Radiotherapy and Oncology 182 (2023) 109494

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

# Prioritizing clinical trial quality assurance for photons and protons: A failure modes and effects analysis (FMEA) comparison

Paige A. Taylor [a,b,*], Elizabeth Miles [v], Lone Hoffmann [c,d], Sarah M. Kelly [e,f,g], Stephen F. Kry [a,b], Ditte Sloth Møller [c,d], Hugo Palmans [h,i], Kamal Akbarov [j], Marianne C. Aznar [k], Enrico Clementel [f], Coreen Corning [f], Rachel Effeney [l], Brendan Healy [m], Alisha Moore [n], Mitsuhiro Nakamura [o], Samir Patel [p], Maddison Shaw [m,q], Markus Stock [h,r], Joerg Lehmann [n,s,t,u,1], Catharine H. Clark [i,v,w,x,1]

[a] Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [b] The Imaging and Radiation Oncology Core, USA; [c] Department of Medical Physics, Aarhus University Hospital, Aarhus, Denmark; [d] Department of Clinical Medicine, Faculty of Health Sciences, Aarhus University, Aarhus, Denmark; [e] SIOP Europe, The European Society for Paediatric Oncology, Clos Chapelle-aux-Champs 30, Brussels, Belgium; [f] EORTC Headquarters, European Organisation for Research and Treatment of Cancer, Avenue E. Mounier 83, Brussels, Belgium; [g] Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium; [h] MedAustron Ion Therapy Center, Wiener Neustadt, Austria; [i] Metrology for Medical Physics, National Physical Laboratory, Teddington, UK; [j] Division of Human Health, Department of Nuclear Sciences and Applications, IAEA, Vienna, Austria; [k] Division of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; [l] Royal Brisbane & Women's Hospital, Brisbane, Australia; [m] Australian Clinical Dosimetry Service, ARPANSA, Melbourne, Australia; [n] TROG Cancer Research, Newcastle, Australia; [o] Department of Advanced Medical Physics, Graduate School of Medicine, Kyoto University, Kyoto, Japan; [p] Division of Radiation Oncology, Department of Oncology, University of Alberta, Edmonton, Alberta, Canada; [q] School of Health and Biomedical Sciences, RMIT University, Melbourne, Australia; [r] Karl Landsteiner University for Health Sciences, Austria; [s] Department of Radiation Oncology, Calvary Mater Newcastle, Newcastle, Australia; [t] School of Information and Physical Sciences, University of Newcastle, Newcastle, Australia; [u] Institute of Medical Physics, University of Sydney, Sydney, Australia; [v] National Radiotherapy Trials Quality Assurance (RTTQA) Group, Mount Vernon Cancer Centre, Northwood, UK; [w] Radiotherapy Physics, University College London Hospital, London, UK; [x] Medical Physics and Bioengineering Department, University College London, London, UK

ARTICLE INFO

*Article history:*
Received 2 December 2022
Received in revised form 13 January 2023
Accepted 18 January 2023
Available online 25 January 2023

*Keywords:*
Quality assurance
Clinical trials
Proton therapy
Photon therapy
FMEA

ABSTRACT

*Background and Purpose:* The Global Clinical Trials RTQA Harmonization Group (GHG) set out to evaluate and prioritize clinical trial quality assurance.
*Methods:* The GHG compiled a list of radiotherapy quality assurance (QA) tests performed for proton and photon therapy clinical trials. These tests were compared between modalities to assess whether there was a need for different types of assessments per modality. A failure modes and effects analysis (FMEA) was performed to assess the risk of each QA failure.
*Results:* The risk analysis showed that proton and photon therapy shared four out of five of their highest-risk failures (end-to-end anthropomorphic phantom test, phantom tests using respiratory motion, pre-treatment patient plan review of contouring/outlining, and on-treatment/post-treatment patient plan review of dosimetric coverage). While similar trends were observed, proton therapy had higher risk failures, driven by higher severity scores. A sub-analysis of occurrence × severity scores identified high-risk scores to prioritize for improvements in RTQA detectability. A novel severity scaler was introduced to account for the number of patients affected by each failure. This scaler did not substantially alter the ranking of tests, but it elevated the QA program evaluation to the top 20th percentile. This is the first FMEA performed for clinical trial quality assurance.
*Conclusion:* The identification of high-risk errors associated with clinical trials is valuable to prioritize and reduce errors in radiotherapy and improve the quality of trial data and outcomes, and can be applied to optimize clinical radiotherapy QA.


Independent, central quality assurance (QA) needs to be an integral component of every clinical trial to ensure that patient treatment is delivered as intended. This is necessary because variations in delivered dose have been consistently shown to significantly impact patient overall survival [1,2]. This impact on outcomes introduces noise into the study, which can limit the ability to achieve the trial's potential by reducing the power of the trial and therefore its ability to answer the trial question reliably [3].

However, there remains a lack of consistent radiotherapy QA (RTQA) in clinical trials [4], particularly when comparing between

*Abbreviations:* FMEA, Failure Modes and Effects Analysis.

* Corresponding author at: UT MD Anderson Cancer Center 1515 Holcombe Blvd. Unit 607 Houston, TX 77030 USA.
*E-mail address:* PATaylor@mdanderson.org (P.A. Taylor).

[1] Co-Senior Authors.

https://doi.org/10.1016/j.radonc.2023.109494
0167-8140/

different international bodies. Moreover, the RTQA that is used in clinical trials has developed organically but has not been rigorously examined in the context of the frequency of errors that are detected or the severity of those errors [5]. Finally, these questions require special attention when evaluating these issues in the context of trials involving novel or emerging technology where more limited clinical experience may further jeopardize consistent and accurate dose delivery. This increased risk of treatment deviations due to treatment novelty likely warrants QA intensification to ensure accuracy in implementation and consistency with existing standards, but little study has been done on this topic to date.

