherapy, both within and outside of clinical trials due to the intra/interobserver variations affecting the TVD quality. Sources of variations such as poor compliance or protocol violation may have adverse effect on treatment outcomes. In this paper, we present and describe the FIELD^{RT} software developed for the ARENA project to improve the quality of TVD through qualitative and quantitative feedbacks and individual and personalized summary of trainee"s performance.

Methods: For each site-specific clinical case included in the FIELD^{RT} software, reference volumes, minimum and maximum "acceptable" volumes and organ at risk were derived by outlines of consultants and senior trainees. The software components currently developed include: (a) user-friendly importing interface (b) analysis toolbox to compute quantitative and qualitative (c) visualiser and (d) structured report generator for personalised feedback. The FIELD^{RT} software was validated by comparing the performance of 63 trainees and by measuring performance over time. In addition, a trainee evaluation day was held in 2019 to collect feedback on FIELD^{RT}.

Results: Results show the trainees' improvement when reoutlining a case after reviewing the feedback generated from the FIELD^{RT} software. Comments and feedback received after evaluation day were positive and confirmed that FIELD^{RT} can be a useful application for training purposes.

Conclusion: We presented a new open-source software to support education in TVD and ongoing continuous professional development for clinical oncology trainees and consultants. ARENA in combination with FIELD^{RT} implements site-specific modules with reference target and organs at risk volumes and automatically evaluates individual performance using several quantitative and qualitative feedbacks. Pilot results suggests this software could be used as an education tool to reduce variation in TVD so to guarantee high quality in radiotherapy.

Advances in knowledge: FIELD^{RT} is a new easy and free to use software aiming at supporting education in TVD and ongoing continuous professional development. The software provides quantitative/qualitative feedback and an exportable report with an individual and personalised summary of trainee's performance.

INTRODUCTION

Radiotherapy (RT) uses ionising radiation to kill cancer cells with the aim of cure or effective palliation and is used in the treatment of over 50% of all patients with cancer.[1] Current practice in RT requires target volume delineation (TVD) usually on a planning CT by a clinician or non-medical outliner (NMO). Macroscopic tumour (defined by the relevant diagnostic investigations) is manually delineated as the "gross tumour volume" (GTV). This is expanded to create the clinical target volume (CTV) to encompass areas of possible microscopic spread surrounding the GTV, and for some tumour sites may include nodal regions at risk. A volumetric expansion is applied to the CTV to create the planning target volume (PTV) to allow for potential patient set-up discrepancies and organ motion.

Organs at risk (OARs) relevant to the particular tumour site, which, if receiving a significant RT dose may lead to significant toxicity for the patient and risk of long-term complications respectively,[2] are also delineated to allow dose to be calculated.

The accuracy of TVD may potentially adversely impact patient outcome through both an effect on normal tissue toxicity and tumour control. "Overdelineation" of target volumes may lead to unnecessary normal issue toxicity whereas "underdelineation" may lead to geographical miss of the target regions.[3,4] Failure to adhere to the protocol has been shown to affect the outcome in several trials and although the focus has historically been on the treatment planning aspects of the protocol, recent interest has focused more on the TVD.[5–8] TVD is an essential step within the RT planning pathway. However, it has been identified as potentially the "weakest link" because it may be affected by intra/interobserver variations.[9]

This variation has been attributed to quality of imaging to assist in TVD, the extent and quality of training in TVD and personal bias.[10,11] To minimise intra/interobserver TVD variation, a number of interventions have been employed.[12,13] One example includes access to a trial protocol and an outlining atlas, which has shown to improve consistency in RT outlining in prostate cancer[14] and rectal cancer,[15] respectively. However, one study in lung TVD demonstrated considerable TVD variation despite a protocol guidance due to participating clinicians reverting to pre-protocol practice. Educational sessions or workshops are another source of TVD training. They have been shown to reduce TVD variation in a range of settings, such as prostate[16] and lung cancer.[17] Examples includes the fellowship in anatomic delineation and contouring educational project (FALCON)[18] developed by the European society for radiotherapy and oncology (ESTRO) which aimed to promote e-learning activities and teaching in addition to providing a platform for contouring workshops[19] organised annually at the ESTRO congress and in the UK, the Royal College of radiologists has also organised a series of outlining workshops using the AQUILAB software (AQUILAB, Loos, France). However, these sessions carry limitations including small number of participants due to logistical reasons and costs[20] and the omission of/inability to finish contouring exercises due to time constraints.[21]

Table 1. Tumour sites with the related volumes and OARs outlined by a team of experts and processed to populate the package of reference clinical sites available in FIELD^{RT}.

Tumour site	Volumes	OARs
Oesophagus	GTV CTVA CTVB CTVC PTV	Aorta Azygous vein Left lung Left main bronchus Liver Pericardium/great vessels Right lung Stomach Vertebra
Prostate	CTVp CTVpsv	Bladder Bowel Left femoral head Penile bulb Rectum Right femoral head

CTV, clinical target volume; OAR, organ at risk; PTV, planning target volume.

