AAPM REPORT NO. 113

Guidance for the Physics Aspects of Clinical Trials

The Report of AAPM Task Group 113

January 2018

This page intentionally left blank.

Guidance for the Physics Aspects of Clinical Trials

The Report of AAPM Task Group 113

Jean M. Moran, Chair
University of Michigan, Ann Arbor, MI

Andrea Molineu
UT MD Anderson Cancer Center, Houston, TX

Jon J. Kruse
Mayo Clinic, Rochester, MN

Mark Oldham
Duke University Medical Center, Durham, NC

Robert Jeraj
University of Wisconsin, Madison, WI

James M. Galvin
Imaging and Radiation Oncology Core, Philadelphia, PA

Jatinder R. Palta
Virginia Commonwealth University, Richmond, VA

Arthur J. Olch, Vice Chair
University of Southern California and Children's Hospital of Los Angeles, Los Angeles, CA

DISCLAIMER: This publication is based on sources and information believed to be reliable, but the AAPM, the authors, and the publisher disclaim any warranty or liability based on or relating to the contents of this publication.

The AAPM does not endorse any products, manufacturers, or suppliers. Nothing in this publication should be interpreted as implying such endorsement.

ISBN: 978-1-936366-59-0
ISSN: 0271-7344

Published by

American Association of Physicists in Medicine
1631 Prince Street
Alexandria, VA 22314

Abstract

The charge of AAPM Task Group 113 is to provide guidance for the physics aspects of clinical trials to minimize variability in planning and dose delivery for external beam trials involving photons and electrons. Several studies have demonstrated the importance of protocol compliance on patient outcome. Minimizing variability for treatments at different centers improves the quality and efficiency of clinical trials. Attention is focused on areas where variability can be minimized through standardization of protocols and processes through all aspects of clinical trials. Recommendations are presented for clinical trial designers, physicists supporting clinical trials at their individual clinics, quality assurance centers, and manufacturers.

Keywords: external beam, quality assurance, clinical trials, protocols, standardization, PACS.

Contents

1. Abbreviations

4DCT	four-dimensional computed tomography
AAPM	American Association of Physicists in Medicine
ACR	American College of Radiology
ACRIN	American College of Radiology Imaging Network
CIRO	Center for Innovation for Radiation Oncology
CT	computed tomography
DICOM-RT	Digital Imaging and Communications in Medicine Standard-Radiation Therapy
DRR	digitally reconstructed radiograph
DVH	dose volume histogram
EUD	equivalent uniform dose
FDG-PET	fluorodeoxyglucose-positron emission tomography
GOG	Gynecologic Oncology Group
ICRU	International Commission on Radiation Units and Measurements
IGRT	image-guided radiation therapy
IHE-RO	Integrating the Healthcare Enterprise in Radiation Oncology
IMRT	intensity-modulated radiation therapy
IOM	Institute of Medicine
IRB	institutional review board
IROC	Imaging and Radiation Oncology Core Group
ITC	Image-guided Therapy Center
MRI	magnetic resonance imaging
MRS	magnetic resonance spectroscopy
MU	monitor unit
NCI	National Cancer Institute
NCTN	National Clinical Trials Network
NRG	National clinical trials group formed from NSABP, RTOG, and GOG (See Table 1)
NSABP	National Surgical Adjuvant Breast and Bowel Project
NTCP	normal tissue complication probability
OAR	organ-at-risk
OSL	optically stimulated luminescent dosimeter
PET	positron emission tomography
PTV	planning target volume
QA	quality assurance
QARC	Quality Assurance Review Center
QIBA	Quantitative Imaging Biomarkers Alliance
QIN	Qualitative Imaging Network
RPC	Radiological Physics Center (see Table x)
RTOG	Radiation Therapy and Oncology Group
SBRT	stereotactic body radiation therapy
SPECT	single photon emission computed tomography
SWOG	South West Oncology Group
TG	task group
TLD	thermoluminescent dosimeter
TPS	treatment planning system
TROG	Trans Tasman Radiation Oncology Group
UPICT	Uniform Protocols for Imaging in Clinical Trials
VMAT	volumetric-modulated arc therapy

2. Introduction and Charge of the Report

There is growing evidence[1–5] on the need for standardization of treatment planning and delivery methods to ensure quality in clinical trials to help support the investigation of new safe and effective treatments or assessment methods in multi-institutional settings. Such standardization will improve the consistency of the radiotherapy received by patients and the radiotherapy data submitted for a given clinical trial. These data are required to validate that all patients in each arm of a given study received the therapy as intended. Violating this assumption can jeopardize the validity of the outcomes reported by the trial group.

The results of the Radiation Therapy Oncology Group (RTOG) head and neck trial (0022) demonstrated the importance of standardized guidelines. This was the first National Cancer Institute (NCI) multi-institutional trial allowing intensity-modulated radiation therapy (IMRT) and required confirmation of the ability to deliver plans with steep dose gradients. Study investigators reported higher clinical failure rates for patients whose treatment plans showed major deviations from the protocol guidelines.[6] The trial required a comparison of calculations and measurements in multiple planes based on irradiation of a custom-designed head and neck phantom[7] and a review of the targets and setup images for the initial cases. The failure rate for the phantom irradiations at the participating institutions was 32% in 2008.[8]

A related consideration which affects overall quality is the ability of those participating in clinical trials to create plans as part of their standard clinical flow that are compliant with protocol specifications. The importance of compliance in trials and the impact on detecting changes in outcome have been demonstrated in a number of trials[1–4,9], such as in TROG 02.02 on advanced head and neck cancer (Figure 1) and meta-analyses of other trials. When designing a trial, the planning guidelines are set to be able to answer the clinical trial questions. However, there may be variation in planning methods, and a planner may not know when a better (such as improved target coverage with reduced dose to normal tissues) plan is reasonably achievable without real-time feedback during the planning process. Knowledge-based planning is a method where the achievable dose volume metrics from previous patients can used to predict each new patient's dose volume histogram (DVH) as well as to identify plans with DVHs that significantly deviate from the model. Moore et al. retrospectively used knowl-

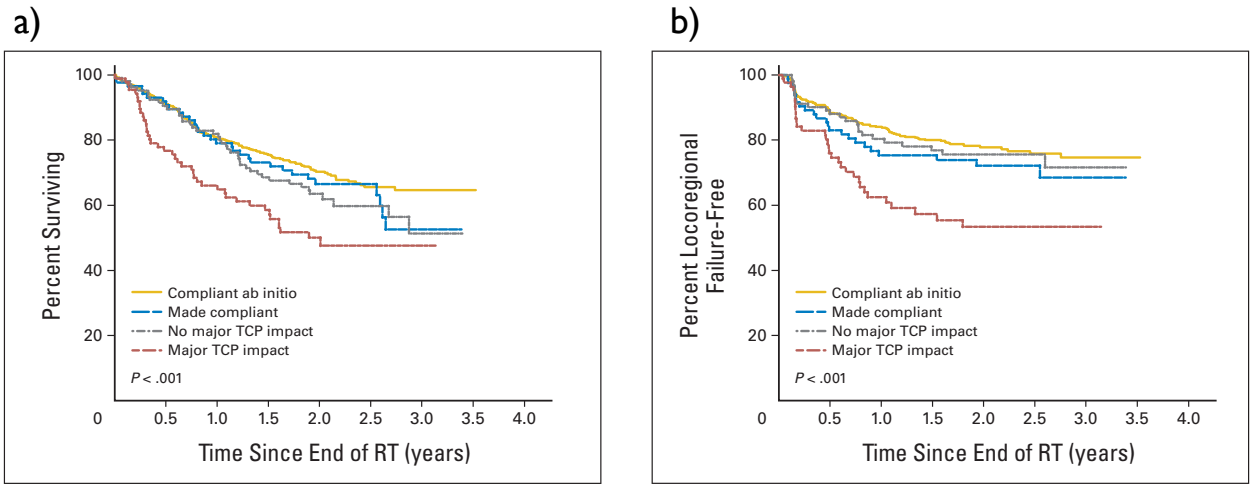

Figure 1. Peters et al. assessed the impact of protocol compliance for TROG 02.02 on advanced head and neck cancer and demonstrated an impact on a) overall survival and b) time to locoregional failure as a function of the deviation status. (Figures 2 and 3 reprinted with permission from Peters et al., JCO, 28: p. 2999.)

edge-based methods for RTOG 0126 by using the best plans to create a model and then identifying plans where improvements could be made when the submitted protocol plan was compared to the prediction from the model.[10] Improved planning tools, such as those with knowledge-based planning, have been needed for some time to provide detailed feedback to institutions on whether or not their treatment plans not only meet the dose volume histogram requirements but are also optimal for use in clinical trials. With respect to quality assurance requirements, there are important ongoing efforts toward global harmonization of quality assurance[9] (such as structure nomenclature addressed by AAPM Task Group (TG) 263[11]) for radiation therapy clinical trials.

The charge of AAPM TG-113 is to:

1. Recommend physics practices for clinical trials involving external photon and electron beam radiation therapy that ensure minimum standards for data quality in clinical trials.
2. Identify opportunities to improve consistency in each part of the planning and delivery process.
3. Provide guidance to QA organizations on how best to support the spectrum of radiotherapy clinical trials, from those with basic to advanced technology.
4. Provide suggestions regarding the credentialing requirements to reduce potential inconsistencies in the radiotherapy process.

The use of protons or brachytherapy in clinical trials is outside of the scope of this document. Throughout the report, recommendations are presented in each section for major areas of the process, from simulation through treatment delivery, in the context of clinical trials. The recommendations are organized by the categories of clinical trial designers, physicists (at the local institution), quality assurance (QA) centers, manufacturers, and advanced technology trials. They are also presented by category in appendices A–D.

3. Clinical Trial Group Restructuring

In response to an Institute of Medicine (IOM) report published in 2010, the clinical trials network and associated QA (Quality Assurance) centers funded by the NCI were restructured. The IOM report emphasized needs such as improving the speed and efficiency of trials, incorporating innovation, and facilitating cooperation to get trials underway.[12] The NCI National Clinical Trials Network (NCTN) consists of the Children's Oncology Group (COG), South West Oncology Group (SWOG), Alliance for Clinical Trials in Oncology, NRG Oncology, and Eastern Cooperative Oncology Group (ECOG)–American College of Radiology Imaging Network (ACRIN), which have centralized functions supporting all trials. This centralized support was an important aspect of the restructuring, which is meant to improve the efficiency of clinical trials. Some centralized support is also provided to the National Cancer Institute of Canada Clinical Trials Group. NTCN also has academic institutions that have a grant which designates the institution as a Lead Academic Participating Site (LAPS). NRG Oncology was formed from the National Surgical Adjuvant Breast and Bowel Project (NSABP), the Radiation Therapy Oncology Group (RTOG), and the Gynecologic Oncology Group (GOG).

As part of the overall restructuring of the clinical trials network, the previous clinical trial QA centers were also restructured to combine radiation therapy and imaging QA into a single core support group called the Imaging and Radiation Oncology Core (IROC) Cooperative.[13] IROC provides support to NCTN for both the radiation oncology and diagnostic imaging to ensure that the appropriate quality assurance measures are available to clinical trials. The current QA center names, sample duties, and the previous organization (when appropriate) are shown in Table 1. Functions previously provided by other QA centers for radiation therapy (QARC, RPC, ITC) and imaging (American College of Radiol-

Table 1. Current and previous names and sample duties of the QA clinical trial groups related to external beam radiation therapy

Current QA Center	Sample Duties and Responsibilities	Previous Name
American College of Radiology (ACR)	Grant holder for IROC (see below); development of software (TRIAD) for data management, storage and retrieval components of the case review process	No name change
Imaging and Radiation Oncology Core (IROC)	Integrated network which oversees each IROC site for protocol development and clinical trial QA for imaging and radiation oncology trials (www.irocqa.org)	No previous organization. The Advanced Technology Consortium had provided a virtual network to support some harmonization of QA efforts for radiation oncology alone.
IROC Houston	Remote dosimetry and phantom programs, unified facility questionnaire, credentialing	Radiological Physics Center (RPC)—NCI funded since 1968; MD Anderson Cancer Center
IROC Rhode Island	Therapy, diagnostic imaging, and data management support	Quality Assurance Review Center (QARC)
IROC St. Louis	Support for digital data submission and efforts toward standardization of data submitted	Image-guided Therapy Center (ITC), Washington University, St. Louis
IROC Philadelphia Radiation Therapy (RT)	Software development for data upload, central remote review services, analysis in support of trials	Radiation Therapy Oncology Group (RTOG) Core Lab at the RTOG QA Center
IROC Philadelphia Radiation Therapy (RT)	Software development and QA for imaging studies	ACR Diagnostic Imaging (DI) Core Laboratory
IROC Ohio	Imaging data analysis for multiple imaging modalities	Imaging Core Laboratory, Ohio State/Wright Imaging Center

ogy Imaging Network (ACRIN) and American College of Radiology (ACR) core lab) have been integrated within IROC. For clarity, this report uses the IROC names. The Center for Innovation in Radiation Oncology (CIRO) consists of radiation oncologists and medical physicists which serves as a resource for incorporating innovation and new technology in radiation oncology NCTN trials (see https://www.nrgoncology.org/Scientific-Program/Center-for-Innovation-in-Radiation-Oncology).

The importance of quality to ensure an adequate test of a trial question must be balanced with demands on resources (such as time and money). The QA centers continually assess the best way to use their own resources to support clinical trials. Resources are often strained at individual institutions, but sufficient resources are needed to perform benchmarks and credentialing activities, to support the enrollment and successful aggregation of quality treatment planning, and to ensure data that meet the protocol requirements.

Throughout this document, we recognize that the NTCN study committees have the important role of determining which data are necessary to collect to adequately answer the question(s) that are being asked by the trial. The data to be collected are determined by the protocol and managed by data management staff. This report provides additional recommendations for the use of advanced technology tools in the context of clinical trials. Advanced technology trials are ones in which a technology is still maturing or where there is complexity in the imaging, planning, or delivery that warrants additional credentialing activities. The utilization of different tools and imaging techniques in the broader community and in the context of clinical trials continues to change as new delivery systems and techniques become available. This report emphasizes that medical physicists play a crucial role in ensuring that treatments are delivered per the protocol.

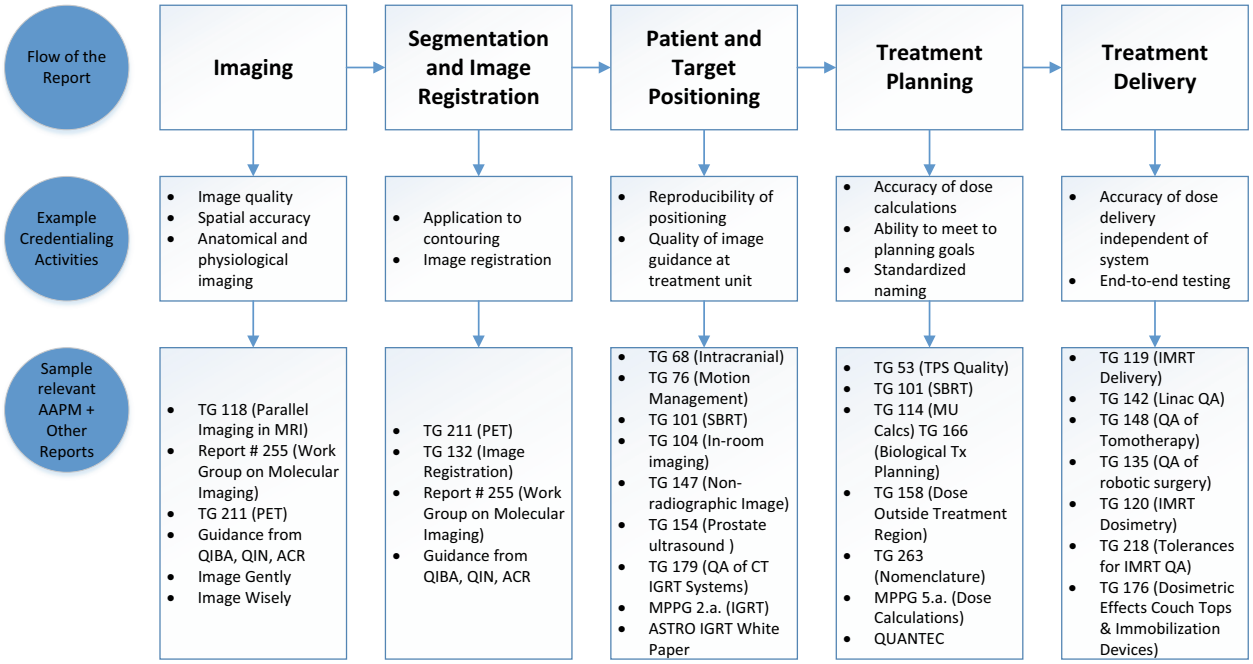

Figure 2. This diagram shows the flow of the AAPM TG-113 report with respect to the patient treatment process (top row). Example credentialing activities related to quality in clinical trials are shown in the middle row. The bottom row shows AAPM and other reports that are related to the areas in the top row.