To better understand, prioritize, harmonise, and manage clinical trial QA, this study undertook three goals within a Failure Modes and Effects Analysis (FMEA) framework. First, we evaluated the relevance and importance of current clinical trial QA methods. That is, we evaluated how current clinical trial QA processes aligned with the needs of the community in terms of the frequency and severity of errors identified in clinical practice. Second, we evaluated the effectiveness of current clinical trial QA methods at identifying current radiotherapy errors. This was done by including the detectability of errors in treatment delivery by current trial QA methods and evaluating the impact of this on the overall priority and ranking scores for different QA tools. Third, we contrasted the importance of specific clinical trial QA processes for standard of care versus novel treatment delivery. As a case study of this, we compared the priority and ranking scores for different QA tools for IMRT versus proton therapy. While proton therapy use is expanding, it is still relatively new in the clinical trial context and currently only available at a small fraction of participating institutions. For example, proton therapy centres make up less than 2 % of RT facilities in the United States [5]. These objectives were examined by the Global Clinical Trials RTQA Harmonization Group (GHG), which has a mission of harmonising clinical trial QA in the global setting [6–7]. This effort represents the first quantitative approach to optimising QA methods for clinical trials.

## Methods

The GHG is comprised of five core groups: European Organisation for Research and Treatment of Cancer (EORTC), Imaging and Radiation Oncology Core (IROC), Japan Clinical Oncology Group (JCOG), UK Radiotherapy Trials Quality Assurance Group (RTTQA), and Trans-Tasman Radiation Oncology Group (TROG) [6]. In addition to these five groups, the following clinical trial QA groups participated in this effort: Aarhus University, MedAustron Ion Therapy Center, and the International Atomic Energy Agency (IAEA).

For our study, a comprehensive list was compiled of every currently implemented clinical trial QA step for radiotherapy. Appendix A contains a glossary of clinical trial QA activities and terminology [8]. Clinical trial RTQA was then evaluated using an FMEA. FMEA is a technique commonly used in radiotherapy to assess the risk of possible failure modes in the radiation treatment process [9]. A comprehensive list of failure modes is identified (ways in which the system can fail), and each mode is given a score for frequency or likelihood of occurrence (O), severity of error (S), and ability of the QA systems to detect the error (D). These scores are on a scale of 1–10, and the combined product ($O \times S \times D$) yields a Risk Priority Number (RPN). The RPN allows stratification of the failure modes by overall risk to the patient. A higher RPN is associated with a higher risk. Any failure mode with an S score greater than or equal to 8 is typically considered to be high risk to the patient, and any mode with an RPN in the top 20th percentile of scores is similarly considered to be high risk to the patient.

Specific failure modes can be hard to tease out, particularly in an external end-to-end audit. For this reason, we did not score failure modes, but rather failures themselves. For example, a QA failure might be failing a moving anthropomorphic phantom test. The actual failure modes might include an error in the 4D simulation, an error with the motion management system, a setup error, etc. However, the same type of risk assessment can be applied to the failures themselves, and this provides information on the QA process in terms of its importance and relevance at detecting such failures.

The qualitative and quantitative metrics for clinical trial QA errors are outlined in Table 1. The severity and detectability scoring were not changed from TG-100 [9]. The occurrence score was adapted from the TG-100 scoring example, since there is a higher frequency of errors observed with clinical trial QA than with many clinical QA programs. For example, an O score of 10 in TG-100 is > 5 %, while an O score of 10 for this study was > 50 %.

One unique aspect of this analysis is that the occurrence and severity are generally under the control of the participating clinic. O and S ask what the occurrence and severity of errors are at institutions participating in clinical trials, i.e., which errors are most important to detect. In contrast, the detectability is under the control of the RTQA office and the QA test being implemented. O and S scores should be largely similar in a global setting, but D scores may be very different depending on how the QA office performs their specific evaluation. Therefore, we first evaluated just $O \times S$ scores to understand what the priority was to perform different QA tests. We then evaluated $O \times S \times D$ to include the impact of how well QA offices (as a whole) perform at managing these scenarios.

The original O, S, and D scores were compiled, based on thorough discussion across all group members, for a clinical trial based on X-ray therapy. The process was then repeated for a trial, and associated trial QA, that used proton therapy. The mean, median,

**Table 1**
Failure mode and effect analysis scoring adapted from AAPM Task Group 100 report [9]. The Occurrence (O) score has been modified to reflect the increased frequency of failures observed in clinical trial quality assurance.

| | **Frequency (Occurrence)** | | | **Severity** | | | **Detectability** |
|---|---|---|---|---|---|---|---|
| O Score | **Qualitative** | **Frequency** | S Score | **Qualitative** | **Categorization** | D Score | **Estimated probability of failure going undetected** |
| **1** | Failure | 0.01 % | **1** | No effect | | **1** | 0.01 % |
| **2** | unlikely | 0.1 % | **2** | Inconvenience | Inconvenience | **2** | 0.2 % |
| **3** | Relatively | 1 % | **3** | | | **3** | 0.5 % |
| **4** | few failures | 2–3 % | **4** | Minor dosimetric error | Suboptimal plan or treatment | **4** | 1 % |
| **5** | | 4–5 % | **5** | Limited toxicity or tumor | Wrong dose, dose distribution, | **5** | 2 % |
| **6** | Occasional | 6 %-10 % | **6** | underdosage | location, or volume | **6** | 5 % |
| **7** | failures | 11 %-20 % | **7** | Potentially serious toxicity or | | **7** | 10 % |
| **8** | Repeated | 21 %-35 % | **8** | tumor underdosage | | **8** | 15 % |
| **9** | failures | 36 %-50 % | **9** | Possible very serious toxicity or tumor underdose | Very wrong dose, dose distribution, location, or volume | **9** | 20 % |
| **10** | Failures inevitable | >50 % | **10** | Catastrophic | | **10** | >20 % |

and standard deviations of the scores were calculated and reviewed for outliers. Scores underwent several iterations after clarifying discussion between the participating RTQA groups.