TVD is a craft and developing these core skills at a training level will inevitably influence future practice and habits. TVD teaching is therefore a key component in clinical/ radiation oncology training. However, in the UK, inadequate RT training is cited as the main training concern for clinical oncology trainees, with only 78 and 61% considering themselves competent in palliative and radical RT planning, respectively.[22] This is attributed to a loss of protected TVD time to other clinical duties and lack of feedback on from consultant supervisors.

Given the potential variation of RT training for trainees across tumour sites and training centres, our group established the ARENA (Assurances in Radiotherapy through EducatioN and Assessment) project in 2017 with the aim of standardising high-quality TVD training for UK clinical oncology trainees, based on experience acquired from RT trials quality assurance.[23] To determine the preferred format for these training packages, in 2018 we surveyed 131 UK clinical oncology trainees across all training grades regarding TVD training quality and preferential format for TVD modules.[24] This survey highlighted the fact that self-directed learning remains one of the most common methods of TVD training, and that most trainees would value a tool supplementing consultant-led TVD teaching in the form of site-specific TVD modules. In response to this survey, the ARENA project team has developed educational packages that comprise of site-specific introductory TVD modules detailing outlining instructions with accompanying interactive cases that can be hosted within appropriate software. Following delineation and

Figure 1. Workflow and internal architecture of the FIELD^{RT} software. User's volumes in DICOM format are exported from the outlining software choosen by the user and imported and processed in the FIELD^{RT} software. Quantitative/qualitative feedback is provided and converted to an exportable document with a personalised summary of the performance.

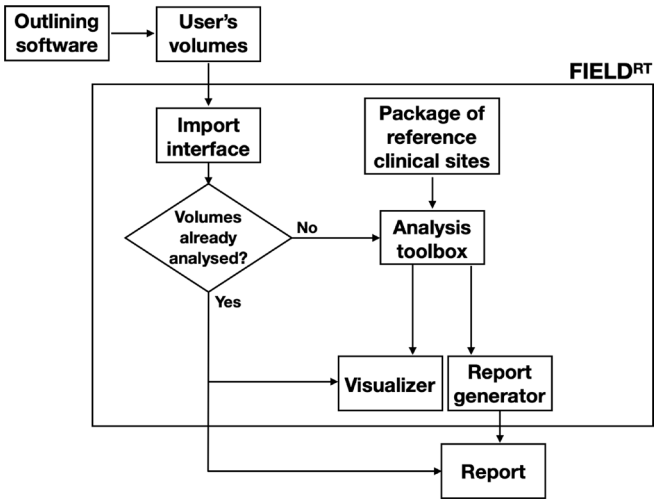
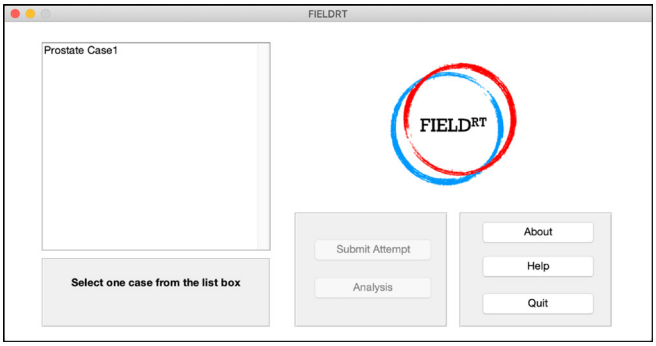


Figure 2. FIELD^{RT} importing interface to select the clinical site to use as a reference for the analysis and to import the user's volumes in DICOM format. The FIELD^{RT} logo at the top left with blue and red contours representing the colour-coded feedback provided by the software.



submission of the case, semi-automated feedback is provided on outlining performance. In this paper, we present and describe the Feedback on Individual Education in anatomicaL Delineation in RadioTherapy (FIELD^{RT}) software developed for the project.

METHODS AND MATERIALS

Site-specific clinical case development

Given the expertise of the ARENA project team, prostate and oesophagus were identified as pilot tumour sites followed by lung, head and neck, rectum and breast. Step-by-step guidance for TVD in a power-point format, based on relevant UK RT trial protocols, were developed. For example, the oesophageal cancer TVD guidance was adapted from the UK SCOPE 2 RT trial.[25] A number of UK prostate RT trial protocols were adapted to create TVD guidance for the prostate cancer case. A clinical vignette was written for each case including anonymised patient and diagnostic information, accompanied by PET-CT and MRI images. The process was then repeated for the other tumour sites, with site-specific leads being identified.

