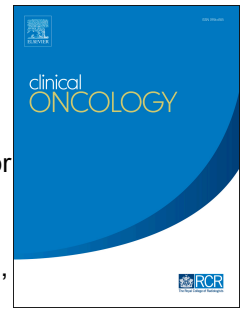


Radiotherapy Quality Assurance in the SCOPE2 trial: what lessons can be learned for the next UK trial in oesophageal cancer?

Jonathan Helbrow, Geraint Lewis, Chris Hurt, Ganesh Radhakrishna, Owen Nicholas, Maria A. Hawkins, Somnath Mukherjee, John Graby, Tom Crosby, Sarah Gwynne

PII: S0936-6555(24)00539-9

DOI: https://doi.org/10.1016/j.clon.2024.103735

Reference: YCLON 103735

To appear in: *Clinical Oncology*

Received Date: 20 August 2024

Revised Date: 24 November 2024

Accepted Date: 13 December 2024

Please cite this article as: Helbrow J, Lewis G, Hurt C, Radhakrishna G, Nicholas O, Hawkins MA, Mukherjee S, Graby J, Crosby T, Gwynne S, Radiotherapy Quality Assurance in the SCOPE2 trial: what lessons can be learned for the next UK trial in oesophageal cancer?, *Clinical Oncology*, https://doi.org/10.1016/j.clon.2024.103735.

Author Contributions

Author	Contribution
Jonathan Helbrow	1, 2, 3, 4, 7, 8
Geraint Lewis	1, 2, 4, 5, 8
Chris Hurt	2, 5, 6, 7, 8
Ganesh Radhakrishna	2, 4, 8
Owen Nicholas	2, 4, 8
Maria A. Hawkins	2, 4, 8
Somnath Mukherjee	2, 4, 8
John Graby	5, 6, 8
Tom Crosby	2, 4, 8
Sarah Gwynne	1, 2, 3, 4, 7, 8

Key

1 guarantor of integrity of the entire study
2 study concepts and design
3 literature research
4 clinical studies
5 experimental studies / data analysis
6 statistical analysis
7 manuscript preparation
8 manuscript editing

Title

Radiotherapy Quality Assurance in the SCOPE2 trial: what lessons can be learned for the next UK trial in oesophageal cancer?

Authors

Jonathan Helbrow[a,b], Geraint Lewis[b,c], Chris Hurt[d], Ganesh Radhakrishna[e], Owen Nicholas[a,f], Maria A. Hawkins[g], Somnath Mukherjee[h], John Graby[i], Tom Crosby[j]* and Sarah Gwynne[a,b,f]*

*joint last authors

Affiliations

a) South West Wales Cancer Centre, Swansea, UK
b) National Radiotherapy Trials Quality Assurance (RTTQA) Group, National Institute for Health and Care Research, UK
c) Department of Medical Physics, Velindre Cancer Centre, Cardiff, UK
d) Southampton Clinical Trials Unit, University of Southampton, Southampton, UK
e) The Christie Hospital NHS Foundation Trust, Manchester, UK
f) Swansea University Medical School, Swansea, UK
g) Department of Medical Physics & Biomedical Engineering, University College London, London, UK
h) Oxford Cancer and Haematology Centre, Oxford, UK
i) Department of Health, University of Bath, Bath, UK
j) Velindre Cancer Centre, Cardiff, UK

Corresponding author

Dr Jonathan Helbrow. FAO Gill Jones, Department of Oncology, South West Wales Cancer Centre, Singleton Hospital, Sketty Lane, Sketty, SA2 8QA, UK. Jonathan.helbrow@nhs.net

Acknowledgements

We would like to thank the following for their contribution to the SCOPE2 RTQA programme: Samantha Cox, Rhiannon Evans, Michael Gray, Gareth Jones, James Fergusson, Aphishek Enchackal, Tom Joslin-Tan and Maeve Smyth. We thank Michael Gray for his contribution to manuscript accuracy. MAH is supported by the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre.

Declaration of Interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Abstract

Background

The SCOPE2 trial evaluates radiotherapy (RT) dose escalation for oesophageal cancer. We report findings from the accompanying RT quality assurance (RTQA) programme and identify recommendations for PROTIEUS, the next UK trial in oesophageal RT.

Methods

SCOPE2's RTQA programme consisted of a pre-accrual and on-trial component. RTQA pre-accrual requirements included acceptable submission of 3D+/-4D benchmark contouring exercise(s) and a high-dose planning case. On-trial requirements for contouring and planning included prospective reviews (PRs) of each centre's first 3D+/-4D patient and all high-dose cases prior to formal safety review. Further PRs were at the RTQA team's discretion. Timely retrospective reviews (TRRs) were also undertaken for a random 10%. Submissions were assessed against pre-defined criteria and RT planning guidance document (RPGD). This study includes initial submissions only; subsequent resubmissions are not included in this analysis.

Results

For contouring, 30/64 (47%) pre-accrual submissions were approved. 38/64 (59%) contained ≥1 target volume (TV) unacceptable variation from protocol (UV), most commonly in CTVB and ITV. Organ-at-risk (OAR) contour review was undertaken in 28/64 (44%); 6/28 (21%) contained ≥1 UV, most commonly in heart and spinal cord. 82/126 (65%) on-trial submissions were approved. 47/126 (37%) contained ≥1 TV UV, most commonly in CTVB, GTV and ITV. For OARs, 30/126 (24%) contained ≥1 UV, most commonly in heart and lungs. On-trial contour submissions were significantly more likely to be approved than pre-accrual (p=0.016).

For planning, 32/43 (79%) pre-accrual plans were approved, those unacceptable were due to PTV coverage/conformity. 118/120 (98%) on-trial plans were approved, the remaining unacceptable were due to PTV coverage/conformity. No UVs in OAR dose constraints were observed.

All on-trial submissions were approved following resubmission where necessary.

Conclusion

Despite RPGD, contouring atlas, and similar contouring protocols from preceding trials, the SCOPE2 RTQA programme demonstrates a high frequency of UVs. Our findings inform recommendations for future oesophageal RT trials.

Key words

Radiotherapy, quality assurance, oesophageal cancer, contouring, planning

Title

Radiotherapy Quality Assurance in the SCOPE2 trial: what lessons can be learned for the next UK trial in oesophageal cancer?