4. The Role of the Physicist in Clinical Trials

Physicists play different roles with respect to clinical trials. At institutional, national, and international levels, physicists may be lead investigators or co-investigators representing clinical and technical components. In the context of clinical trial groups, physicists may lead or co-design a clinical trial. For national trials supported at individual institutions, physicists play a key role with physicians in ensuring protocol compliance. Other perspectives include physicist roles in QA centers and as employees of a manufacturer whose products are being used to support clinical trials.

The physics support for clinical trial groups can vary, and it is affected by each group's resources. In 2004, AAPM TG-86, Quality Assurance for Clinical Trials: A Primer for Physicists, reviewed the structure and activities of clinical trial groups.[14] In that report, recommendations were made throughout the report relating to the major areas of the patient simulation and treatment process, as well as the credentialing of institutions.

TG-113 considers the entire process—from simulation to planning and treatment delivery—to improve the consistency for clinical trials, whether trials are funded by NCI, industry, or other entities. Many AAPM task group reports are relevant to the work of TG-113. Figure 2 shows an overview of the major areas involved once a patient is enrolled in a clinical trial. For each area, both sample relevant task group reports, as well as credentialing types, are noted. Many of the referenced task group reports are ones that are already relevant to the practice of clinical medical physics in radiation therapy which then have an impact on the treatment of patients enrolled in clinical trials. Therefore, minimal additional references are made to task group reports throughout this report.

5. Image Acquisition for Target Definition and Treatment Assessment

Image quality is paramount to many clinical trials for both target definition and treatment assessment. This section makes recommendations to facilitate consistent and accurate volume definition for clinical trials. Numerous collaborative efforts are focused on standardization of imaging, including quantitative applications. Formed in 2008, the Quantitative Imaging Biomarkers Alliance (QIBA) involves drug and equipment companies and imaging societies and has a charge to develop and advance standards for the use of volumetric computerized tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI) in clinical trials. QIBA has created validated datasets (including ones that can be used for evaluating lung nodules[15]) and phantom datasets that are used to validate analytical tools such as dynamic contrast-enhanced MRI image datasets.[16] The Uniform Protocols for Imaging in Clinical Trials (UPICT) initiative has created a protocol for trials involving imaging with FDG-PET/CT.[17] Several groups within the AAPM are actively advancing the use of quantitative imaging information, and guidance will continue to evolve in this area.

Some clinical trials require credentialing or a central imaging review by QA centers that have expertise in quantitative imaging, such as IROC Ohio, IROC Philadelphia (DI), and IROC Rhode Island. Credentialing may evaluate characteristics such as image quality, spatial integrity, and contrast; the requested characteristics will depend on the role of imaging within a given trial. For example, considerations with respect to understanding uncertainties in molecular imaging have been described.[18]

The primary goal of imaging for treatment planning is to define, as accurately as possible, the extent (with margins) of diseased tissue to be treated (the gross tumor and clinical target volumes) and healthy tissue to be avoided. Significant advances in the accuracy of volume definition for gross disease are achieved in many sites through technological developments in image quality, increased use of multi-modality imaging, and new strategies and algorithms for determining tissue boundaries and margins.

Many trials involve multiple image datasets. Clinical trial designers should ensure that all protocols involving imaging include development and review by imaging experts. They should also provide a standardized procedure for imaging that can be used to create a standard operating procedure at participating institutions with respect to the image acquisition, reconstruction, processing, and analysis.

The clinical trial designers should make sure that the treatment team at each institution receives a template with a full description of any specific patient preparations (e.g., diet guidelines and timelines for drinking and voiding to enable treatment with full or empty bladder, respectively). Other considerations for clinical trials include the use of accredited scanners (such as by the ACR), phantom testing that is specific to the imaging technology being used, following profiles established by organizations such as QIBA and the Quantitative Imaging Network (QIN) to ensure that quantitative data are valid, and having image acquisitions led by knowledgeable personnel, including an imaging scientist.

QIBA has led the creation of a number of profiles for clinical investigators to assess the accuracy of imaging biomarkers. These profiles are used to communicate with vendors and users to facilitate collaborative work. QIBA profiles are well described at: https://www.rsna.org/QIBA-Profiles-and-Protocols/. The current statistical methodology for evaluating quantitative imaging techniques includes the use of phantoms and digital reference images.[19] These considerations are crucial for clinical trials which incorporate quantitative biomarkers in multi-institutional trial settings.

The clinical trial designers are responsible for identifying the key imaging information, including modality, contrast agents, pulse sequences for MRI and magnetic resonance spectroscopy (MRS), timing of imaging with respect to the start of treatment, and the anatomical extent. It may be necessary to translate parameters between equipment from different manufacturers. For CT, the AAPM has created a lexicon and some standard imaging protocols for diagnostic applications.[20] Guidelines should also be given for minimal requirements for spatial resolution (in-plane and slice thickness) of the acquired

image data. PET and MRS have an inherent spatial resolution less than that of CT or MRI. In these cases, resolution limits should be specified to preserve consistency of data among trial investigators. AAPM TG-174 addresses standardizing FDG-PET protocols, including patient preparation and image acquisition parameters, in order to improve SUV consistency. These are aspects that impact inter-institution translatability. Modalities such as PET and MRS may require supplemental credentialing using benchmark datasets by a QA center to ensure that the department's systems for contouring are capable of representing that data adequately in support of the clinical trial. There may be significant discrepancies between the volumes defined by images from different imaging modalities, which should be addressed by the designers of clinical trials working with imaging experts.

5.1 Imaging Modalities

CT continues to be an integral part of radiation therapy planning for clinical trials for volume definition and treatment planning. Several additional imaging techniques are also used in clinical trials, such as MRI and magnetic resonance spectroscopy (MRS), PET, and single photon emission computed tomography (SPECT). Innovation continues in these and other areas regarding the use of imaging in clinical trials for target definition and the assessment of treatment response. Therefore, collaboration at individual institutions is essential between specialists in radiation oncology, radiology, and nuclear medicine, as appropriate.

MRI and MRS play a crucial and continually expanding role in clinical trials. MR may be used to improve the accuracy of both targets and organs-at-risk. Spatial and volumetric accuracy are important in order to use the information quantitatively. Imaging sequences should be provided to participating institutions by imaging experts who aid in the design of the clinical trial. Also, MR and other modalities aid in assessing tumor response to score outcomes. Following protocols in order to have high-quality acquisition of these scans is relevant to radiotherapy physicists for the success of the trial. High quality is needed for any additional scoring required for the protocol. The manufacturer-specific differences in hardware and software affect the equipment capabilities and need to be taken into consideration.

PET has been used both to determine the extent of disease for primary and nodal targets and to assess tumor response to treatment. PET and other modalities may also be used to determine tumor stage or a patient's trial eligibility. The primary PET tracer in use is 2-[18F] fluoro-2-deoxy-d-glucose (FDG), a reporter analogue of glucose that indicates hyper-metabolic regions and may indicate active tumor tissue. PET may also provide information on vascularization, cell proliferation, apoptosis, oxygenation, and receptor expression.[21–23] FDG is relatively non-specific for cancer and, thus, image interpretation can be confounded by numerous issues, such as local inflammation and tissue repair.

PET/CT has greatly improved the accuracy of interpretation of the PET image, as the tracer distribution can be overlaid onto the anatomy in the CT image. The primary sources of uncertainties for PET/CT are in variations in patient biology related to the uptake of FDG, the scanning protocol, the acquisition (such as patient positioning or patient motion during scanning), reconstruction variability, analysis, and the impact of registration errors and motion on target delineation, as shown in Table 2.[18] The long scan times encountered in PET compared to CT can lead to significant differences between the PET and CT images on the fused PET/CT images. When the patient is in the treatment position with immobilization equipment or positioned similarly, uncertainties are reduced in image registration for target and normal tissue definition.[24]

To determine the appropriate imaging, the recommendations of Image Wisely (www.imagewisely.org) and Image Gently (www.imagegently.org) can be followed to minimize dose due to imaging, including for pediatric patients. Imaging protocols continue to be developed and shared via those efforts.

Table 2. Uncertainties and Quality Control Measures for PET/CT in Radiation Therapy Planning (Reproduced from Jeraj et al.[18])

Category	Procedure	Uncertainties	Quality Control
Scanning protocol	Patient preparation	Metabolism levels (^{18}F-FDG)	Limit physical activity
		Blood glucose levels (^{18}F-FDG)	Measure fasting blood glucose with exclusion criteria
		Bowel size/positioning	Use fasting protocol
	Radiotracer injection	Residual activity in syringe	Measure/correct for residual activity
		Decay correction errors	Synchronize scanner clock
Acquisition	Patient positioning	Spatial offset between PET and treatment planning CT	Ensure consistent patient positioning using identical positioning devices
		Quantitative uncertainties from attenuating objects	Avoid placing objects outside image field of view
	Scanning	Patient motion	Implement motion management strategies
		Attenuation correction uncertainties from iodine contrast material	Acquire separate low-dose CT scan or apply corrections
		Equipment failure or electronic drift	Calibrate detector and equipment frequently
		Increased SUV because of longer uptake period	Apply strict protocol for uptake period
Reconstruction	Reconstruction	Selection of optimal image reconstruction method/parameters	Benchmark algorithms using phantoms (task-specific)
		Randoms, scatter, attenuation, detector sensitivity, and partial-volume effect	Apply appropriate calibrations and corrections
Analysis	Segmentation	Differentiation of normal tissue and tumor uptake	Know radiotracer's normal biodistribution
		Segmentation uncertainties	Develop segmentation protocol; benchmark algorithms with phantoms
		Limited spatial resolution and sensitivity	Include margins
	Quantification	Quantitative accuracy	Calibrate PET scanner to dose calibrator
		Selection of relevant quantitative measures	Compare semiquantitative metrics with kinetic analysis–derived parameters; consult literature
		Quantitative differences between scanners/institutions	Quantitatively harmonize scanners
Treatment planning	Target definition	Registration errors	Benchmark algorithms using physical or digital phantoms; crop images
		Motion	Use same motion management method as was used during imaging

5.2 Recommendations Regarding Image Acquisition

5.2.1 Recommendations for Clinical Trial Designers

1. Determine if imaging-specific credentialing is required through a review by imaging experts (such as the imaging organizations within IROC) and whether or not variability in techniques or variations in commercial scanner technology need to be considered.

2. Design a standard operating procedure for imaging, incorporating expertise of imaging physicists/scientists where appropriate.
    a. Specify the extent of anatomy to be imaged, including whole organs when required for dose volume analyses.
    b. Specify any timing requirements of the acquisition in relation to treatment start for all imaging data for treatment planning and assessment. Be explicit regarding patient preparations for imaging.
    c. Keep image acquisition, reconstruction, and analysis procedures consistent when multiple imaging sessions for a patient are required.
    d. Ensure consistent patient setup and immobilization between different imaging modalities and treatment (see sections 5.2 and 8.1) through credentialing of multi-modality image registration.
    e. Specify which contrast agents are permitted and provide details on the timing and amount of the agent to be used.
    f. Provide guidelines on basic imaging parameters for trials permitting different modalities, such as MRI, MRS, or PET/CT, to account for the variability of different scanners.
    g. Develop imaging benchmarks when modalities such as PET and MRS are used to ensure that the department's systems for contouring are capable of representing that data adequately in support of the clinical trial.

5.2.2 Recommendations for Physicists at the Local Institution

1. Train and work with the appropriate personnel to implement the protocol-specified imaging standard operating procedures for image acquisition, reconstruction, processing, and analysis.
2. Review patient imaging scans regularly to ensure compliance to the standard operating procedure.
3. Consider utilization of immobilization and setup methods and devices that are compatible with all imaging modalities used in the trial to reproduce the setup for the treatment planning CT.

5.2.3 Recommendation for QA Centers

Specify if an existing imaging benchmark would be beneficial for ensuring that enrolling institutions would be able to acquire scans of the appropriate quality to support the trial.

6. Image Registration in Clinical Trials

6.1 Considerations Regarding Image Registration

Clinical studies that require multiple image datasets need to use image registration software. When multiple image modalities are used for treatment planning, the protocol designers should consider providing specific recommendations for internal or external landmarks that can validate the adequacy of the registration for treatment planning.

If the accuracy of the image registration for each patient affects the quality of the trial (such as in defining the target volume), the protocol designers and QA centers should require credentialing of the image registration software by using phantoms of known geometry and should follow the guidance of AAPM TG-132.[24] The physician directive should specify the goals of the image registration, the method, and what anatomical region was emphasized in the registration.[24]

Credentialing may require an image registration benchmark. For on-line imaging modalities at the treatment unit, protocol designers should specify how the allowable image registration techniques—

such as the approved techniques and the alignment—will be judged.[24] For any applications of image registration in a trial, the protocol designers should specify which methods are allowed (rigid only, deformable) and any additional constraints. The guidance of AAPM TG-132 should be followed, and the trials designers should determine if it is necessary to distinguish between applications for target and normal tissue definition compared to daily on-line treatment guidance. Image registration considerations may also differ if there is a mid-course plan adaptation and dose accumulation methods are utilized.[25]

6.2 Recommendations Regarding Image Registration in Clinical Trials

6.2.1 Recommendations for Clinical Trial Designers

1. For any applications of image registration in a trial, the protocol designers should specify which methods are allowed (rigid only, deformable), and any additional constraints.
2. Guidance should be provided about how the quality of an image registration is judged which should distinguish between applications for target and normal tissue definition compared to daily on-line treatment guidance. This information should be considered when image registrations are evaluated as part of credentialing for a given trial.

6.2.2 Recommendations for Physicists at the Local Institution

1. Evaluate the ability of the institution to follow protocol guidelines for segmentation and image registration.
2. Follow recommendations of AAPM TG-132 with respect to image registration.[24]
3. Adjust monitors for adequate resolution and properly calibrate for contrast and brightness to ensure consistency in target delineation.[26] Note minimum settings in the standard operating procedure.

6.2.3 Recommendations for QA Centers

1. Develop imaging benchmarks as needed, including when modalities such as PET and MRS are used, to ensure that the department's systems for contouring are capable of representing that data adequately in support of the clinical trial.
2. Develop credentialing methods incorporating deformable image registration following the recommendations of AAPM TG-132.[24]

6.2.4 Recommendations for Manufacturers

1. For a given registration, develop methods to extract the primary goals of the image registration (e.g., target evaluation or organ-at-risk, such as from a structured field) and the goodness of the registration (see TG-132 recommendations).[24]
2. In image registration software, provide the ability to export necessary data for QA centers to be able to assess the quality of a registration (quantitative and qualitative) and export the needed information for straightforward review by those credentialing for clinical trials and investigators for patients enrolled in clinical trials.