Reference volumes

For each case, up to six UK clinical oncologists comprising of both site-specific consultants and senior trainees (post Fellowship of the Royal College of Radiologists (FRCR), developing a subspeciality interest in the respective tumour site) delineated the relevant case target volumes on the clinical cases identified (Table 1), following the delineation guidance in the TVD module. Before importing the pre-defined reference cases in the FIELD^{RT} software, all outlines were processed using CERR[26,27] (Computational Environment for Radiological Research), an open-source platform custom built for the purpose of analysing and sharing RT data. Using the same approach adopted by the Radiation Therapy Oncology Group (RTOG) when developing contouring atlases for prostate[28] and rectal cancer,[29] a single contour was created from the constituent outlines using the Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm available in CERR with a 95% confidence.[30] As with the RTOG groups, as a final step, this contour was reviewed again by the site-specific lead consultant for the respective tumour site and

Figure 3. Qualitative and quantitative feedback of the user's performance in the FIELD^{RT} software: user's and GS volumes with maximum/minimum acceptable outlining areas (a), user's volumes and OARs with a 'red flag' for inappropriately contoured volumes including encroachment into OAR (b) and user's and GS volumes with (c) or without (d) over/undercontoured areas highlighted. OAR, organ at risk.

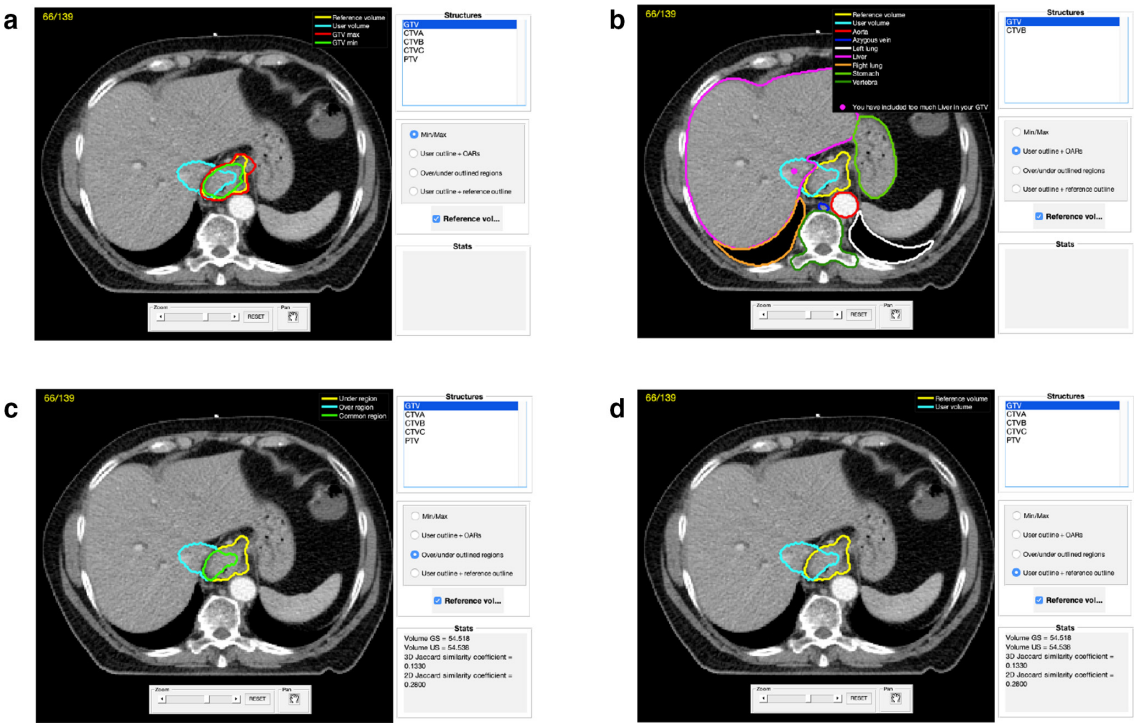
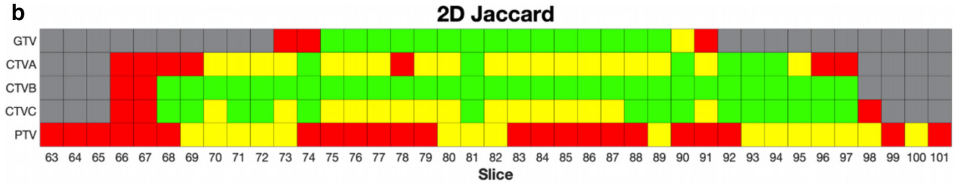


Figure 4. Personalised summary of quantitative feedback of the trainee's performance with whole volume (a) and slice by slice (b) indexes highlighted in red, yellow and green if Jaccard value is <0.50, 0.50–0.75 and >0.75, respectively. Slices not included in the reference volume are highlighted in dark grey.

a

	Reference volume [ml]	User volume [ml]	3D Jaccard
GTV	15.26	15.62	0.77
CTVA	22.97	14.36	0.55
CTVB	72.57	76.70	0.81
CTVC	59.16	47.52	0.71
PTV	180.58	81.33	0.45

b



agreed to be the reference volume, hereafter referred to as gold -standard (GS).

Minimum and maximum "acceptable" volumes

Given the variation in TVD even among experts, along with the difficulties in defining the "ground truth", we also developed a minimum and maximum "acceptable" volume for the relevant target outlines made up of the constituent outlines as shown in S1 Fig. These were edited by the site-specific leads, omitting "incorrect" and "nonsense" volumes to create a final minimum and maximum "acceptable" volume, with the intention that a trainee contour is within this range would be considered to be clinically acceptable.