Abbreviations

3DCRT	3D conformal radiotherapy
CI	Chief Investigator
CRT	Chemoradiotherapy
CTV	Clinical target volume
FRCR	Fellowship of the Royal College of Radiologists
GHG	Global Harmonization Group
GOJ	Gastro-oesophageal junction
GTV	Gross tumour volume
ICR	Individual case review
IMRT	Intensity-modulated radiotherapy
ITV	Internal target volume
NIHR	National Institute for Health and Care Research
OAR	Organ-at-risk
OG	Oesophago-gastric
PBT	Proton beam therapy
PI	Principal Investigator
PRV	Planning Organ at Risk Volume
PTV	Planning target volume
QA	Quality Assurance
RPGD	Radiotherapy planning guidance document
RT	Radiotherapy
RTQA	Radiotherapy quality assurance
RTTQA	National Radiotherapy Trials Quality Assurance Group
VMAT	Volumetric modulated arc therapy

Introduction

Chemoradiotherapy (CRT) has an important role in the curative treatment of oesophageal carcinoma [1]. Advances in this treatment method have contributed to improved outcomes, with a modern benchmark 5-year overall survival approaching 50% in patients with locally advanced disease receiving multi-modal therapy [2]. In the UK, the SCOPE trials (SCOPE1 [3], NeoSCOPE [4] and SCOPE2 [5]) have formed the backbone of radiotherapy (RT) studies for oesophageal cancer in the modern era, recruiting more than 700 patients between 2008 and 2024. In recognition that RT standards impact patient outcome [6-10], integral to this trial series has been a robust, educational RT quality assurance (QA) programme which has evolved though an iterative and collaborative process and led to national improvements in RT delivery and technique [11]. Table 1 illustrates these developments [3-5,11-13], which have been a consequence of a reflective and analytical learning process following each trial.

SCOPE2 was a phase 2/3 randomised controlled trial of definitive CRT which explored dose escalation (50Gy vs 60Gy in 25 fractions) and PET-CT guided systemic therapy adaptation [5], completing recruitment in January 2024 with 439 patients. Seeking to continue to the ethos of evolving and improving RTQA processes in oesophago-gastric (OG) cancer, findings from the SCOPE2 RTQA programme- and how these will inform PROTIEUS- the next UK OG RT trial- are reported here.

	SCOPE1 [3]	**NeoSCOPE [4]**	**SCOPE2 [5]**
Pre-accrual benchmark case (contouring and planning)	Yes (mid-third only)	Yes (mid- and lower-third)	Yes (mid- and lower-third, GTV provided)*
Additional 4D benchmark case	NA**	Yes	Yes
Reference volume for pre-accrual case	Yes (defined by CI and radiologist)	Yes (defined by consensus of TMG members)	Yes (defined by consensus of TMG members)
Protocol variations pre-defined	Yes (GTV only)	Yes	Yes
Radiotherapy protocol	Yes	Yes (iterative process with feedback from users)	Yes (iterative process with feedback from users)
Worked examples provided	Yes	Yes	Yes
Outlining atlas provided	No	Yes	Yes
Outlining workshop	No	Yes	Yes
On-trial review	*Retrospective*- First patient from each	*Prospective*- First 20 cases, first patient	*Prospective*- First 3D +/- 4D case, all high

	centre (achieved in 33/36 centres)	from each centre *Timely retrospective* - Remainder	dose cases prior to formal safety review (20 cases), at reviewers' discretion *Timely retrospective*- Random 10%
RTQA reviewer(s)	Chief investigator and RTQA physicist	A group of clinical oncologists from recruiting centres & on TMG, and trial RTTQA clinical scientist	A group of clinical oncologists from recruiting centres & on TMG, and trial RTTQA clinical scientist

*Table 1. Comparing RTQA processes for the SCOPE trials. *If not NeoSCOPE-approved (i.e. not streamlined) **4DCT not used in SCOPE1. CI=Chief Investigator, TMG=Trial Management Group, RTTQA=National RT Trials Quality Assurance Group*

Methods

Protocol and Radiotherapy Planning Guidance Document

The SCOPE2 RTQA programme was developed in conjunction with the National Institute for Health and Care Research (NIHR) RT Trials Quality Assurance (RTTQA) group [14].

There was a comprehensive Radiotherapy Planning Guidance Document (RPGD) developed by a subgroup of the Trial Management Group, informed by the SCOPE1 and NeoSCOPE RPGDs, SCOPE2 trial launch meeting discussions (2015), the EORTC-ROG guidelines for neoadjuvant RT of adenocarcinomas of the gastro-oesophageal junction (GOJ) and stomach [15], UK and US patterns of failure data [16,17] and dose escalation modelling work [18]. The RPGD was subject to an iterative process and additions/modifications made in collaboration with, and in response to, queries raised by participating centres during the lifetime of the trial [19-21]. Definitions of the target volumes are listed in Table 2 and supplementary materials.

Review team/criteria for assessment

Pre-accrual and on-trial (individual case review, ICR) contouring reviews were undertaken by one of six consultant clinical oncologists from the SCOPE2 Trial Management Group practising at recruiting centres, supported by senior (post-FRCR) research fellows under supervision. Reviews were allocated on a rota basis. Each target volume domain was deemed 'acceptable- per protocol', 'acceptable variation' or 'unacceptable variation' as per Global Harmonization Group (GHG) terminology [22], with reference to the RPGD and pre-defined criteria (see Table 2). Organ-at-risk (OAR) contour review was undertaken on a discretionary basis for pre-accrual but was mandatory for ICRs. Though OAR contour guidance was contained within the RPGD, there were no pre-defined criteria for assessing OAR variations, and the significance/grading of any variation was at the discretion of the RTQA reviewer.

The submission, after discussion with the submitting centre if appropriate, was then classified as overall acceptable and approved, or overall unacceptable and amendment +/- resubmission requested. Written feedback, in the form of a standardised feedback report (accompanied by