7. Motion Assessment and Management

For many treatment sites, physiological motion must be assessed to determine if management of that motion is necessary for segmentation and treatment delivery. The AAPM Task Group 76 report, published in 2006, provides guidance for considerations at simulation and for treatment planning.[27] Since that time there have been technological changes and new knowledge has been gained on successful

assessment and management of motion. Therefore, efforts are underway to update that report with guidance needed today for clinic care and clinical trials. In 2017, several members of the Medical Physics Committee of NRG Oncology reviewed guidance in the context of stereotactic body radiation therapy for thoracic and upper abdominal tumors.[28] The group addressed the major categories of target definition, imaging studies to measure and evaluate motion, breathing consistency, different methods of addressing motion, planning considerations, and, finally, use of imaging for localizing the target position.[28] While the emphasis of their report is on respiratory motion, the technological considerations apply to the assessment and management of other physiologic motions and are relevant for many clinical trials, not only NRG SBRT trials.

7.1 Motion Assessment

Motion should be assessed at the time of simulation to determine if it will impact the accuracy of treatment delivery.[27] Imaging methods for motion assessment of targets and organs at risk include the use of 4DCT, breath hold scans at inhale and exhale, fluoroscopy, and other methods.[27,28] For example, 4DCT scans can be used to create the internal target volume (ITV) or internal gross tumor volume (IGTV). Respiratory-correlated PET/CT may provide additional advantages for radiation therapy trials[18,29] but the technology may only be available at a limited number of centers. Variations between 3D and 4D PET/CT scans have been assessed in a number of studies, with 4D providing enhanced information about target motion and having an impact on the standard uptake values.[30]

Regardless of the technique used for the motion assessment, the target motion should be documented with specific actions to take depending on the amount of motion, the technology in the clinic for motion management, and patient-specific factors (such as ability to use a breath hold method). For treatment sites where the impact of motion may be crucial[27], it is recommended that the QA centers develop guidance that can be easily followed by participating centers. Trial designers should recommend the acceptable techniques for motion assessment when they may have an impact on the quality of the trials. They may also specify when a particular type of technique should be used, such as when motion exceeds a given action level. This information should be reported along with the treatment plan for patients enrolled on a specific trial.

7.2 Motion Management

Some clinical trials will require motion assessment and management for patients. As noted above, Brandner et al. report on different techniques used for motion management specific to respiratory motion.[28] The techniques range from gating based on imaging or other patient-specific information, to the use of breath hold techniques. AAPM TG-76 recommends the use of motion management when motion exceeds 0.5 cm as long as a center has a method in place and the patient can tolerate that method. When devices are not available or a patient cannot tolerate using a given system, larger planning margins should be used.

The reproducibility of that motion should be assessed and documented when patients are freely breathing or when a motion management technique is used. For clinical trial designers, a determination should be made regarding whether or not the reproducibility should be reported for patients enrolled on a specific trial.

Another critical component of accurate targeting is the use of motion management methods to account for physiological motions during treatment. Motion management techniques must be investigated for situations where the motion of a target is large for two situations: (1) range of motion is greater than 0.5 cm, or (2) significant normal tissue sparing can be achieved.[27] Protocol designers for clinical trials should specify the proper margins for the trial given the type of immobilization equipment, motion management, localization techniques permitted, and the data that should be submitted to the QA center. Some trials that require motion management may also require passage of credentialing tests related to that functionality.

7.3 Recommendations for Motion Assessment and Management

7.3.1 Recommendations for Clinical Trial Designers

1. For relevant body sites, specify that the degree of target motion should be assessed at the time of simulation. For treatment sites where the impact of motion can be crucial, it is recommended that QA centers develop guidance with respect to the acceptable imaging techniques to assess motion, documentation of that motion for a given patient, and how the information should be incorporated for creating target volumes.
2. Incorporate guidance on motion management techniques when the range of motion is greater than published limits (or significant normal tissue sparing can be achieved through their use). For trials when target motion may be ≥5 mm and delivery of a high daily dose (e.g., SBRT), institutions should be required to document the assessment and follow formal guidance, such as that provided by AAPM TG-76[27] or other organizations such as NRG[28] to ensure motion assessment and management information is accurately captured for patients enrolled on the trial.
3. For protocols involving monitoring of intra-fraction motion, provide information regarding the acceptable technologies for monitoring and the thresholds for evaluation. Information should be provided as to whether intra-fraction monitoring is required and the acceptable methods.

7.3.2 Recommendations for Physicists at the Local Institution

1. Confirm that the motion assessment and management guidance specified in the protocol is followed whenever the range of motion meets published guidance limits.
2. Ensure that the contoured IGTV is reasonable considering the measured motion for a given protocol patient.

7.3.3 Recommendation for QA Centers

Determine if a motion benchmark is required and if so, if an existing benchmark will meet the needs of a given trial.

7.3.4 Recommendations for Manufacturers

1. Provide on-line 4D tools such as 4D CBCT capability at the treatment machine to support protocol motion management requirements.
2. Provide tools to document and export the range of motion data for different imaging platforms.

8. Patient Immobilization, Target Definition, and Treatment Guidance

Patient and target positioning is affected by immobilization and the frequency and type of image guidance used at the treatment unit. The accuracy of immobilization and patient positioning affects the size of the treatment margins. Variability in image registration for trials involving multi-modality imaging may also have an impact on the target definition for planning and subsequent analyses involving an assessment of dose response.

8.1 Immobilization

In the context of clinical trials, the type of recommended immobilization described or required in a particular trial depends on (1) the available and acceptable equipment in potentially accruing clinics, (2) the accuracy required by the protocol, and (3) the frequency and accuracy of the treatment guidance methods that may be recommended during patient treatment. Trial designers should determine if a given trial requires specific immobilization, such as for stereotactic radiosurgery or stereotactic body

radiation therapy. Patient immobilization is relevant during assessment of treatment response with imaging: correlation to previous scans (single- or multiple-scan sessions) is essential for accurate interpretation.

Immobilization equipment and methods should be used consistently. The type of equipment used depends on the body site and required degree of reproducibility. Some immobilization systems can position patients repeatedly within one to three millimeters (e.g., brain, head and neck), whereas larger planning margins (≥3 mm) may be required for targets located in other parts of the body (e.g., breast and abdomen). The required reproducibility and tumor location need to be considered in combination with the image guidance strategy for treatment. Protocol authors should check the demonstrated accuracy of commercial immobilization systems in the literature and include realistic requirements in the guidelines: these guidelines need to be appropriate for the goals of the protocol and achievable by the participating institutions.

The physicist in each department should participate at an early stage of review (before the institutional review board, or IRB) has approved the protocol prior to opening a protocol. Then the physicist can determine if the department will need to purchase any additional immobilization equipment to support the protocol. The proposed immobilization device should be determined prior to the time of CT-simulation and other imaging scans. The physicist should confirm that the immobilization equipment works correctly, meets the desired or required accuracy for the trial, and is comfortable and safe for the patient. Table tops for simulation equipment (CT and MR) that are flat and have indexing permit fixation of the immobilization device to the table top in a specific location for imaging and all treatments. This consistent location of the immobilization device aids in the reproducibility of the treatment setup and should be encouraged in the protocol guidelines for treatment and imaging. Also, it should be determined if fiducials, such as linear gold coils or multi-modality radiological markers, will be used to aid in the registration of image datasets.[24]

8.2 Target Definition and Treatment Guidance

Clinical trials should specify the criteria used to define the target volumes and corresponding margins. The concepts for defining treatment margins in radiotherapy are described by the International Commission on Radiation Units and Measurements' reports #50, #62, and #83.[32–34] For clinical trials, trial designers should be specific on the acceptable methods for target definition of protocol patients so that there is consistency throughout the trial.

Protocols should be specific with respect to the type and frequency of image guidance. The relationship between localization methods and the appropriate PTV margin[35] should be considered in the design of all clinical trials. For example, a trial involving treatment of breast cancer may involve weekly portal imaging, whereas a trial involving SBRT may require daily volumetric imaging. The designers of clinical trials should be specific with respect to the recommendations for intra- and inter-treatment margins in a given trial for consistency and reproducibility.

The doses delivered to patients from imaging are affected by both the imaging techniques used and the frequency of imaging. This must be balanced with the value of improved target localization. Physicists in each clinic should monitor setup reliability for a given protocol as data are generated to ensure compliance with the protocol or to be able to make adjustments as soon as possible.

8.3 Recommendations for Patient Immobilization, Target Definition, and Treatment Guidance

8.3.1 Recommendations for Clinical Trial Designers

1. The clinical trial design should survey the literature, including relevant AAPM task group reports, to determine the type of immobilization suitable to meet aims of the clinical trial.

2. Consult with physicist(s) at a lead institution and other possible participating institutions to ensure that the proposed accuracy limits are achievable at a number of centers.
3. Clearly specify which immobilization equipment is required for the trial (where a preliminary assessment of equipment availability in the community could be done via the IROC Houston facility questionnaire if needed) or if certain types of equipment are not permitted.
4. Use the most up-to-date terminology to specify definitions of target volumes in the trial design (e.g., ICRU #83 at time of publication).
5. Review data in the literature to define acceptable PTV margins related to the technology used for simulation (such as 4DCT) and the frequency and type of imaging for the anatomical site.
6. Provide explicit guidance on the contouring of targets and necessary expansions.
7. If a protocol requires an evaluation of target margins mid-treatment, the clinical trial designers should specify the frequency and methods of evaluation in the clinical trial design. For example, how to address changes in tumor physiology or shape, such as changes to targets in the lung or head and neck region due to shrinkage or growth of the tumor.

8.3.2 Recommendations for Physicists at the Local Institution

1. Determine that the institution's immobilization equipment is appropriate for the clinical trial before IRB submission.
2. Ensure consistency of equipment for planning and treatment, e.g., flat table tops for diagnostic scanners and the use of compatible immobilization equipment for imaging scans, when possible.
3. Confirm the accuracy of the immobilization method used in the clinic for the protocol.
4. Ensure personnel are adequately trained to support the process.
5. For each protocol, understand how target margins are specified and make sure the margins are reasonable for the department's imaging, immobilization, planning, delivery, and treatment guidance process for the patients enrolled on the trial.
6. For each protocol, monitor the effectiveness of the patient localization method for the patients enrolled on the trial.

8.3.3 Recommendations for QA Centers

1. Confirm that the precision of commercial immobilization systems and field experiences indicate that the proposed techniques realistically can meet the accuracy requested in the protocol.
2. Ensure the appropriateness of the margin for a given trial.
3. Determine credentialing methods for new techniques, such as those requiring intra-fraction monitoring.

8.3.4 Recommendations for Manufacturers

1. Make immobilization devices that enhance reproducibility of patient setup over time so serial images can be used for quantitative treatment assessment and subsequent treatment planning.
2. Incorporate interchangeable fiducials in the immobilization devices to facilitate merging the scans from two or more types of instruments, such as MRI, CT, and PET.
3. Develop tools to quantitatively review localization images with field outline and anatomy contours exported from the treatment management system.

4. Develop tools to quantitatively monitor daily setup correction trends for patient positioning, such as from on-board imaging or other methods.

9. Segmentation and Recommendations for Clinical Trial Designers

9.1 Segmentation

Important technical sources of variation in segmentation include variable window and level settings, the use and sensitivity of auto-segmentation algorithms to input parameters, and inappropriate margin expansion algorithms. For example, inappropriate window and level parameters can lead to significant bias and errors in volume definition. One study found variations up to 42%, which were then reduced by using a standard protocol.[36] AAPM TG-211 is developing recommendations on applications and limitations of auto-segmentation for contouring. The consistency of contours created on image datasets can be improved by ensuring that workstation monitors are properly calibrated with respect to visualization (e.g., brightness and contrast).[26] Also, physicians can create contours from single versus multiple imaging modalities, but significant volume or boundary differences may arise.[37–39] Therefore, guides for investigators should be available, either as appendices to protocols or in more general atlases.[40–43] Studies have demonstrated the potential for large variability in volumes defined by different physicians when experts use the same image datasets for contouring, with many groups demonstrating improved consistency of contouring or segmentation when atlases are created and used.[43–47] Improvements in the consistency of contours are seen when pretreatment reviews of contoured structures are performed by protocol principal investigators. Training, such as via workshops or webinars, should be provided to physicians and other personnel for a given trial if there could be significant variability in the delineation of structures.

For organs which will be evaluated with dose volume histograms (DVHs), the protocol should specify how much of the organ must be contoured. For example, it may be appropriate to specify a region of spinal cord to be contoured with respect to the superior and inferior borders of the PTV. Structures with mean dose objectives should be contoured in their entirety. For structures where the entirety may not be included within the planning scan, the protocol should specify dose limits in absolute volume (cc) instead of relative volume (%).

Clinical trial designers should specify the preferred methods for handling artifacts on CT simulation data, such as those due to metal hip implants or dental artifacts. Methods include the use of metal artifact reduction software at simulation as well as simple methods, such as contouring of artifact regions and assigning a reasonable value for the density.

9.2 Recommendations for Clinical Trial Designers

1. Specify window and level values, when appropriate, for consistent visualization and segmentation.
2. Refer investigators to published consensus atlases for target and organ at risk delineation as a reference when appropriate.
3. Provide training to physicians for a given trial if there could be significant variability in the delineation of structures among physicians.
4. Provide guidelines to physicians, physicists, and dosimetrists on how to address imaging artifacts that interfere with target or normal tissue segmentation (e.g., scatter from metal or the presence of contrast on a CT simulation scan).
5. For organs which will be evaluated with DVHs, the protocol should specify how much of the organ must be contoured for structures such as the spinal cord.

It is crucial that protocol designers provide explicit guidance in how structures are defined, especially when multiple structures are involved. Significant differences have been shown in dosimetric parameters for lung cancer for different definitions of normal lung, the gross tumor volume, clinical target volume, internal target, or planning target volume.[48] Variability of such definitions in a clinical trial would have a significant detrimental impact on the ability of the trial to resolve the study question. It may also lead to inconsistency in the application of dose goals if the same dose goals are used but with different definitions from one trial to another. Therefore, definitions and dose goals across trials to the same body site should be standardized as much as possible with the expectation of evolution of care over time. Additionally, the protocol should specify any additional limits to doses to organs outside the treatment field.[49] A final critical concern is that some systems ignore the volume of an organ outside the dose calculation grid when reporting dose-volume parameters. For such systems, the dose-grid should cover the entire organ of interest so that derived dose volume parameters used for treatment planning represent the entire organ.

10. Treatment Planning Considerations

This section provides guidance on aspects of TPSs for external beam therapy that affect the consistency of clinical trial data among institutions. Considerations regarding planning in clinical trials are included.

There are numerous publications emphasizing the improved dose calculation accuracy of model-based algorithms over pencil beam algorithms.[50,51] Yet, this continues to be a challenge in some situations, such as trials where the primary trial question may be focused more on a pharmaceutical rather than radiation therapy or trials with simple delivery techniques. More accurate model-based algorithms, rather than pencil beam algorithms, should be used for planning for patients in clinical trials. A consideration for the model-based algorithms is that they include many approximations for modeling IMRT beams that have been found to affect dosimetric benchmarks.[8,52] The use of many overlapping small fields for IMRT, including VMAT, can increase the monitor units for the field by a factor of 2.5 to 3 or more. Most current protocols require dose heterogeneity calculations in which CT Hounsfield units are converted to electron density.

With respect to clinical trial requirements, the physicist should understand how the TPS handles overlapping structures: for example, whether the intersection of two or more overlapping structures belongs to one structure or to two distinct structures. This interpretation can significantly affect the resulting dose volume histograms (DVHs), particularly for small structures. Clinical protocol designers should specify that investigators and QA centers provide guidelines for accurately merging overlapping structures and delineating distinct, but close, structures.

The extent of the dose grid should be specified by the protocol designers. High resolution is required to accurately represent the dose distribution in regions of high dose gradients. For example, a spatial resolution of 0.2 cm or less is desirable and feasible for cranial radiosurgery treatments.[53] For most extra-cranial protocols, a 0.25 cm dose grid may be appropriate.[53] Clinical trial designers should indicate when a finer grid is required. When optimizing IMRT or VMAT plans, it may be necessary to specify a resolution in the beamlet grid or the number of degrees between dose calculations for arc delivery methods.