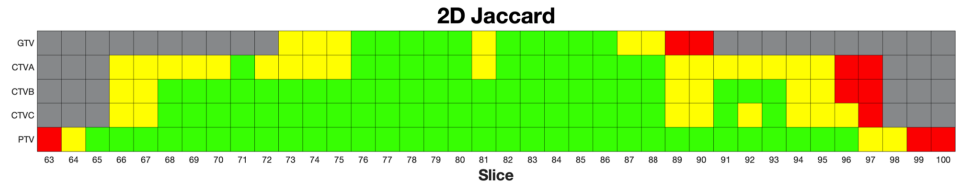
Organs at risk

On the basis of feedback from the UK trainee survey that qualitative feedback was most useful for training,[31] we wanted to include descriptive feedback on outlining performance, within

Figure 5. Summary of the quantitative feedback of all 63 trainees' performance for oesophagus (a) prostate (b) with slice by slice and whole volume and global indexes highlighted in red, yellow and green if Jaccard value is <0.50, 0.50–0.75 and >0.75. Slices not included in the reference volume are highlighted in dark grey. CTV, clinical target volume; GTV, gross tumour volume; PTV, planning target volume.

a

	Reference volume [ml]	User volume [ml]	3D Jaccard
GTV	15.26	15.75	0.63
CTVA	22.97	22.62	0.69
CTVB	72.57	79.84	0.70
CTVC	59.16	64.46	0.71
PTV	180.58	191.83	0.76



b

	Reference volume [ml]	User volume [ml]	3D Jaccard
CTVp	33.08	43.36	0.70
CTVpsv	42.05	54.37	0.64

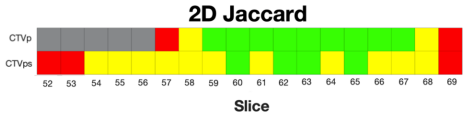
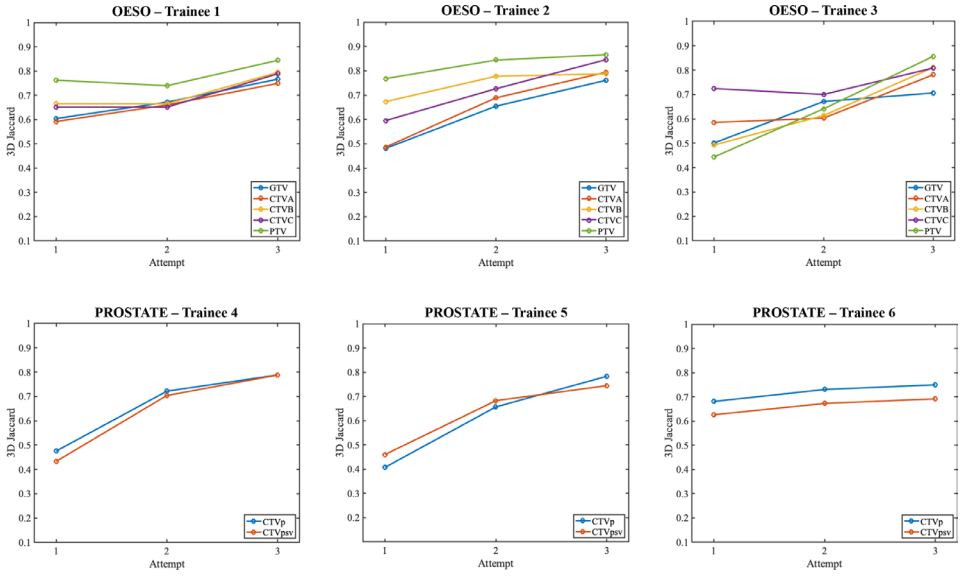


Figure 6. Improvement of six trainees when reoutlining the same case after reviewing the quantitative and qualitative feedback generated from the FIELD^{RT} software. For each attempt, the mean 3D Jaccard value computed by the FIELD^{RT} software for each volume is reported.



the limitations of an "offline" feedback platform. Due to the critical nature of the incorrect inclusion/extension to surrounding OARs, we developed a "flag" tool within our software, based on pre-outlined OARs for each case (Table 1). "Dummy" OARs were created by reducing the OAR contours by 3 mm circumferentially.[32] If a trainee contour extended beyond this 3 mm limit this was considered unacceptable and would be included in the feedback report.

Architecture and workflow of FIELD^{RT} software

In order to provide a platform for the interactive cases to be analysed, we developed software components including the importing interface, analysis computation toolbox, visualiser and report generator. Figure 1 shows the workflow and the internal architecture of the FIELD^{RT} software.

Importing interface

The FIELD^{RT} software uses the same routine and file format as CERR to import and store RT treatment planning and contour data to be compared with the corresponding reference volumes for each clinical case. The software also includes a module that keeps track of previous attempts. This allows the user to directly view their analysed performances using either the visualizer or personalised performance report without the need to re-compute quantitative and qualitative analyses. Figure 2 shows the FIELD^{RT} importing interface to upload and analyse user's volumes in DICOM format.

