Radiotherapy and Oncology 137 (2019) 1–8

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Review Article

Impact of deviations in target volume delineation – Time for a new RTQA approach?

Samantha Cox [a,*], Anne Cleves [b], Enrico Clementel [c], Elizabeth Miles [d], John Staffurth [e], Sarah Gwynne [a,f]

[a] *South West Wales Cancer Centre, Singleton Hospital, Swansea;* [b] *Velindre NHS Trust Library, Velindre Cancer Centre, Cardiff, UK;* [c] *EORTC Headquarters, Brussels, Belgium;* [d] *Mount Vernon Cancer Centre, Northwood;* [e] *School of Medicine, Cardiff University and Velindre Cancer Centre; and* [f] *Swansea University Medical School, Swansea, UK*

ARTICLE INFO

Article history:
Received 21 January 2019
Received in revised form 3 April 2019
Accepted 7 April 2019
Available online 28 April 2019

Keywords:
Radiation Therapy Quality Assurance
Target volume delineation
Clinical trials

ABSTRACT

The international radiotherapy community has recognised that non-adherence to RT protocols can influence trial endpoints. However this conclusion is based on studies predominantly assessing the impact of deviations in dosimetric or treatment delivery protocol parameters rather than target volume delineation (TVD). This review evaluates the assessment of TVD within Radiation Therapy Quality Assurance (RTQA) programmes in clinical trials and the clinical impact of TVD protocol deviations. The implications for RTQA programmes are discussed. MEDLINE, PreMEDLINE, Embase, Cochrane Library, Web of Science, OpenGrey, WHO International Clinical Trials Registry Platform portal and ClinicalTrials.gov were searched. Full-length articles and conference abstracts were included to avoid publication bias. 5864 abstracts were screened for relevance; 94 full-length articles were reviewed and 5 relevant trials identified. Various classification systems were used to assess protocol deviations; 'unacceptable' or 'major' deviations in TVD occurred in 2.9–13.4% of assessed RT plans (when reported). It was often not possible to establish deviation rates specifically related to TVD as these were frequently combined with other types of protocol deviations. Details on the nature of unacceptable deviations was also not routinely reported and difficulties in establishing a 'consensus' for appropriate TVD for on-trial patients highlighted. Results suggest that deviations in TVD were associated with poorer outcomes for overall survival, local control and treatment-related toxicity; however the data were heterogeneous. RTQA of TVD was retrospective and feedback on the quality of TVD to recruiting centres was not standard. In summary, few trials have published outcomes on the impact of assessing the quality of TVD in trials. We propose that a new approach is now required. Unacceptable TVD deviations must be clearly defined at the time of protocol development to minimise interobserver variation, thereby promoting consistency in RTQA feedback. Prospective TVD reviews should be implemented for trials involving novel or complex RT techniques to identify deviations that require modification prior to treatment delivery. Furthermore, the consistent reporting of RTQA programme outcomes, both within and across trial groups, is of paramount importance to accelerate the evidence-base for the best RTQA approach when assessing TVD and to enable the impact on clinical outcomes within RT trials to be assessed.

Radiation Therapy Quality Assurance (RTQA) is now an integral part of clinical trials involving radiotherapy (RT). The aim is to ensure adherence to RT trial protocols, thereby minimising variations in the planning and delivery of RT between recruiting centres which may subsequently influence trial results [1].

In recent years RT has become more complex with significant technological advances such as intensity-modulated RT (IMRT), image-guided RT (IGRT) and adaptive RT. This improvement in conformality has led to reduction in size of margins around target volumes and steeper dose gradients, necessitating more accurate target volume delineation (TVD). Developments in both diagnostic and treatment simulation imaging modalities (which can now be readily co-registered and fused) have improved the accuracy of TVD. This has been complemented by the widespread publication of site-specific consensus guidelines and atlases [2], often developed from RT planning guidance documents within trial protocols. Furthermore the increasing use of peer-review may also improve the overall standard of TVD within clinical departments, acting as a routine QA tool [3]. However, interobserver variation in TVD remains both within and outside of clinical trials [4,5] and it has been argued that the magnitude of these delineation inaccuracies may now exceed that of geometric systematic errors [6,7].

In 2010, the Global Quality Assurance of Radiation Therapy Clinical Trials Harmonisation Group (GHG) was established to standardise and develop evidence-based RTQA procedures used

* Corresponding author.
E-mail address: samantha.cox@wales.nhs.uk (S. Cox).

https://doi.org/10.1016/j.radonc.2019.04.012
0167-8140/

by the various clinical trial groups across the world, thereby facilitating intergroup collaborations [8]. The group identified the need for a standardised RTQA naming convention [9], as well as consistent reporting of trial quality factors and robust data analysis.

The GHG has recognised that non-adherence to RTQA programmes directly correlates with patient outcomes. This conclusion was based on a systematic review by Weber et al. [10] which included the widely cited TROG 02.02 HeadSTART trial of radiotherapy in advanced head and neck cancer that demonstrated a significant reduction in overall survival (OS) and locoregional failure-free rates in patients with major deficiencies in their RT treatment plans [11]. However, the majority of trials focused on deviations related to the prescription criteria, dosimetry or technical issues with RT planning (such as 90% isodose not encompassing the planning target volume (PTV), dose delivered not within ±10% of the prescribed dose, extended overall treatment time, variations in dose per fraction), highlighting the impact of sub optimally delivered RT on patient outcomes. Indeed the resulting impact of inaccurate TVD on clinical outcomes has not been fully investigated within the trial setting [2,6].

This systematic review evaluates the assessment of TVD within RTQA programmes and the subsequent clinical impact of TVD protocol deviations. The implications of the results for RTQA programmes are discussed.