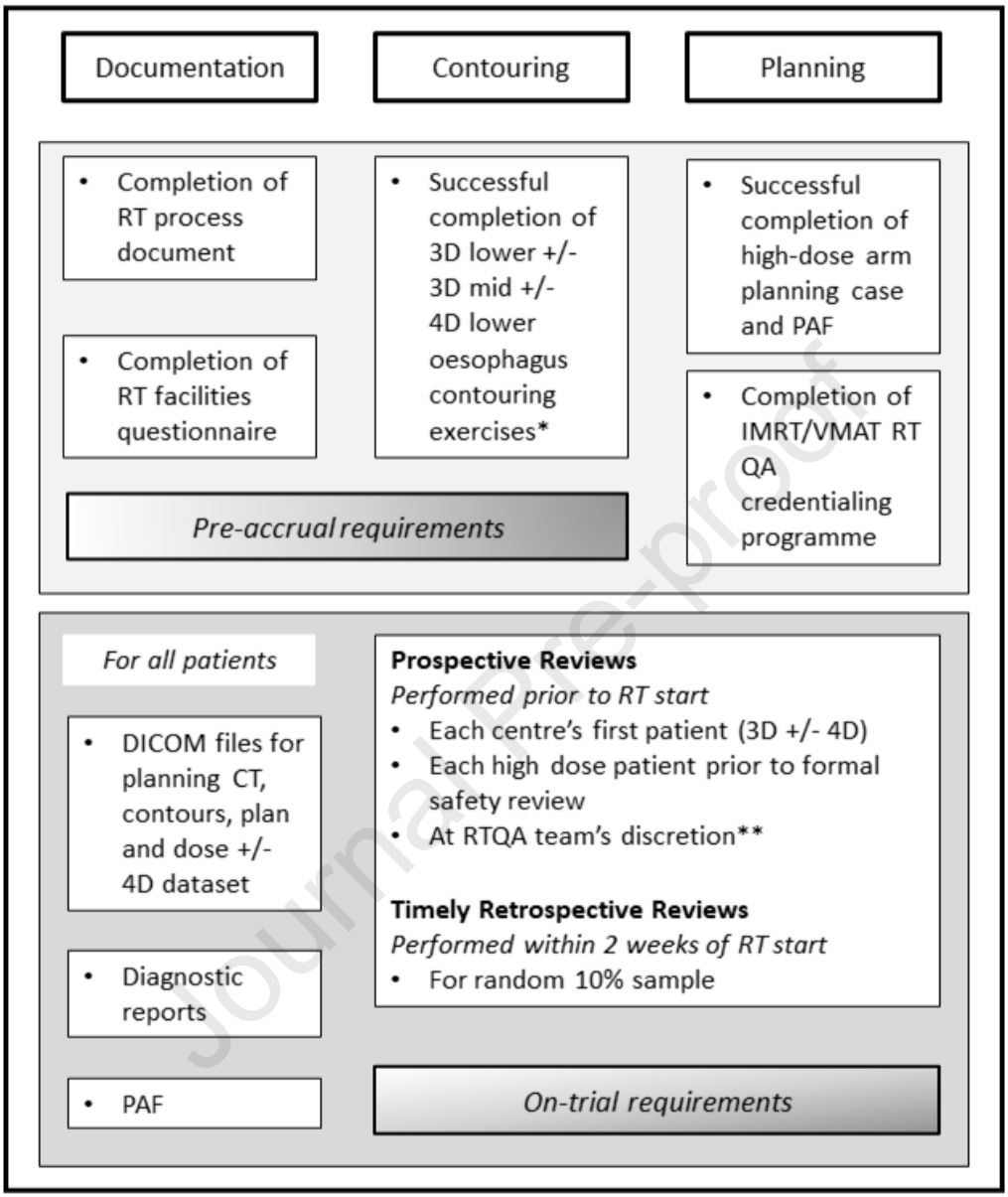
screenshots where applicable) was provided to submitting centres. Reviewers were not blinded to submitting centre/clinician.

RT treatment plans were reviewed by a RTTQA clinical scientist, and categorised as overall acceptable and approved, or overall unacceptable and resubmission required. These plan reviews consisted of a qualitative check of the dose distribution, including coverage of the target and dose fall-off for OARs and normal tissue (PTV coverage/conformity), and an independent check of dose-volume statistics against the optimal and mandatory dose constraints published in the RPGD. This qualitative assessment of the dose distribution was not based on pre-defined acceptability criteria but was performed using clinical judgement of PTV coverage within the context of each case, such as the need for centres to compromise PTV coverage to meet OAR dose constraints, or loss of PTV coverage over regions of low-density tissue in the lungs and trachea.

Queries that required medical input were addressed, in the first instance, to the same consultant clinical oncologist performing the contour review, with opinions sought from clinical colleagues on the Trial Management Group as needed.

Pre-accrual RTQA

Pre-accrual RTQA documentation, contouring and planning requirements are outlined in Figure



1.

*Figure 1. Pre-accrual and on-trial RTQA requirements. *If prior NeoSCOPE approval not achieved **Most due to issues with prior submissions. PAF=Plan Assessment Form*

Contouring

Pre-accrual benchmarking of contouring was required for each centre's Principal Investigator (PI). In order to reduce workload for recruiting centres, a streamlining approach was adopted where PIs who were already approved for the preceding NeoSCOPE trial [4] were exempt from repeating the benchmark contouring requirements. For 3D, this consisted of successful completion of the 3D NeoSCOPE benchmarking exercise, and for 4D, either successful completion of the 4D NeoSCOPE benchmarking exercise or attendance at a 4D contouring workshop.

Those without NeoSCOPE approval were required undertake the 3D SCOPE2 lower-third benchmarking exercise. PIs were asked to submit contours (CTVs and PTVs, see supplementary materials) for a lower-third case using provided GTVs. For centres wishing to use 4DCT, PIs were also required to outline volumes using the 4D dataset on the same 3D SCOPE2 benchmarking case. The requirement for an additional middle-third contouring case was omitted during the lifetime of the trial.

Pre-accrual contours were reviewed in reference to the consensus reference ("gold standard") contour created by the RTQA group members using a STAPLE algorithm approach [23,24]. A centre was not able to open to recruitment until approval was granted.

In the event of a change in local PI, satisfactory completion of the pre-accrual contouring case was required for the new PI. Some centres submitted additional pre-accrual cases undertaken by other clinicians at their centre.