Analysis toolbox

To assess the user's performance a range of quantitative and qualitative evaluations are computed to assess the comparison between their contours and the expert contours. The analysis includes operations such as intersection, union and difference between the user volumes and the corresponding reference volume as shown in S2 Fig. Any overlap between the user volumes and the OARs is also detected and reported. In addition, a comprehensive quantitative analysis of the user volumes is performed in terms of 2D (slice by slice) and 3D (whole volume) comparison against the reference volume using the Jaccard index. The Jaccard similarity index (also known as Jaccard similarity coefficient) compares two sets or volumes to determine the extent of their overlap with a range from 0% (two distinct and separate volumes) to 100% (volumes completely overlapping) and is calculated by:

$$Jaccard\,(US,\ GS) = \frac{|US \cap GS|}{|US \cup GS|} \qquad (1)$$

where $US \cap GS$ and $US \cup GS$ are the intersection and union of the user's volume with the corresponding reference volume (GS). A more detailed description of Jaccard and other measures of conformity be found in the work Gwynne et al[33].

Visualiser

The interface of the FIELD^{RT} visualiser shows the CT planning scan and target volumes in different anatomical planes and related quantitative and qualitative feedback. Qualitative evaluations of the FIELD^{RT} visualiser include a visual display of:

- user's and reference volumes with maximum/minimum acceptable contours (Figure 3a);
- user's volumes and OARs with a 'red flag' for inappropriately contoured volumes including unacceptable extension into OAR (Figure 3b);
- over- and undercontoured areas defined as regions present and absent in the GS reference volume, respectively (Figure 3c);
- user's and reference volumes (Figure 3d);

For each user's contours, the volume and 2D/3D Jaccard conformity indexes are also provided and displayed as quantitative feedback (Figure 3c and d).

Individual report

The FIELD^{RT} report generator processes both quantitative and qualitative feedback to create an exportable document with an individual summary of the user's performance. The report includes flagged slices in which performance is below a threshold (2D Jaccard <0.50) or includes inappropriately contoured volumes that excessively overlap with OARs (S3 Fig). In addition, a table containing quantitative feedback (Figure 4a) and a colour-coded graph of the local performance of the user (Figure 4b) are included in the report. Whole volume and slice by slice JCIs are colour-coded with in red, yellow and green corresponding to values of <0.50, 0.50–0.75 and >0.75, respectively. These predetermined levels were set based on previous work undertaken by the group.[34]

Additional measures of performance

In addition to the analyses discussed above, we also compared the performance of a candidate against both their own and their peers' performance over time. To develop such features, we analysed the mean 3D Jaccard values computed by the FIELD^{RT} software when processing 63 oesophagus and prostate cancer contours submitted by national and international trainees who attended a FRCR part 2B preparation course in Cardiff in September 2019.

RESULTS

Validation of FIELD^{RT}

The FIELD^{RT} software successfully analysed and generated personalised quantitative and qualitative feedback of the oesophageal and prostate volumes outlined by the 63 candidates. The processing time required to import each set of user volumes, analyse them and produce the exportable document with a summary of the user's performance was approximately 4 min on a conventional machine (Intel Core e5-1620, 32 GB RAM, 3.50 GHz). The table containing quantitative feedback of the slice by slice and whole volume performance of the trainees are shown in Figure 5. The mean values for volume and 3D Jaccard obtained by each trainee for the oesophageal and prostate cases are showed in S1 and S2 Tables.

To perform a longitudinal validation of the FIELD^{RT} software, three randomly selected candidates were requested to reoutline the same oesophagus case after reviewing the quantitative and qualitative feedback generated from the FIELD^{RT} software at the end of each outlining session. A further three randomly selected candidates were asked to do the same for the prostate case. For the oesophagus case, the overall mean percent change of 3D Jaccard of all attempts from baseline (first attempt) for the three candidates was +10.24%, +17.61% and +22.17%, respectively. For the prostate case, a mean percent change of 3D Jaccard of +33.84%, +33.95% and +4.86% was observed for the three candidates, respectively. The percent changes for each candidate's attempt are showed in S3 and S4 Tables. The improvement in the performance based on whole volume assessment is showed in Figure 6. For the two oesophagus and prostate cases, variations of 2D Jaccard index for slices in the reference range and the offset in the selection of the lower and upper slices outlined by the trainees with respect to the GS are provided in S4, S5 and S6 Figs.

Evaluation of FIELD^{RT}

As part of the ongoing development of the material for the ARENA project and the FIELD^{RT} software, we held a trainee evaluation day in 2019 with seven local trainees, who were asked to outline the oesophagus and prostate cases in advance of the day. The cases were then analysed in real time and the trainees asked to provide a score from 1 to 10 and comments for different components present in the FIELD^{RT} software. The overall score, expressed as mean ± standard deviation (SD), was 8.71 ± 1.70, 8.77 ± 0.25 and 8.57 ± 0.74 for the importing interface, the visualiser and the individual report, respectively. A table with the specific scores for each tool present in the FIELD^{RT} software is provided in Supplementary Material 1 (S5 Table). The comments provided by the trainees were very positive. All seven trainees found the interface very easy to use and navigate, commenting on software that was "self-explanatory and simple". Five out of seven found it useful to have their contour and the volume on the same screen and liked the minimum/maximum "acceptable" volume tool, enhancing understanding of the fact that there can be a range of acceptable volumes. Highlighting the areas where there was unacceptable extension into OARs was appreciated, and trainees stated they would bear this in mind when attempting future outlining. All trainees (seven out of seven) appreciated the quantitative assessment of their performance to assess the conformity of their outlining and valued the PDF report, commenting that it provided good background information on conformity indices and could serve as an essential part of their RT training logs. All the seven trainees appreciated being informed of areas of concern where conformity was low and found it useful being able to see screenshots of their contours.