The physicist at each institution should work with the lead physician and other personnel to create simulation and planning directives that are specific to the clinical protocol. These documents can be used to ensure that the proper operating procedures are followed and that members of the department's treatment team have the information they need to ensure creation of a treatment plan that is compliant with the protocol.

Protocol designers and manufacturers should consider creating templates and tools that can be used to support the uniform implementation of clinical trial guidelines. CIRO creates templates for NCTN trials (https://www.nrgoncology.org/Scientific-Program/Center-for-Innovation-in-Radiation-Oncology). Templates may include information such as the preferred beam arrangement, energy considerations, and optimization suggestions if a protocol may involve use of IMRT/VMAT. These tools may include structure templates that work on multiple vendor platforms (such as following the nomenclature recommendations of AAPM TG-263), specification of optimization criteria customized to a given protocol, and advanced planning tools that aid in meeting the dosimetric requirements of a protocol. For example, a dosimetric model could be developed for knowledge-based planning or a script could be created with standard input—such as the beam energy, beam arrangement, and modality—to best meet a given protocol. Tools or scripts that can be shared and then used at the local institution to assess protocol compliance are invaluable.

10.1 Treatment Plan Evaluation Tools

The DVH is used to assess the quality of a treatment plan by computing dose and volume statistics for a specific anatomical structure or region of interest. DVH results have been shown to depend on the TPS[54], and sources of dosimetric uncertainties have been assessed[55]. Whiles 3D doses are not reconstructed, DVHs are recalculated at QA centers based on the submitted DICOM-RT data. This ensures uniform assessment of any secondary analyses of target volume and normal tissue doses, and compliance with the protocol. The calculated dose metrics vary for small structures as a function of the calculation grid size, bin size, and the volume of interest.

When appropriate, Clinical trial designers should specify the dose and volume limits for targets and organs at risk within a given protocol. AAPM TG-263 will provide guidance on how DVH limits are specified, especially for overlapping structures. Consensus efforts such as QUANTEC[56], which recommend normal tissue limits, are excellent resources for protocol designers.

The DVH is also the basis of computing tumor control probabilities and normal tissue complication probabilities.[57] Specifications will be needed by protocol designers for tools such as normal tissue complication probability (NTCP) and equivalent uniform dose (EUD) as they are incorporated into multi-institutional clinical trials.[58] Since these models may not be in widespread use, additional credentialing requirements will be necessary and will need to be supported by QA centers considering guidance from AAPM TG-166[58].

10.2 Tools to Improve Consistency between Planned and Delivered Dose

Clinical trial designers and protocol guideline writers should assess and specify irradiation delivery technologies with well-enough established QA practices to permit their use in the trial. This decision in part considers whether radiation therapy is part of the trial question and whether sufficient credentialing tests exist to assure consistency of practice. Tools are being maintained by the NRG Oncology Center for Innovation in Radiation Oncology that may be beneficial for trials by other organizations (see https://www.nrgoncology.org/Scientific-Program/Center-for-Innovation-in-Radiation-Oncology). Credentialing is appropriate, along with monitoring the use of techniques that harbor potential for variation in the implementation in clinical trials. For example, the NCI has published guidelines on the use of IMRT that required credentialing for patients treated with IMRT on NCI-sponsored clinical trials.[31]

New devices and techniques often offer so much promise for improving care that the use of such devices is incorporated into clinical trials concurrently or in advance of a consensus on standards for commissioning and quality assurance of these new techniques. As AAPM task group reports are published on QA of various new delivery devices, standardization of QA practices is expected to increase over time. When guidance is lacking, the clinical trial designers should work with the appropriate

quality assurance center to ensure that the dosimetric delivery is adequately tested. The QA requirements will depend on the technology used for delivery.

Additional factors can influence the agreement between the planned and delivered dose. The surface dose can change with the type of delivery technique and immobilization material used. For example, the maximum surface dose for head and neck cancer varied from 82% for IMRT to 71% for tomotherapy to 69% for bilateral fields.[59] An early IMRT study by Lee et al. (2002) found that the surface dose for head and neck cancer increased from 84% with opposed lateral fields to 100% with an IMRT technique.[60] Many clinical trials that involve volumes which may be near the surface specify that contours should be at least 3–5 mm from the external surface to minimize the impact of uncertainties in the dose calculation. This restriction could be reinvestigated for the treatment sites to provide that appropriate guidance is provided. The protocol designers should consider how to best collect the dose data in these regions, while taking into account the limitations in the dose calculation accuracy.

Tabletops and immobilization devices have been demonstrated to impact the delivered dose to patients.[61] The skin dose can be a concern due to the buildup effect of the couchtop, as well as due to attenuation by those devices, affecting the delivered target dose.

10.3 Automated Planning Tools

Advances are being made in the use of automated tools for planning and for assessing the consistency of a treatment plan with respect to previous trials. This development has important implications for clinical trials, both for secondary analyses and for more robustly assessing plan quality during the accrual phase of a trial. For advanced technology trials, there may be a learning curve for new institutions with respect to planning, especially when creating new plans over the course of treatment. The ability to improve plan quality using knowledge-based methods was evaluated for RTOG 0126 where predictive DVHs showed that further sparing of normal tissues was achievable with a group of plans (Figure 3).[10] Figure 3e demonstrates that plans which were defined as “low-quality” had significant improvements with respect to the predicted rectal toxicity based on the calculated normal tissue complication probability values for each plan. Such tools will be valuable both for the teams at the institution performing treatment planning for protocol patients, as well as for the analysis of plan quality at the QA centers (https://www.nrgoncology.org/Scientific-Program/Center-for-Innovation-in-Radiation-Oncology).

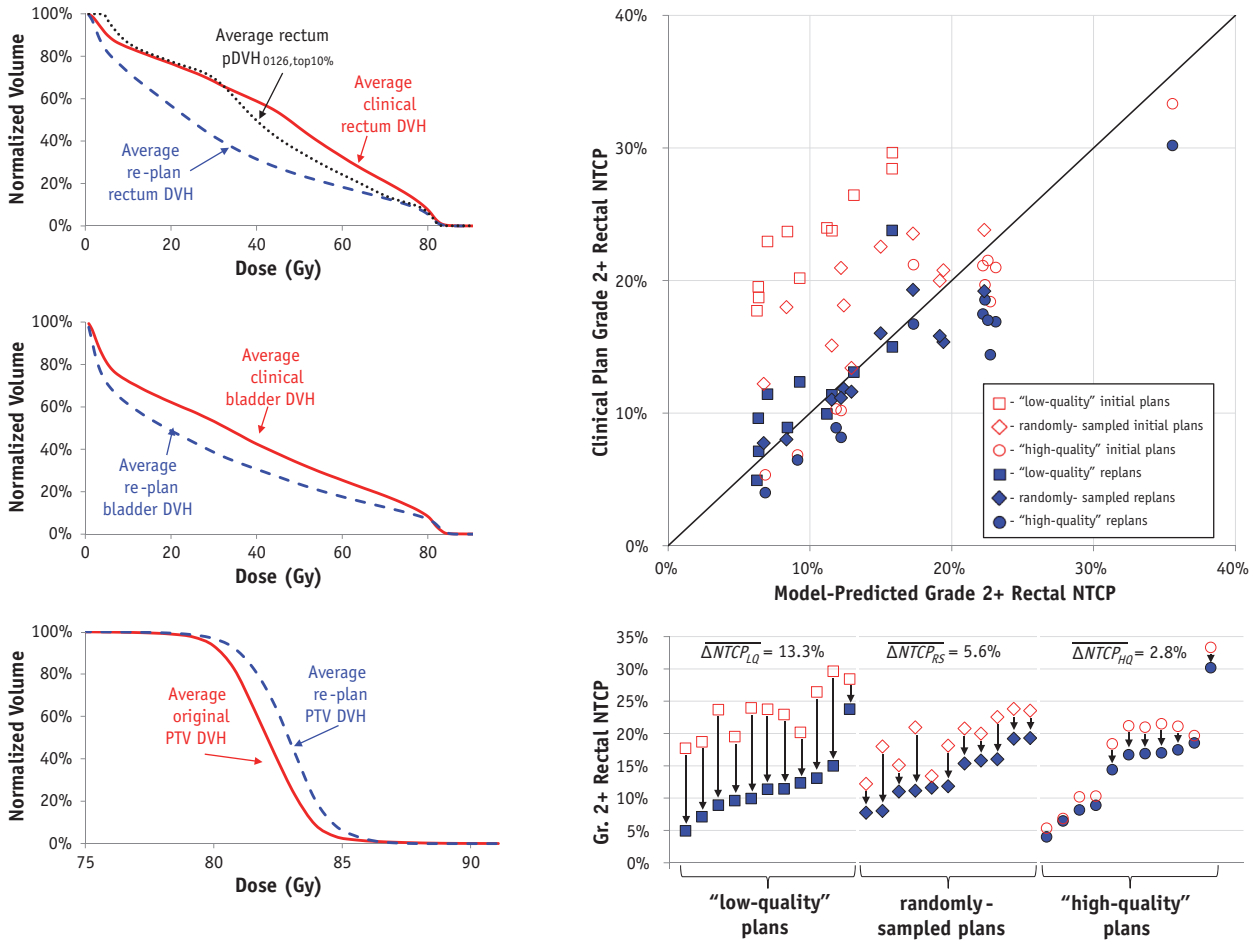

Figure 3. Moore et al. retrospectively evaluated the impact of knowledge-based methods, such as calculation of a predicted DVH (pDVH), on the overall IMRT plan quality for RTOG 0126 for prostate cancer and the resulting predicted grade 2 rectal normal tissue complication probability (NTCP). (Figure courtesy of Moore et al.[10].)

10.4 Compatibility with Adaptive Therapy and Re-irradiation

Emerging new technologies in radiation treatment planning and image guidance will place additional requirements on the capabilities of the TPS. Investigators and manufacturers are developing tools to better support adaptive therapy, such as deformable image registration and the creation of a model based on the accumulated dose to a patient.[62] Many of these considerations are beneficial for patients who are retreated, which may also be a component of a clinical trial. Adaptive therapy requires the ability to replan based on dose calculated using images at various time points during the course of therapy. It should be possible to robustly combine multiple fractional doses given with the same or different treatment plans into a composite delivered dose presented on a single CT dataset that is a reasonable representation of the treatment delivered over the course of therapy. Deformable registration and fusion algorithms are currently being investigated and should ultimately be included in the software tool set available at individual institutions and at QA centers. These algorithms are an integral part of accurately assessing and reporting the dose given to the patient throughout the course of therapy. To fully appreciate the impact of anatomical changes for case review in a clinical trial, the com-

posite delivered dose would be best, but if not available, multiple imaging studies, their time sequences, and all treatment plans should be submitted to the QA center.

10.5 Additional Requirements

Treatment planning systems should be capable of submitting anonymized data for dry run benchmark tests and for protocol patient data. The anonymization should include fields such as the patient name, date of birth, contact information, and registration number. Several tools are freely available if this functionality is not available from the TPS manufacturer. For those participating in NRG Oncology trials, anonymization tools are available in the TRIAD system, which is valuable for institutions with planning systems that do not yet support anonymization.

Plan data should be submitted via the Digital Imaging and Communications in Medicine (DICOM)-RT format.[63] The minimum exported data should include the planning CT, the contours of the required volumes, the 3D dose matrix, and its associated plan file.

All TPS manufacturers should comply with the standards developed as part of Integrating the Healthcare Enterprise in Radiation Oncology (IHE-RO). IHE-RO is a consortium of health professionals and manufacturers that ensures uniform interpretation of data communication standards by all vendors to improve interconnectivity, interoperability of radiotherapy planning, and delivery systems.[64]

10.6 Recommendations Regarding Treatment Planning

10.6.1 Recommendations for Clinical Trial Designers

1. Specify standard structure names that must be used for the clinical trial (follow consensus guidance when available) such as provided by AAPM TG-263 or other appropriate ontologies.
2. Use published information on normal tissue limits, such as through consensus efforts as appropriate, when specifying the limits to normal tissues.
3. For organs which will be evaluated with DVHs, the protocol should specify how much of the organ must be contoured for structures such as the spinal cord.
4. Specify spatial resolution requirements for dose and DVH calculations that are commensurate with target and organ-at-risk (OAR) sizes.
5. Specify the use of 3D treatment planning for all clinical trials (possibly excluding special procedures, such as total body irradiation or total skin electron treatments).
6. Require the use of more accurate algorithms (such as convolution/superposition, Monte Carlo) for trials where tissue heterogeneities may be significant.
7. Develop credentialing approaches for new applications of the TPS, such as biological treatment planning[58]. Credentialing may include intercomparison of results using standardized datasets.
8. Specify the dose-volume constraints for organs-at-risk and consider any special concerns, such as the buildup region or structures outside the treatment area.
9. Specify the minimization of the integral dose or total dose to other normal tissues that may not be contoured in trials which allow the use of dose optimization techniques.

10.6.2 Recommendation for Physicists Involved in Clinical Trials

1. Ensure that the TPS is capable of meeting protocol requirements by:
    a. Use of a model-based algorithm such as convolution/superposition, Monte Carlo, or deterministic methods.

   b. Accurate modeling of beams and output factors, especially for small fields and IMRT techniques.
   c. Validating the dose-volume histogram and analysis algorithms.
2. Ensure that 3D volumetric information can be exported to the Clinical Trial QA Center in DICOM-RT format.
3. Implement templates in your treatment planning system to use the standard names for targets and structures as specified by the clinical trial designers.
4. Coordinate an end-to-end dry run of the protocol at his or her center on one of their patient's dataset(s). (Note that this requires support from the department's administration for this valuable effort.)
5. Determine the degree of attenuation by immobilization equipment, and determine whether the attenuation should be accounted for in monitor units (MU) calculations.

10.6.3 Recommendations for QA Centers

1. Enable as much automation of data submission as possible.
2. Continue validation and cross-comparison of the performance of different dose algorithms with other QA centers, and revise requirements as appropriate.
3. Work with manufacturers to design interfaces that can be customized for electronic submission of all necessary protocol data.
4. Provide the clinical trial groups with a template of standard target and structure names across clinical protocols consistent with the nomenclature of AAPM TG-263.
5. Develop mechanisms to share scripts or other tools (such as Excel spreadsheets with macros enabled) to aid the institutional teams in assessing whether or not protocol guidelines are met prior to submission to the QA center. Tools could potentially be developed on multiple TPS platforms.

10.6.4 Recommendations for Manufacturers

It is recommended that manufacturers provide tools that support clinical trials locally and nationally, such as:

1. Include DICOM-RT export in the base purchase of a TPS rather than an add-on option with the ability to export coded ID cases to the QA centers (including image datasets, plans, structures, and dose).
2. Provide standard target and structure names as provided by the QA centers or allow upload of files with the names of the structures (as defined in AAPM TG-263).
3. Enable the use of protocol-specific scripts, including standard target and structure names (AAPM TG-263).
4. Create interfaces that import the necessary standard names, beam arrangement (if appropriate), and other information for treatment planning.
5. Create the appropriate software to allow automatic anonymization with coded ID labels of patients and plans.
6. Develop and make available a straightforward export of information to QA centers.
7. Make treatment planning systems IHE-RO compliant.

8. Enable the creation of tools or scripts that can be shared and then used at the local institution to assess protocol compliance are invaluable.

11. Treatment Delivery Documentation

Treatment management systems permit verification that the correct energy, beam modifiers, monitor units, treatment dates, and number of fractions were used for individual patient treatments. A summary of this information should be exportable in a standard format for a clinical trial. Patients may have poorer outcomes as a result of missed radiation therapy treatments.[65] Missed treatments may also impact the interpretation of the effectiveness of a clinical trial if not documented and considered. Clinical trial groups should consider the implications of missed treatments and how best to collect the information.

Recommendations for Clinical Trial Designers

1. Determine which aspects of the treatment history should be required as part of the data submission.
2. Require a record of missed treatments as part of the data submission.