Materials and methods

Search strategy

Electronic searches were performed on Medline, Medline In Process & Epubs ahead of Print, Embase, Cochrane Library, Web of Science, OpenGrey, World Health Organisation International Clinical Trials Registry Platform (ICTRP) and ClinicalTrials.gov. All databases and trial registers were searched from 2005 to present for English language material only. The searches were last performed on the 23rd January 2018. The search strategies included relevant subject terminology and text words around the terms "radiotherapy planning", "interobserver variation", "target volume delineation", "radiotherapy trial quality assurance", "radiation therapy quality assurance", "randomised controlled trial", "individual case review", "protocol non-adherence", and "protocol non-compliance". Manual searches of reference lists were also conducted.

Studies relating to the clinical impact of clinician non-adherence to TVD protocols within RT trials were included. To avoid publication bias, both full-length articles and conference abstracts were deemed eligible. 'Clinician' was defined as a clinical oncologist or radiation oncologist participating in the RT trial. Only trials using external beam RT techniques including 3D conformal RT and IMRT were included; brachytherapy trials were excluded. Trials using anatomical landmarks or 2D field borders were excluded as it was felt these techniques do not reflect current clinical practice and that both TVD and coverage of GTV and CTV could not be accurately assessed within such studies.

Studies were deemed eligible according to the following criteria: (1) TVD had been included as part of the RTQA process in a phase II or III clinical trial; (2) the impact of any TVD protocol deviations had been assessed and categorised; (3) the impact of any TVD protocol deviations was correlated with a clinically meaningful and/or relevant patient outcome. Two authors (SC, SG) screened the abstracts for relevance, which was followed by assessment of eligibility of the full-length articles. Authors of conference abstracts were contacted via email in all instances to identify whether any updated results were available and if a manuscript was being drafted or had been submitted for publication in a peer-reviewed journal; a period of 28 days to allow a response was permitted. Conference abstracts initially identified by the literature search that were later published as a journal paper were not included. When necessary, authors of papers were contacted via email to expand on methods used in assessing the protocol deviations and to clarify any results; a period of 28 days to allow a response was permitted. In cases where an email response was not received from authors of abstracts or full-length articles, the studies were still included in the review to avoid publication bias; missing data were recorded in tables as 'not reported' or 'not available' (NA).

Results

Overall, the results of 5 trials were included in this review, which relate to 7 records identified by the literature search (Fig. 1; Table 1). The outcomes of 3 trials were reported as articles in peer-reviewed journals [11–14] whilst the outcomes of the other 2 trials were published as conference abstracts [15–17]; it was noted that different analyses of the NRG/RTOG 0522 trial were reported as two separate abstracts [16,17].

3DCRT was used in all studies and 2 trials permitted the use of IMRT [15–17]; tumour sites included head and neck, pancreas, prostate and small cell lung cancer. The number of recruiting centres in the trials ranged from 23 to 151 centres and were open to patient randomisation from 2002 to 2013. Major or unacceptable deviations in TVD occurred in 2.9–13.4% of RT plans (when reported) (Table 2). Various outcomes were assessed including OS [15–17], failure-free survival [11], time to locoregional failure [11], local control [12,16,17], distant metastases [11,16,17] and treatment-related toxicity [14].

Definition of protocol deviations

In 2014, the GHG published standardised RTQA naming conventions. A 'major' deviation was defined as a significant variation from the protocol that may affect the validity of interpretation of trial results and 'minor' deviations were those which require education or modification to prevent major deviations from occurring [9]. This has now been refined to categorise RT treatment as either 'acceptable-per protocol' or in the event of any departure from the protocol, variations are classified as either 'acceptable' or 'unacceptable' according to the likely impact on trial outcomes (Table 2) [18]. Various definitions were used to assess protocol deviations in the 5 studies reviewed (Table 1). It was assumed that unacceptable or major deviations were equivalent in terms of their impact.

Assessment of protocol deviations

Quantitative assessment of TVD was performed in the majority of studies and related to growth of margins around target volumes as stipulated within the protocol [12–17]. Qualitative assessment of the TVD was performed in all of the trials (Table 1); two reported on the accuracy of GTV delineation [11,14]; both GTV and CTV were reviewed in the TROG 03.04 RADAR and NRG/RTOG 0522 trials [12,13,16,17]; limited details were available for CONVERT trial with 'target volumes' reviewed [15]. RTQA team access to diagnostic imaging is required to enable a complete and accurate assessment of GTV delineation, however this practice was only reported in one trial [11].

RTQA outcomes for OAR contouring and dosimetric or treatment delivery parameters were also recorded in most instances. Deviations in these protocol items were often combined with the frequencies of TVD deviations. It was therefore not possible to establish the number of deviations specifically relating to TVD in some of the trials and these were only reported separately for three of the studies (Table 3) [11,14,17]. The details on what specific