Target Volume	Brief Description	Acceptable variation	Unacceptable variation
GTV	Primary tumour, involved lymph nodes, intervening oesophagus, whole circumference of oesophagus at level of GTV	>1.0 - ≤1.8cm longer, >0.7- ≤1.0cm shorter	>1.8cm longer, >1.0cm shorter
CTVA Length/margin	GTV + 20mm sup-inf along axis of oesophagus	≤3mm deviation from protocol	>3mm deviation from protocol, grown below GOJ
CTVA Gross appearance	Whole circumference of oesophagus for the necessary length of CTVA	Reviewer discretion	No attempt to include whole oesophageal circumference, not grown along axis of oesophagus
CTVC Length/margin	CTVA + 5mm circumferentially	≤3mm deviation from protocol	>3mm deviation from protocol
CTVB Length/margin	CTVA + 10mm circumferentially + CTVn	≤3mm deviation from protocol	>3mm deviation from protocol
CTVB Editing for normal structures	CTVB manually edited off normal structures e.g. heart, lungs, bone, great vessels, but not in to CTVC	Reviewer discretion, but minor inconsistent or incomplete editing of CTVB permitted	CTVB edited to exclude some of GTV or ELNI areas e.g. subcarinal LNs, absent editing CTVB edited in to CTVC
CTVB Elective lymph node irradiation*	CTVB manually expanded to include gastro-hepatic space and mucosa of gastric lesser curve	Reviewer discretion, but greater/lesser extent of inclusion of lesser curve and gastro-hepatic fat compared to atlas permitted	Whole stomach included, no ELNI included, lesser curve not included
ITV (4DCT only)	Composite of CTVBs on Reference, Maximum Inspiration and Maximum Expiration scans, manually expanded on all respiratory phases	Reviewer discretion	Range of CTVB movement not included on all respiratory phases

PTV_5000	Proximal: CTVB + 5mm circumferentially + 10mm sup-inf Distal 3D: CTVB + 5mm circumferentially + 10mm sup + 15mm inf Distal 4D: ITV + 5mm isotropically	≤3mm deviation from protocol	>3mm deviation from protocol
PTV_6000	GTV primary + 5mm isotropically	No deviation from protocol	Any deviation from protocol

*Table 2. Overview description of target volume delineation guidance plus pre-defined criteria for acceptable/unacceptable target volume variations. *For lower-third tumours. Upper-third tumours with node-positive disease within the SCF required elective SCF nodal irradiation. See supplementary materials for more detailed descriptions of target volumes. Sup=superiorly, inf=inferiorly, ELNI=elective lymph node irradiation*

Planning

All centres were required to successfully complete the IMRT/VMAT credentialing programme through the National RTTQA group or equivalent.

Centres were also required to satisfactorily complete a planning exercise for the high-dose arm (i.e. 6000cGy, including PTV_6000) using a provided, pre-contoured 3D DICOM dataset, accompanied by a plan assessment form. A standardised feedback report was completed for each submission. Centres were required to submit a further planning exercise if treatment technique or planning system changed during the trial.

On-trial (ICR) RTQA

On-trial QA consisted of ICR of both contours and plans according to the criteria set out in Figure 1. ICRs were either prospective reviews (performed prior to RT start) or timely-retrospective reviews (performed within two weeks of RT start). Requirements for on-trial RTQA documentation are also detailed in Figure 1. Centres occasionally requested a contouring review outside of the mandatory requirements.

Statistical analysis

All analyses were conducted using STATA SE 18. Chi square tests were used to compare proportions. All comparisons were pre-defined and all results have been reported. Due to the exploratory nature of this work, no correction for multiple testing has been made.

Resubmissions were excluded from this analysis.

Results

Pre-accrual contouring

After streamlining, 34/39 participating centres submitted at least one pre-accrual benchmark case. There was a total of 64 pre-accrual contour submissions: 14/64 (22%) 3D middle-third, 25/64 (39%) 3D lower-third and 25/64 (39%) 4D lower-third. 30/64 (47%) submissions were approved: 11/14 (79%) 3D mid-third, 14/25 (56%) 3D lower-third and 9/25 4D (36%) (p=0.036). 38/64 (59%) submissions contained at least one, with 24 (38%) containing more than one, target volume unacceptable variation, with a total of 81 identified. These were most frequently in CTVB (commonly

due to variation in contouring of the elective lymph node volume and in editing for normal structures) and, for 4D submissions, in ITV (failure to manually expand the CTVB in all phases of the respiratory cycle), as described in Table 3.

Domain	Pre-accrual	ICR	p value
% Acceptable submissions	*30/64 (47%)*	*82/126 (65%)*	*0.016*
Total number of target volume unacceptable variations	**81**	**82**	
GTV	- †	19/82 (23%)	
CTVA	5/81 (6%)	6/82 (7%)	0.771
CTVB	39/81 (48%)	34/82 (41%)	0.391
CTVC	4/81 (5%)	1/82 (1%)	0.169
ITV	12/44* (27%)	6/23* (26%)	0.917
PTV5000	9/81 (11%)	7/82 (9%)	0.581
PTV6000	12/81 (15%)	8/57** (14%)	0.898

*Table 3. Target volume unacceptable variations for pre-accrual and ICR. †GTV provided *number of unacceptable variations within 4D cases: pre-accrual=44, ICR=23 **number of unacceptable variations within high-dose cases: ICR=57.*

OAR contour review was undertaken in 28/64 (44%) submissions. 6/28 (21%) submissions contained at least one, with 2/28 (7%) containing more than one, OAR unacceptable variation, with a total of 8 identified. These were most frequently in heart (4/8, 50%) and spinal cord +/- PRV (2/8, 25%). For heart, unacceptable variation was most frequently at the superior border (defined where the pulmonary trunk and pulmonary artery are first seen as separate structures) and incomplete inclusion of the pericardium, and for spinal cord, not including the entire OAR at levels defined by the protocol (2cm superior and inferior to PTV).

ICR contouring

126/439 patients underwent an ICR for contouring. Of these, 82/126 (65%) were approved. 47/126 (37%) submissions contained at least one, with 23/126 (18%) containing more than one, target volume unacceptable variation, with a total of 82 identified. These were most frequently in GTV (commonly due variations in discrepancy between delineated volume and investigation reports and failure to include whole circumference of the oesophagus) and CTVB (as per pre-accrual). For 4D cases, unacceptable variations were also frequently seen in ITV (as per pre-accrual), as described in Table 3. The remaining ICR unacceptable variation (n=1, not included in Table 3) was in using the 'proximal' rather than the 'distal' delineation guidance for a lower-third case.

OAR contour review was undertaken for all 126 submissions. 30/126 submissions (24%) contained at least one, with 10/126 (8%) containing more than one, OAR unacceptable variation, and a total of 47 were identified. These were most frequently identified in heart (18/47), lungs (8/47), liver (7/47) and spinal cord +/- PRV (7/47). For heart and spinal cord, unacceptable variation was most frequently

seen in the same components as pre-accrual. For lungs and liver, unacceptable variation was most frequently due to incomplete inclusion of the structure, or for the former, inclusion of central airways/bowel.

Table 4 evaluates factors which may influence ICR outcome; only review type showed a significant difference, with 46/82 (56%) prospective reviews and 36/44 (82%) timely-retrospective reviews approved (p=0.004).