Typically, FMEA does not consider the number of patients affected by a given error. It is thought that an error to any one patient is an error too many. However, within the context of clinical trials, the goal of ensuring consistency in radiotherapy delivery relates to validity of trial end-point analysis and therefore severity associated with numbers of patients differs. For clinical trial QA, an error that affects a single patient has less of an effect on the clinical trial data than an error that affects all patients enrolled by a participating institution. For this reason, we performed a sub-analysis that introduced a severity scaler based on how many clinical trial patients would be affected by the error. The scaler increased the S score if more than a few patients could be affected by the error (Table 2). An example of an error with a scaler might be a systematic beam output calibration error, which would affect all patients treated on that machine for the trial, as opposed to an error in contouring that would only affect a single patient on the trial.

## Results

The comprehensive list of clinical trial QA steps that was compiled by the GHG QA groups is shown in Table 3. Further descriptions of the terms can be found in the glossary in Appendix A.

**Table 2**
Severity (S) score scalers that were added to the S score to account for the proportion of patients potentially impacted by a clinical trial QA failure.

| Severity (S) Scalers | |
|---|---|
| **Value** | **Comment** |
| **+0** | Affects only 1 or a few patients |
| **+2** | Affects all patients of a certain disease site |
| **+4** | Affects all patients treated on that machine/facility |

While the specifics of a QA test may be implemented differently between photons and protons, the combined list was ultimately appropriate for either modality. Therefore, this is the list of tests that was ultimately scored according to the FMEA analysis.

The compilation and discussion of our clinical trial QA processes established that most RTQA groups used similar dosimeters and processes for RTQA in clinical trials. Between photon and proton therapy, many QA tests can be conducted using the same framework (e.g., an anthropomorphic phantom can use the same type of dosimeters for both modalities). There were a few exceptions for CT number vs relative linear stopping power or electron density evaluation (due to physical differences in dose deposition), treatment plan evaluation (due to differences in planning target volume definition and coverage), and robust optimization review (not yet universally performed for photons).

To ensure the consistency of D scores among RTQA group and modalities, as well as to examine which tests pose the highest risk from a clinical QA perspective, the median O $\times$ S scores were calculated. Fig. 1 shows the median O $\times$ S score for photons. These scores prioritize where we can focus our detection efforts. The analysis of the O $\times$ S scores showed similar trends of high scores to the full RPN (Fig. 2). Detectability did not have a big impact on the ranking of failures. Four out of five top photon scores and three out of five top proton scores overlapped.

The FMEA scores from each QA group were combined and the median scores are reported in Fig. 2. Photon results showed the highest risk tests were: general anthropomorphic end-to-end tests, phantoms assessing SABR/stereotactic treatment, phantoms assessing motion, pre-treatment review of contouring/outlining, and post-treatment review. The lowest risk tests for photons were: facility questionnaire, knowledge assessment, remote output check, on-site dosimetric output check, and robust optimization review. For proton therapy, the highest risk tests were: general anthropomorphic end-to-end tests, phantoms assessing heterogeneity, phantoms assessing motion, pre-treatment review

**Table 3**
Clinical trial quality assurance tests performed by members of the Global Clinical Trials RTQA Harmonization Group (GHG). Every test was performed for both proton and photon therapy, but the high occurrence (O), high severity (S), and high RPN scores are highlighted. An O score $\geq$ 6 corresponds to $\geq$ 6 % likelihood of failure. An S score $\geq$ 8 corresponds to a potentially serious or catastrophic normal tissue toxicity or tumour underdose.

| QA Category | Test | O $\geq$ 6 | | S $\geq$ 8 | | Top 20 % Median or Average RPN | |
|---|---|---|---|---|---|---|---|
| | | Photon | Proton | Photon | Proton | Photon | Proton |
| End-to-End Anthropomorphic Phantom Test | Anthropomorphic Phantom End-to-End (general) | | | | | x | x |
| | Phantom Assessing SABR/Stereo | x | | | x | x | |
| | Phantom Assessing Motion | x | x | | | x | x |
| | Phantom Assessing Heterogeneity | | | | | | x |
| | Phantom Assessing Multiple Targets | | | | | | |
| Dosimetric Measurements | Remote Output Check (out of tolerance) | | | | x | | |
| | Dosimetric Measurements - General | | | | x | | |
| | Dosimetric Measurements - Output | | | | x | | |
| | Dosimetric Measurements - Range | | | | | | |
| | Dosimetric Measurements - Lateral Profile | | | | | | |
| | IGRT/Radiation Coincidence | | | x | x | | |
| | CT Number vs RLSP/Electron Density Test | | | | | | |
| Programmatic Review | Facility Questionnaire | | | | | | |
| | Treatment Planning Evaluation | | x | | | | |
| | QA Program Evaluation | | x | | | | |
| | IGRT Credentialing | | | | | | |
| | Robust Optimization Review | | | | | | |
| Protocol Compliance | Knowledge Assessment | | | | | | |
| | Pre-trial Benchmark Contouring/Outlining | x | x | | | | |
| | Pre-trial Benchmark Plan/DVH Evaluation | x | | | | | |
| | Pre-treatment Review Contouring/Outlining | x | x | | | x | x |
| | Pre-treatment Review Plan/DVH Evaluation | x | | | | | |
| | On-treatment Review | | x | | | | |
| | Post-treatment Review | x | x | | x | x | x |
| | Post-trial feedback (e.g. deviations, accrual) | x | x | | | | |