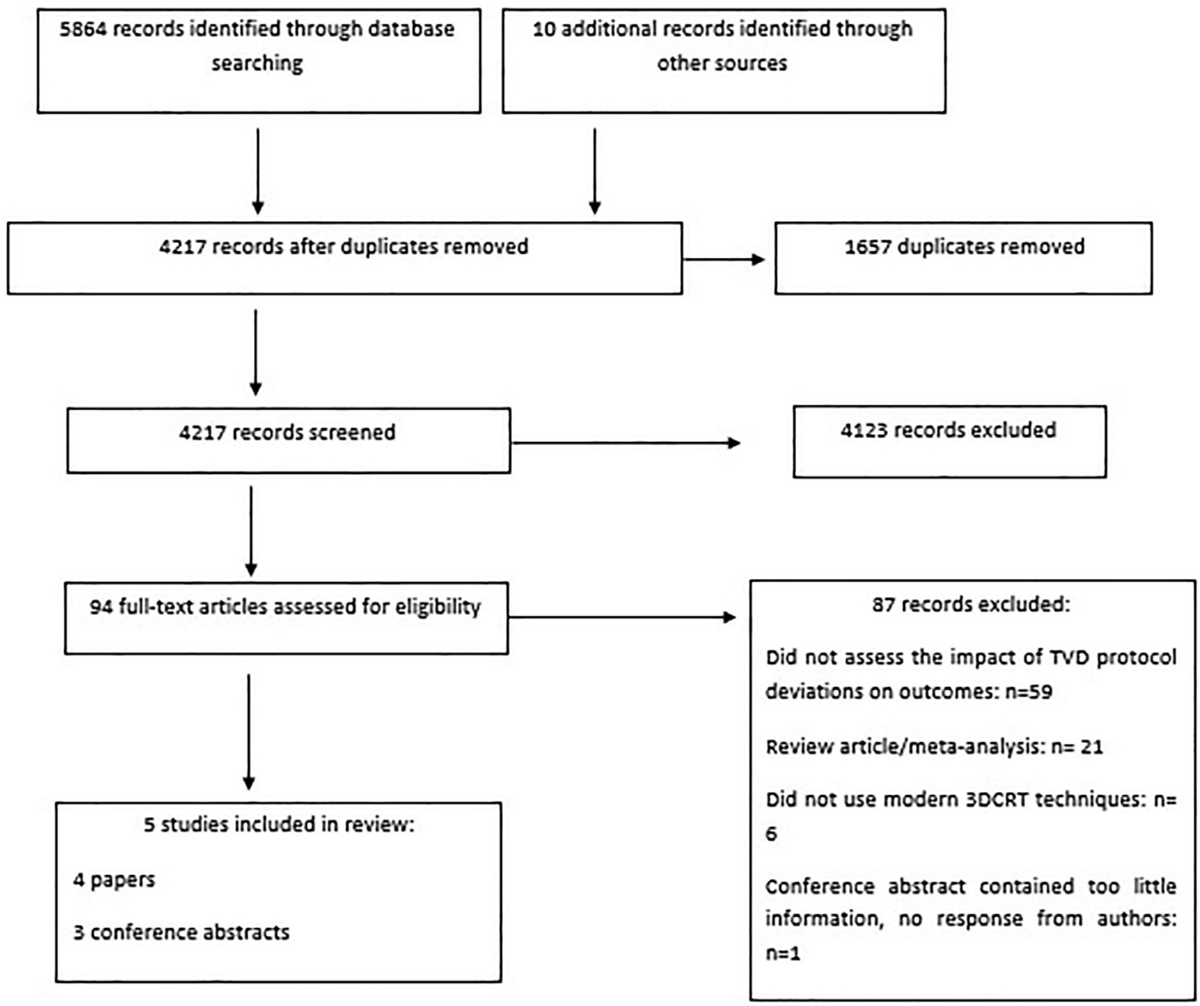

Fig. 1. Systematic search outcomes.

deviations in TVD were deemed to be 'unacceptable' was not routinely reported.

Timing of RTQA process and feedback to recruiting centres

On-trial RTQA is often conducted using an individual case review (ICR) approach which can be either prospective or retrospective. The benefits of the former are that any unacceptable deviations can be identified and corrected prior to the patient commencing treatment [19]. However this approach is time and labour intensive for both the recruiting centre and RTQA team, with a lack of funding for the clinical review of TVD in some trial groups. RTQA processes should aim not to delay the start of treatment however the often short turnaround time between full data set submission and start date remains a major issue, particularly if amendments are suggested. Timely retrospective reviews, performed within the first few days of treatment, have more recently been introduced which may reduce the time pressures on RTQA teams to turn-around the reviews without causing treatment delays [20]; this allows protocol deviations to be corrected (and compensated for), however the frequency of reviews with this approach is still dependent on the recruitment rate to the trial which may therefore not reduce the overall burden of RTQA processes.

All of the studies in this review performed retrospective RTQA, usually via central review by a panel of radiation oncologists and radiation therapists or physicists (Table 1). The type of feedback provided to centres was reported in three trials. In the CONVERT trial, feedback on a random selection of treatment plans from each centre was provided, with agreement from the centre to address any issues identified in future submissions [21].

A timely retrospective 'interventional review' of 687 (80.5%) RT plans prior to the end of the first week of treatment was performed in TROG 02.02 HeadSTART, with feedback given within 4 days of receipt [11]. 197 of these plans were identified as requiring modification with a turnaround time to final RTQA feedback of 15 days (median; range −4 to 72 days) after the commencement of RT, however the recommended changes were not made in 54.8% of cases. A secondary review of 820 (96.1%) RT plans for protocol compliance was conducted by a central panel following completion of treatment; compliance was correlated to the outcome of the initial timely retrospective review conducted during the first week of RT treatment [11]. 208 patients were identified as having received noncompliant radiotherapy, of which 97 RT plans were graded as having protocol deviations with a predicted major adverse impact on tumour control (as determined by a single RTQA reviewer). The majority related to inadequate dose coverage of gross disease (41/97; 42.3%) followed by incorrect dose prescription (25/97; 25.8%), incorrect TVD (24/97; 24.7%) and prolonged treatment time (7/97; 7.2%).

The TROG 03.04 RADAR trial reported feedback to centres within one 1 week of plan submission [13]. However the authors

Table 1
Summary of the trials.