12. Credentialing and Institutional Preparation

Credentialing for clinical trials is the performance and documentation of specific processes by an institution and its team to demonstrate their ability to accurately plan and treat patients for a particular protocol or treatment modality. In addition, a part of credentialing verifies that the institution is capable of submitting the required datasets to the QA center. The credentialing process is designed to ensure that all participating institutions can faithfully apply the protocol guidelines and deliver comparable doses in a clinical trial. This improves the ability of the trial to detect outcome differences as a function of either dose or volume. IROC Houston has reported on the failure rate of its various phantom benchmark tests[66,67], demonstrating that it has taken years for that rate to reduce to less than 20% for most tests. That centers continue to fail the benchmarks is further justification for the need for them. When large variability exists, clinical trial groups should work with physicists and other members of the institutional team to understand the causes of the variation, how to minimize them, and how to share the knowledge with the broader community to improve the quality of the clinical trial and to benefit other patients treated with similar technology. In some situations, it may be necessary to work closely with manufacturers since the QA centers supporting clinical trials have a unique database with both single-vendor and multiple-vendor solutions represented by institutional members.

Clinical trial groups face a challenge in determining the safest way to adopt and incorporate new technologies in both existing and newly developed clinical trials. When incorporating new or less uniformly applied technologies in clinical trials, the results of credentialing tests aid in discovering and correcting variable, outlier, or noncompliant performance by participating institutions, and this helps to lessen the variability in protocol performance across all institutions. The test can consist of a combination of questionnaires, benchmark plans, dry run digital data submissions, and phantom irradiations. If the institution passes the test, then it is approved for enrollment of patients for the pertinent protocol and the specified treatment modality.

12.1 Purpose and Types of Benchmarks

The acceptable variation in components of a trial must be determined at the time the protocol is designed. The recommendations in this section address aspects of credentialing that can help minimize variability in the treatment delivery to the greatest extent possible.

Many benchmarks are designed as end-to-end tests to evaluate the chain of events from CT simulation to delivery with a dosimetric measurement after delivery of a treatment plan. For example, the IROC Houston head and neck phantom benchmark involves CT simulation, treatment planning, IMRT quality assurance, localization, and delivery. Identification of the source of error introduced by using the advanced technology to deliver the treatment may require testing parts of the simulation, planning, and delivery process separately. For example, if the protocol assesses the effect of reducing margins when using image-guidance technology, then using a test that isolates the effectiveness of this feature may be appropriate. Other benchmarks may evaluate protocol compliance, such as contouring and treatment planning rather than including a phantom irradiation. Another consideration is whether a particular protocol is asking a radiation therapy question compared to trials that utilize equivalent radiation therapy delivered to patients in all arms of the study. In the sections that follow, the recommendations are meant to improve both physics practices for clinical trials and the ability of QA centers to assure treatment quality and compliance with a protocol.

It is important to note the acceptable variations of the performance of a benchmark test are expected to be smaller than the variations in the actual patient treatments because of the inherent variability of humans compared to rigid phantoms.

The task group understands that the quality assurance centers and cooperative groups incur a cost in performing benchmarks, and support for this is part of their funded mandate by the NCI. Depending on the type of credentialing, there may be an impact on the initial patient accrual to clinical trials. An analysis performed by IROC Houston and AAPM demonstrated that the head and neck phantom cost is reasonable in the context of quality for head and neck trials.[68] Also, there are situations where a phantom irradiation with independent dosimetry is much more valuable compared to tests that rely on the institution's dose measurement and analysis of the agreement between the plan and the delivery.[68] Such phantom benchmarks, when appropriately utilized, provide a consistent QA baseline to assess comparability between participating institutions in support of minimizing clinical trial data uncertainty.

12.2 Credentialing Techniques

Participation in any NCI protocol requires, at a minimum, monitoring the output of each institution's radiation beams in reference conditions by the IROC Houston's remote monitoring program. Simple radiation protocols frequently require submission of a questionnaire by the participating institution. Questionnaires typically require both a description of an institution's facility and knowledge of the staff in meeting the protocol requirements. The questions include information about the radiotherapy equipment and treatment planning system at the institution and information about the experience of the team with the steps in the protocol. Questionnaires maintained by the QA center should be updated at least annually by personnel from each institution. In addition to questionnaires and patient case review, advanced technology protocols may require additional QA efforts, such as a dry run planning exercise, verification of accuracy of volume definition, or a phantom irradiation.

A "dry run" planning case tests the ability of the facility to plan the treatment for an actual eligible patient according to protocol guidelines. The dry run planning case(s) using data for the protocol or a patient dataset chosen by the institution is then uploaded to the QA center.

The treatment planning data are transferred to the QA center and reviewed to ensure that the institution is able to meet the protocol specifications. The QA center provides feedback to the institutional team on its performance and requires a repeat of the benchmark if the plan does not pass. The QA center may also perform an independent dose calculation and compare the results to the institution's calculation. Dry runs are recommended for trials involving complex protocol dosing and contouring guidelines. Also, the dry run is important for assessing a team's performance to assure that any differences in technologies do not introduce excessive variability in meeting the protocol.

Benchmarks that test an institution's ability to accurately define volumes are increasingly important for clinical trials that evaluate radiation dose or volume effects. The protocol designers should explicitly state which structures must be delineated or approved by a physician rather than other personnel. These contouring benchmarks test the ability of the institution's staff to define the required structures that are relevant to the trial. The volumes of interest are specific to each protocol. Some of these volumes (such as the cochlea, pituitary, and certain lymph nodes) may be difficult to contour without training or additional information. As noted previously, atlases are invaluable in improving the consistency of contouring. In many situations, the use of contouring benchmarks is valuable in reducing per patient effort spent by QA centers to validate volumes in data submissions. When a benchmark includes volume definition, then the ability of different centers to meet the planning requirements can be directly evaluated.[69]

Treatment delivery verification often includes irradiation of a phantom. Phantoms can be geometric (cylindrical, rectangular) or anthropomorphic. The choice of phantom geometry should be based on the application. Some examples of the features of dose delivery and image guidance that might be tested are:

- Evaluation of the delivery of steep dose gradients. Film dosimetry with 2D analysis in one or more planes should be a feature of the benchmark to adequately measure dose in the gradient region. Consideration should be given to 3D measurements when they become available.
- Evaluation of delivery of stereotactic irradiation of small targets. The benchmark should test the precision of patient positioning. For example, the test could require institutions to image a phantom using orthogonal views or cone-beam CT to establish the difference between the planned and actual position of the patient. The delivered dose should also be assessed.
- Evaluation of delivery to moving structures, such as lung or liver tumors. Tests should use a realistic moving phantom with an appropriate dosimeter to assess the institution's ability to accurately treat these targets.
- Evaluation of target localization accuracy. The anatomy used to guide the image registration at the treatment unit should be specified by the protocol designer.

As new technologies become available, QA centers should assess the best way to measure the inter-institutional variability and work with the clinical trial designers in assessing the impact on the study, especially if the focus is on radiotherapy. NRG Oncology has been gaining more experience in addressing the QA needs for complex trials. For example, NRG Oncology's BR001 trial for SBRT treatment of multiple metastasis involved multiple credentialing steps that could be performed in a time-efficient approach.[69] Due to the complexity of the protocol, credentialing included planning (benchmark plan), delivery (phantom irradiation), patient localization (image-guided radiation therapy credentialing), and a pretreatment case review to comprehensively evaluate the team at each institution.[69]

12.3 Phantom Considerations

Anthropomorphic phantoms in humanoid shapes are used for credentialing of particular advanced technology clinical trials so that the planning and dose delivery to the phantom is as realistic as possible. The phantom should be relevant to the anatomical region described in the clinical trial protocol where appropriate: for example, prostate protocols would employ a pelvic phantom, while a head and neck protocol would use a phantom with head and neck features. Anthropomorphic phantoms are useful where inhomogeneities need to be included in a realistic way. Also, anthropomorphic phantoms generally provide more information about the accuracy of the facility's treatment planning and delivery process.

IROC Houston has designed anthropomorphic phantoms specifically for credentialing the use of advanced technologies at institutions. A critical point is that IROC Houston requests that the team at

each institution treats the phantom as they would a patient. Careful adherence to the instruction from the QA center is critical to the process. Also, institutions should follow the imaging instructions for the protocol. Because this is an independent test, all sources of potential dosimetric error, including linac calibration, are included in the dosimetric analysis. In 2016, IROC Houston is the only organization that makes, distributes, and analyzes anthropomorphic phantoms for NCI clinical trial quality assurance. The suite of IROC Houston phantoms can be found at http://irochouston.mdanderson.org.

To evaluate the impact of motion on dose delivery, some of the IROC Houston's phantoms can be placed on an oscillating platform that simulates breathing motion. A phantom irradiation represents the optimal conditions for localization accuracy since current phantoms are typically rigid and cannot be used to assess deformation over time.

12.4 Pretreatment and On-treatment Review

Pretreatment review of patient cases is often a good alternative to some types of credentialing exercises, in particular the dry run planning exercise or the contouring benchmark. Depending on the protocol, if the treatment plan review is conducted pretreatment, the institution will have sufficient time to revise the plan as needed. The review may include target definition, treatment planning technique and dose distributions, or dose volume histogram analysis of target or normal organ doses. Other protocols may require review of patient volumes and treatment plans within the first few days of treatment. Post-treatment reviews are not an alternative to credentialing, but they are needed to ensure compliance with the protocol and can be useful in checking and improving the performance of individual institutions if performed early in the life of a protocol. This is accomplished through the processes of protocol amendment, institution feedback, and institution mentoring.

The pre- or on-treatment review process is used to ensure that the volumes and plans are consistent with the protocol requirements. Changes may be requested by the principal investigator or the designated reviewer, if needed, to ensure protocol compliance. For the partial breast protocol NSABP B-39/RTOG 0413, the first five cases of each institution were reviewed to provide feedback to the on-site physicians prior to the start of patient treatment (see the protocol, which is available at www.ctsu.org). When additional mentoring is performed, it may be possible to improve compliance with the protocol in real time. As more institutions complete the credentialing process NRG Oncology BR-001, the efficacy of the additional mentoring can be assessed to determine if this model should be used for future clinical trials. This report recommends that the trial designers should specify in the protocol who reviews the case (QA center staff, study principal investigators and co-investigators, or other designated reviewers), the number of cases from each center to be reviewed (e.g., the first two patients enrolled from a given center or based on compliance), the type and timing of the review, and whether or not the credentialing should be for each participating physician or the institution as a whole. When designing the trial, it should also be determined if on-treatment imaging, such as for localization, should be reviewed.

12.5 Considerations Regarding Benchmarks

In general, the dose and distance-to-agreement criteria when comparing calculations and measurements for a given benchmark should be customized to the needed accuracy and comparability between institutions for the question in a given clinical trial. Additional sources of variation include the uncertainties of the detectors used in the benchmark. The following considerations are recommended for absolute and relative dose comparisons for benchmark phantom tests. These are typically end-to-end tests that begin with simulation and end with treatment delivery.

The acceptable absolute dose for a given protocol encompasses the errors that are present at each of the individual steps. IAEA Report 31 addresses the accuracy requirements and uncertainties in radiation therapy. This report is aimed at reducing these uncertainties in the context of clinical trials. The

IROC Houston photon beam output TLD and OSL check has a stated uncertainty of 5% at the 90% confidence level for 3 TLD or OSL per measurement point. For this check, 5% dose agreement is considered acceptable by IROC Houston and the QA centers. The 5% tolerance for the beam output check does not include delivery errors for treatments more complicated than open 10×10 fields. IROC Houston uses the same TLD or OSLs and measurement methods in their anthropomorphic phantoms that they use for the remote dosimetry monitoring program of accelerator output. For benchmark measurements with these detectors, the QA center should assess the measurement uncertainty and its impact on determining the clinically required dose accuracy for a given type of credentialing.

When new planning and delivery technologies are first introduced into clinical trials, this task group recommends that a group of facilities irradiate the phantom before widespread distribution to determine if there are any changes that need to be made to the instructions or phantom itself. This is also an opportunity to confirm that a given technique can be reliably and accurately delivered with different personnel, software, and hardware. If a QA center finds substantial variation among institutions with a new technique, this task group recommends that the QA center and the clinical trial designers determine if the technique is mature enough for use in a multi-institutional clinical trial. We recommend that the criteria for dosimetric agreement should not be broadened solely to accommodate the large variation found among the first set of institutions performing a benchmark, which is what had been done for the first IMRT trials in head and neck[7]. In such situations, it may be necessary to wait for a technology to mature, to provide supplemental education, or to limit participants based on reasonable acceptance criteria.

In the high-dose, high-gradient regions, a distance-to-agreement metric should be specified for comparing measured and calculated doses. A large difference in the value for distance-to-agreement may identify an important discrepancy between the planned and delivered doses.[52] Benchmark calculations and measurements should be evaluated using multiple tools, including dose difference, distance-to-agreement, and Gamma[70] (or a similar metric) to analyze composite treatment doses.

Benchmark tests are often created for a particular protocol or treatment site and sometimes for a particular cooperative group. This task group recommends that the QA centers encourage the use of benchmarks that can be used for multiple protocols by body site or delivery technique. For example, if an institution passes a site-specific anthropomorphic phantom-based IMRT benchmark, it may be reasonable to allow this institution to participate in protocols using IMRT for other sites. One important consideration is whether the benchmark that was passed previously is for a situation of similar complexity as the situation being considered. Also, benchmarks should be created so that all cooperative groups can utilize them for their applicable protocols.

The QA centers should have established criteria for re-credentialing for clinical trials that are regularly reviewed. For example, when VMAT was initially introduced, IMRT re-credentialing of the planning and delivery was required. Once the results were found to be comparable to those for static-gantry IMRT, no further distinction between the two treatment modalities was determined necessary in terms of credentialing.

When necessary for a protocol, documentation of the IGRT alignment should be provided. The IGRT treatment images should be made available by manufacturers in DICOM-RT format,[63] with linac coordinates encoded in the images. The images could be reregistered to the reference set at the QA centers with window and level control, as well as split screen or fused image displays. The registration results for a given fraction, such as the shifts applied for on-line image guidance determined by the QA center, can then be compared to those reported by the personnel at the institution to verify the adequacy of the institution's technique. Some variation in alignment results may occur due to aspects of the image acquisition (slice thickness) and the region of interest used, as shown by Cui et al.[71] For NRG Oncology's BR001 trial, the guidance was explicit regarding when a single isocenter could be

used compared to when multiple isocenters were needed.[69] The IGRT credentialing addressed this situation explicitly in the overall credentialing design.

The designers of benchmark tests for image registration should consider the recommendations of guidance reports such as AAPM TG-132.[24] Image registration benchmarks provide much-needed information that aids QA centers in assessing the treatment center's ability to perform the necessary tasks for a given clinical trial. As the available technology (both software and hardware) and capabilities expand, such benchmarks may need to be modified. The QA centers can design two types of benchmarks: 1) a standard benchmark with the images provided to assess the capabilities of the institution's protocol team in using the information (and to compare performance across multiple individuals/institutions) and 2) submission of an image dataset generated by the institution.

Some clinical trials require the treatment facility to export the images used for alignment, such as digitally reconstructed radiographs (DRRs) to visualize the beam's-eye-views and reference CT images for alignment based on volumetric imaging. QA centers should consider requiring a review of sample alignments for trials requiring high-precision treatments (<0.3 cm). Regarding quality assurance for image acquisition, there has been growing recognition of the value of imaging benchmarks for clinical trial applications. IROC-Houston has maintained a robust phantom program emphasizing evaluation of the treatment planning and delivery chain. Credentialing of imaging for therapeutic clinical trials needs to expand beyond the ACR accreditation, which is largely concerned with diagnostic aspects of imaging rather than quantitative aspects required for therapeutic trials. For this reason, for example, American College of Radiology Imaging Network (ACRIN), now part of the ECOG-ACRIN cooperative group, requires extended credentialing of scanners that is more rigorous than ACR accreditation. Clinical trials should begin incorporating phantom benchmarks to capture and manage the known variability in different imaging systems, especially because of the critical impact on how imaging data are processed to aid in identifying the tumor and its response to therapy. Investigators have been active in this area at their local institutions. For example, apparent diffusion coefficient (ADC) values in MRI are known to vary by scanner, and phantoms have been investigated at individual institutions for assessing the ADC both with a cost-effective phantom[72] and for evaluation of motion effects[73]. The QA needs of clinical trials, such as the reproducible assessment of biomarkers from MR, differ from those addressed by the American College of Radiology accreditation program.

12.6 Institutional Preparation for a Clinical Trial

To participate in clinical trials, a team-based approach is best when it acknowledges the explicit effort required by the physicist and other team members to prepare for the trial. When cooperative group trials are opened at an institution, the physicist and dosimetrist should also receive a copy of the protocol prior to submission of the protocol to the IRB so that the team can prepare for the protocol. This allows the physicist to determine if his or her institution has the necessary equipment for all aspects of the protocol. This is also an important time for the physician principal investigator, physicist, administrator, and other team members to discuss and determine if the institution has the appropriate resources available for the credentialing steps, including the completion of any benchmarks that are necessary. For any NCTN trial, the physicist should submit a web-based request to IROC Houston to determine whether or not the institution is eligible for the trial based on previously completed benchmarks. QA centers should continue to develop facile methods to determine eligibility, as well as the credentialing requirements for a given trial. At this time, physicists or other personnel at an institution can request their credentialing status from the IROC Houston website.

In addition to performing any necessary benchmarks for credentialing, custom simulation and planning directives should be created to be consistent with the protocol. The physicist should refer to a detailed physics section of the trial, provided by the protocol designers, to create standard operating procedures for the simulation, planning, and delivery (including imaging). The simulation directive

should include details related to imaging and positioning following the protocol guidelines. The planning directive should include information regarding the volumes to be defined, the margins for the clinical target volume and planning target volumes, as well as the organs at risk to be contoured. Dose limits should be specified for all structures that should be considered. For other structures, the doses may be scored but may not include a limit. As noted earlier, the trial designers should ensure that structures that behave serially (spinal cord) have explicit contouring specifications or, preferably, the dose limits are specified in cc. The type and frequency of image guidance should also be specified to be consistent with the care needed for the protocol.

A kick-off meeting is recommended with the appropriate research staff, clinical trials coordinator, principal investigator, physicist, dosimetrist, and a therapist before patients are enrolled on the protocol. At the meeting, the protocol, procedures, work flow, and supplemental documentation for the protocol should be reviewed. The roles and responsibilities of the team members should be clearly defined, including if specific staff are required to be present for the simulation or first treatment and to identify who will manage the upload of the required data to the QA center. The required effort for the protocol should also be discussed, especially if a rapid plan evaluation will be necessary or QA processes need to be added for the protocol. An approximate time line for each portion of the process should also be discussed. If a planning benchmark was not required, it may be helpful for the physicist to coordinate an end-to-end test run of the clinical protocol for a mock patient that includes contouring of datasets by the physician, planning by a dosimetrist, and delivery by a therapist. Such a test would likely improve their compliance with the protocol. It should be noted that as the complexity of the trial increases, the required effort by the members of the institutional team increases as well. For example, trials that require multi-modality imaging require additional dosimetry and physics effort for image registration and additional physician effort for contouring requiring more hand-offs throughout the patient care process. Clear communication, reasonable expectations, and teamwork are essential.

12.7 Recommendations Regarding Credentialing

12.7.1 Recommendations for Clinical Trial Design

1. For credentialing, explicitly state which structures must be delineated by a physician rather than other personnel.
2. Work with QA center staff to determine the type of credentialing and if existing benchmarks or other credentialing tests are appropriate before designing new tests.
3. Require a credentialing process with pre- or on-treatment review for at least the first few cases, and perhaps for all cases, prior to treatment for trials that are dependent upon consistent contouring of target and normal structures, adherence to strict margin expansions, dose-volume constraints, and novel treatment techniques.
4. Require credentialing of technologies which may be susceptible to significant inter-institutional variability.
5. Confirm with physicist stakeholders (such as the NRG Medical Physics group and the AAPM Work Group on Clinical Trials), physicians, and administrators, when necessary, to assess whether there are enough centers with adequate equipment and personnel available to meet the specifications, guidelines, benchmarks, and credentialing requirements (by the center) in a timely manner (estimate time of needed training/s).
6. Require QA centers to confirm that the submitted treatment plan of a benchmark irradiation meets the specified requirements for the phantom plan, not only that the measurements and calculations are in agreement.

7. For applicable treatment sites, require a benchmark test that assesses the accuracy of image fusion, IGRT, or other methods critical to the outcome of the trial performed by the institutional personnel routinely planning and treating patients in the clinical trial.
8. The protocol should specify who reviews the case (QA center staff, study principal investigators and co-investigators, or other designated reviewers), the number of cases from each center to be reviewed (e.g., the first two patients enrolled from a given center or based on compliance), the type and timing of the review, and whether or not the credentialing should be for each participating physician or the institution as a whole.

12.7.2 Recommendations for Physicists at the Local Institution

1. Repeat the credentialing benchmark if a major change is made that may affect the quality in the clinical trial. Changes such as to the dose calculation algorithm may only require a resubmission of calculation data results rather than a re-irradiation.
2. Read the protocol and become familiar with the protocol guidelines and credentialing requirements to serve as the institutional expert on the planning and delivery details of each protocol that involves radiotherapy.
3. Complete the Credentialing Status Inquiry (CSI) form, and request the credentialing phantom for a particular trial, if needed. Treat the phantom as a patient, including involvement of the appropriate personnel. Return the phantom to the QA center in a timely manner.
4. Work with the institutional team, including the physician, to ensure a kick-off meeting for the protocol and to create protocol-specific simulation and planning directives to ensure protocol compliance.
5. Coordinate, develop, and perform an end-to-end test for a given protocol where each team member does his or her part to test drive and make corrections to the process before the first protocol patient is enrolled.

12.7.3 Recommendations for QA Centers

1. Regarding data format:
    a. Have a methodology for anonymization of patient data if appropriate for a benchmark planning study. For example, TRIAD (NRG Oncology) includes an anonymization function.
    b. When needed for a study, the image format should be DICOM or DICOM RT (as appropriate) for CT, MR, PET, portal, simulator, and DRR images.
    c. When needed for a study, the structure set, plan, and dose files should be in DICOM RT format.
    d. Supplemental data that needs to be submitted to QA centers should be able to be electronically submitted.
2. For new protocols, determine if an existing benchmark would meet the testing needs of the clinical trial.
3. Develop benchmarks that are applicable across cooperative groups.
4. Annually review facility questionnaires for all institutions participating in clinical trials.
5. Determine when re-credentialing is necessary.
6. Provide appropriate benchmark phantoms for each trial that requires them, as resources permit. Existing phantoms should be assessed for suitability before new ones are made.

7. Determine benchmark acceptability based on reasonable clinical practice for the radiation treatment convolved with the 90% confidence limit of the dose measurements by the QA center.
8. Make information available to team members at an institution to determine eligibility for a given trial based on past credentialing efforts.
9. When new planning and delivery techniques are introduced, evaluate the consistency with a subset of centers. This information should aid in assessing the appropriateness and need of a phantom irradiation.
10. When large variability exists in benchmark results, work with key stakeholders to identify causes and methods to minimize dosimetric discrepancies. This may include working with physicists at local institutions as well as with manufacturer representatives.
11. Develop with imaging experts a suite of benchmark phantoms and a robust program for image acquisition QA with different systems.

13. Summary

It has been shown that the quality and consistency of the trial impacts patient outcomes.[1–5] This report identifies physics and other team member practices that specifically improve the treatment planning and delivery data for clinical trials. It provides benchmark and other quality assurance recommendations for groups that design and conduct clinical trials that can be used by participating institutions to minimize inconsistencies in the radiotherapy processes and treatment. The consistency and efficiency of individual institutions and QA centers can be improved by the development of standardization tools and the ability to export trial-specific optimization and dose response protocols to support participation in clinical trials. The recommendations in this report are reformatted in the appendices for the clinical trial designers (Appendix A), physicists at individual institutions (Appendix B), QA centers (Appendix C), and manufacturers (Appendix D). Recommendations are numbered sequentially in the appendix by category.

There are unique challenges posed by advanced technology trials in a multi-institutional setting. To achieve the desired level of statistical power in a clinical trial, the QA center must verify that the technology is implemented uniformly in multiple settings. The QA centers have had to adapt quickly as new technology becomes available and is implemented into clinical practice. Such trials also require extra effort from the physicists and other members of the treatment planning and delivery team to perform the designated benchmark tests. Other guidance will need to be developed as current advanced technologies mature and others develop.

With technological advancements, manufacturers play a role in the development of improved technology and in providing updates to software tools to enhance the conduct of clinical trials. For example, developers can create mechanisms for the distribution of structure, plan, and reporting templates for a given clinical trial to all users, or they can provide easy-to-use comprehensive digital export tools in the treatment management system. For imaging, manufacturers may make it feasible to distribute templates with sequence definitions to all centers participating in a given trial. Such tools improve the quality of the trial and provide additional support in settings where there may otherwise be inconsistencies or deficiencies in imaging techniques. The utilization of tools will be updated through ongoing international cooperative efforts toward harmonization of clinical trials. Important work has been ongoing in harmonization of credentialing for clinical trials that the NCI has advocated along with other changes[74]. Quality for NCI-funded clinical trials continues to be supported by the IROC infrastructure.

Finally, successful clinical trials involve a partnership relationship among all of those involved.[75] Multi-institutional trials assess the applicability of a new treatment in many settings. Improved consistency in the design and performance of the physics aspects of clinical trials will help ensure that the data is of high integrity and can be used to answer the clinical trial questions and ultimately affect clinical practice.

14. References

1. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, Bourhis J, Yuen K, Fisher R, and Rischin D. (2010). "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* 28(18):2996–3001.
2. Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, Konski AA, Benson AB, Macdonald JS, Rich TA, and Willett CG. (2012). "Failure to adhere to protocol specified radiation therapy guidelines was associated with decreased survival in RTOG 9704—a phase III trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas." *Int J Radiat Oncol Biol Phys* 82(2):809–16.
3. Fairchild A, Straube W, Laurie F, and Followill D. (2013). "Does quality of radiation therapy predict outcomes of multicenter cooperative group trials? A literature review." *Int J Radiat Oncol Biol Phys* 87(2):246–60.
4. Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, and Showalter TN. (2013). "Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials." *J Natl Cancer Inst* 105(6):387–93.
5. Weber DC, Hurkmans CW, Melidis C, Budach W, Langendijk JH, Peters LJ, Grégoire V, Maingon P, and Combescure C. (2014). "Outcome impact and cost-effectiveness of quality assurance for radiotherapy planned for the EORTC 22071–24071 prospective study for head and neck cancer." *Radioth Oncol* 111(3):393–99.
6. Eisbruch A, Harris J, Garden AS, Chao CK, Straube W, Harari PM, Sanguineti G, Jones CU, Bosch WR, and Ang KK. (2010). "Multi-institutional trial of accelerated hypofractionated intensity-modulated radiation therapy for early-stage oropharyngeal cancer (RTOG 00-22)." *Int J Radiat Oncol Biol Phys* 76(5):1333–38.
7. Molineu A, Followill DS, Balter PA, Hanson WF, Gillin MT, Huq MS, Eisbruch A, and Ibbott GS. (2005). "Design and implementation of an anthropomorphic quality assurance phantom for intensity-modulated radiation therapy for the Radiation Therapy Oncology Group." *Int J Radiat Oncol Biol Phys* 63(2):577–83.
8. Ibbott GS, Followill DS, Molineu HA, Lowenstein JR, Alvarez PE, and Roll JE. (2008). "Challenges in credentialing institutions and participants in advanced technology multi-institutional clinical trials." *Int J Radiat Oncol Biol Phys* 71(1 Suppl):S71–75.
9. Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ulin K, Ishikura S, Followill D, Galvin J, Haworth A, Besuijen D, Clark CH, Miles E, Aird E, Weber DC, Hurkmans CW, and Verellen D. (2014). "Global harmonization of quality assurance naming conventions in radiation therapy clinical trials." *Int J Radiat Oncol Biol Phys* 90(5):1242–49.
10. Moore KL, Schmidt R, Moiseenko V, Olsen LA, Tan J, Xiao Y, Galvin J, Pugh S, Seider MJ, Dicker AP, Bosch W, Michalski J, and Mutic S. (2015). "Quantifying unnecessary normal tissue complication risks due to suboptimal planning: A secondary study of RTOG 0126." *Int J Radiat Oncol Biol Phys* 92(2):228–35.
11. Mayo CS, Moran JM, Bosch W, Xiao Y, Mcnutt T, Popple R, Michalski J, Feng M, Marks LB, Fuller CD, Yorke E, Palta J, Gabriel PE, Molineu A, Matuszak MM, Covington E, Masi K, Rich-

ardson SL, Ritter T, Morgas T, Flampouri S, Santanam L, Moore JA, Purdie TG, Miller R, Hurkmans C, Adams J, Wu Q-R J, Fox CJ, Siochi RA, Brown NL, Verbakel W, Archambault Y, Chmura SJ, Eagle DG, Fitzgerald TJ, Dekker AL, Hong T, Kapoor R, Lansing B, Jolly S, Napolitano M, Percy J, Rose MS, Saddiqui S, Schadt C, Simon WE, Straube WL, St. James ST, Ulin K, Yom SS, and Yock T. "AAPM TG-263: Standardizing nomenclatures in radiation oncology." Alexandria, VA: American Association of Physicists in Medicine, 2018. https://www.aapm.org/pubs/reports/detail.asp?docid=171.

12. IOM. A National Cancer Clinical Trials System for the 21st Century Reinvigorating the NCI Cooperative Group Program. (2010). http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2010/A-National-Cancer-Clinical-Trials-System-for-the-21st-Century-Reinvigorating-the-NCI-Cooperative/NCI%20Cancer%20Clinical%20Trials%202010%20%20Report%20Brief.pdf. Accessed 04/10/2017.
13. Fitzgerald TJ, Bishop-Jodoin M, Bosch WR, Curran WJ, Followill DS, Galvin JM, Hanusik R, King SR, Knopp MV, Laurie F, O'Meara E, Michalski JM, Saltz JH, Schnall MD, Schwartz L, Ulin K, Xiao Y, and Urie M. (2013). "Future vision for the quality assurance of oncology clinical trials." *Front Oncol* 3:31.
14. Olch AJ, Kline RW, Ibbott GS, Anderson JR, Deye J, FitzGerald TJ, Followill D, Gillin MT, Huq MS, Palta JR, Purdy JA, and Urie MM. *Quality Assurance for Clinical Trials: A Primer for Physicists.* Madison, Wisconsin: Medical Physics Publishing, 2004.
15. Mulshine JL, Gierada DS, Armato SG 3rd, Avila RS, Yankelevitz DF, Kazerooni EA, McNitt-Gray MF, Buckler AJ, and Sullivan DC. (2015). "Role of the quantitative imaging biomarker alliance in optimizing CT for the evaluation of lung cancer screen-detected nodules." *J Am Coll Radiol* 12(4):390–95.
16. Smith DS, Li X, Arlinghaus LR, Yankeelov TE, and Welch EB. (2015). "DCEMRI.jl: a fast, validated, open source toolkit for dynamic contrast enhanced MRI analysis." *PeerJ.* 3:e909.
17. Graham MM, Wahl RL, Hoffman JM, Yap JT, Sunderland JJ, Boellaard R, Perlman ES, Kinahan PE, Christian PE, Hoekstra OS, and Dorfman GS. (2015). "Summary of the UPICT protocol for 18F-FDG PET/CT imaging in oncology clinical trials." *J Nucl Med* 56(6):955–61.
18. Jeraj R, Bradshaw T, and Simoncic U. (2015). "Molecular imaging to plan radiotherapy and evaluate its efficacy." *J Nucl Med* 56(11):1752–65.
19. Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, Barnhart HX, Jackson EF, Giger ML, Pennello G, Toledano AY, Kalpathy-Cramer J, Apanasovich TV, Kinahan PE, Myers KJ, Goldgof DB, Barboriak DP, Gillies RJ, Schwartz LH, Sullivan DC. (2015). "Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res* 24(1):68–106.
20. AAPM. AAPM CT Lexicon. 2012.
21. Gregoire V, Haustermans K, Geets X, Roels S, and Lonneux M. (2007). "PET-based treatment planning in radiotherapy: a new standard? *J Nucl Med* 48 Suppl 1:68S–77S.
22. Grigsby PW. (2007). "The role of FDG-PET/CT imaging after radiation therapy." *Gynecol Oncol.* 107(1 Suppl 1):S27–29.
23. Perez CA, Bradley J, Chao CK, Grigsby PW, Mutic S, and Malyapa R. (2002). "Functional imaging in treatment planning in radiation therapy: a review." *Rays* 27(3):157–73.
24. Brock KK, Mutic S, McNutt TR, Li H, and Kessler ML. (2016). "Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Submitted to Med Phys*.
25. Meng X, Frey K, Matuszak M, Paul S, Ten Haken R, Yu J, and Kong FM. (2014). "Changes in functional lung regions during the course of radiation therapy and their potential impact on lung dosimetry for non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 89(1):145–51.

26. ACR. ACR-AAPM-SIIM Technical standard for electronic practice of medical imaging. 2014.
27. Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, Kapatoes JM, Low DA, Murphy MJ, Murray BR, Ramsey CR, Van Herk MB, Vedam SS, Wong JW, and Yorke E. (2006). "The management of respiratory motion in radiation oncology report of AAPM Task Group 76." *Med Phys* 33(10):3874–900.
28. Brandner ED, Chetty IJ, Giaddui TG, and Xiao Y. (2017). "Motion management strategies and technical issues associated with stereotactic body radiotherapy of thoracic and upper abdominal tumors: A review from NRG Oncology." *Med Phys* TBD: in press.
29. Nyflot MJ, Lee TC, Alessio AM, Wollenweber SD, Stearns CW, Bowen SR, and Kinahan PE. (2015). "Impact of CT attenuation correction method on quantitative respiratory-correlated (4D) PET/CT imaging." *Med Phys* 42(1):110–20.
30. Sindoni A, Minutoli F, Pontoriero A, Iati G, Baldari S, and Pergolizzi S. (2016). "Usefulness of four dimensional (4D) PET/CT imaging in the evaluation of thoracic lesions and in radiotherapy planning: Review of the literature. *Lung Cancer* 96:78–86.
31. NCI. "ATC guidelines for the use of IMRT (including intra-thoracic treatments)." (2006). http://atc.wustl.edu/home/NCI/NCI_IMRT_Guidelines_2006.pdf.
32. Measurements ICoRUa. *ICRU Report 50: Prescribing, recording, and reporting photon beam therapy.* Bethesda, MD: ICRU, 1993.
33. Measurements ICoRUa. *ICRU Report 62: Prescribing, recording and reporting photon beam therapy (Supplement to ICRU report 50).* Bethesda, MD: ICRU, 1999.
34. Measurements ICoRUa. *ICRU Report 83: Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT).* Bethesda, MD: ICRU, 2010.
35. Jaffray DA, Langen KM, Mageras G, Dawson LA, Yan D, Edd RA, Mundt AJ, and Fraass B. (2013). "Safety considerations for IGRT: Executive summary. *Pract Radiat Oncol* 3(3):167–70.
36. Bowden P, Fisher R, Mac Manus M, Wirth A, Duchesne G, Millward M, McKenzie A, Andrews J, and Ball D. (2002). "Measurement of lung tumor volumes using three-dimensional computer planning software." *Int J Radiat Oncol Biol Phys* 53(3):566–73.
37. Geets X, Daisne JF, Arcangeli S, Coche E, De Poel M, Duprez T, Nardella G, and Grégoire V. (2005). "Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI." *Radiother Oncol* 77(1):25–31.
38. Villeirs GM, Van Vaerenbergh K, Vakaet L, Bral S, Claus F, De Neve WJ, Verstraete KL, and De Meerleer GO. (2005). "Interobserver delineation variation using CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer." *Strahlenther Onkol* 181(7):424–30.
39. Li XA, Tai A, Arthur DW, Buchholz TA, Macdonald S, Marks LB, Moran JM, Pierce LJ, Rabinovitch R, Taghian A, Vicini F, Woodward W, and White JR. (2009). "Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study." *Int J Radiat Oncol Biol Phys* 73(3):944–51.
40. Lawton CA, Michalski J, El-Naqa I, Buyyounouski MK, Lee WR, Menard C, O'Meara E, Rosenthal SA, Ritter M, and Seider M. (2009). "RTOG GU Radiation oncology specialists reach consensus on pelvic lymph node volumes for high-risk prostate cancer." *Int J Radiat Oncol Biol Phys* 74(2):383–87.
41. Michalski JM, Lawton C, El Naqa I, Ritter M, O'Meara E, Seider MJ, Lee WR, Rosenthal SA, Pisansky T, Catton C, Valicenti RK, Zietman AL, Bosch WR, Sandler H, Buyyounouski MK, and Ménard C. (2010). "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* 76(2):361–68.

42. Myerson RJ, Garofalo MC, El Naqa I, Abrams RA, Apte A, Bosch WR, Das P, Gunderson LL, Hong TS, Kim JJ, Willett CG, and Kachnic LA. (2009). "Elective clinical target volumes for conformal therapy in anorectal cancer: a radiation therapy oncology group consensus panel contouring atlas." *Int J Radiat Oncol Biol Phys* 74(3):824–30.
43. Feng M, Moran JM, Koelling T, Chughtai A, Chan JL, Freedman L, Hayman JA, Jagsi R, Jolly S, Larouere J, Soriano J, Marsh R, and Pierce LJ. (2011). "Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer." *Int J Radiat Oncol Biol Phys* 79:10–18.
44. Cheng G, Yang X, Wu N, Xu Z, Zhao H, Wang Y, and Liu T. (2013). "Multi-atlas-based segmentation of the parotid glands of MR images in patients following head-and-neck cancer radiotherapy." *Proc SPIE Int Soc Opt Eng* 8670.
45. Conson M, Cella L, Pacelli R, Comerci M, Liuzzi R, Salvatore M, and Quarantelli M. (2014). "Automated delineation of brain structures in patients undergoing radiotherapy for primary brain tumors: from atlas to dose-volume histograms." *Radiother Oncol* 112(3):326–31.
46. Genovesi D, Perrotti F, Trignani M, Di Pilla A, Vinciguerra A, Augurio A, Di Tommaso M, Caulo M, Savastano M, Tartaro A, Cotroneo AR, and Cèfaro GA. (2015). "Delineating brachial plexus, cochlea, pharyngeal constrictor muscles and optic chiasm in head and neck radiotherapy: a CT-based model atlas. *Radiol Med* 120(4):352–60.
47. Sun Y, Yu XL, Luo W, Lee AW, Wee JT, Lee N, Zhou GQ, Tang LL, Tao CJ, Guo R, Mao YP, Zhang R, Guo Y, and Ma J. (2014). "Recommendation for a contouring method and atlas of organs at risk in nasopharyngeal carcinoma patients receiving intensity-modulated radiotherapy." *Radiother Oncol* 110(3):390–97.
48. Wang W, Xu Y, Schipper M, Matuszak MM, Ritter T, Cao Y, Ten Haken RK, and Kong FM. (21013). "Effect of normal lung definition on lung dosimetry and lung toxicity prediction in radiation therapy treatment planning." *Int J Radiat Oncol Biol Phys* 86(5):956–63.
49. Kry S, Bednarz B, Howell RM, Dauer L, Followill D, Klein E, Paganetti H, Wang B, Wuu CS, and George Xu X. (2017). "AAPM Task Group No. 158: Measurements and calculations of doses outside the treatment volume from external beam radiation therapy." *Med Phys* 44(10):e391–e429. Epub Aug 20, 2017.
50. Davidson SE, Popple RA, Ibbott GS, and Followill DS. (2008). "Technical note: Heterogeneity dose calculation accuracy in IMRT: study of five commercial treatment planning systems using an anthropomorphic thorax phantom." *Med Phys* 35(12):5434–39.
51. Chetty IJ, Curran B, Cygler JE, DeMarco JJ, Ezzell G, Faddegon BA, Kawrakow I, Keall PJ, Liu H, Ma CM, Rogers DW, Seuntjens J, Sheikh-Bagheri D, and Siebers JV. (2007). "Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning." *Med Phys* 34(12):4818–53.
52. Cadman P, Bassalow R, Sidhu NP, Ibbott G, and Nelson A. (2002). "Dosimetric considerations for validation of a sequential IMRT process with a commercial treatment planning system." *Phys Med Biol* 47(16):3001–10.
53. Benedict SH, Yenice KM, Followill D, Galvin JM, Hinson W, Kavanagh B, Keall P, Lovelock M, Meeks S, Papiez L, Purdie T, Sadagopan R, Schell MC, Salter B, Schlesinger DJ, Shiu AS, Solberg T, Song DY, Stieber V, Timmerman R, Tomé WA, Verellen D, Wang L, and Yin FF. (2010). "Stereotactic body radiation therapy: the report of AAPM Task Group 101." *Med Phys* 37(8):4078–101.
54. Straube W, Matthews J, Bosch W, and Purdy J. (2005). DVH Analysis: Consequences for quality assurance of multi-institutional clinical trials (abstract). *Med. Phys.* 32(6):2021. http://irocstl.wustl.edu/publications/AAPM2005/ITC.Poster.aapm.2005.pdf.
55. Weinberg R, Kaurin DG, Choy H, Curran WJ Jr, MacRae R, Kim JS, Kim J, Tucker SL, Bonomi PD, Belani C, and Starkschall G. (2004). "Dosimetric uncertainties of three-dimensional

dose reconstruction from two-dimensional data in a multi-institutional study." *J Appl Clin Med Phys* 5(4):15–28.
56. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, Ten Haken RK, and Yorke ED. (2010). "Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues." *Int J Radiat Oncol Biol Phys* 76(3 Suppl):S3–9.
57. Warkentin B, Stavrev P, Stavreva N, Field C, and Fallone B. (2004). "A TCP-NTCP estimation module using DVHs and known radiobiological models and parameter sets. *J Appl Clin Med Phys* 5(1):50–63.
58. Allen Li X, Alber M, Deasy JO, Jackson A, Ken Jee KW, Marks LB, Martel MK, Mayo C, Moiseenko V, Nahum AE, Niemierko A, Semenenko VA, and Yorke ED. (2012). "The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM." *Med Phys* 39(3):1386–409.
59. Higgins PD, Han EY, Yuan JL, Hui S, and Lee CK. (2007). "Evaluation of surface and superficial dose for head and neck treatments using conventional or intensity-modulated techniques." *Phys Med Biol.* 52(4):1135–46.
60. Lee N, Chuang C, Quivey JM, Phillips TL, Akazawa P, Verhey LJ, Xia P. (2002). "Skin toxicity due to intensity-modulated radiotherapy for head-and-neck carcinoma." *Int J Radiat Oncol Biol Phys* 53(3):630–37.
61. Olch AJ, Gerig L, Li H, Mihaylov I, and Morgan A. (2014). "Dosimetric effects caused by couch tops and immobilization devices: report of AAPM Task Group 176." *Med Phys* 41(6):061501.
62. Thor M, Andersen ES, Petersen JB, Sørensen TS, Noe KØ, Tanderup K, Bentzen L, Elstrøm UV, Høyer M, and Muren LP. (2014). "Evaluation of an application for intensity-based deformable image registration and dose accumulation in radiotherapy. *Acta Oncol* 53(10):1329–36.
63. MITA. *Digital Imaging and Communications in Medicine (DICOM), Parts 1–16.* Rosslyn, VA: National Electrical Manufacturers Association (NEMA, 2015.
64. Rengan R, Curran B, Able C, Abdel-Wahab M, Ravi A, Albuquerque K, Field C, Abdul S, Kapoor R, Tripuraneni P, and Palta J. (2011). "Addressing connectivity issues: The Integrating the Healthcare Enterprise-Radiation Oncology (IHE-RO) initiative." *Pract Radiat Oncol* 1(4):226–31.
65. Ohri N, Rapkin BD, Guha C, Kalnicki S, and Garg M. (2016). "Radiation therapy noncompliance and clinical outcomes in an urban academic cancer center. *Int J Radiat Oncol Biol Phys* 95(2):563–70.
66. Carson ME, Molineu A, Taylor PA, Followill DS, Stingo FC, and Kry SF. (1016). "Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. *Med Phys* 43(12):6491.
67. Molineu A, Hernandez N, Nguyen T, Ibbott G, and Followill D. (2013). "Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom." *Med Phys* 40(2):022101.
68. Science Council and Therapy Physics Committee of the American Association of Physicists in Medicine (AAPM). (2015). "Independent phantom irradiation is both necessary and cost effective for clinical trial credentialing." *Int J Radiat Oncol Biol Phys* 92:501–03.
69. Al-Hallaq HA, Chmura S, Salama JK, Winter KA, Robinson CG, Pisansky TM, Borges V, Lowenstein JR, McNulty S, Galvin JM, Followill DS, Timmerman RD, White JR, Xiao Y, and Matuszak MM. (2016). "Rationale of technical requirements for NRG-BR001: The first NCI-sponsored trial of SBRT for the treatment of multiple metastases." *Pract Radiat Oncol* 6(6):E291–98.
70. Low DA and Dempsey JF. (2003). "Evaluation of the gamma dose distribution comparison method." *Med Phys* 30(9):2455–64.

71. Cui Y, Galvin JM, Straube WL, Bosch WR, Purdy JA, Li XA, and Xiao Y. (2011). "Multi-system verification of registrations for image-guided radiotherapy in clinical trials." *Int J Radiat Oncol Biol Phys* 81(1):305–12.
72. Pullens P, Bladt P, Sijbers J, Maas AI, and Parizel PM. (2017). "Technical note: A safe, cheap, and easy-to-use isotropic diffusion MRI phantom for clinical and multicenter studies." *Med Phys* 44(3):1063–70.
73. Kwee TC, Takahara T, Koh DM, Nievelstein RAJ, and Luijten PR. (2008). "Comparison and reproducibility of ADC measurements in breathhold, respiratory triggered, and free-breathing diffusion-weighted MR imaging of the liver." *J Magn Reson Imaging* 28(5):1141–48.
74. Bekelman JE, Deye JA, Vikram B, Bentzen SM, Bruner D, Curran WJ Jr, Dignam J, Efstathiou JA, FitzGerald TJ, Hurkmans C, Ibbott GS, Lee JJ, Merchant TE, Michalski J, Palta JR, Simon R, Ten Haken RK, Timmerman R, Tunis S, Coleman CN, and Purdy J. (2012). "Redesigning radiotherapy quality assurance: opportunities to develop an efficient, evidence-based system to support clinical trials—report of the National Cancer Institute Work Group on Radiotherapy Quality Assurance." *Int J Radiat Oncol Biol Phys* 83(3):782–90.
75. FitzGerald TJ, Urie M, Ulin K, Laurie F, Yorty J, Hanusik R, Kessel S, Jodoin MB, Osagie G, Cicchetti MG, Pieters R, McCarten K, and Rosen N. (2008). "Processes for quality improvements in radiation oncology clinical trials." *Int J Radiat Oncol Biol Phys* 71(1):S76–79.

15. Acknowledgments

The authors gratefully acknowledge the expert contributions of Marcia M. Urie, Robert E. Drzymala, and James A. Purdy who participated in the task group at the beginning of its work. The authors also acknowledge the valuable feedback of members of the AAPM Work Group on Clinical Trials, Quality Assurance Subcommittee, Therapy Physics Committee, and Science Council. Special thanks go to Robin Stern who shepherded the report through multiple review stages.

16. Appendices

These appendices duplicate information in the text but present them in one place for the user.

Appendix A

Recommendations for Clinical Trial Designers

Image Acquisition

1. Determine if imaging-specific credentialing is required through a review by imaging experts (such as the imaging organizations within IROC) and whether or not variability in techniques or variations in commercial scanner technology need to be considered.
2. Design a standard operating procedure for imaging, incorporating expertise of imaging physicists/scientists where appropriate.
    a. Specify the extent of anatomy to be imaged, including whole organs when required for dose volume analyses.
    b. Specify any timing requirements of the acquisition in relation to treatment start for all imaging data for treatment planning and assessment. Be explicit regarding patient preparations for imaging.
    c. Keep image acquisition, reconstruction, and analysis procedures consistent when multiple imaging sessions for a patient are required.
    d. Ensure consistent patient setup and immobilization between different imaging modalities and treatment (see sections 5.2 and 8.1) through credentialing of multi-modality image registration.
    e. Specify which contrast agents are permitted and provide details on the timing and amount of the agent to be used.
    f. Provide guidelines on basic imaging parameters for trials permitting different modalities, such as MRI, MRS, or PET/CT, to account for the variability of different scanners.
    g. Develop imaging benchmarks when modalities such as PET and MRS are used, to ensure that the department's systems for contouring are capable of representing that data adequately in support of the clinical trial.

Image Registration

1. For any applications of image registration in a trial, the protocol designers should specify which methods are allowed (rigid only, deformable), and any additional constraints.
2. Guidance should be provided about how the quality of an image registration is judged which should distinguish between applications for target and normal tissue definition compared to daily on-line treatment guidance. This information should be considered when image registrations are evaluated as part of credentialing for a given trial.

Motion Assessment and Management

1. For relevant body sites, specify that the degree of target motion should be assessed at the time of simulation. For treatment sites where the impact of motion can be crucial, it is recommended that QA centers develop guidance with respect to the acceptable imaging techniques to assess motion, documentation of that motion for a given patient, and how the information should be incorporated for creating target volumes.
2. Incorporate guidance on motion management techniques when the range of motion is greater than published limits (or significant normal tissue sparing can be achieved through their use). For trials when target motion may be ≥5 mm and delivery of a high daily dose (e.g., SBRT), institutions should be required to document the assessment and follow formal guidance, such as that provided by AAPM TG-76[27] or other organizations such as NRG28 to ensure motion assessment and management information is accurately captured for patients enrolled on the trial.

3. For protocols involving monitoring of intra-fraction motion, provide information regarding the acceptable technologies for monitoring and the thresholds for evaluation. Information should be provided as to whether intra-fraction monitoring is required and the acceptable methods.

Patient Immobilization, Target Definition, and Treatment Guidance

1. The clinical trial design should survey the literature, including relevant AAPM task group reports, to determine the type of immobilization suitable to meet aims of the clinical trial.
2. Consult with physicist(s) at a lead institution and other possible participating institutions to ensure that the proposed accuracy limits are achievable at a number of centers.
3. Clearly specify which immobilization equipment is required for the trial (where a preliminary assessment of equipment availability in the community could be done via the IROC Houston facility questionnaire if needed) or if certain types of equipment are not permitted.
4. Use the most up-to-date terminology to specify definitions of target volumes in the trial design (e.g., ICRU #83 at time of publication).
5. Review data in the literature to define acceptable PTV margins related to the technology used for simulation (such as 4DCT) and the frequency and type of imaging for the anatomical site.
6. Provide explicit guidance on the contouring of targets and necessary expansions.
7. If a protocol requires an evaluation of target margins mid-treatment, the clinical trial designers should specify the frequency and methods of evaluation in the clinical trial design. For example, how to address changes in tumor physiology or shape, such as changes to targets in the lung or head and neck region due to shrinkage or growth of the tumor.

Segmentation

1. Specify window and level values, when appropriate, for consistent visualization and segmentation.
2. Refer investigators to published consensus atlases for target and organ at risk delineation as a reference when appropriate.
3. Provide training to physicians for a given trial if there could be significant variability in the delineation of structures among physicians.
4. Provide guidelines to physicians, physicists, and dosimetrists on how to address imaging artifacts that interfere with target or normal tissue segmentation (e.g., scatter from metal or the presence of contrast on a CT simulation scan).
5. For organs which will be evaluated with DVHs, the protocol should specify how much of the organ must be contoured for structures such as the spinal cord.

Treatment Planning Considerations

1. Specify standard structure names that must be used for the clinical trial (follow consensus guidance when available) such as provided by AAPM TG-263 or other appropriate ontologies.
2. Use published information on normal tissue limits, such as through consensus efforts as appropriate, when specifying the limits to normal tissues.
3. For organs which will be evaluated with DVHs, the protocol should specify how much of the organ must be contoured for structures such as the spinal cord.
4. Specify spatial resolution requirements for dose and DVH calculations that are commensurate with target and organ-at-risk (OAR) sizes.

5. Specify the use of 3D treatment planning for all clinical trials (possibly excluding special procedures, such as total body irradiation or total skin electron treatments).
6. Require the use of more accurate algorithms (such as convolution/superposition, Monte Carlo) for trials where tissue heterogeneities may be significant.
7. Develop credentialing approaches for new applications of the TPS, such as biological treatment planning[58]. Credentialing may include intercomparison of results using standardized datasets.
8. Specify the dose-volume constraints for organs-at-risk and consider any special concerns, such as the buildup region or structures outside the treatment area.
9. Specify the minimization of the integral dose or total dose to other normal tissues that may not be contoured in trials which allow the use of dose optimization techniques.

Treatment Planning Delivery Documentation

1. Determine which aspects of the treatment history should be required as part of the data submission.
2. Require a record of missed treatments as part of the data submission.

Credentialing and Institutional Preparation

1. For credentialing, explicitly state which structures must be delineated by a physician rather than other personnel.
2. Work with QA center staff to determine the type of credentialing and if existing benchmarks or other credentialing tests are appropriate before designing new tests.
3. Require a credentialing process with pre- or on-treatment review for at least the first few cases, and perhaps for all cases, prior to treatment for trials that are dependent upon consistent contouring of target and normal structures, adherence to strict margin expansions, dose-volume constraints, and novel treatment techniques.
4. Require credentialing of technologies which may be susceptible to significant inter-institutional variability.
5. Confirm with physicist stakeholders (such as the NRG Medical Physics group and the AAPM Work Group on Clinical Trials), physicians, and administrators, when necessary, to assess whether there are enough centers with adequate equipment and personnel available to meet the specifications, guidelines, benchmarks, and credentialing requirements (by the center) in a timely manner (estimate time of needed training/s).
6. Require QA centers to confirm that the submitted treatment plan of a benchmark irradiation meets the specified requirements for the phantom plan, not only that the measurements and calculations are in agreement.
7. For applicable treatment sites, require a benchmark test that assesses the accuracy of image fusion, IGRT, or other methods critical to the outcome of the trial performed by the institutional personnel routinely planning and treating patients in the clinical trial.
8. The protocol should specify who reviews the case (QA center staff, study principal investigators and co-investigators, or other designated reviewers), the number of cases from each center to be reviewed (e.g., the first two patients enrolled from a given center or based on compliance), the type and timing of the review, and whether or not the credentialing should be for each participating physician or the institution as a whole.

Appendix B

Recommendations for Physicists at the Local Institution

Image Acquisition

1. Train and work with the appropriate personnel to implement the protocol-specified imaging standard operating procedures for image acquisition, reconstruction, processing, and analysis.
2. Review patient imaging scans regularly to ensure compliance to the standard operating procedure.
3. Consider utilization of immobilization and setup methods and devices that are compatible with all imaging modalities used in the trial to reproduce the setup for the treatment planning CT.

Image Registration

1. Evaluate the ability of the institution to follow protocol guidelines for segmentation and image registration.
2. Follow recommendations of AAPM TG-132 with respect to image registration.[24]
3. Adjust monitors for adequate resolution and properly calibrate for contrast and brightness to ensure consistency in target delineation.[26] Note minimum settings in the standard operating procedure.

Motion Assessment and Management

1. Confirm that the motion assessment and management guidance specified in the protocol is followed whenever the range of motion meets published guidance limits.
2. Ensure that the contoured IGTV is reasonable considering the measured motion for a given protocol patient.

Patient Immobilization, Target Definition, and Treatment Guidance

1. Determine that the institution's immobilization equipment is appropriate for the clinical trial before IRB submission.
2. Ensure consistency of equipment for planning and treatment, e.g., flat table tops for diagnostic scanners and the use of compatible immobilization equipment for imaging scans, when possible.
3. Confirm the accuracy of the immobilization method used in the clinic for the protocol.
4. Ensure personnel are adequately trained to support the process.
5. For each protocol, understand how target margins are specified and make sure the margins are reasonable for the department's imaging, immobilization, planning, delivery, and treatment guidance process for the patients enrolled on the trial.
6. For each protocol, monitor the effectiveness of the patient localization method for the patients enrolled on the trial.

Treatment Planning Considerations

1. Ensure that the TPS is capable of meeting protocol requirements by:
    a. Use of a model-based algorithm such as convolution/superposition, Monte Carlo, or deterministic methods.

   b. Accurate modeling of beams and output factors, especially for small fields and IMRT techniques.
   c. Validating the dose-volume histogram and analysis algorithms.
2. Ensure that 3D volumetric information can be exported to the Clinical Trial QA Center in DICOM-RT format.
3. Implement templates in your treatment planning system to use the standard names for targets and structures as specified by the clinical trial designers.
4. Coordinate an end-to-end dry run of the protocol at his or her center on one of their patient's dataset(s). (Note that this requires support from the department's administration for this valuable effort.)
5. Determine the degree of attenuation by immobilization equipment, and determine whether the attenuation should be accounted for in monitor units (MU) calculations.

Credentialing and Institutional Preparation

1. Repeat the credentialing benchmark if a major change is made that may affect the quality in the clinical trial. Changes such as to the dose calculation algorithm may only require a resubmission of calculation data results rather than a re-irradiation.
2. Read the protocol and become familiar with the protocol guidelines and credentialing requirements to serve as the institutional expert on the planning and delivery details of each protocol that involves radiotherapy.
3. Complete the Credentialing Status Inquiry (CSI) form, and request the credentialing phantom for a particular trial, if needed. Treat the phantom as a patient, including involvement of the appropriate personnel. Return the phantom to the QA center in a timely manner.
4. Work with the institutional team, including the physician, to ensure a kick-off meeting for the protocol and to create protocol-specific simulation and planning directives to ensure protocol compliance.
5. Coordinate, develop, and perform an end-to-end test for a given protocol where each team member does his or her part to test drive and make corrections to the process before the first protocol patient is enrolled.

Appendix C

Recommendations for QA Centers

Image Acquisition

Specify if an existing imaging benchmark would be beneficial for ensuring that enrolling institutions would be able to acquire scans of the appropriate quality to support the trial.

Image Registration

1. Develop imaging benchmarks as needed, including when modalities such as PET and MRS are used to ensure that the department's systems for contouring are capable of representing that data adequately in support of the clinical trial.
2. Develop credentialing methods incorporating deformable image registration following the recommendations of AAPM TG-132.[24]

Motion Assessment and Management

Determine if a motion benchmark is required and if so, if an existing benchmark will meet the needs of a given trial.

Patient Immobilization, Target Definition, and Treatment Guidance

1. Confirm that the precision of commercial immobilization systems and field experiences indicate that the proposed techniques realistically can meet the accuracy requested in the protocol.
2. Ensure the appropriateness of the margin for a given trial.
3. Determine credentialing methods for new techniques, such as those requiring intra-fraction monitoring.

Treatment Planning Considerations

1. Enable as much automation of data submission as possible.
2. Continue validation and cross-comparison of the performance of different dose algorithms with other QA centers, and revise requirements as appropriate.
3. Work with manufacturers to design interfaces that can be customized for electronic submission of all necessary protocol data.
4. Provide the clinical trial groups with a template of standard target and structure names across clinical protocols consistent with the nomenclature of AAPM TG-263.
5. Develop mechanisms to share scripts or other tools (such as Excel spreadsheets with macros enabled) to aid the institutional teams in assessing whether or not protocol guidelines are met prior to submission to the QA center. Tools could potentially be developed on multiple TPS platforms.

Credentialing and Institutional Preparation

1. Regarding data format:
    a. Have a methodology for anonymization of patient data if appropriate for a benchmark planning study. For example, TRIAD (NRG Oncology) includes an anonymization function.
    b. When needed for a study, the image format should be DICOM or DICOM RT (as appropriate) for CT, MR, PET, portal, simulator, and DRR images.

   c. When needed for a study, the structure set, plan, and dose files should be in DICOM RT format.
   d. Supplemental data that needs to be submitted to QA centers should be able to be electronically submitted.
2. For new protocols, determine if an existing benchmark would meet the testing needs of the clinical trial.
3. Develop benchmarks that are applicable across cooperative groups.
4. Annually review facility questionnaires for all institutions participating in clinical trials.
5. Determine when re-credentialing is necessary.
6. Provide appropriate benchmark phantoms for each trial that requires them, as resources permit. Existing phantoms should be assessed for suitability before new ones are made.
7. Determine benchmark acceptability based on reasonable clinical practice for the radiation treatment convolved with the 90% confidence limit of the dose measurements by the QA center.
8. Make information available to team members at an institution to determine eligibility for a given trial based on past credentialing efforts.
9. When new planning and delivery techniques are introduced, evaluate the consistency with a subset of centers. This information should aid in assessing the appropriateness and need of a phantom irradiation.
10. When large variability exists in benchmark results, work with key stakeholders to identify causes and methods to minimize dosimetric discrepancies. This may include working with physicists at local institutions as well as with manufacturer representatives.
11. Develop with imaging experts a suite of benchmark phantoms and a robust program for image acquisition QA with different systems.

Appendix D

Recommendations for Manufacturers

Image Acquisition

1. For a given registration, develop methods to extract the primary goals of the image registration (e.g., target evaluation or organ-at-risk, such as from a structured field) and the goodness of the registration (see TG-132 recommendations).[24]
2. In image registration software, provide the ability to export necessary data for QA centers to be able to assess the quality of a registration (quantitative and qualitative) and export the needed information for straightforward review by those credentialing for clinical trials and investigators for patients enrolled in clinical trials.

Motion Assessment and Management

1. Provide on-line 4D tools such as 4D CBCT capability at the treatment machine to support protocol motion management requirements.
2. Provide tools to document and export the range of motion data for different imaging platforms.

Treatment Planning Considerations

1. Include DICOM-RT export in the base purchase of a TPS rather than an add-on option with the ability to export coded ID cases to the QA centers (including image datasets, plans, structures, and dose).
2. Provide standard target and structure names as provided by the QA centers or allow upload of files with the names of the structures (as defined in AAPM TG-263).
3. Enable the use of protocol-specific scripts, including standard target and structure names (AAPM TG-263).
4. Create interfaces that import the necessary standard names, beam arrangement (if appropriate), and other information for treatment planning.
5. Create the appropriate software to allow automatic anonymization with coded ID labels of patients and plans.
6. Develop and make available a straightforward export of information to QA centers.
7. Make treatment planning systems IHE-RO compliant.
8. Enable the creation of tools or scripts that can be shared and then used at the local institution to assess protocol compliance are invaluable.

Patient Immobilization, Target Definition, and Treatment Guidance

1. Make immobilization devices that enhance reproducibility of patient setup over time so serial images can be used for quantitative treatment assessment and subsequent treatment planning.
2. Incorporate interchangeable fiducials in the immobilization devices to facilitate merging the scans from two or more types of instruments, such as MRI, CT, and PET.
3. Develop tools to quantitatively review localization images with field outline and anatomy contours exported from the treatment management system.
4. Develop tools to quantitatively monitor daily setup correction trends for patient positioning, such as from on-board imaging or other methods.