Factor	*Acceptable*	*Unacceptable*	*Total*	*% Acceptable*	*p value*
*Benchmarking method**					
Streamlined	52	25	77	68%	0.469
Non-streamlined	30	19	49	61%	
*Centre volume***					
Low	43	28	71	61%	0.227
High	39	16	55	71%	
Planning modality					
3DCT	64	35	99	65%	0.845
4DCT	18	9	27	66%	
Planned dose					
5000cGy	37	16	53	70%	0.342
6000cGy	45	28	73	62%	
Review type					
Prospective	46	36	82	56%	0.004
Timely-retrospective	36	8	44	82%	
All	*82*	*44*	*126*	*65%*	

*Table 4. Factors which may influence ICR contour approval. *Streamlined=contouring clinician approved using prior NeoSCOPE benchmarking, non-streamlined=contouring clinician approved using SCOPE2 benchmarking. **High-volume centre defined as recruiting ≥20 patients to SCOPE2*

Comparing pre-accrual and ICR contouring

The distribution of unacceptable variations was similar between pre-accrual and ICR (Table 3). ICRs were significantly more likely to be approved than pre-accrual submissions, however (respectively 82/126 (65%) vs 30/64 (47%) p=0.016). Additionally, fewer ICRs had at least one target volume unacceptable variation than pre-accrual submissions (respectively 47/126 (37%) vs 38/64 (59%), p=0.004).

Pre-accrual planning

43 plans were submitted from 39 centres. Of these 35/43 (81%) were VMAT, 7/43 (16%) IMRT and 1/43 (3%) tomotherapy. 32/43 (79%) were approved. The 9/42 (21%) categorised unacceptable were due to PTV coverage/conformity. Mandatory OAR dose constraints were met in all cases.

ICR Planning

120/439 patients underwent an ICR for planning. Of these, 50/120 (42%) were 5000cGy and 70/120 (58%) were 6000cGy plans, 77/120 (64%) were prospective and 43/120 (36%) timely-retrospective reviews, and 29/120 (24%) were 4D. 118 (98%) were approved; the two unacceptable were due to PTV coverage/conformity. Mandatory OAR dose constraints were met in all cases.

Discussion

We have described the SCOPE2 RTQA programme, detailing the protocol variations in contouring and planning at both pre-accrual and ICR. Below, we explore our findings, compare with published literature, and share strategies to further improve RTQA for the PROTIEUS trial.

Contouring

Despite an RPGD, contouring atlases, and similar protocols in preceding trials, a high frequency of unacceptable variations from protocol persist.

At the pre-accrual stage, less than half of submissions were approved. This is notably lower than the 74% reported in the PRODIGE-26 trial, also a phase 2/3 trial of RT dose escalation in oesophageal cancer (6600cGy in 33 fractions vs 5000cGy in 25 fractions) [25]. This difference may in part be explained by SCOPE2's use of 4DCT, which was associated with more unacceptable submissions than the 3D mid- and lower-third cases, reflecting the additional complexity of 4D contouring. PRODIGE-26's major/minor deviations also only included variations that led to a smaller-than-per-protocol CTV/PTVs, unlike SCOPE2, where all variations were recorded.

In the ICR setting, an improvement from pre-accrual was observed, with almost two-thirds of submissions approved. This demonstrates the value of both components of the RTQA process [9]. These findings are comparable with those of TOPGEAR, a large phase 3 study of neoadjuvant CRT in cancer of the stomach or GOJ, who undertook on-trial RTQA review for 203 patients from 53 centres [26] and report an approval rate of 72%. However, given in our study, more than one-third of ICRs contained unacceptable variations in target volume and one-quarter in OAR delineation, further improvement is required.

Timely-retrospective reviews were much more likely to be approved than prospective reviews. This may be explained by the chronological distribution of review type, with prospective reviews inherently conducted earlier in the trial timeline, closer to the start of the trial learning curve. An alternative explanation is that, for timely-retrospective reviews, the threshold for an unacceptable outcome may be higher as a result of the additional logistical and practical implications introduced when treatment has already commenced. The national RTTQA group are working in conjunction with the GHG to determine if criteria for acceptable and unacceptable variation should in fact be different for prospective and timely-retrospective reviews. There was also a trend suggesting approval was more likely for high-volume versus low-volume centres, although this was statistically non-significant. Similar findings were reported in TOPGEAR (81% vs 68%, p=0.078) [26], and this warrants further investigation in future RTQA programmes.

Planning

Overall, the quality of planning in SCOPE2 was very high. For pre-accrual, we report comparable rates of variance in PTV coverage and conformity to that in PRODIGE-26. However, despite SCOPE2's more stringent OAR dose constraints (e.g. spinal cord/D0.1cc PRODIGE-26=4800cGy vs SCOPE2=4200cGy, lungs V20 35% vs 25%, liver V30 60% vs 30%), we found fewer variations in meeting constraints [25]. This is likely due to SCOPE2's use of IMRT in all cases, compared to 58% in PRODIGE-26; the remaining 42% were 3D conformal RT (3DCRT), which were associated with a significantly higher rate of major deviations in meeting OAR constraints (3DCRT vs IMRT, 44% vs 8% p=0.009) [25]. Other potential contributing factors include a higher escalated dose in PRODIGE-26 (6600cGy vs SCOPE2=6000cGy), and larger margins applied in the cranio-caudal direction (GTV to CTV expansion: PRODIGE-26=30-50mm vs SCOPE2=20mm). In ICRs, 98% of submissions were approved. Unacceptable variations in PTV coverage/conformity were seen in the remaining 2%.

Informing RTQA for next OG trial

Based on the findings of the SCOPE2 RTQA programme, we have identified strategies to take forward into the next UK OG RT trial, PROTIEUS. This is a phase 2 study randomising patients to neoadjuvant, 4DCT-mandated, hypofractionated CRT (4005cGy/15#) to RT with either IMRT/VMAT or proton beam therapy (PBT). This trial commenced recruitment in June 2024 [27,28]. The relevant findings, together with the accompanying recommendations incorporated into the PROTIEUS RTQA programme, are described in greater detail below and summarised in Table 5.

Focus on domains of greatest variation in target volume and OAR delineation

Given CTVB and ITV are the most frequent domains of target volume unacceptable variation, these have been prioritised as areas for improvement. The PROTIEUS pre-accrual benchmark case was selected to include T4 disease (pleura) and a small (reactive) pericardial effusion which led to more challenging CTVB editing, and the 4D component necessitated manual expansion of ITV to account for the motion seen on other phases. The PROTIEUS RPGD provides more detailed support for contouring these domains, including worked examples and the addition of a simplified method for ITV creation [21] which is hypothesised to reduce variation.

