UWA Research Publication

Kearvell, R., Haworth, A., Ebert, M. A., Murray, J., Hooton, B., Richardson, S., Joseph, D. J., Lamb, D., Spry, N. A., Duchesne, G. and Denham, J. W. (2013), Quality improvements in prostate radiotherapy: Outcomes and impact of comprehensive quality assurance during the TROG 03.04 'RADAR' trial. *Journal of Medical Imaging and Radiation Oncology*, 57: 247–257. doi: 10.1111/1754-9485.12025.

© 2013 The Authors. *Journal of Medical Imaging and Radiation Oncology* © 2013 The Royal Australian and New Zealand College of Radiologists

This is the peer reviewed version of the following article: Kearvell, R., Haworth, A., Ebert, M. A., Murray, J., Hooton, B., Richardson, S., Joseph, D. J., Lamb, D., Spry, N. A., Duchesne, G. and Denham, J. W. (2013), Quality improvements in prostate radiotherapy: Outcomes and impact of comprehensive quality assurance during the TROG 03.04 'RADAR' trial. *Journal of Medical Imaging and Radiation Oncology*, 57: 247–257. doi: 10.1111/1754-9485.12025, which has been published in final form at <http://dx.doi.org/10.1111/1754-9485.12025>. This article may be used for non-commercial purposes in accordance With [Wiley Terms and Conditions for self-archiving](#).

This version was made available in the UWA Research Repository on 1 April 2014 in compliance with the publisher's policies on archiving in institutional repositories.

Use of the article is subject to copyright law.

Quality improvements in prostate radiotherapy: Outcomes and impact of comprehensive quality assurance during the TROG 03.04 'RADAR' trial

Journal:	<i>Journal of Medical Imaging and Radiation Oncology</i>
Manuscript ID:	JMIRO-12-0209.R1
Manuscript Type:	Radiation Oncology Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Kearvell, Rachel; Sir Charles Gairdner Hospital, Radiation Oncology Haworth, Annette; Peter MacCallum Cancer Centre, Physical Sciences; Ebert, Martin; Sir Charles Gairdner Hospital, Radiation Oncology; University of Western Australia, School of Physics Murray, Judy; University of Otago, Pathology and Molecular Medicine Hooton, Ben; Sir Charles Gairdner Hospital, Radiation Oncology Richardson, Sharon; Sir Charles Gairdner Hospital, Radiation Oncology Joseph, David; Radiation Oncology, SCGN Lamb, David; University of Otago, Pathology and Molecular Medicine Spry, Nigel; Sir Charles Gairdner Hospital, Radiation Oncology Duchesne, Gillian; Director of Radiation Oncology, Peter MacCallum Cancer Institute Denham, James
Keywords:	Radiation Oncology, Prostate cancer, Quality Assurance, Clinical trials

SCHOLARONE™
Manuscripts

Quality improvements in prostate radiotherapy: Outcomes and impact of comprehensive quality assurance during the TROG 03.04 ‘RADAR’ trial

Rachel Kearvell¹, PGradDipHlthAdmin

Annette Haworth^{2,3}, PhD

Martin A Ebert^{1,4}, PhD

Judy Murray⁵, PGDipHealSci

Ben Hooton¹, Grad Dip Comp Sci

Sharon Richardson¹, BSc (Honours) Therapeutic Radiology

David J Joseph^{1,6}, FRANZCR

David Lamb⁵, FRANZCR

Nigel A Spry^{1,7}, FRANZCR

Gillian Duchesne^{8,9,10}, FRANZCR

James W Denham¹¹, FRANZCR

¹ Department of Radiation Oncology, Sir Charles Gairdner Hospital, Western Australia, Australia

² Department of Physical Sciences, Peter MacCallum Cancer Centre, Victoria, Australia

³ School of Applied Sciences, RMIT University, Victoria, Australia

⁴ School of Physics, University of Western Australia, Western Australia, Australia

⁵ Department of Pathology and Molecular Medicine, University of Otago, New Zealand

⁶ School of Surgery, University of Western Australia, Western Australia, Australia

⁷ School of Medicine and Pharmacology, University of Western Australia, Western Australia, Australia

⁸ Department of Radiation Oncology, Peter MacCallum Cancer Centre, Victoria, Australia

⁹ Department of Pathology, University of Melbourne, Victoria, Australia

¹⁰ Departments of Biochemistry and Radiation Sciences, Monash University, Victoria, Australia

¹¹ School of Medicine and Public Health, University of Newcastle, New South Wales, Australia

Number of pages: 21

Number of figures: 4

Number of tables: 2

Supporting online material to be included.

Correspondence address:

Dr Martin A Ebert

Department of Radiation Oncology

Sir Charles Gairdner Hospital

Hospital Ave

Nedlands, Western Australia 6009

Tel: 08 9346 4931

Fax: 08 9346 3402

Email: Martin.Ebert@health.wa.gov.au

Running Title: Technical QA for the RADAR trial

ABSTRACT

Introduction: The Trans-Tasman Radiation Oncology Group (TROG) 03.04 'RADAR' multicentre prostate cancer trial examined the optimal duration of androgen deprivation in combination with dose-escalated radiotherapy. Rigorous quality assurance (QA) processes were undertaken to ensure the validity and reliability of the RT treatment plan data.

Method: QA processes included a planning benchmarking exercise and a periodic audit of target and normal tissue delineation. Centralised electronic review of digital plan data for external beam radiotherapy was undertaken to detect protocol variations. The impact of clinical factors and feedback to submitting centres during the trial on variation rates was investigated.

Results: 23 centres across Australia and New Zealand recruited 1071 participants to the trial. Treatment plans for 754 participants receiving external-beam treatment alone were reviewed. From these, 1185 minor and 86 major variations were identified leading to feedback to treating centres to reduce variations for subsequent patients' treatment and plans, suggesting improvement in treatment quality through these QA programs. Participant anatomical factors (delineated CTV and rectal volume) and treatment planning factors (beam energy, beam definition, patient position orientation) were found to significantly impact variation rates. The dummy-run demonstrated disagreement in identification of the base of the prostate and the superior extent of the rectum. Feedback from the periodic audit led to a change of practice at five contributing centres.

Conclusion: The application of a suite of complementary QA activities allows the quality of trial data to be optimised and quantified, and can provide a catalyst for reforming treatment practices.

Key words: prostate cancer; radiotherapy; quality assurance; clinical trials

For Peer Review

INTRODUCTION

In multicentre radiotherapy clinical trials, lack of consistency in trial protocol interpretation, treatment planning and delivery can reduce the ability of the trial to generate significant results (1-6). Several collaborative trials groups have adopted procedures for quality assurance (QA) of treatment planning data for assessing centres' abilities to meet protocol requirements or for review of individual trial participants' plans (7, 8). The complexity of such reviews can vary from simple visual inspection of hardcopy information (9), to more automated reviews based on electronic submission and review systems (10-13). The importance of rigorous QA of radiotherapy was highlighted by the results of the TROG 02.02 'HeadSTART' trial for head and neck chemoradiotherapy, where it was shown that the influence of radiotherapy quality could be greater than that of the intervention being evaluated in the trial. (5)

The Trans-Tasman Radiation Oncology Group (TROG) 03.04 'RADAR' (Randomised Androgen Deprivation and Radiotherapy) trial was principally aimed at evaluation of variable-duration androgen deprivation (AD) for localised prostate carcinoma.(14) The trial also examined the role of three-dimensional conformal radiation therapy (3DCRT) including dose escalation. RADAR built on the TROG 96.01 prostate trial which required a dose prescription of 66 Gy and was carried out in the era of early computed tomography (CT) planning (< 60% of plans were based on CT imaging). Late Grade 2 rectal toxicity was consistent with that reported by Pollack et al (15, 16) who demonstrated that rectal toxicity will significantly increase as the dose prescription is increased unless rectal sparing can be achieved via stringent conformal techniques.

To facilitate the generation of quality data for the RADAR trial and provide a safe environment for the introduction of techniques that were novel to many centres, the technical requirements were audited through a series of complementary QA studies. These were aimed at ensuring protocol compliance and allowed quantification of the variability between centres and participants in planning/treatment methodology. The ability of participating centres to deliver accurate radiotherapy was assessed by means of (previously published) setup accuracy (SUA) (17) and phantom-based dosimetric (18-21) studies. These two studies provided a quality framework for the RADAR trial, establishing minimum requirements for spatial and dosimetric accuracy, quantifying the related underlying uncertainties and challenging contributing centres to meet the trial protocol requirements. This paper describes three additional QA activities around treatment planning which provided complementary and ongoing quality assessment during the trial. The overall influence of these activities was examined via their impact on the quality of planned radiotherapy delivery for trial participants and general radiotherapy practice at participating centres. Future publications will report on the non-technical QA activities associated the RADAR trial.

METHODS AND MATERIALS

Trial Overview

The RADAR trial (TROG 03.04, NIH trial identifier NCT00193856) was conducted between 2003 and 2008 and accrued 1071 participants. The trial aimed to determine the optimal duration of AD for intermediate (T2a) and high (T2b+) risk prostate cancer patients undergoing radiotherapy. Radiotherapy dose escalation (66 Gy, 70 Gy, 74 Gy, 78 Gy) was a key feature of the trial, as was the alternative to provide external beam

radiotherapy (46 Gy) combined with a high dose rate (HDR) brachytherapy boost (3 fractions to total dose 19.5 Gy). Twenty-three sites accrued participants across Australia and New Zealand following approval of the trial protocol and consent form by local Human Research Ethics Committees.

The trial protocol provided external beam radiotherapy (EBRT) treatment planning guidelines largely based on the recommendations of a consensus workshop (22). For doses above 66 Gy, participation in the QA program described in this paper was mandatory.

Development of Electronic Plan Review Software (SWAN)

Plan review software, 'SWAN', was conceived to allow electronic treatment plan review, via RTOG [format of the North American Radiation Therapy Oncology Group, based on AAPM Report #10 (23)] and DICOM [Digital Imaging and Communication in Medicine (24)] formatted plan exports, for multicentre trials in Australasia (25, 26). RADAR plans were exported from SWAN to a relational database, allowing manually-reported treatment information to be directly linked to each treatment plan. It also allowed retrieval of plans from the database directly into SWAN, and the ability to perform operations on the stored data without requiring that it be re-imported in full. The database can be queried using standard Structured Query Language (SQL) operations. Functionality was incorporated to enable specific treatment fields and dosimetric parameters to be compared with protocol requirements (26). The resultant treatment plan's protocol conformity was then summarised and provided to the submitting institution. The variability in dose-volume histogram (DVH) calculation

between treatment planning systems (27) was overcome using independent DVH calculations in SWAN.

Treatment Planning QA Exercises

Benchmarking Exercise

A benchmarking exercise ('dummy-run') was undertaken to identify ambiguities in protocol recommendations and to collect plan data in electronic format for validation of the QA process. Participation in the benchmarking exercise was encouraged, though was not mandatory for trial participation. Participating centres were asked to contour one intermediate and one high risk case (using supplied DICOM CT data) according to the protocol. In addition, sites were required to plan the high risk case to a dose of 74 Gy prescribed to the ICRU reference point (28).

Treatment Plan Review and Post-Treatment Verification

Centres exported the electronic treatment plans for submission to the radiotherapy review team (RTRT), together with a report that ensured that reviewed data were consistent with those from the original treatment planning system (TPS). Electronic plan submission was not mandatory for patients receiving the combined EBRT/HDR boost treatment approach. All submitted EBRT plans were reviewed by an expert RTRT radiation therapist, in addition to automated review using the auto-report function in SWAN (see Table 1). 'Variation' of items relative to protocol were categorised as either:

- 'Minor' – outside protocol recommendations but unlikely to influence clinical outcome;

- ‘Major’ - outside protocol recommendations and may influence clinical outcome.

The definitions of these specific to each reviewed item are provided in Table 1.

Periodic Contour Audit

A separate QA program was developed to audit contouring practices. The audit team included 8 experienced radiation oncologists (RO) and 4 radiation therapists (RT) from 4 institutions. At each of 5 periodic rounds a small number of submitted treatment plans were reviewed by all auditors and the remaining treatment plans equally shared.

The auditors rated the planning target volumes (PTVs) according to their assessment of the margins applied to the clinical target volume (CTV). Any plan scored “Feedback Required”, or “Equivocal” was subsequently reviewed by at least one additional audit RO before feedback to the treating centre. Those with systematic errors requiring feedback had subsequent plans audited to confirm practice change.

RESULTS

Benchmarking Exercise

Thirteen centres participated. For centres that did not participate, submission of data for their first accrued participant was used as an opportunity to validate the centre’s ability to submit correctly formatted, complete and compliant data.

One centre submitted 5 sets of contours from 5 clinicians. Variability was seen in contouring practices particularly in the superior-inferior direction – in a coronal plane

through the centre of the prostate, the prostatic base (superior extent of prostate) varied 23 mm and 15 mm across observers for the intermediate and high risk cases respectively, the apex varying 33 mm and 24 mm respectively). Further quantitative results for the two test cases are provided in online supporting material, demonstrating that for the high-risk case the greatest variation in contouring is at the base, with some variation also at the apex and posterior prostate. There was reasonable agreement in defining the mid-gland CTV. Similar variability was observed with the intermediate case which did not require the seminal vesicles to be contoured. Although rectal contouring was more consistent due to better definition on CT, there was considerable variation in the extent of rectum to be contoured. For the high risk case, definition of the inferior extent of the rectum varied by 24 mm and the superior extent by 39 mm, with a resulting variation in rectal volume from 39.8 cm³ to 67.6 cm³. As a result, rectal contouring for many trial cases was considered inconsistent with protocol definition (see [Figure 2](#)~~Figure 2~~) and corrected prior to re-calculating the DVH using the SWAN analysis tools.

Treatment Plan Review and Post-Treatment Verification

Figure 1 summarises trial accrual and treatment plan reviews, including issues that prevented plan archive review (no plans were reviewed/archived for participants receiving a HDR boost). All 23 participating centres submitted plans for review. Although plan review and feedback was provided within 1 week of receipt of a plan, submission of the plan relative to RT treatment commencement was variable (median 70 days, range -286 to 1158 days). A total of 1185 minor and 86 major variations were identified across 14,326 reviewed items spanning 754 plans, distributed by item as shown in [Figure 2](#)~~Figure 2~~. [Figure 3](#)~~Figure 3~~ shows the number of participants accrued

by each centre (reverse-ordered by accrual number), together with the number of plan reviews undertaken per centre and the rate of protocol variations. The variation rate (minor, major or either) does not correlate with the number of participants accrued per centre, and no significant cut-point in accrual number could be identified.

As shown in [Figure 4](#), variation rate decreased as the trial progressed. Increasing compliance with the margin definitions of the PTVs and 95% coverage of the PTVs (both automatically reviewed items) made a significant contribution to this pattern, as did increasing compliance with definition of the femoral head (manually reviewed). When arranged by the time from which each centre accrued their first participant (assuming this reflects the time from which each centre began receiving review feedback), a similar pattern is seen (see [Figure 4a](#)) with an initial rapid decrease in variation rates. [Figure 4b](#) shows the mean variation rate per review calculated according to the number of previous plans submitted for a centre, the variation rate consistently decreasing with increasing centre experience. (Note that the delay between plan submission and feedback to centre has not been included in this analysis). Trends for three sample centres (low, medium and high accruing) are also shown in [Figure 4b](#). Quantitative data on review of margins are provided in Table e1 of the supporting online material.

It has been suggested that treatment planning factors including plan complexity and patient anatomy may weigh heavily on variation rates (29). As such, other factors were investigated for their impact on compliance as summarised in Table 2. In order to attempt to identify which aspects of the protocol were leading to differences in variation

rate, for dichotomised items in Table 2 showing a significant difference between the groups, the entire list of review items for individual participants was reviewed in order to determine which review items most contributed to the observed difference.

Periodic Contour Audit

19 of the 23 sites were audited. Sites were audited according to accrual numbers, with the highest being in the first round. Of the other four centres, two treated only two participants with EBRT alone and two centres (accruing 13 participants each) did not provide any treatment plan data until the trial had closed.

A systematic error was detected as part of the audit at seven centres. In round 1, one centre consistently failed to have the 95% isodose line encompass the PTV in the superior-inferior direction. Round 2 identified one centre that consistently failed to apply a PTV margin superiorly or inferiorly. Round 3 identified 3 centres that had difficulty determining the superior and inferior extents of the CTV. Round 4 identified 2 centres with systematic errors: one centre consistently treated to 74 Gy in one phase with margins smaller than that stated in the protocol (i.e. $< 10\text{mm}$) and the other consistently failed to delineate a CTV at all. Five out of the seven centres changed their practice as a direct result of the feedback provided from the audit.

DISCUSSION

Benchmarking Exercise

The aim of the benchmarking exercise was to identify ambiguities in the protocol and test the plan review process prior to recruitment of a large number of participants. Based

on earlier publications (30, 31), some variation in contouring the prostate was expected. The amount by which the contouring practices for the target volume and PTV varied was more than expected (41.2 cc – 158.8 cc for the gross tumour volume (GTV) or CTV and 144.4 cc – 435.5 cc for the PTV). In cases where it was clear that the planning-target margin had not been applied according to the protocol, feedback was provided to the centre. For the remaining data sets, the review team were faced with considerable difficulty in reaching a consensus on classifying acceptability of the GTV/CTV. The variation in prostate contouring revealed by the benchmarking exercise highlighted the need for a periodic contour audit to minimise the risk of extreme contouring practices.

As the definition of the inferior rectum was linked to the definition of the prostatic apex, a large variability of the inferior extent of the rectum was expected. The variability in defining the superior extent was not expected, though in some cases this was because the centre had defined the superior rectum to be 15 mm above the prostate, rather than using the anatomical definition.

Overall the benchmarking exercise for this trial served to alert the RTRT that variations in contouring were likely and that centres would need to be reminded to correctly apply PTV margins and follow the protocol definition for rectal length. Generally, treatment planning complied with the protocol.

Treatment Plan Review and Post-Treatment Verification

Review of treatment plans prior to treatment was not possible for this trial as submission was on compact disk requiring postal delivery. Major variations were

identified in 86 cases and delays in plan submission meant losing the opportunity for plan modification. However, it was still possible to influence a change in practice for future participants and the incidence of major and minor variations over time did decrease, as shown in [Figure 4](#)~~Figure 4~~. When the variation rate is viewed according to time from first accrual for each centre, a rapid initial decrease is seen, suggesting an initial impact of feedback to submitting centres. In [Figure 4](#)~~Figure 4b~~, a decrease in protocol variation rate is seen with increasing experience of submitting centres. With multiple centres submitting smaller numbers of plans (see [Figure 3](#)~~Figure 3~~), it is difficult in such visualisation to separate the impact of the resources and experience of a contributing centre from the influence of review feedback.

The SWAN system enabled multiple treatment and dosimetric parameters to be quickly assessed and passed, without transposition errors, to the trial database. The automated functionality of the system enabled multiple variables to be objectively assessed for 93% of participants treated using EBRT alone in this trial, providing reliable information on the quality of plan data submitted and a quantitative measure of protocol conformity.

The patient anatomical factors shown to impact on protocol compliance (Table 2) reflect the difficulty to meet protocol requirements with increasing prostate and rectal volumes. The investigated treatment planning factors suggest that the use of lower beam energies, MLCs for field definition and prone patient orientation will lead to a plan of poorer quality. This is only according to the definitions within the RADAR protocol however. These factors will be prime candidates for assessment against toxicity and progression

rates. The factors listed in Table 2 were not the subject of the trial and were accordingly not randomised.

The significant results displayed in Table 2 should be considered in light of potential underlying correlations with other treatment factors. For the dichotomised factors showing significant difference in mean variation rates between groups for example, it was determined that the impact of the factors 'beam energy' and 'rectal filling protocol' was due to difficulty in meeting dosimetric constraints. For the factors 'field definition' and 'patient setup orientation' however, some main influencing review items (e.g., applied margin and immobilisation technique) would not likely be influenced by those factors and more than likely the association is simply via the contributing centre.

With 754 treatment plans stored in the SWAN database, comparing dosimetric parameters with patient outcomes will provide definitive correlation. For example, the planned rectal contours varied greatly. Using the SWAN software, rectal contours have subsequently been amended. These modified volumes will be used for correlation of a range of rectal toxicity endpoints with the dose volume relationship using analysis methods similar to those reported in the MRC RT-01 trial (32, 33).

Periodic Contour Audit

Auditing CTV contouring practices presented a challenge for the audit team as no clear guidelines existed for the definition of major and minor protocol variations. Even among reviewers consensus of opinion was difficult to achieve. Feedback to trial centres was therefore limited to identifying outliers such as CTVs that clearly extended inferiorly into the urogenital diaphragm, which did not include the proximal seminal

vesicles in high risk cases, or had failed to have the correct margins applied when creating the PTV. Since the RADAR trial opened, there has been much discussion on variations in contouring practices and strategies for limiting contouring variations [eg., (34)]. With the use of magnetic resonance imaging (MRI) to guide delineation of the prostate along with credentialing of the audit team, it is expected that a tighter control on contouring variations can be achieved in future trials. Ensuring all participating centres complete a benchmarking or credentialing exercise prior to accrual is also useful.

Impact of QA Program

The results of the RADAR trial are still maturing and the clinical impact of protocol non-conformity is yet to be tested. We have found that rigorous pre-trial credentialing provides an opportunity to identify and correct for systematic planning deficiencies before they have the opportunity to impact on trial outcomes. To this we add that it also provides an ideal opportunity for the development of consensus guidelines for plan review and for centres to become familiar with the plan submission process. Furthermore, electronic systems such as SWAN that allow for collection, semi-automated review and analysis of treatment plan data will facilitate rapid feedback and hence may increase protocol conformity. Reviewing departmental practice under the auspices of clinical trial QA also provides an effective and non-confrontational means of introducing new technologies and techniques safely and effectively into treatment centres. It has recently been suggested that new technologies are most appropriately introduced in the context of a multicentre trial with associated rigorous trials QA (35). This was certainly the experience for the RADAR trial in Australasia.

The set-up accuracy study previously reported (17) resulted in more than 65% of centres changing their patient positioning and/or immobilisation protocols. This further demonstrates the benefit of QA studies conducted within clinical trials. In most cases position verification was based on matching the bony anatomy in the post-treatment image with the DRR. Whilst this method is rapidly becoming obsolete in the prostate cancer patient as real-time image guidance becomes practical and feasible, QA protocols remain useful in identifying centres where the quality of radiotherapy has the potential to impact on clinical outcomes.

RADAR was a high-recruitment multicentre trial that placed stringent requirements on the use of CT guided conformal therapy. The associated QA program successfully reduced and rectified inconsistencies in rectum identification though revealed an unresolved variability in prostate definition. We speculate that with the inherent deficiencies in CT identification of prostatic boundaries, the QA program is likely to have had a greater impact on reducing morbidity than increasing efficacy. The morbidity results of the RADAR trial are maturing and eagerly anticipated to confirm or refute this hypothesis. Preliminary analysis from RADAR indicates no increase in rectal and urinary morbidity with increasing prescription dose (from outcomes analyses currently underway by author JD), which supports the hypothesis that increasing QA requirements can minimise the impact of treatment variations.

ACKNOWLEDGEMENTS

We acknowledge funding from Cancer Australia and the Diagnostics and Technology Branch of the Australian Government Department of Health and Ageing (grant 501106 supporting SWAN development), the National Health and Medical Research Council

(grants 300705, 455521, 1006447), the Hunter Medical Research Institute, the Health Research Council (New Zealand), the University of Newcastle and the Calvary Mater Newcastle, Abbott Laboratories, Novartis Pharmaceuticals. We gratefully acknowledge the support of the Sir Charles Gairdner Hospital, the 'Elvis' study team, the contour audit team, participating RADAR centres, the Trans-Tasman Radiation Oncology Group, Jean Ball, Chantelle Wilcox, Cate Sproston, Sean Hall, Celia Gordon, Dr Sarat Chander, A/Prof Scott Williams, Dr Mahesh Kumar, Prof Sean Bydder, Rhonda Coleman.

References

1. Bydder S, Spry N. Chemotherapy for pancreatic cancer. *New Eng J Med*. 2004;350(26):2713.
2. Deasy JO, Niemierko A, Herbert D, Yan D, Jackson A, Ten Haken RK, et al. Methodological issues in radiation dose-volume outcome analyses: summary of a joint AAPM/NIH workshop. *Med Phys*. 2002;29(9):2109-27.
3. Roos DE, Davis SR, Turner SL, O'Brien PC, Spry NA, Burmeister BH, et al. Quality assurance experience with the randomized neuropathic bone pain trial (Trans-Tasman Radiation Oncology Group, 96.05). *Radioth Oncol*. 2003;67(2):207-12.
4. Pettersen MN, Aird E, Olsen DR. Quality assurance of dosimetry and the impact on sample size in randomized clinical trials. *Radioth Oncol*. 2008;86(2):195-9.
5. 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 Jun;28(18):2996-3001.
6. Rischin D, Young RJ, Fisher R, Fox SB, Le QT, Peters LJ, et al. Prognostic Significance of p16(INK4A) and Human Papillomavirus in Patients With Oropharyngeal Cancer Treated on TROG 02.02 Phase III Trial. *J Clin Oncol*. 2010 Sep;28(27):4142-8.
7. Ibbott GS, Followill DS, Molineu A, Lowenstein JR, Alvarez PE, Roll JE. Challenges in credentialing institutions and participants in advanced technology multi-institutional clinical trials. *Int J Radiat Oncol Biol Phys*. 2008 2008;71(1):S71-S5.
8. 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. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2012;103(3):279-86.
9. Steigler A, Mameghan H, Lamb D, Joseph D, Matthews J, Franklin I, et al. A quality assurance audit: Phase III trial of maximal androgen deprivation in prostate cancer (TROG 96.01). *Aus Radiol*. 2000;44(1):65-71.
10. Palta JR, Frouhar VA, Dempsey JF. Web-based submission, archive, and review of radiotherapy data for clinical quality assurance: a new paradigm. *Int J Radiat Oncol Biol Phys*. 2003;57(5):1427-36.
11. Purdy JA, Harms WB, Michalski J, Bosch WR. Initial experience with quality assurance of multi-institutional 3D radiotherapy clinical trials. A brief report. *Strahlen Onkol*. 1998;174 Suppl 2:40-2.
12. Matzinger O, Poortmans P, Giraud JY, Maingon P, Budiharto T, van den Bergh ACM, et al. Quality assurance in the 22991 EORTC ROG trial in localized prostate cancer: Dummy run and individual case review. *Radioth Oncol*. 2009;90(3):285-90.
13. Purdy JA. Quality assurance issues in conducting multi-institutional advanced technology clinical trials. *Int J Radiat Oncol Biol Phys*. 2008 2008;71(1):S66-S70.
14. TROG. TROG Clinical Trials Summary. TROG 03.04 - Randomised trial investigating the effect on survival and PSA control of different durations of adjuvant androgen deprivation in association with definitive radiation treatment for localised carcinoma of the prostate (RADAR). TROG; 2005; Available from: <http://clinicaltrials.gov/ct/show/NCT00193856?order=5>.

15. Pollack A, Zagars GK, Starkschall G, Antolak JA, Lee JJ, Huang E, et al. Prostate cancer radiation dose response: Results of the M. D. Anderson phase III randomized trial. *Int J Radiat Oncol Biol Phys*. 2002;53(5):1097-105.
16. Christie D, Denham J, Steigler A, Lamb D, Turner S, Mameghani H, et al. Delayed rectal and urinary symptomatology in patients treated for prostate cancer by radiotherapy with or without short term neo-adjuvant androgen deprivation. *Radioth Oncol*. 2005;77(2):117-25.
17. Haworth A, Kearvell R, Greer PB, Hooton B, Denham JW, Lamb D, et al. Assuring high quality treatment delivery in clinical trials - Results from the Trans-Tasman Radiation Oncology Group (TROG) study 03.04 "RADAR" set-up accuracy study. *Radiother Oncol*. 2009;90(3):299-306.
18. Ebert MA, Harrison K, Howlett S, Cornes D, Bulsara M, Hamilton CS, et al. Dosimetric intercomparison for multicentre clinical trials using a patient-based anatomic pelvic phantom. *Med Phys*. 2011;38(9):5167-75.
19. Ebert MA, Harrison K, Cornes D, Howlett S, Joseph DJ, Kron T, et al. A comprehensive Australasian multi-centre dosimetric intercomparison - issues, logistics and recommendations. *J Med Imag Radiat Oncol*. 2008;86:200-10.
20. Ebert MA, Howlett SJ, Harrison K, Cornes D, Hamilton CS, Denham JW. Linear-accelerator X-ray output: a multicentre chamber-based intercomparison study in Australia and New Zealand. *Aust Phys Eng Sci Med*. 2008;31(4):255-66.
21. Harrison K, Ebert MA, Kron T, Howlett SJ, Cornes D, Hamilton C, et al. Design, manufacture and evaluation of an anthropomorphic pelvic phantom purpose-built for radiotherapy dosimetric intercomparison. *Med Phys*. 2011;38(10):5330-7.
22. Skala M, Berry M, Duchesne G, Gogna K, Tai KH, Turner S, et al. Australian and New Zealand three-dimensional conformal radiation therapy consensus guidelines for prostate cancer. *Aus Radiol*. 2004 2004-Dec;48(4):493-501.
23. Baxter BS, Hitchner LE, Maguire Jr GQ. A standard format for digital image exchange. Madison WI: AAPM; 1982.
24. NEMA. Digital Imaging and Communications in Medicine (DICOM) Standard. Washington DC: Office of Publications; 2001.
25. Ebert MA, Blight J, Price S, Haworth A, Hamilton C, Cornes D, et al. Multicentre analysis of treatment planning information: technical requirements, possible applications and a proposal. *Aus Radiol*. 2004;48(3):347-52.
26. Ebert MA, Haworth A, Kearvell R, Hooton B, Coleman R, Spry NA, et al. Detailed review and analysis of complex radiotherapy clinical trial planning data: Evaluation and initial experience with the SWAN software system. *Radioth Oncol*. 2008;86:200-10.
27. Ebert MA, Haworth A, Kearvell R, Hooton B, Hug B, Spry NA, et al. Comparison of DVH data from multiple radiotherapy treatment planning systems. *Phys Med Biol*. 2010 Jun;55(11):N337-N46.
28. ICRU. Prescribing, recording and reporting photon beam therapy. Bethesda MD: International Commission on Radiological Units; 1999. Report No.: Report 62.
29. Weber DC, Tomsej M, Melidis C, Teglas V, Hurkmans CW. QA makes a trial stronger: evidence-based medicine in radiation therapy. *Radioth Oncol*. 2012;103(Suppl 1):S187-S8.
30. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radioth Oncol*. 1998 Jun;47(3):285-92.

31. Rasch C, Barillot I, Remeijer P, Touw A, van Herk M, Lebesque JV. Definition of the prostate in CT and MRI: a multi-observer study. *Int J Radiat Oncol Biol Phys.* 1999;43(1):57-66.
32. Gulliford SL, Foo K, Morgan RC, Aird EG, Bidmead AM, Critchley H, et al. Dose-volume constraints to reduce rectal side effects from prostate radiotherapy: Evidence from MRC RT01 trial ISRCTN 47772397. *Int J Radiat Oncol Biol Phys.* 2010;76(3):747-54.
33. Gulliford SL, Partridge M, Sydes MR, Andreyev J, Dearnaley DP. A comparison of dose-volume constraints derived using peak and longitudinal definitions of late rectal toxicity. *Radioth Oncol.* 2010 Feb;94(2):241-7.
34. Michalski JM, Lawton C, El Naqa I, Ritter M, O'Meara E, Seider MJ, et al. Development of RTOG consensus guidelines for the definition of the clinical target volume for postoperative conformal radiation therapy for prostate cancer. *Int J Radiat Oncol Biol Phys.* 2010;76(2):361-8.
35. James H, Beavis A, Budgell G, Clark C, Convery D, Mott J, et al. Guidance for the clinical implementation of intensity modulated radiation therapy, IPEM;2008. Report No.: 96.

TABLE CAPTIONS

Table 1: Treatment plan review parameters and variation criteria

Table 2: Factors investigated for their impact on review variations. For dichotomised factors, variations for individual participants were further investigated to determine which review items most contributed to (affected) the difference in variation rate.

For Peer Review

FIGURE CAPTIONS

Figure 1: Summary of participant accrual according to treatment groups, treatment plan reviews and issues preventing final review and archive of treatment plans

Figure 2: Distribution of protocol item variations

Figure 3: Accrual, reviews and variation rates (major and minor) ordered by centre accrual.

Figure 4: Variation in variation rate for reviewed items with a) time throughout trial (time from start of trial and time from each centre's first accrual, cut into 9 equal time intervals), and b) number of previous plans reviewed per centre (with sample results for 3 individual centres). Scores have been categorised simply into protocol compliance and non-compliance (minor or major variation). Only centres contributing > 15 treatment plans have been included.

Table 1: Treatment plan review parameters and variation criteria

<i>Review Item</i>	<i>Criteria</i>	<i>Minor Variation</i>	<i>Major Variation</i>
Automated review	Planned prescription dose To be consistent with dose (estimated from total dose to nominated at randomisation centre of PTV)	Within $\pm 5\%$	Outside $\pm 5\%$
Dose per fraction	2 Gy per week-day fraction	Dose per fraction between 1.9 Gy – 1.96 Gy, or between 2.04 Gy – 2.1 Gy	Dose per fraction < 1.9 Gy or > 2.1 Gy
Margins applied to each Planning Target Volume (PTV)	PTV1 = CTV + 1-1.5 cm, posterior margin 0.5-1.0 cm PTV2 = CTV + 0-1.0 cm, posterior margin ≤ 0.5 cm	PTV differs from allowable margin range by 5% - 10% by volume	PTV differs from allowable margin range by > 10% by volume
95% coverage of each PTV	95% isodose should encompass PTV	95% - 98% coverage	< 95% coverage
CT slice thickness	≤ 5 mm	N/A	Slice thickness > 5 mm
Number of treatment fields	≥ 3	N/A	All treatment phases used less than 3 fields

	Beam energy	≥ 6 MV	N/A	One or more beams < 6 MV
	Conformity Index (CI) (28) for each PTV	≤ 1.5	CI > 1.58% (allowing 5% calculation uncertainty)	N/A
	Rectal dose constraints	65 Gy to maximum 40% 70 Gy to maximum 30% 75 Gy to maximum 5%	N/A	DVH outside constraints (allowing 2% calculation uncertainty)
	Femoral head dose constraints	35 Gy to maximum 100% 45 Gy to maximum 60% 60 Gy to maximum 30%	N/A	DVH outside constraints (allowing 2% calculation uncertainty)
	Maximum dose to PTV	Complies with ICRU definition (28)	N/A	Maximum dose > 107% or < 95% prescription dose
Expert review	Rectal wall contouring	Sigmoid flexure to 15mm inferior to prostate apex	Minor deviations according to subjective assessment reviewer	Blatantly misjudged rectal contour according to subjective assessment reviewer
	Left femoral head contouring	Acetabulum to inferior border of field.	Minor deviations according to subjective assessment reviewer	Blatantly misjudged femoral head contour according to subjective assessment reviewer

Digitally Reconstructed DRRs for patient set-up	DRRs not created	N/A
Radiographs (DRR) created	verification	
Prescription point	Complies with ICRU definition (28)	Prescription point does not comply with ICRU definition
Immobilisation device appropriate	Should be consistent with that used in SUA study	N/A
Percentage isodose encompassing rectum	$\leq 50\%$	N/A

Table 2: Factors investigated for their impact on review variations. For dichotomised factors, variations for individual participants were further investigated to determine which review items most contributed to (affected) the difference in variation rate (N = sample/group size).

Investigated Factor	Impact on Variation Rate	Most Affected Review Items (dichotomised factors only)
Rectal distension (Ant- Post width at centre of PTV)	No significant impact [†]	N/A
PTV-Rectum separation	No significant impact [†]	N/A
PTV Volume	Variation rate increased with increasing volume (coefficient 0.016 %/cm ³ , p = 0.0002) ^{†,‡}	N/A
Rectum volume	Variation rate increased with increasing volume (coefficient 0.024 %/cm ³ , p = 0.0006) ^{†,‡}	N/A
Body mass index (BMI)	No significant impact ^{†,‡}	N/A
Bladder filling protocol	No significant impact [§]	N/A
Rectal filling protocol	Higher variation rate when laxatives are given prior to CT (N = 33, mean 18.0 %) than when a bulking agent is used (N = 700, mean = 7.5 %)	<ul style="list-style-type: none">• Femoral head dose constraints• CI PTV1• CI PTV2

($p < 10^{-19}$)^{§,††}

Prescription dose (by dose group)	No significant impact [§]	N/A
Number of beams	No significant impact [†]	N/A
Beam energy	Higher variation rate for beams ≤ 10 MV (N = 252, mean 10.4 %) than for beams > 10 MV (N = 497, mean 7.0 %) ($p < 10^{-11}$) [§]	<ul style="list-style-type: none"> • Rectal DVH constraints • CI PTV1 • CI PTV2
Field definition	Higher variation rate for beams defined by multileaf collimators (MLCs) (N = 481, mean 9.3 %) than by collimators/custom blocks (N = 260, mean 5.8 %) ($p < 10^{-11}$) [§]	<ul style="list-style-type: none"> • Immobilisation device appropriate • 95% coverage of PTV1 • Margin applied to PTV2
Patient setup orientation	Higher variation rate for prone orientation (N = 64, mean 11.6 %) than supine orientation (N = 684, mean 7.8 %) ($p < 10^{-5}$) ^{§,‡‡}	<ul style="list-style-type: none"> • CI PTV1 • CI PTV2 • Margin applied to PTV2 • 95% coverage of PTV1 • Immobilisation device appropriate

[†] Assessed via linear regression

[‡] Note that rectum volume and PTV volume are correlated with BMI ($p = 0.004$ and $p = 0.002$ respectively), though PTV volume and rectal volume are poorly correlated ($p = 0.07$)

[§] Assessed via 2-sided t-test for difference in means between groups

^{††} Note that only three centres contributed patients where laxatives had been used prior to planning CT

^{‡‡} Note that only three centres contributed plans for patients to be treated with a prone orientation

Figure 1

Figure 1: Summary of participant accrual according to treatment groups, treatment plan reviews and issues preventing final review and archive of treatment plans
189x274mm (200 x 200 DPI)

Figure 2

Figure 2: Distribution of protocol item variations
189x274mm (200 x 200 DPI)

Figure 3

Figure 3: Accrual, reviews and variation rates (major and minor) ordered by centre accrual.
189x274mm (144 x 144 DPI)

Figure 4

Figure 4: Variation in variation rate for reviewed items with a) time throughout trial (time from start of trial and time from each centre’s first accrual, cut into 9 equal time intervals), and b) number of previous plans reviewed per centre (with sample results for 3 individual centres). Scores have been categorised simply into protocol compliance and non-compliance (minor or major variation). Only centres contributing > 15 treatment plans have been included.
189x274mm (200 x 200 DPI)