Study	Phase Years of randomisation Tumour site	Number of recruiting centres	Benchmark case	Timing of individual case review relative to RT treatment	Feedback provided to recruiting centres	Classification of protocol deviations	Protocol items assessed relating to TVD	Clinical outcome measure
TROG 02.02 HeadSTART [11] Cisplatin-based definitive chemoradiotherapy (70 Gy in 35 fractions over 7 weeks) ± tirapazamine	Phase III 2002–2005 Locally advanced H&N	82	Not reported	Retrospective	Yes	Compliant Non-compliant but no predicted major adverse impact on TCP Non-compliant with predicted major adverse impact on TCP	GTV delineation (other items included OAR contouring, treatment planning and delivery parameters)	OS at 2 years FFS at 2 years[a] TTLRF at 2 years[a]
TROG 03.04 RADAR [12,13] ADT (18 months vs. 6 months) + definitive RT (66 Gy, 70 Gy or 74 Gy in 2 Gy per fraction) ± zoledronic acid	Phase III 2003–2008 Prostate	23	Non-mandatory benchmark case (completed by 56.6% of recruiting centres)	Prospective and retrospective Central review process using SWAN software platform	Yes	Minor – outside protocol recommendations but unlikely to influence clinical outcome Major – outside protocol recommendations and may influence clinical outcome	GTV and CTV delineation PTV margins (other items included OAR contouring, dosimetric and treatment planning parameters)	LC progression[b]
RTOG 0411 [14] Capecitabine based chemoradiotherapy (50.4 Gy in 28 fractions) followed by maintenance gemcitabine ± bevacizumab	Phase II 2005–2006 Unresectable localised pancreas cancer	36	Not reported	Retrospective Central review process	Not reported	Major -GTV >5 cm (in any dimension) greater than actual tumour size -inability to contour GTV -use of larger margins	GTV delineation PTV margins	Acute grade 3 or greater GI toxicity during chemoradiotherapy and maintenance chemotherapy
CONVERT [15] Cisplatin-etoposide based chemoradiotherapy; twice-daily RT (45 Gy in 30 fractions) vs. once daily 66 Gy in 33 fractions	Phase III 2008–2013 Limited stage small cell lung cancer	73	No standard benchmark case Centres selected previously treated case, submitted re-planned for each treatment arm	Retrospective Central review process using VODCA software system	Yes	Acceptable Acceptable variation Unacceptable variation	Target delineation Margins (other items included OAR, treatment planning and delivery parameters)	OS
NRG/RTOG 0522 [16,17] Cisplatin-based chemoradiotherapy (54 Gy in 30 fractions with a boost of 18 Gy in 12 fractions to the GTV and clinically involved nodes) ± cetuximab	Phase III 2005–2011 Locally advanced H&N	151	Not reported	Retrospective Central review process	Not reported	Per-protocol Variation acceptable Deviation unacceptable	Target volume contouring PTV margins (other items included OAR contouring and treatment planning parameters)	OS, distant metastases and local recurrence at 2 years

Abbreviations: ADT, androgen deprivation therapy; CR, complete response; CTV, clinical target volume; FFS, failure-free survival; GI, gastrointestinal; H&N, Head and neck squamous cell carcinoma; LC, local-composite; OAR, organs at risk; OS, overall survival; PTV, planning target volume; RTQA, radiation therapy quality assurance; TCP, tumour control probability; TTLRF, time to locoregional failure; TVD, target volume delineation; VODCA, Visualisation and Organisation of Data for Cancer Analysis.

[a] Failure defined as persistent disease in primary site, disease progression in neck in patients not undergoing neck dissection, residual disease post-dissection, locoregional relapse after CR, or distant metastases.

[b] Post-treatment occurrence of either local or clinical failure (as confirmed by CT imaging or digital rectal examination following biochemical failure) or PSA progression, with a PSA doubling time between 6 and 100 months).

Table 2
Definition of protocol deviations.

	GHG definition [18]	Applicability to RTQA of TVD
Acceptable – per protocol	RT planned/delivered to the patient according to the protocol specifications and meets the per protocol criteria as defined by the protocol	Within the consensus range as defined by RTQA team
Acceptable variation	RT not planned/delivered according to all protocol specifications No major clinical impact expected	Remains within consensus range If prospective or timely retrospective review: RT treatment can proceed/continue but feedback to centre required May consider reviewing future submission(s)
Unacceptable variation	RT treatment not in accordance with the protocol Deviation(s) could have a clinical impact on the patient Trial outcome could be affected and true trial outcome obscured if the patient's data were included in the final analysis Chief investigator/RTQA team to decide whether to exclude patient from the trial and any other remedial action	Outside of consensus range Feedback of issues and how to improve TVD required If prospective or timely retrospective review: may need to resubmit TVD before proceeding to planning/continuing treatment Future submission(s) should be reviewed to ensure issues not recurring

Abbreviations: RT, radiotherapy; RTQA, radiotherapy quality assurance; TVD, target volume delineation.

identified that the date of RT plan submission relative to the date of RT commencement varied widely (median time of 70 days, range −286 to 1158 days). A separate audit process was undertaken periodically, during which a central panel of 12 individuals reviewed the margins applied to CTV to generate PTV in a sample of submitted RT plans [13]; any plans that were scored as 'requiring feedback' or 'equivocal' were then reviewed by a second member of the panel to ensure consistency before feedback was provided to the treating centre. In total, 5 rounds of auditing were conducted in 19 of the 23 centres; the authors did not report why the delineation in 4 centres was not assessed. 7 centres were identified as having systematic errors which required the review of subsequent RT plan submissions to ensure a change in practice in line with the protocol. Protocol variations included absence of superior and inferior PTV margins, smaller margins than specified in the protocol and a difference between the outlined CTV and the panel's opinion of what should have been delineated. However the authors highlighted that there was difficulty in assessing CTV contouring due to a lack of consensus amongst the reviewers as to what would constitute a major or minor variation in TVD. Furthermore, the total number of plans reviewed in this audit process was not reported and the results were not included in the subsequent substudy assessing the impact of RTQA and disease progression by Marcello et al. [12].

Impact of protocol deviations on clinical outcomes

Four of the trials reported the impact of protocol deviations, including those related to TVD, on survival or local control. In the TROG 02.02 HeadSTART trial, patients with protocol deviations predicted to have a major adverse impact on tumour control had significantly worse OS at 2 years compared to those with initially compliant RT (50% vs 70%, respectively; hazard ratio (HR), 1.99; $p < 0.001$); freedom from locoregional failure was also worse (54% vs 78%, respectively; HR 2.37; $p < 0.001$) [11]. This effect remained highly significant in multivariate analysis and was independent of treatment arm allocation.

In the TROG 03.04 RADAR trial, major protocol deviations were associated with local composite (LC) progression (defined as the post-treatment occurrence of either local or clinical failure (as confirmed by CT imaging or digital rectal examination following biochemical failure) or PSA progression, with a PSA doubling time between 6 and 100 months). However this was only seen in low-risk disease patients receiving 70 Gy ($p = 0.01$, borderline significant) and was not significant in multivariate analysis ($p = 0.66$) [13].

The CONVERT trial used the frequency of protocol deviations to calculate a multiplicative factor (F) for each treatment plan, whereby a low number of protocol deviations equated to a low F score [15]. Poorer OS outcomes were associated with both the nature and frequency of protocol deviation however specific details were not reported in the abstract.

Two feasibility studies used data from the NRG/RTOG 0522 trial to demonstrate that it is possible to include RTQA results in outcome modelling [16,17]. Local control rates were significantly better for patients whose target volume and OAR delineation were completed as per-protocol compared with those with either acceptable or unacceptable deviations (log rank testing p value = 0.004 and 0.043, respectively); there was no association with OS and distant control rates [16]. Incorporating RTQA results into Cox modelling for local control improved the model performance as assessed by C-index calculation. A separate substudy confirmed that the deviations in TVD were associated with local recurrence rates on log-rank testing ($p = 0.01$) [17]. As there was no evidence of an association with two-year OS, this parameter was not carried forward into the logistic regression model; the only RTQA parameter identified as affecting OS was the assessment of dose–volume coverage of targets.

The RTOG 0411 study was the only study to assess the impact of TVD deviations on treatment-related toxicity. Rates of acute grade 3 or greater GI toxicity were significantly higher in patients with major deviations compared to those without, both during chemoradiotherapy (45% vs 18%, adjusted odds ratio 3.7; 95% CI 0.98–14.1, $p = 0.05$) and on maintenance chemotherapy (45% vs 13%, adjusted odds ratio 5.7; 95% CI 1.45–22.8, $p = 0.01$) [14]. Retrospective review of TVD identified that all cases of major (unacceptable) deviations in the trial related to over-contouring of the GTV by at least 5 cm greater than the actual tumour size; there were no deviations relating to the inability to include the primary tumour in the GTV or inappropriate margins. The authors argued that the higher rates of GI toxicity may therefore have occurred due to the inclusion of normal gastrointestinal mucosa within the GTV which could be avoided in future trials by prospective RTQA and pancreatic-specific diagnostic imaging protocols.

Discussion

This systematic review summarises the phase II-III RT trials published since 2005 which have assessed the clinical impact of non-adherence to TVD protocols. Approximately 40% of the RT plans that underwent RTQA had some form of protocol deviation, with major or unacceptable TVD deviations occurring between 2.9% and 13.4% of RT plans. Published results to date suggest that such deviations may impact OS, local control and treatment-related toxicity outcomes. However, very few trials have published data on this issue.

To our knowledge, this is the first review to specifically assess the clinical impact of variations in TVD as identified by RTQA programmes. Weber et al. concluded from their review of 9 prospec-

Table 3
Type and frequency of protocol deviations.

Study reference	Number of patients receiving RT within the trial whose plans underwent RTQA n (%)	Protocol deviations (all types)			Deviations specifically relating to TVD			Summary of the impact of protocol deviations on clinical outcome
		Acceptable/per-protocol/compliant n (%)	Acceptable/minor deviations n (%)	Unacceptable/major deviations n (%)	Acceptable/per-protocol n (%)	Acceptable/minor deviations n (%)	Unacceptable/major deviations n (%)	
TROG 02.02 HeadSTART [11]	820/853 (96.1%)	612/820 (74.6%)	109/820 (13.3%)	97/820 (11.8%)	NA	NA	24/820 (2.9%)[a]	2 years OS and TTLRF significantly worse in H&N cancer patients with major protocol deviations compared to those with initially compliant RT
TROG 03.04 RADAR [12,13]	754/813 (92.7%)	NA	1185/14326[b] (8.3%)	86/14326[b] (0.6%)	NA	NA	NA	Protocol deviations only associated with LC progression in sub-group of prostate cancer patients with low-risk disease
RTOG 0411 [14]	82/82 (100%)	NA	NA	NA	NA	NA	11/82 (13.4%)[c]	Major protocol deviations associated with grade 3 or greater GI toxicity during treatment for pancreatic cancer patients
CONVERT [15]	94/489 (19.2%)	57.6%	23.3%	19.1%	NA	NA	NA	OS in small cell lung cancer patients affected by the number and type of protocol deviation
NRG/RTOG 0522 [16,17]	896/896 (100%)	NA	NA	NA	472/896[d] (52.7%)	331/896 (36.9%)	50/896 (5.6%)	Target contouring quality grade was associated with local recurrence but not OS or distant metastases in H&N cancer patients

Abbreviations: GI, gastrointestinal; H&N, Head and Neck; LC progression, post-treatment occurrence of either local or clinical failure (as confirmed by CT imaging or digital rectal examination following biochemical failure) or PSA progression, with a PSA doubling time between 6 and 100 months; OS; overall survival; NA, not available or not reported; TTLRF, time to locoregional failure; TVD, target volume delineation.

[a] 2 not evaluated

[b] Deviation rates expressed as a frequency across the total number of protocol items reviewed (14326 protocol items assessed in the 754 plans reviewed).

[c] All related to excessive inclusion of normal tissue within the GTV.

[d] 43 (4.8%) non-evaluated/missing.

tive phase II-III clinical trials that there was undisputed evidence that RT protocol deviations do have a detrimental impact on patient outcomes, with major deviations occurring in 11% to 48% (mean 34.2%) of RT plans [10]. However only 22% of the studies reported major protocol deviations relating specifically to TVD. A meta-analysis of 8 trials by Ohri et al. identified that the frequency of RT deviations ranged from 8% to 71% and that protocol deviations were associated with increased risks of treatment failure (HR = 1.79, 95% CI 1.28–2.35; $p < 0.001$) and overall mortality (HR = 1.79, 95% CI 1.14–2.78; $p = 0.009$) [22]. Although both of these papers included the TROG 02.02 trial, the majority of the included studies focused on treatment planning and delivery protocol parameters rather than TVD. Furthermore, many were conducted between the 1970s and early 2000s using what would now be classed as outdated RT techniques, such as the use of field margins and borders in relation to anatomical landmarks rather than assessments of TVD in the context of current advanced radiotherapy planning.

This review is limited by the small number of studies to date that have conducted an analysis of TVD deviations on patient outcomes. Two trials have only reported in abstract form with limited results, however to avoid publication bias we felt it appropriate to include these studies. It is also difficult to assess for any bias in the feedback process itself for the trials included in our review as none of the trials reported on whether the names of the recruiting centre and clinician responsible for TVD were anonymised prior to the central panel review; if the RTQA reviewers were affiliated with either the centre or clinician, this knowledge could have influenced the grading of any deviations. Although this review suggests that non-adherence in TVD may have negative clinical consequences, interpretation of the results is limited. Different classification systems to define protocol deviations are in use and reporting of data remains heterogeneous. The use of standardised nomenclature for grading deviations is therefore recommended. Furthermore the type and frequency of protocol deviations should be routinely recorded according to whether they relate to TVD or treatment planning and delivery.

RTQA performed by a central review panel not directly involved with the management of an individual patient remains challenging. Assessment of TVD in the reviewed studies was retrospective and recommended changes in RT plans as highlighted by the RTQA team were not always instigated by recruiting centres. Processes for feedback on the quality of TVD to recruiting centres was also not standardised. Many RTQA groups have now introduced centralised uploading and review information technology (IT) software platforms to improve the review process [23–26]. This is particularly useful in large multi-institutional or international trials, enabling remote reviewing of RT plans by members of the RTQA team who may be geographically separated. There is also the possibility of automated reviews for some aspects of the RTQA process such as the use of standardised structure nomenclature, appropriateness of applied margins and compliant dose distributions [27]. Although it is technically possible to have access to diagnostic imaging and multidisciplinary meeting discussions to facilitate the review of TVD, a lack of consensus amongst the review panel regarding both the classification and clinical relevance of observed contouring deviations within a trial can persist [13,14].

Whilst clear and unambiguous RT planning guidance documents within trials can serve as educational tools to reduce uncertainty in TVD, interobserver variation amongst reviewers needs to be addressed. This could be achieved by developing a consensus at the time of protocol development as to which specific deviations in TVD are of an unacceptable nature as well as criteria against which any deviations in TVD can be judged. Pictorial examples of unacceptable deviations could be included within radiotherapy planning guidance documents, in addition to step-by-step written TVD instructions and case studies in order to further educate clinicians. Standardised proforma documents could be used to report feedback to recruiting centres with reference made to the specific planning CT slice number of any deviations; in our experience it is also useful to include screenshots and graphical annotations highlighting the area(s) of concern.

The turnaround time for prospective reviews of TVD by a central panel is reliant on the timely submission of the full RT plan datasets by recruiting centres. Unacceptable deviations in TVD requiring RT plan modification should be resubmitted for a further review and ideally not delay the start of treatment. We therefore suggest an optimal turnaround time of 72 h, which is now feasible via computerised central review platforms. With improvements to RTQA informatic capabilities, it would also be possible for diagnostic and follow-up imaging, and non-DICOM data (such as imaging reports and patient history) to be anonymised, uploaded and stored. Although there are significant resource implications, it would facilitate a more accurate review of TVD in the event of any clinical queries; it would also enable a more accurate analysis of the clinical impact of any TVD deviations in the case of treatment failures [28,29]. Furthermore, if correlations between any such deviations and patterns of disease relapse were to be identified, this could be addressed in future RT trial designs [25,30].

Recruiting centres should be made aware of practices in TVD which are likely to be deemed unacceptable to further facilitate protocol adherence. This may be supported by pre-trial benchmark cases which provide education and feedback, highlighting any issues requiring attention prior to patient recruitment. Benchmark cases must be carefully selected to ensure they are representative of the study population and that they adequately assess the clinician's understanding of the protocol [6]. Onerous pre-accrual TVD benchmark exercises may deter clinician participation in a trial or delay a centre opening to accrual if multiple resubmissions are required. Nonetheless, their value has been demonstrated in reducing the number of errors seen on-trial which in turn may avoid treatment delays [31]. Only two trials reported the use of a benchmark case in this current review [13,15]. The completion of the case was non-mandatory in the TROG 03.04 RADAR study but did improve familiarity with the protocol and identified planning deficiencies before patient recruitment started [13]. In the CONVERT trial, clinicians were asked to select an appropriate case previously treated in their centre that met the eligibility criteria for the study; the case was replanned for following the protocol for each treatment arm and the two plans were reviewed by the RTQA team [21]. With more efficient IT systems to process prospective reviews, trial groups may consider removing benchmark cases to minimise the time taken for centres to open to recruitment. However, there is a potential for the 'educational' component of such exercises to be lost. Robust mechanisms are also needed to ensure that high-quality RT continues to be delivered in centres that have successfully passed an ICR, for example selecting a random sample of treated patients from the trial cohort for timely retrospective reviews or recredentialing at set time points [25,32].

It is important that any unacceptable deviations are defined according to their clinical relevance. Although deviations were not correlated with survival outcomes in the ProtecT trial, a retrospective review of TVD in 108 prostate cancer patients identified an unacceptable deviation rate of 12% of RT plans [28]. However on further inspection this was only felt to be of possible clinical significance in 3 patients (1.4%), all relating to poor prostate outlining, with the remaining 10.6% of cases felt to be unlikely to have a clinical impact. We suggest that secondary analyses of the association between deviations and clinical outcome should be planned for at the time of trial conception. Trial groups should be encouraged to publish RTQA outcomes alongside the trial endpoints, as few have done so since the results of the TROG 02.02 trial were published. At the time this was hailed as a 'contemporary benchmark' in terms of RTQA programmes demonstrating the principle of evidence-based medicine in RT [10], and set the precedent for future trials. However, in our experience it can be challenging to successfully publish trial RTQA-related papers in peer-reviewed journals, in which case RTQA outcomes could be published in a supplementary report or made available on the trial website. This could raise awareness of the importance of RTQA, develop an evidence-base for the best RTQA approach and allow meaningful interpretation of RT trial outcomes. However, this will require ongoing support from clinical trial sponsors and cooperation from recruiting centres until mature data is available; data access for subsequent analyses also needs to be addressed.

Table 4
High-risk trial design factors requiring assessment of TVD accuracy in RT trials.

High-risk trial design factors requiring assessment of TVD accuracy in RT trials
Reduced RT margins
Dose escalation or hypofractionation
Concurrent use of novel systemic therapies in early phase RT trials
Multimodal trials e.g. surgery & RT
New techniques for image-guided RT or adaptive RT

Abbreviations: TVD, target volume delineation; RT, radiotherapy.

It is accepted that RTQA strategies should be trial-specific although there is a lack of evidence regarding which are the most effective RTQA programmes [25]. Ideally, prospective RTQA of TVD would be done for every patient recruited to a trial involving RT. Whilst this may be possible with improved automation in QA, it may be necessary to be trial selective given the considerable resource implications of such an approach. More intensive RTQA of TVD is likely to be required for trials in which there is a change in TVD from current standards of care or in the presence of factors for which incorrect TVD could conceal the benefit of the treatment being investigated. Innovative prospective trial designs are now required to establish an evidence base for the impact of RTQA [25,32]. This is likely to become increasingly important as trials incorporate other treatments and technologies, such as surgery and biomarkers, which in themselves pose significant QA challenges. The use of 'risk-adjusted' RTQA should therefore be explored (Table 4), similar to systems being introduced for surgical QA [1]. It is also important to consider the experience of recruiting centres in the proposed methods of TVD as the rate of protocol deviations and subsequent quality of RT is associated with individual centre-accrual rates and prior trial participation [10,11,15]. Finally, robust mechanisms are required to confirm that any required modifications in TVD, as identified by RTQA, are instigated prior to treatment starting.

Conclusion

The results of this review suggest that deviations in TVD are associated with poorer outcomes for OS, local control and treatment-related toxicity. However few trials have published outcomes on the impact of assessing the quality of TVD in trials. We propose that a new approach is now required. Unacceptable TVD deviations must be clearly defined at the time of protocol development to minimise interobserver variation, thereby promoting consistency in RTQA feedback. Prospective TVD reviews should be implemented for trials involving novel or complex RT techniques to identify deviations that require modification prior to treatment delivery. Furthermore, the consistent reporting of RTQA programme outcomes, both within and across trial groups, is of paramount importance. This will develop an evidence-base for RTQA processes and enable an assessment of the impact of deviations in TVD on clinical outcomes within radiotherapy trials.

Declaration of interest

EC is supported by Fonds Cancer (FOCA), Belgium. The other authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