Unacceptable variations in OAR contouring were present in approximately one-quarter of submissions. It is important to note, however, that there were no pre-defined approval criteria for OAR variations. This may have led to greater variation between reviewers, which in turn, may have led to overestimation of this variation. For PROTIEUS, we have developed pre-defined criteria for OARs, considering the GHG guidelines on OAR definitions [29]. This will enable more consistent, objective assessment and allow more detailed and transparent feedback for submitting centres.

We have also incorporated strategies to address OAR variation. For spinal cord, an atlas has been added to the PROTIEUS RPGD. Despite the SCOPE2 RPGD already containing an atlas to aid heart delineation, the heart was the most frequent OAR unacceptable variations the ICR setting. PROTIEUS mandates prospective ICRs for the first three patients recruited from each centre, enabling the RTQA team to prospectively highlight specific OAR variation, provide individualised education and promote awareness of GHG OAR definitions [29]. We anticipate this will complement the pre-existing atlas and newly updated feedback pro forma.

The increasing use and advancing standards of AI-generated OAR delineation may further promote consistency, improve quality, and enhance workflow [30,31]. However, the need for clinical evaluation of AI-generated contours along with ongoing education and training will remain of paramount importance as AI is further integrated into both clinical practice and future RT trials.

RTQA programme should include both pre-accrual and ICR components

We have shown a reduction in target volume unacceptable variations between pre-accrual and ICR. This improvement is likely to be an underestimate; GTV was provided and thus not evaluated for the former, and accounted for almost a quarter of unacceptable variations in the latter.

The pre-accrual benchmark case is typically undertaken in a less time-critical setting, with support and feedback from the trial RTQA team. Its aim is to assess and increase the understanding of, and adherence to, the RPGD, thereby reducing contouring variation within the trial itself [24]. It also facilitates identification of any RPGD ambiguities or errors that can be addressed prior to trial recruitment [24]. A recent meta-analysis by Brooks et al [9] included eight studies seeking to evaluate the relationship with pre-accrual and ICR assessments. Results from five of these studies, including RTQA results from the SCOPE1 trial [12], support the inclusion of both components within an RTQA programme. Together with our findings, we recognise the importance of a pre-accrual component in RTQA for oesophageal RT trials, and this is reflected in the requirements for PROTIEUS.

Streamlining does not affect ICR performance

Alongside the benefit of a pre-accrual component, we have also shown that streamlining is safe and effective in this setting. No difference in ICR approval was observed between those who had successfully completed a SCOPE2 pre-accrual case versus those who were exempt through NeoSCOPE benchmarking (i.e. streamlined). This streamlining approach, which increases efficiency in trial set up [32] and reduces clinician workload, has been adopted for the PROTIEUS trial, with clinicians who have successfully completed the SCOPE2 4D pre-accrual case not required to complete the PROTIEUS pre-accrual case. However, given the time that has elapsed between trials, those who have completed the NeoSCOPE 4D case, and not the SCOPE2 4D case, will not be streamlined.

Ensure consistency in methodology and terminology for feedback

GHG criteria refers to three categories of variation: acceptable- per protocol, acceptable variation and unacceptable variation from protocol [22]. In SCOPE2, reviewers did not always distinguish between the two acceptable categories. This limits the ability to compare SCOPE2 contouring with other published trials in the context of acceptable variation/minor deviation from protocol. A new RTQA feedback form, developed from the national RTTQA template and using the GHG criteria, has been introduced for PROTIEUS in attempt to improve this, and we encourage the international community to strive for greater consistency in terminology when reporting their findings.

Planning constraints can be tightened

In both the pre-accrual and ICR setting, mandatory OAR dose constraints were met for every submission, suggesting that tighter OAR constraints could be applied. This would reduce dose to normal tissues and potentially lower toxicity rates. Preliminary work using SCOPE2 cases by our group [33] shows more stringent constraints are achievable, and these have been introduced for PROTIEUS.

SCOPE2 findings	PROTIEUS interventions
Most frequent target volume unacceptable variations seen in	• Pre-accrual benchmark case selected to highlight challenges in CTVB contouring (including T4 pleural

CTVB and ITV	disease and a small, reactive pericardial effusion) and ITV (warranting expansion on all respiratory phases) • Sections added to RPGD to provide further support, including a simplified method for ITV creation
Unacceptable variations in OAR contouring seen in one-quarter of cases	• Pre-defined approval criteria for OARs, informed by GHG criteria, introduced • Spinal cord atlas added • Heart contour variations to be discussed in the mandatory, prospective ICRs with individualised feedback provided • AI-generated contours to continue to undergo RTQA review
Reduction in target volume unacceptable variations between pre-accrual and ICR observed	• RTQA programme to contain pre-accrual and ICR RT QA components
Streamlining does not impact approval	• Streamlining strategy in place for those with relevant prior SCOPE2 approval
Acceptable variations not always documented	• New, standardised feedback pro forma developed from RTTQA template introduced
Mandatory OAR dose constraints achieved in all pre-accrual and ICR submissions	• Tighter OAR constraints introduced

Table 5. SCOPE2 findings and corresponding interventions for PROTIEUS

Further work

The heterogeneity of the SCOPE2 cases (e.g. mix of 3D/4DCT, standard dose/high dose, proximal versus distal disease) did not allow a meaningful assessment of the quality of intra-clinician contouring over the lifetime of the trial. Such information may enable quantification of the benefit of ongoing involvement in the RTQA programme on contouring performance. We hope to explore this in the PROTIEUS trial, given cases are likely to be more homogenous, with 4DCT mandated, and most cases likely to be lower-third. Further assessment of submission approvals from high- and low-volume centres will also be undertaken.

In addition, given the recognised association between protocol non-compliance and outcomes [6-10], we will seek to explore this in our cohort once data are mature.

Conclusion

The incidence of unacceptable variations in the SCOPE2 RTQA programme highlights the need for ongoing, robust RTQA in OG RT trials, allowing significant issues to be identified and addressed whilst developing strategies to reduce variation.

The ongoing ethos of learning and development within the UK OG community embodied by the SCOPE trial series is crucial to continuing national improvement, and the SCOPE2 RTQA team are grateful to participating centres for their receptive, open and collaborative attitude.