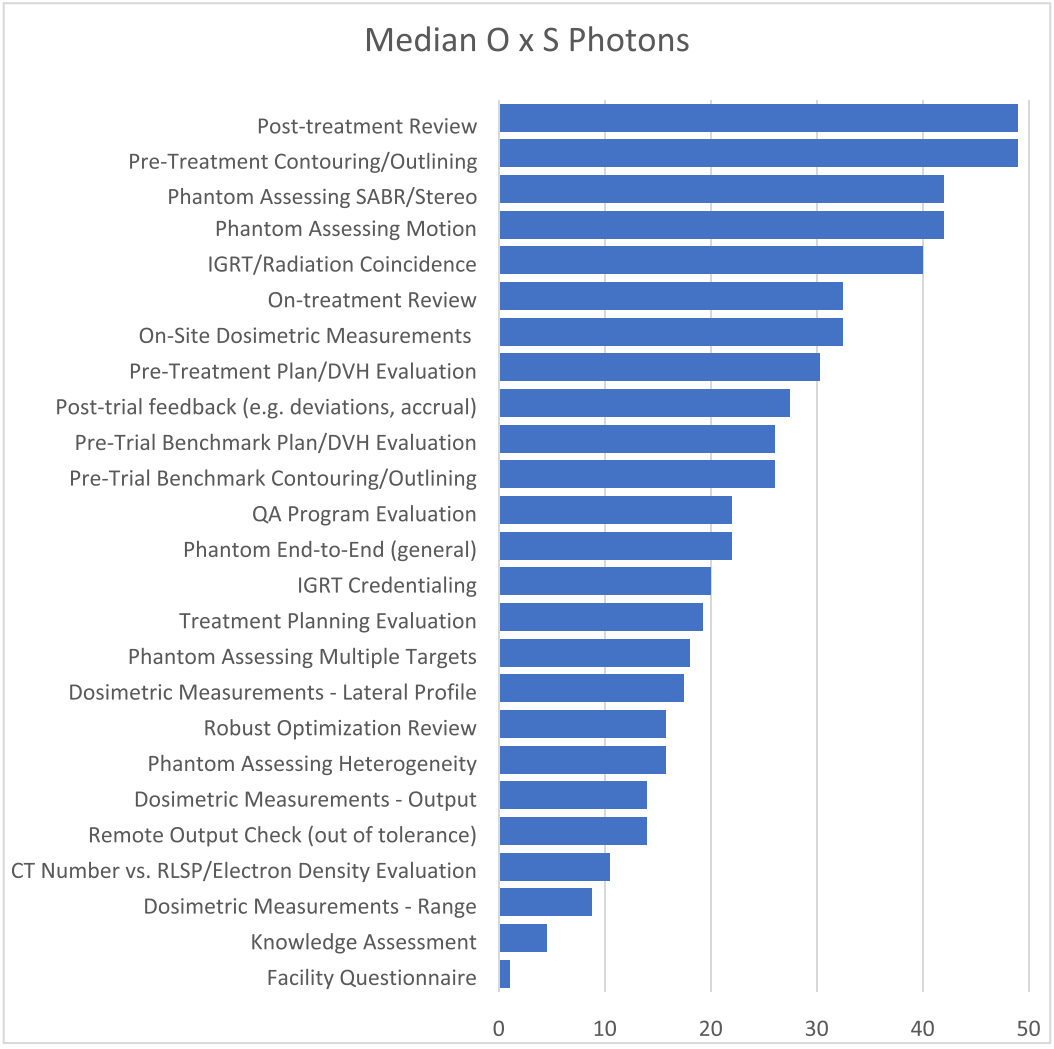


**Fig. 1.** The median O × S score for photon RTQA, ranked from highest score to lowest.

of contouring/outlining, and post-treatment review. The lowest risk tests for protons were: facility questionnaire, knowledge assessment, on-site dosimetric range check, on-site dosimetric output check, and IGRT/radiation isocenter coincidence.

Photon and proton QA showed similar top and low scorers. Proton RPN scores were higher than photon RPN scores for 15 of the 25 items assessed. This was largely driven by higher S scores for proton therapy. According to AAPM TG-100 [9], a severity score $\geq$ 8 is considered high risk. For photon QA, just one out of the 25 QA procedures had a median or average S score $\geq$ 8, while six out of 25 proton QA procedures had a median or average S score $\geq$ 8. Table 3 highlights which tests had high O and S scores.

The scaled RPN scores showed similar trends to the unscaled scores (Fig. 2). Four out of five photon and proton top scores overlapped between the scaled and non-scaled scores. The one test that was elevated to the top 20th percentile in the scaled scores was QA program evaluation, and the post-treatment patient plan review dropped out of the top 20th percentile for the scaled scores.

## Discussion

Radiotherapy QA in prospective clinical trials is correlated with improved patient outcomes. A reduction in major violations has been reported in studies with mandatory RTQA [4]. Because many violations occur during the treatment planning process, implementation of an efficient and streamlined real-time (pre-treatment) RTQA procedure can have substantial benefit to the trial. In a systematic review of 42 randomized controlled trials of RT since 1994, only 69 % of trials mandated RTQA, 45 % required institutional credentialing, and 50 % published protocol deviation outcomes [4]. These data support the inclusion of high-quality RTQA procedures, preferably with real-time central RTQA review, in proton and photon clinical trials.

Proton therapy is still new in some regions of the world – QA needs to rigorously test new technologies, then later be adapted as they become standard of care and the associated risk of errors reduces. For example, IROC performs on-site audits for every proton therapy centre as it is considered a "novel" technique, but only for a small subset of photon centres as it is the standard of care. At this time, many errors are still identified during on-site proton QA audits [10], but as our community knowledge of proton therapy increases and global guidelines and standards are adopted, these differences in clinical trial QA will diminish.