DISCUSSION

The ARENA project, using FIELD^{RT} software, aims to facilitate a standardised approach for TVD training. The key features of the FIELD^{RT} software include: (a) training packages which include a site-specific RT TVD document and corresponding clinical cases for users to practice TVD; (b) a user-friendly platform and interface to import and upload attempts at TVD which are automatically analysed and assessed; (c) image analysis package to compute quantitative (*i.e.* conformity metrics such as volumes, 2D/3D Jaccard) and qualitative evaluation of the user's performance (user outline *vs* reference volume, maximum and minimum acceptable volumes, over- and undercontoured regions and a "red flag" for volumes inappropriately including OARs); (d) a user-friendly viewer to display CT planning images and contoured structures in axial, coronal, and sagittal planes; (e) a structured report generator that produces a personalised summary of the quantitative and qualitative feedback of the user's performance.

The base FIELD^{RT} package (*i.e.* importing interface, analysis computation toolbox, visualiser, report generator) with a prostate case (including the expert contours and dummy OARs) is freely available at the following address https://githubcom/concettapiazzese/FIELDRT-GitHub. In addition to the desktop version that requires a MATLAB license, the FIELD^{RT} software is also available as a standalone executable that includes all the required runtime libraries.

The main differences between the FIELD^{RT} software and other contouring software (*i.e.* eContour,[35] EduCase,[36] WorkflowBox,[37] ProKnow[38] and other contouring tools[39] are the following: (a) FIELD^{RT} software and the cases are opensource and freely accessible (the software is available for both Macintosh and Windows operating systems), (b) new educational material and cases can be created by any user and uploaded into the software, (c) the quantitative and qualitative feedback created by software are saved locally and can reviewed within the visualizer, (d) an individual report is generated as an exportable document with some unique features for qualitative and quantitative feedback (*i.e.* colour-coded graph of the local performance of the user, flagged slices in which performance is below a threshold or inappropriately contoured volumes that excessively overlap with OARs) and (e) FIELD^{RT} software doesn't include any contouring tools as we decided at the start of the project to not replicate the high quality treatment planning systems already used in clinical practice and with which trainees and consultants are more familiar.

From the work achieved so far, we have demonstrated that this software is a platform that is able to successfully upload, analyse and provide useful qualitative and quantitative analysis on TVD performance for training purposes. Experience and feedback received so far show that the FIELD^{RT} software can be a useful application to prepare training material, evaluate trainees' delineations and provide useful qualitative and quantitated of information on TVD performance for training purposes.

CONCLUSIONS

We developed a software to support education in TVD and ongoing continuous professional development for clinical oncology trainees and consultants. The ARENA project in combination with the FIELD^{RT} software implements site-specific modules with reference target and OARs volumes and automatically evaluates individual performance using several quantitative and qualitative feedbacks. Pilot results and quantitative improvement of trainee performance suggest this software could be used as an education tool to reduce variation in TVD so to improve the quality of RT. The availability of the FIELD^{RT} software as an opensource software will enable the wider RT community to create their own educational material and cases to be hosted locally or shared with others to create a large library of educational material in TVD.

HIGHLIGHTS

- ARENA project, using the FIELD^{RT} software, aims at supporting TVD education and ongoing continuous professional development.
- The software is easy and free to use, and it quickly and automatically evaluates the trainee's performance.
- Quantitative/qualitative feedback is provided through the user-friendly viewer.
- An exportable report with an individual and personalised summary of trainee's performance is generated.

FUNDING

This work was funded by Velindre UNHS Trust Advancing Radiotherapy Fund and the Moondance Foundation.

DATA SHARING

The base FIELD^{RT} package (i.e., importing interface, analysis computation toolbox, visualizer, report generator) with a prostate case (including the expert contours and dummy OARs) is freely available at the following address https://github.com/concettapiazzese/FIELDRT-GitHub. The authors do not own the contours submitted by trainees from UK and overseas who attended a Fellowship of the Royal College of Radiologists (FRCR) part 2B preparation course in Cardiff in September 2019. For this reason, authors are not permitted to share them in the original form.