[1] Members of EORTC QAC WGoSE. Multidisciplinary quality assurance and control in oncological trials: Perspectives from European Organisation for Research and Treatment of Cancer (EORTC). Eur J Cancer 2017;86:91–100.
[2] 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.
[3] Brundage M, Foxcroft S, McGowan T, Gutierrez E, Sharpe M, Warde P. A survey of radiation treatment planning peer-review activities in a provincial radiation oncology programme: current practice and future directions. BMJ Open 2013;3:e003241.
[4] Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy – Are they relevant and what can we do about them? Radiother Oncol 2016;50:254–62.
[5] Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. Radiother Oncol 2016;121:169–79.
[6] 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.
[7] Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. J Med Imaging Radiat Oncol 2010;54:401–10.
[8] Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ishikura S, et al. Radiation therapy quality assurance in clinical trials – global harmonisation group. Radiother Oncol 2014;111:327–9.
[9] Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ulin K, et al. Global harmonization of quality assurance naming conventions in radiation therapy clinical trials. Int J Radiat Oncol Biol Phys 2014;90:1242–9.
[10] Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy. Radiother Oncol 2012;105:4–8.
[11] 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.
[12] Marcello M, Ebert MA, Haworth A, Steigler A, Kennedy A, Bulsara M, et al. Association between measures of treatment quality and disease progression in prostate cancer radiotherapy: an exploratory analysis from the TROG 03.04 RADAR trial. J Med Imaging Radiat Oncol 2017;62:248–55.
[13] Kearvell R, Haworth A, Ebert MA, Murray J, Hooton B, Richardson S, et al. Quality improvements in prostate radiotherapy: Outcomes and impact of comprehensive quality assurance during the TROG 03.04 'RADAR' trial. J Med Imaging Radiat Oncol 2013;57:247–57.
[14] Crane CH, Winter K, Regine WF, Safran H, Rich TA, Curran W, et al. Phase II study of bevacizumab with concurrent capecitabine and radiation followed by maintenance gemcitabine and bevacizumab for locally advanced pancreatic cancer: Radiation Therapy Oncology Group RTOG 0411. J Clin Oncol 2009;27:4096–102.
[15] Groom N, Wilson E, Faivre-Finn C. Analysis of radiotherapy quality assurance data for the convert trial-does non-compliance to protocol affect survival? J Thorac Oncol 2017;12:S1745.
[16] Wang J, Zhong H, Van Soest J, Rosenthal DI, Axelrod R, Galvin JM, et al. A feasibility study on incorporating clinical trial quality assurance parameters into outcome prediction in head and neck radiation therapy treatment. Int J Radiat Oncol Biol Phys 2017;99:E380.
[17] Zhong H, Wang J, Van Soest J, Geng H, Huang M, Cheng C, et al. The evidence driven dosimetric constraints from outcome analysis of H&N patients' data from NRG oncology RTOG 0522 trial. Int J Radiat Oncol Biol Phys 2017;99:S137.
[18] Global Quality Assurance of Radiation Therapy Clinical Trials Harmonisation Group. GHG variations from protocol definitions. URL https://rtqaharmonization.com/publications/2019; [accessed on 21/01/19].
[19] Gondi V, Cui Y, Mehta MP, Manfredi D, Xiao Y, Galvin JM, et al. Real-time pretreatment review limits unacceptable deviations on a cooperative group radiation therapy technique trial: quality assurance results of RTOG 0933. Int J Radiat Oncol Biol Phys 2015;91:564–70.
[20] Evans E, Jones G, Radhakrishna G, Hawkins MA, Mukherjee S, Maggs R, et al. On-trial radiotherapy quality assurance in NeoSCOPE: a randomised phase II trial of chemoradiotherapy in oesophageal cancer. J Clin Oncol 2016;34:S119.
[21] Faivre-Finn C, Falk S, Ashcroft L, Bewley M, Lorigan P, Wilson E, et al. Protocol for the CONVERT trial-Concurrent ONce-daily VErsus twice-daily RadioTherapy: an international 2-arm randomised controlled trial of concurrent chemoradiotherapy comparing twice-daily and once-daily radiotherapy schedules in patients with limited stage small cell lung cancer (LS-SCLC) and good performance status. BMJ Open 2016;6:e009849.
[22] Ohri N, Shen X, Dicker A, Doyle L, Harrison A, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials. J Clin Oncol 2012;30.
[23] Ebert MA, Haworth A, Kearvell R, Hooton B, Coleman R, Spry N, et al. Detailed review and analysis of complex radiotherapy clinical trial planning data: evaluation and initial experience with the SWAN software system. Radiother Oncol 2008;86:200–10.
[24] Fairchild A, Aird E, Fenton PA, Gregoire V, Gulyban A, Lacombe D, et al. EORTC Radiation Oncology Group quality assurance platform: establishment of a digital central review facility. Radiother Oncol 2012;103:279–86.
[25] FitzGerald TJ, Bishop-Jodoin M, Bosch WR, Curran WJ, Followill DS, Galvin JM, et al. Future vision for the quality assurance of oncology clinical trials. Front Oncol 2013:00031.
[26] Gwynne S, Jones G, Maggs R, Eaton D, Miles E, Staffurth J, et al. Prospective review of radiotherapy trials through implementation of standardized multicentre workflow and IT infrastructure. Br J Radiol 2016;89:20160020.
[27] Santanam L, Hurkmans C, Mutic S, Van Vliet-Vroegindeweij C, Brame S, Straube W, et al. Standardizing naming conventions in radiation oncology. Int J Radiat Oncol Biol Phys 2012;83:1344–9.
[28] Mason MD, Moore R, Jones G, Lewis G, Donovan JL, Neal DE, et al. Radiotherapy for Prostate Cancer: is it 'what you do' or 'the way that you do it'? A UK perspective on technique and quality assurance. Clin Oncol (R Coll Radiol) 2016;28:e92–e100.
[29] Weber DC, Poortmans PMP, Hurkmans CW, Aird E, Gulyban A, Fairchild A. Quality assurance for prospective EORTC radiation oncology trials: the challenges of advanced technology in a multicenter international setting. Radiother Oncol 2011;100:150–6.
[30] Button M, Morgan C, Croydon E, Roberts S, Crosby T. Study to determine adequate margins in radiotherapy planning for esophageal carcinoma by detailing patterns of recurrence after definitive chemoradiotherapy. Int J Radiat Oncol Biol Phys 2009;73:818–23.
[31] Fairchild A, Collette L, Hurkmans CW, Baumert B, Weber DC, Gulyban A, et al. Do results of the EORTC dummy run predict quality of radiotherapy delivered within multicentre clinical trials? Eur J Cancer 2012;48:3232–9.
[32] Bekelman JE, Deye JA, Vikram B, Bentzen SM, Bruner D, Curran Jr WJ, et al. Redesigning radiotherapy quality assurance: opportunities to develop an efficient, evidence-based system to support clinical trials – report of the national cancer institute work group on radiotherapy quality assurance. Int J Radiat Oncol Biol Phys 2012;83:782–90.