The RPN can help focus which tests are most critical to perform, and help improve workload, efficiency, and streamlining for RTQA groups and recruiting clinics. Knowledge assessments, for example, have a low RPN and could potentially be eliminated from a clinical trial QA program. Similarly, it could be useful to develop QA tests

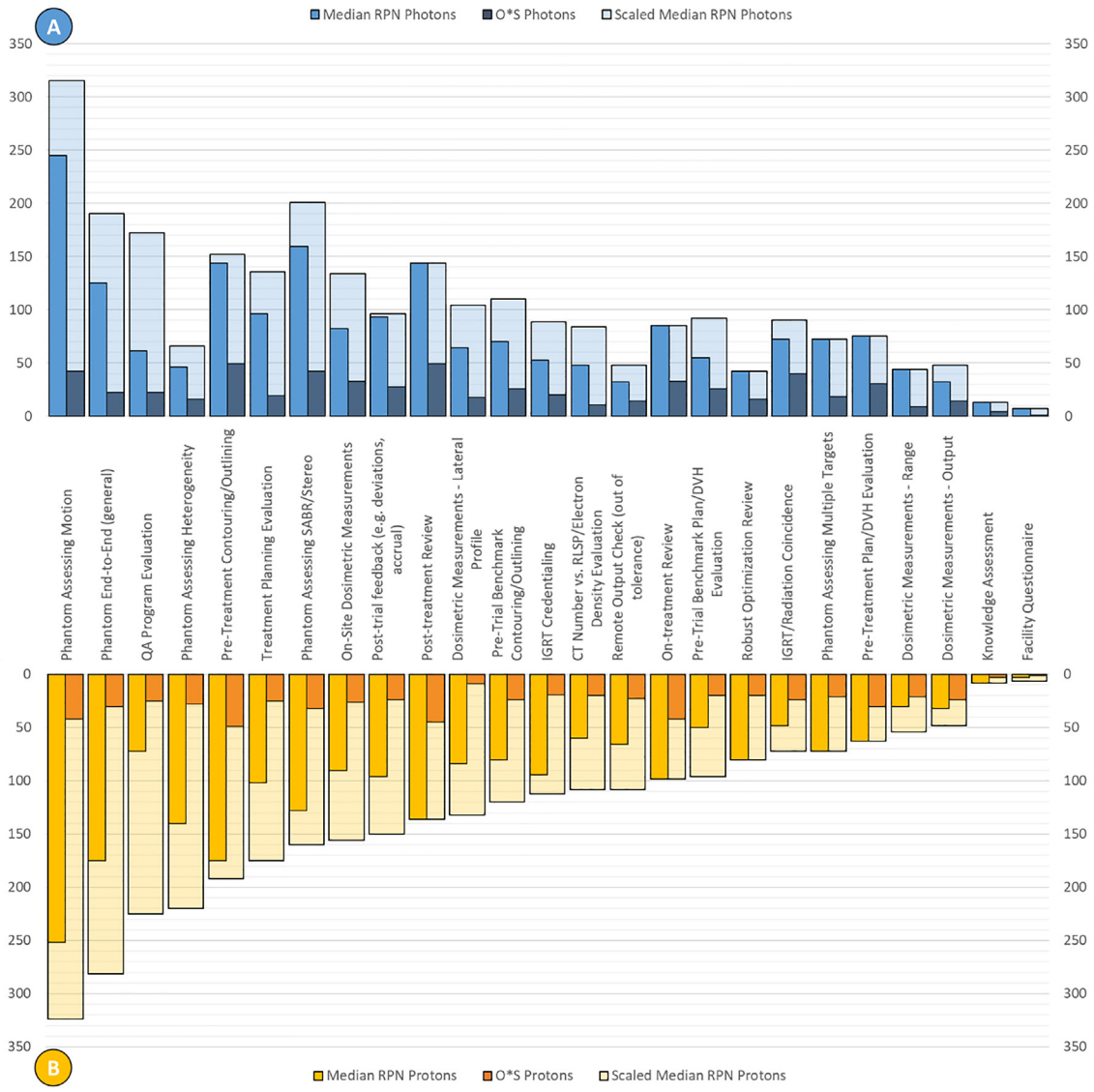


**Fig. 2.** Median Risk Priority Number, scaled RPNs, and O*S scores for both photon (A) and proton (B) modalities.

that are designed to test multiple high-scoring facets simultaneously, for example a phantom that assesses both heterogeneity and motion.

The RPN scores could also help QA groups focus on improving detectability for certain tests. For example, the median D score for proton anthropomorphic phantoms assessing motion was 6, and the RPN was in the top 20 %. A D score of 6 means that the probability of an error going undetected by the RTQA test is 5 %. For an RTQA program that performs 200 phantom audits assessing motion, as many as 10 failures might be missed. This suggests that efforts could be made to improve the accuracy of the end-to-end assessment in order to catch more errors. This could be done by improving the precision of the detectors or tightening the criteria used for assessment, for example. The O × S scores can be a useful tool for clinical QA, highlighting high risk failures that should be rigorously tested with institutional QA protocols, such as stereotactic treatment, motion management, and plan contouring. While clinical QA detectability may differ from clinical trial QA groups, studies show that routine patient QA practices lack sensitivity to detect many errors [11–13]. Clinical QA can improve detectability of high-risk failures.

The difference in S scores between photon and proton therapy highlight that proton errors are more likely to have a severe impact on patient treatment than photon errors. This is particularly concerning in the context of clinical trials. If there are more potential severe error modes in proton therapy, the clinical trial data could be more likely to show adverse outcomes as a result. For randomized photon vs proton trials, this could have a real impact on the ability to fairly compare both modalities. This highlights why it is critical, both at the clinical level and the trial QA level, that high-quality QA be performed to ensure minimal errors.

Thorough QA of local adaptive procedures might require site visits and continuous dialogue with the treating centres during the period of patient enrollment. In the thorax and oesophagus, anatomical changes have a very different impact on photon and proton therapy and the absence of treatment adaptation may lead to detrimental outcomes in one arm, and minor changes in the other arm [14–16]. Thus, continuous monitoring of dosimetric differences between treatment arms is necessary and may lead to changes in statistical power calculations if the expected dosimetric differences are not maintained.

These differences between photon and proton therapy and the concurrent high-risk failures for treatment with motion and heterogeneity suggest that clinical trials for new and evolving radiotherapy techniques should initially focus on more straightforward treatment disease sites. A randomized photon vs proton lung trial, for example, might not be the best place to start comparing new techniques. The RPN scores can help guide the development of clinical trials, as well as the rigor of associated clinical trial QA. For more complicated treatment sites, prospective and realistic QA (e.g. QA that challenges the variety of treatment scenarios with motion, heterogeneity, adaptive treatment, etc.) is essential to discover deficiencies in RTQA practices before patients are enrolled on trial. Clinical trials of advanced radiotherapy should, at a minimum, require anthropomorphic end-to-end phantom tests for credentialing. Evaluation of contouring and treatment planning prior to patient treatment is also highly recommended.

In summary of our three aims, we evaluated the relevance of current clinical trial QA methods by estimating the RPN for each failure. High RPN scores reinforced the relevance and importance of particular QA procedures, such as end-to-end phantom tests. Low RPN shed light on QA tests that may not be as important for clinical trial QA, such as knowledge assessments. We evaluated the effectiveness of current clinical trial QA methods by looking at the detectability of RTQA tests, especially in relation to the frequency and severity of errors. This demonstrated a need for improvement in detectability of errors like those associated with motion. Lastly, we contrasted the importance of specific clinical trial QA processes for standard of care versus novel treatment delivery. We established that the severity scores (and overall RPN) were higher for protons than for photons. The large difference in the RPN for phantom assessing heterogeneity is indicative that for certain sites with a lot of heterogeneity, such as head & neck and thorax, the auditing effort on this particular test may differ quite significantly between both modalities to ensure the same plan delivery quality. How this can be done in a balanced way, i.e., avoiding bias between both arms is an open question that needs further investigation.

The GHG's collaboration and findings highlight opportunities for shared learning across the radiotherapy community. One limitation of the study is that there is still relatively little international experience with proton therapy clinical trials. Many RTQA groups are still developing their practices for clinical trial QA. The GHG will continue to develop RTQA for proton therapy clinical trials and re-assess its status in about 5 years. RTQA organisations which are developing their practices for proton clinical trials QA will be able to find help and support from the more experienced groups.

The authors have learned much from each other throughout this process, which has identified steps to optimize each individual group's practice. Moving forward, the RTQA groups will prioritize pre/prospective QA, particularly for contouring, motion mitigation and IGRT assessment, as well as continuing to keep dosimetry audit and delineation assessment as priorities. Additionally, these data will allow us to adjust our allocation of human and material resources. Real time review is staff and time intensive (often working to very short turnaround times) and our RTQA groups will be investigating potential solutions for automation of the review process to facilitate wider adoption across trials. Lastly, for RTQA centers that are in the development stages for proton QA, the results of the study will inform future areas of focus for clinical trial RTQA. TROG, for example, is yet to develop an RTQA program for a proton therapy focused clinical trial, with the first Australian proton therapy facility currently under construction. The high-risk failures for proton therapy that have been highlighted by this study will critically inform RTQA program development in the future, as well as the design of ACDS' proton therapy dosimetry audit service.

Collaborations like these can help develop QA reciprocity programs among the RTQA groups to enable international recruitment without increasing the QA burden. The EORTC and QUARTET are currently reliant on certification from external credentialing providers. The results of this analysis will be considered when reviewing EORTC credentialing procedures and certification requirements for external partners. Similarly, clinical trials have a unique infrastructure to foster learning across multiple institutions. Some ideas to continue community-wide learning include cross-training through webinars or workshops, round-table discussions between investigators at different clinics, and meetings to compare treatment contouring and planning. These learning opportunities can reduce treatment errors and improve clinical trial data quality.

## Conclusion

An FMEA approach to evaluating RTQA in clinical trials identifies high-risk failures that occur in radiotherapy treatment. While there is higher risk associated with many proton failures, there are many failures for both modalities that warrant rigorous QA to prevent these errors. RTQA groups can use these findings to focus trial QA efforts, and clinical centres can similarly focus clinical QA programs on high-risk errors. These combined efforts will not only help improve clinical trial data but have the potential to improve outcomes for radiotherapy patients worldwide.

## Conflict of Interest

The authors declare there are no conflicts of interest.

## Acknowledgements

The National Institute for Health Research (NIHR) fund the RTTQA Group. The EORTC Radiation Oncology Group acknowledge contributions from the Radiotherapy Quality Assurance and Radiation Therapy Technologists committees. The activities of QUARTET are supported by the Fondatioun Kriibskrank Kanner, Luxembourg. IROC is supported by grant 2U24CA180803-06 from the National Cancer Institute. The Trans Tasman Radiation Oncology Group is supported by funding from Cancer Australia through the Support For Cancer Clinical Trials Program Grant.

## Appendix A. RT QA Terminology

Terminology descriptions for the various radiotherapy quality assurance tests performed by the Global Clinical Trials RTQA Harmonization Group (GHG).

| QA Terminology/Activity | Purpose |
|---|---|
| Facility questionnaire | Demographics, workload, equipment, techniques, and procedures of the site |
| Beam output audit | Beam output measurement under reference conditions [17,18] |
| Dosimetry audit | Independent machine output measurement and dosimetric measurements of a treatment plan. Can be performed as a site visit or remote process using a phantom [10,19–21] |
| Anthropomorphic phantom | A phantom simulating human anatomy [22–29] |
| Knowledge assessment | A questionnaire assessing the reader's comprehension of the clinical trial protocol requirements |
| Benchmark case (pre-trial) | Planning and/or delineation test on a common CT dataset |
| IGRT credentialing | QA of image matching accuracy and adaptive decision making |
| Individual case review (on trial) | Review of the delineation and plan for individual trial patients<br>● Pre-treatment: in real time prior to a patient starting treatment.<br>● On-treatment: during the course of a patient's treatment<br>● Post-treatment - after the patient has been treated |
| Review of patients' treatment records | Retrospective review of the actual treatment plan and delineation of trial patients and protocol compliance [30] |
| Case report forms | Review of retrospectively submitted forms by each site on all trial patients concerning their RT treatment |
| Local QA program evaluation | Observations of the entire treatment process for a trial patient, local QA procedures, and documentation and review of treatment records |
| Credentialing | Granting an institution a QA approval to participate in a clinical trial |
| Robust optimization | Optimization of a radiotherapy treatment plan to account for various types of uncertainty in beam modeling, patient setup, and treatment delivery |

## Appendix B. The following table shows the median scores for the failure modes and effects analysis (FMEA) for photon and proton clinical trial QA.

| QA Category | Test | Photon | | | | Proton | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | O | S | D | RPN | O | S | D | RPN |
| End-to-End Anthropomorphic Phantom Test | Anthropomorphic Phantom End-to-End | 4 | 6 | 5 | 125 | 5 | 6 | 5 | 175 |
| | Phantom Assessing SABR/Stereo | 6 | 7 | 4 | 160 | 4 | 8 | 4 | 128 |
| | Phantom Assessing Motion | 6 | 7 | 6 | 245 | 6 | 7 | 6 | 252 |
| | Phantom Assessing Heterogeneity | 4 | 5 | 3 | 46 | 4 | 7 | 5 | 140 |
| | Phantom Assessing Multiple Targets | 3 | 6 | 4 | 72 | 3 | 7 | 4 | 72 |
| Dosimetric Measurements | Remote Output Check (out of tolerance) | 2 | 7 | 4 | 32 | 3 | 8 | 4 | 66 |
| | Dosimetric Measurements - General | 5 | 7 | 3 | 82 | 4 | 8 | 4 | 90 |
| | Dosimetric Measurements - Output | 2 | 7 | 2 | 32 | 3 | 8 | 2 | 32 |
| | Dosimetric Measurements - Range | 3 | 4 | 3 | 44 | 3 | 7 | 2 | 30 |
| | Dosimetric Measurements - Lateral Profile | 5 | 4 | 3 | 64 | 3 | 3 | 4 | 84 |
| | IGRT/Radiation Coincidence | 5 | 8 | 3 | 72 | 3 | 8 | 3 | 48 |
| | CT Number vs RLSP/Electron Density Evaluation | 3 | 4 | 4 | 48 | 4 | 5 | 3 | 60 |
| Programmatic Review | Facility Questionnaire | 1 | 1 | 1 | 7 | 1 | 1 | 1 | 3 |
| | Treatment Planning Evaluation | 4 | 6 | 5 | 96 | 6 | 5 | 5 | 102 |
| | QA Program Evaluation | 4 | 6 | 5 | 61 | 5 | 5 | 5 | 72 |
| | IGRT Credentialing | 4 | 5 | 3 | 53 | 3 | 7 | 4 | 95 |
| | Robust Optimization Review | 4 | 5 | 4 | 42 | 5 | 4 | 4 | 80 |
| Protocol Compliance | Knowledge Assessment | 3 | 2 | 3 | 13 | 3 | 1 | 2 | 8 |
| | Pre-trial Benchmark Contouring/Outlining | 7 | 4 | 4 | 70 | 6 | 4 | 4 | 80 |
| | Pre-trial Benchmark Plan/DVH Evaluation | 7 | 4 | 2 | 55 | 5 | 4 | 2 | 50 |
| | Pre-treatment Review Contouring/Outlining | 7 | 7 | 4 | 144 | 7 | 7 | 4 | 175 |
| | Pre-treatment Review Plan/DVH Evaluation | 6 | 6 | 3 | 75 | 5 | 6 | 3 | 63 |
| | On-treatment Review | 5 | 7 | 3 | 85 | 6 | 7 | 3 | 98 |
| | Post-treatment Review | 7 | 7 | 3 | 144 | 6 | 8 | 4 | 136 |
| | Post-trial feedback (e.g. deviations, accrual) | 6 | 5 | 4 | 93 | 6 | 4 | 5 | 96 |

## References


[1] Peters LJ, O'Sullivan B, Giralt 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. Journal of Clinical Oncology. Published online 2010. doi:10.1200/JCO.2009.27.4498.

[2] 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. Journal of the National Cancer Institute 2013;105:387–93. https://doi.org/10.1093/jnci/djt001.

[3] Pettersen MN, Aird E, Olsen DR. Quality assurance of dosimetry and the impact on sample size in randomized clinical trials. Radiotherapy and Oncology 2008;86. https://doi.org/10.1016/j.radonc.2007.07.001.

[4] Corrigan KL, Kry S, Howell RM, et al. The radiotherapy quality assurance gap among phase III cancer clinical trials. Radiotherapy and Oncology 2022;166. https://doi.org/10.1016/j.radonc.2021.11.018.

[5] Maroongroge S, Wallington DG, Taylor PA, et al. Geographic Access to Radiation Therapy Facilities in the United States. International Journal of Radiation Oncology, Biology, Physics 2022;112. https://doi.org/10.1016/j.ijrobp.2021.10.144.

[6] Group Members – Global Quality Assurance of Radiation Therapy Clinical Trials Harmonization Group. Accessed November 28, 2022. https://rtqaharmonization.org/group-members/.

[7] Terms of Reference – Global Quality Assurance of Radiation Therapy Clinical Trials Harmonization Group. Accessed November 28, 2022. https://rtqaharmonization.org/terms-of-reference/.

[8] Melidis C, Bosch WR, Izewska J, et al. Radiation therapy quality assurance in clinical trials – Global harmonisation group. Radiotherapy and Oncology 2014;111:327–9. https://doi.org/10.1016/J.RADONC.2014.03.023.

[9] Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management. Medical Physics 2016;43:4209–62. https://doi.org/10.1118/1.4947547.

[10] Taylor PA, Lowenstein J, Followill D, Kry SF. The Value of On-Site Proton Audits. International Journal of Radiation Oncology*Biology*Physics 2022;112:1004–11. https://doi.org/10.1016/J.IJROBP.2021.10.145.

[11] Pulliam KB, Followill D, Court L, et al. A six-year review of more than 13,000 patient-specific IMRT QA results from 13 different treatment sites. Journal of Applied Clinical Medical Physics 2014;15:196–206. https://doi.org/10.1120/jacmp.v15i5.4935.

[12] Kry SF, Molineu A, Kerns JR, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. Int J Radiat Oncol Biol Phys. Published online 2014. doi:10.1016/j.ijrobp.2014.08.334.

[13] Kry S, Glenn M, Peterson C, et al. Independent recalculation outperforms traditional measurement-based IMRT QA methods in detecting unacceptable plans. Medical Physics 2019;46:3700–8. https://doi.org/10.1002/MP.13638.

[14] Hoffmann L, Alber M, Jensen MF, Holt MI, Møller DS. Adaptation is mandatory for intensity modulated proton therapy of advanced lung cancer to ensure target coverage. Radiotherapy and Oncology 2017;122. https://doi.org/10.1016/j.radonc.2016.12.018.

[15] Møller DS, Poulsen PR, Hagner A, et al. Strategies for Motion Robust Proton Therapy With Pencil Beam Scanning for Esophageal Cancer. International Journal of Radiation Oncology, Biology, Physics 2021;111. https://doi.org/10.1016/j.ijrobp.2021.04.040.

[16] Shamshad M, Møller DS, Mortensen HR, Ehmsen ML, Jensen MF, Hoffmann L. Bone versus soft-tissue setup in proton therapy for patients with oesophageal cancer. Acta Oncol (Madr) 2022;61:994–1003. https://doi.org/10.1080/0284186X.2022.2091949/SUPPL_FILE/IONC_A_2091949_SM9880.DOCX.

[17] Alvarez P, Kry SF, Stingo F, Followill D. TLD and OSLD dosimetry systems for remote audits of radiotherapy external beam calibration. Radiation Measurements 2017;106:412–5. https://doi.org/10.1016/j.radmeas.2017.01.005.

[18] Kry SF, Peterson CB, Howell RM, et al. Remote beam output audits: A global assessment of results out of tolerance. Phys Imaging Radiat Oncol 2018;7. https://doi.org/10.1016/j.phro.2018.08.005.

[19] Kry SF, Dromgoole L, Alvarez P, et al. Radiotherapy deficiencies identified during on-site dosimetry visits by the IROC Houston QA Center. International Journal of Radiation Oncology*Biology*Physics 2017.

[20] Carlino A, Palmans H, Gouldstone C, et al. Results of an independent dosimetry audit for scanned proton beam therapy facilities. Zeitschrift für Medizinische Physik 2021;31:145–53. https://doi.org/10.1016/J.ZEMEDI.2021.01.003.

[21] Nishio T, Nakamura M, Okamoto H, et al. An overview of the medical-physics-related verification system for radiotherapy multicenter clinical trials by the Medical Physics Working Group in the Japan Clinical Oncology Group-Radiation Therapy Study Group. Journal of Radiation Research 2020;61:999–1008. https://doi.org/10.1093/jrr/rraa089.

[22] Nishio T, Tachibana H, Kase Y, et al. Liver phantom design and dosimetric verification in participating institutions for a proton beam therapy in patients with resectable hepatocellular carcinoma: Japan Clinical Oncology Group trial (JCOG1315C). Radiotherapy and Oncology 2019;140:98–104. https://doi.org/10.1016/J.RADONC.2019.06.005.

[23] Clark CH, Hurkmans CW, Kry SF. The role of dosimetry audit in lung SBRT multi-centre clinical trials. Physica Medica 2017;44:171–6. https://doi.org/10.1016/J.EJMP.2017.04.003.

[24] Molineu A, Hernandez N, Nguyen T, Ibbott G, Followill D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. Medical Physics 2013;40. https://doi.org/10.1118/1.4773309.

[25] Kry S, Alvarez P, Molineu A, Amador C, Galvin J, Followill D. Algorithms used in heterogeneous dose calculations show systematic differences as measured with the Radiological Physics Center's anthropomorphic thorax phantom used for RTOG credentialing. International Journal of Radiation Oncology, Biology, Physics 2013;85. https://doi.org/10.1016/J.IJROBP.2012.08.039.

[26] Lewis DJ, Taylor PA, Followill DS, et al. A New Anthropomorphic Pediatric Spine Phantom for Proton Therapy Clinical Trial Credentialing. Int J Part Ther 2018;4:20–7. https://doi.org/10.14338/ijpt-17-00024.1.

[27] Branco D, Taylor P, Zhang X, Li H, Guindani M, Followill D. An Anthropomorphic Head and Neck Quality Assurance Phantom for Credentialing of Intensity-Modulated Proton Therapy. *Int J Part Ther* Published online 2017. https://doi.org/10.14338/ijpt-17-00005.1.

[28] Taylor PA, Kry SF, Alvarez P, et al. Results from the Imaging and Radiation Oncology Core Houston's Anthropomorphic Phantoms Used for Proton Therapy Clinical Trial Credentialing. Int J Radiat Oncol Biol Phys. Published online 2016. doi:10.1016/j.ijrobp.2016.01.061.

[29] Taylor PA, Kry SF, Followill DS. Pencil Beam Algorithms Are Unsuitable for Proton Dose Calculations in Lung. International Journal of Radiation Oncology, Biology, Physics 2017;99. https://doi.org/10.1016/j.ijrobp.2017.06.003.

[30] Ford E, Conroy L, Dong L, et al. Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275. Medical Physics 2020;47. https://doi.org/10.1002/mp.14030.