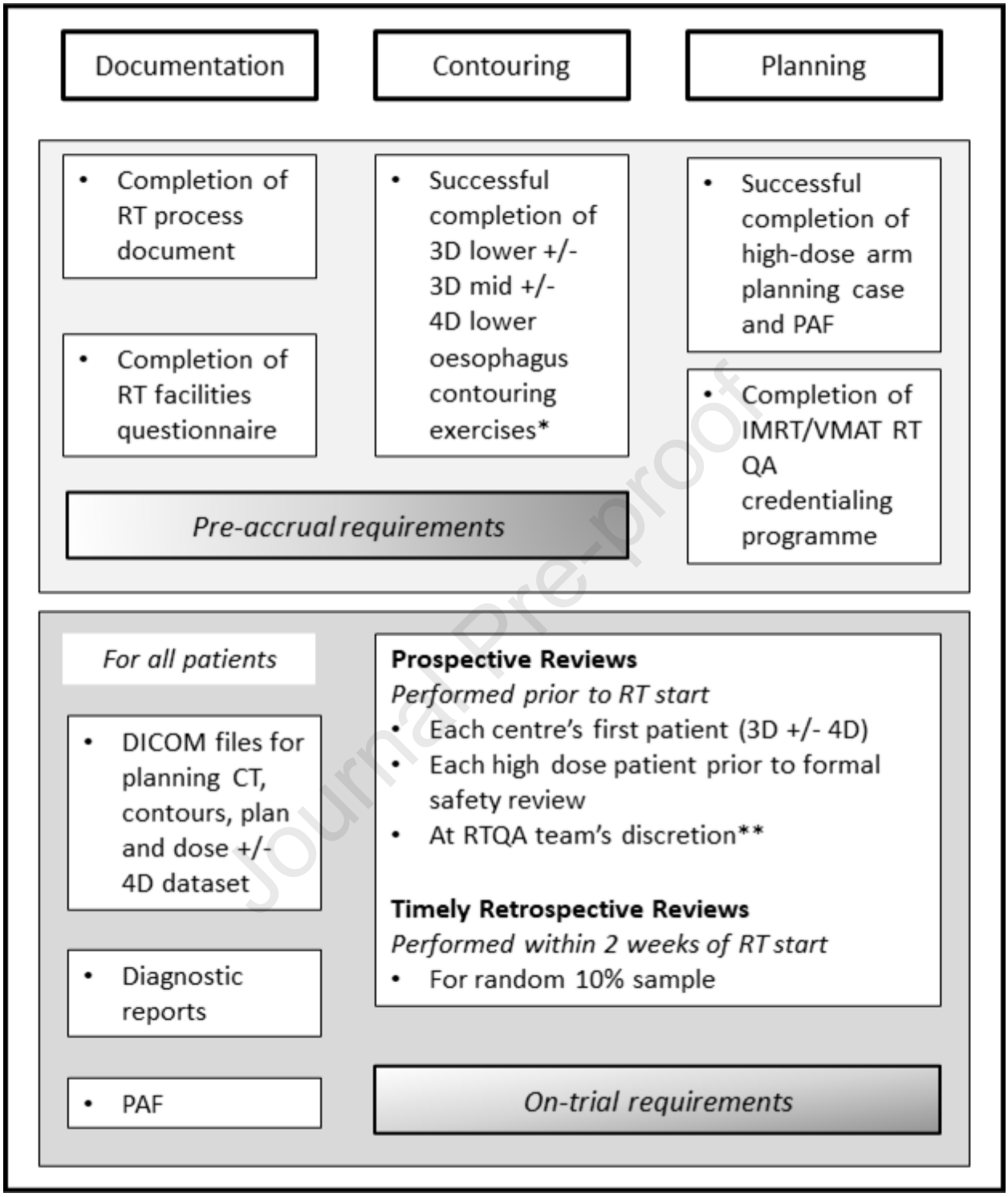
Consistent with this, our findings have informed a number of strategies to take forward to the next UK OG RT trial and beyond, and we are confident RTQA will continue to advance through ongoing partnership with both the UK and wider OG community.

References

1. Obermannová R, Alsina M, Cervantes A, Leong T, Lordick F, Nilsson M, et al. Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology*.2022;33(10):992-1004. doi:10.1016/j.annonc.2022.07.003
2. Eyck BM, van Lanschot JJB, Hulshof M, van der Wilk BJ, Shapiro J, van Hagen P, et al. Ten-Year Outcome of Neoadjuvant Chemoradiotherapy Plus Surgery for Esophageal Cancer: The Randomized Controlled CROSS Trial. *J Clin Oncol*.2021;39(18):1995-2004. doi:10.1200/jco.20.03614
3. Crosby T, Hurt CN, Falk S, Gollins S, Mukherjee S, Staffurth J, et al. Chemoradiotherapy with or without cetuximab in patients with oesophageal cancer (SCOPE1): a multicentre, phase 2/3 randomised trial. *Lancet Oncol*.2013;14(7):627-37. doi:10.1016/s1470-2045(13)70136-0
4. Mukherjee S, Hurt CN, Gwynne S, Bateman A, Gollins S, Radhakrishna G, et al. NEOSCOPE: a randomised Phase II study of induction chemotherapy followed by either oxaliplatin/capecitabine or paclitaxel/carboplatin based chemoradiation as pre-operative regimen for resectable oesophageal adenocarcinoma. *BMC Cancer*.2015;15:48. doi:10.1186/s12885-015-1062-y
5. Bridges S, Thomas B, Radhakrishna G, Hawkins M, Holborow A, Hurt C, et al. SCOPE 2 – Still Answering the Unanswered Questions in Oesophageal Radiotherapy? SCOPE 2: a Randomised Phase II/III Trial to Study Radiotherapy Dose Escalation in Patients with Oesophageal Cancer Treated with Definitive Chemoradiation with an Embedded Phase II Trial for Patients with a Poor Early Response using Positron Emission Tomography/Computed Tomography. *Clinical Oncology*.2022;34(7):e269-e280. doi:https://doi.org/10.1016/j.clon.2022.03.019
6. Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: evidence-based medicine in radiation therapy. *Radiother Oncol*.2012;105(1):4-8. doi:10.1016/j.radonc.2012.08.008
7. Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials. *J Natl Cancer Inst*.2013;105(6):387-93. doi:10.1093/jnci/djt001
8. 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(18):2996-3001. doi:10.1200/jco.2009.27.4498
9. Brooks C, Miles E, Hoskin PJ. Radiotherapy trial quality assurance processes: a systematic review. *Lancet Oncol*.2024;25(3):e104-e113. doi:10.1016/s1470-2045(23)00625-3
10. 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. *International Journal of Radiation Oncology*Biology*Physics*.2012;82(2):809-816. doi:https://doi.org/10.1016/j.ijrobp.2010.11.039
11. 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(1):26. doi:10.1186/s13014-019-1225-0
12. Wills L, Maggs R, Lewis G, Jones G, Nixon L, Staffurth J, et al. Quality assurance of the SCOPE 1 trial in oesophageal radiotherapy. *Radiat Oncol*.2017;12(1):179. doi:10.1186/s13014-017-0916-7
13. Evans E, Jones G, Rackley T, Maggs R, Radhakrishna G, Mukherjee S, et al. PO-0769: NeoSCOPE RTTQA: pre-accrual and on-trial review of all patients in a UK oesophageal RT trial. *Radiotherapy and Oncology*.2018;127:S397. doi:https://doi.org/10.1016/S0167-8140(18)31079-X
14. Miles E, Wadsley J, Diez P, Patel R, Gwynne S. The National Radiotherapy Trials Quality Assurance Group - Driving up Quality in Clinical Research and Clinical Care. *Clin Oncol (R Coll Radiol)*.2024;36(5):273-277. doi:10.1016/j.clon.2024.01.017