REFERENCES

1. Baskar R, Lee KA, Yeo R, Yeoh K-W. Cancer and radiation therapy: current advances and future directions. *Int J Med Sci* 2012; **9**: 193–9. doi: https://doi.org/10.7150/ijms.3635
2. The Royal College of Radiologists.*Radiotherapy target volume definition and peer review RCR guidance*. London: The Royal College of Radiologists; 2017. pp. 1e35. https://www.rcr.ac.uk/system/files/publication/field_publication_files/bfco172_peer_review_outlining.pdf.
3. Hamilton CS, Ebert MA. Volumetric uncertainty in radiotherapy. *Clin Oncol* 2005; **17**: 456–64. doi: https://doi.org/10.1016/j.clon.2005.03.014
4. Van de Steene J, Linthout N, de Mey J, Vinh-Hung V, Claassens C, Noppen M, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiother Oncol* 2002; **62**: 37–49. doi: https://doi.org/10.1016/s0167-8140(01)00453-4
5. Willett CG, Moughan J, O'Meara E, Galvin JM, Crane CH, Winter K, et al. Compliance with therapeutic guidelines in radiation therapy Oncology Group prospective gastrointestinal clinical trials. *Radiother Oncol* 2012; **105**: 9–13. doi: https://doi.org/10.1016/j.radonc.2012.09.011
6. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, 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**: 2996–3001. doi: https://doi.org/10.1200/JCO.2009.27.4498
7. Cox S, Cleves A, Clementel E, Miles E, Staffurth J, Gwynne S. Impact of deviations in target volume delineation - Time for a new RTQA approach? *Radiother Oncol* 2019; **137**: 1–8. doi: https://doi.org/10.1016/j.radonc.2019.04.012
8. Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, et al. Failure to adhere to protocol specified radiation therapy guidelines was associated with decreased survival in RTOG 9704--a phase III trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas. *Int J Radiat Oncol Biol Phys* 2012; **82**: 809-16. doi: https://doi.org/10.1016/j.ijrobp.2010.11.039
9. Gwynne S, Spezi E, Sebag-Montefiore D, Mukherjee S, Miles E, Conibear J, et al.

Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. *Br J Radiol* 2013; **86**: 20120398. doi: https://doi.org/10.1259/bjr.20120398
10. Njeh CF. Tumor delineation: the weakest link in the search for accuracy in radiotherapy. *J Med Phys* 2008; **33**: 136–40. doi: https://doi.org/10.4103/0971-6203.44472
11. Sundar S, Symonds RP. Diagnostic radiology for radiotherapist: the case for structured training in cross-sectional imaging (CT and MRI. *Clin Oncol* 2002; **14**: 413–4. doi: https://doi.org/10.1053/clon.2002.0109
12. van Mourik AM, Elkhuizen PH, Minkema D, Duppen JC. Dutch young boost Study Group, van Vliet-Vroegindeweij C. Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines. *Radiother Oncol* 2010; **94**: 286–91.
13. Machiels M, Jin P, van Hooft JE, Gurney-Champion OJ, Jelvehgaran P, Geijsen ED, et al. Reduced inter-observer and intra-observer delineation variation in esophageal cancer radiotherapy by use of fiducial markers. *Acta Oncol* 2019; **58**: 943–50. doi: https://doi.org/10.1080/0284186X.2019.1588991
14. Mitchell DM, Perry L, Smith S, Elliott T, Wylie JP, Cowan RA, et al. Assessing the effect of a contouring protocol on postprostatectomy radiotherapy clinical target volumes and interphysician variation. *Int J Radiat Oncol Biol Phys* 2009; **75**: 990–3. doi: https://doi.org/10.1016/j.ijrobp.2008.12.042
15. Fuller CD, Nijkamp J, Duppen JC, Rasch CRN, Thomas CR, Wang SJ, et al. Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting. *Int J Radiat Oncol Biol Phys* 2011; **79**: 481–9. doi: https://doi.org/10.1016/j.ijrobp.2009.11.012
16. Khoo ELH, Schick K, Plank AW, Poulsen M, Wong WWG, Middleton M, et al. Prostate contouring variation: can it be fixed? *Int J Radiat Oncol Biol Phys* 2012; **82**: 1923-9. doi: https://doi.org/10.1016/j.ijrobp.2011.02.050
17. Dewas S, Bibault J-E, Blanchard P, Vautravers-Dewas C, Pointreau Y, Denis F, et al. Delineation in thoracic oncology: a prospective study of the effect of training on contour variability and dosimetric consequences. *Radiat Oncol* 2011; **6**: 118. doi: https://doi.org/10.1186/1748-717X-6-118
18. Eriksen JG, Salembier C, Rivera S, De Bari B, Berger D, Mantello G, et al. Four years with FALCON - an ESTRO educational project: achievements and perspectives. *Radiother Oncol* 2014; **112**: 145–9. doi: https://doi.org/10.1016/j.radonc.2014.06.017
19. ESTRO.ESTRO – Education pre-meeting courses – Foundations of Leadership in Radiation Oncology. 2020. Available from: https://www.estro.org/Congresses/ESTRO-38/Pre-meeting-courses/Education-pre-meeting-course-Foundations-of-Leader.
20. Chang ATY, Tan LT, Duke S, Ng W-T. Challenges for quality assurance of target volume delineation in clinical trials. *Front Oncol* 2017; **7**: 221. doi: https://doi.org/10.3389/fonc.2017.00221
21. Iqbal MS, Hothi A, Evans ES, Gilson D, Laws K, Saleem W, et al. An evaluation report on radiotherapy contouring workshops during the Royal College of radiologists' annual meeting 2018. *Clin Oncol* 2020; **32**: 276. doi: https://doi.org/10.1016/j.clon.2019.09.060
22. Casswell G, Shakir R, Macnair A, O'Leary B, Smith F, Rulach R, et al. Uk training in clinical oncology: the trainees' viewpoint. *Clin Oncol* 2018; **30**: 602e604. doi: https://doi.org/10.1016/j.clon.2018.06.009
23. Gwynne S, Gilson D, Dickson J, McAleer S, Radhakrishna G. Evaluating target volume delineation in the era of precision radiotherapy: FRCR, revalidation and beyond. *Clin Oncol* 2017; **29**: 436–8. doi: https://doi.org/10.1016/j.clon.2017.01.045
24. Evans E, Piazzese C, Spezi E, Staffurth J, Gwynne S. Arena: improving training in target volume delineation for radiotherapy. *Radiotherapy and Oncology* 2019; **133**: 896–7.
25. Gwynne S, Higgins E, Poon King A, Radhakrishna G, Wills L, Mukherjee S, et al. Driving developments in UK oesophageal radiotherapy through the scope trials. *Radiat Oncol* 2019; **14**: 26. doi: https://doi.org/10.1186/s13014-019-1225-0
26. Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. *Med Phys* 2003; **30**: 979–85. doi: https://doi.org/10.1118/1.1568978
27. Apte AP, Iyer A, Crispin-Ortuzar M, Pandya R, van Dijk LV, Spezi E, et al. Technical note: extension of CERR for computational radiomics: a comprehensive Matlab platform for reproducible radiomics research. *Med Phys* 2018;13 Jun 2018. doi: https://doi.org/10.1002/mp.13046
28. Michalski JM, Lawton C, El Naqa I, Ritter M, O'Meara E, Seider MJ, et al. Development of RTOG consensus guidelines for the definition of the clinical target volume for postoperative conformal radiation therapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2010; **76**: 361–8. doi: https://doi.org/10.1016/j.ijrobp.2009.02.006
29. Myerson RJ, Garofalo MC, El Naqa I, Abrams RA, Apte A, Bosch WR, et al. Elective clinical target volumes for conformal therapy in anorectal cancer: a radiation therapy Oncology Group consensus panel contouring atlas. *Int J Radiat Oncol Biol Phys* 2009; **74**: 824–30. doi: https://doi.org/10.1016/j.ijrobp.2008.08.070
30. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004; **23**: 903–21. doi: https://doi.org/10.1109/TMI.2004.828354
31. Evans E, Piazzese C, Spezi E, Staffurth J, Gwynne S. Arena: improving training in target volume delineation for radiotherapy. *Radiotherapy and Oncology* 2019; **133**: 896–7.
32. International Commission on Radiation Units and Measurements. ICRU report 83 prescribing, recording, and reporting Photon-beam intensity-modulated radiation therapy (IMRT. *Journal of the ICRU* 2010; **10**.
33. Gwynne S, Spezi E, Hurt C, Falk S, Gollins S et al. eds. *Importance of the reference volume in assessing outlining performance for the purpose of training and revalidation. NCRI Cancer Conference*. UK: Liverpool; 2012 .
34. Gwynne S, Spezi E, Wills L, Nixon L, Hurt C, Joseph G. Inter-Observer variation in Outlining of Pre-trial test case in the SCOPE1 trial: a United Kingdom definitive chemoradiotherapy trial for esophageal cancer. *International Journal of Radiation Oncology Biology Physics* 2012; **84**: 1037–42.
35. Sherer MV, Lin D, Puri K, Panjwani N, Zhang Z, Murphy JD, et al. Development and Usage of *eContour*, a Novel, Three-Dimensional, Image-Based Web Site to Facilitate Access to Contouring Guidelines at the Point of Care. *JCO Clin Cancer Inform* 2019; **3**: 1–9. doi: https://doi.org/10.1200/CCI.19.00041
36. EduCaseTM.Welcome to FALCON. 2021. Available from: https://estro.educase.com/.
37. Xu H, Arsene Henry A, Robillard M, Amessis M, Kirova YM. The use of new delineation tool "MIRADA" at the level of regional lymph nodes, step-by-step development and first results for early-stage breast cancer patients. *Br J Radiol* 2018; **91**: 20180095. doi: https://doi.org/10.1259/bjr.20180095
38. ProKnow.ProKnow is high-impact, low-cost investment in quality, consistency, and improved patient care. 2021. Available from: https://proknowsystems.com.
39. Duke SL, Tan LT, Eminowicz G, Park WH, Wharrad H, Patel R, et al. Rapid radiotherapy contouring practice: pilot study of a novel web-based tool enabling automated individualized feedback. *International Journal of Radiation Oncology• Biology• Physics* 2019; **105**: E147.