15. Matzinger O, Gerber E, Bernstein Z, Maingon P, Haustermans K, Bosset JF, et al. EORTC-ROG expert opinion: radiotherapy volume and treatment guidelines for neoadjuvant radiation of adenocarcinomas of the gastroesophageal junction and the stomach. *Radiother Oncol*.2009;92(2):164-75. doi:10.1016/j.radonc.2009.03.018
16. Button MR, Morgan CA, Croydon ES, Roberts SA, Crosby TD. 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(3):818-23. doi:10.1016/j.ijrobp.2008.04.062
17. Welsh J, Settle SH, Amini A, Xiao L, Suzuki A, Hayashi Y, et al. Failure patterns in patients with esophageal cancer treated with definitive chemoradiation. *Cancer*.2012;118(10):2632-40. doi:10.1002/cncr.26586
18. Warren S, Partridge M, Carrington R, Hurt C, Crosby T, Hawkins MA. Radiobiological determination of dose escalation and normal tissue toxicity in definitive chemoradiation therapy for esophageal cancer. *Int J Radiat Oncol Biol Phys*.2014;90(2):423-9. doi:10.1016/j.ijrobp.2014.06.028
19. Cox S, Jones GW, Radhakrishna G, Mukherjee S, Hawkins MA, Crosby T, et al. PO-1080: 4DCT oesophageal tumour delineation in SCOPE2 – how is radiotherapy quality assurance beneficial? *Radiotherapy and Oncology*.2018;127
20. Radhakrishna G, Banner R, Mukherjee S, Hawkins M, Crosby T, Gwynne SH. PO-1037: A new nodal delineation protocol for upper third oesophageal cancers in the SCOPE 2 trial. *Radiotherapy and Oncology*.2020;152:S552. doi:10.1016/S0167-8140(21)01054-9
21. Lewis G, Nicholas O, Gwynne S, Crosby T, Chand A, Thomas B, et al. PO-1371 A simplified method for delineation of target volumes in 4D oesophageal cancer radiotherapy. *Radiotherapy and Oncology*.2023;182:S1100-S1101. doi:https://doi.org/10.1016/S0167-8140(23)66286-3
22. Global Harmonization Group. Protocol variations definitions. https://rtqaharmonization.org/resources/. 2019. [Accessed 30.07.2024]
23. 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(7):903-21. doi:10.1109/tmi.2004.828354
24. 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(1024):20120398. doi:10.1259/bjr.20120398
25. Boustani J, Rivin Del Campo E, Blanc J, Peiffert D, Benezery K, Pereira R, et al. Quality Assurance of Dose-Escalated Radiation Therapy in a Randomized Trial for Locally Advanced Oesophageal cancer. *Int J Radiat Oncol Biol Phys*.2019;105(2):329-337. doi:10.1016/j.ijrobp.2019.06.2542
26. Lukovic J, Moore AJ, Lee MT, Willis D, Ahmed S, Akra M, et al. The Feasibility of Quality Assurance in the TOPGEAR International Phase 3 Clinical Trial of Neoadjuvant Chemoradiation Therapy for Gastric Cancer (an Intergroup Trial of the AGITG/TROG/NHMRC CTC/EORTC/CCTG). *Int J Radiat Oncol Biol Phys*.2023;117(5):1096-1106. doi:10.1016/j.ijrobp.2023.06.011
27. NHS Health Research Authority. PROTIEUS Trial Summary. https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/protieus/. 2024. [Accessed 16.7.2024]
28. PROTIEUS Trial Management Group. PROTIEUS: A Randomised Phase 2 Trial Comparing Proton versus Photon Based Neoadjuvant Chemoradiation, followed by Adjuvant Immunotherapy, in Oesophageal Cancer Trial Protocol. 1.1 2023.
29. Mir R, Kelly SM, Xiao Y, Moore A, Clark CH, Clementel E, et al. Organ at risk delineation for radiation therapy clinical trials: Global Harmonization Group consensus guidelines. *Radiotherapy and Oncology*.2020;150:30-39. doi:10.1016/j.radonc.2020.05.038
30. Huynh E, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, et al. Artificial intelligence in radiation oncology. *Nature Reviews Clinical Oncology*.2020;17(12):771-781. doi:10.1038/s41571-020-0417-8

31. Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. *Radiotherapy and Oncology*.2020;153:55-66. doi:https://doi.org/10.1016/j.radonc.2020.09.008
32. Tyler J, Baines H, Clark C, Eyles H, Miles E, Nabi R, et al. Impact of radiotherapy clinical trial participation and quality assurance on head and neck practice. *Radiotherapy and Oncology*.2024;194(1):s2684-s2668. doi:https://doi.org/10.1016/S0167-8140(24)02138-8
33. Nicholas O, Lewis G, Gwynne S. Optimising dose constraints for the PROTIEUS trial. Personal communication. 2024.



Research Highlights

- For SCOPE2, frequent variations from protocol in target volume contouring were seen
- The role of both pre-accrual and on-trial QA has been confirmed
- Streamlining was shown to be a safe and effective method of pre-accrual QA
- Opportunity for tightening OAR constraints for future trials was identified
- Areas for improvement in target volume and OAR contouring were also shown

*Figure 1. Pre-accrual and on-trial RTQA requirements. *If prior NeoSCOPE approval not achieved **Most due to issues with prior submissions. PAF=Plan Assessment Form*

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: