TG113: Physics Practice Standards for Clinical Trials

Jean M. Moran, Ph.D.

Chair of AAPM TG113

March 19, 2013

Challenges

- Challenges to multi-institutional clinical trials
  - Accrual, time, expense
  - Variability in technology and its implementation at different centers
- Physics issues are not always explicitly included in the design of clinical trials
- These exclusions may dilute the quality of data from a clinical trial
  - Treatment plan quality
  - Imaging
  - Patient setup details
  - Technique variations

Goals

1. To present the goals of an AAPM Task Group (113) on Physics Practice Standards for Clinical Trials
2. To describe mechanisms for ensuring quality in RT trials, especially those involving advanced technologies
   - Dose delivery
   - Treatment planning system implementation
   - Protocol testing – dry runs and interventional reviews
3. To illustrate the clinical impact of compliance with clinical trials

AAPM Task Group 113:
Physics Practice Standards for Clinical Trials

- The goal of the report is to increase the consistency of the physics aspects in each part of the treatment planning and delivery process.

Members

- Robert Dryzmala
- Jon Kruse
- Jean Moran (Chair)
- Art Olch
- Mark Oldham
- Robert Jeraj

Liaisons

- James Galvin – RTOG
- Andrea Molineu – RPC
- Jatinder Palta – TG100
- James Purdy – ITC
- Marcia Urie – QARC

Charge of AAPM Task Group 113

- Identify physics practice standards that impact the quality of data for clinical trials and the treatment of patients in the imaging, planning, and delivery chain
- Propose achievable standards of accuracy for each part of the chain based on published reports
- Provide guidance to physicists, QA organizations, and those who design clinical trials on addressing issues in radiotherapy that are most likely to cause inconsistencies in treatment

Expected Users of TG113

Physicists & Others

Implement trials at department level

Involved in all parts of the treatment planning and delivery process

Can significantly improve data consistency if guidelines are available

QA Organizations

Credentialing

Benchmarks - Dry run, phantom

Cooperative Groups

Trial Design

Make trials more specific with respect to physics aspects

Vendors

Data export: Dosimetric, imaging, localization

Dose calculation quality – heterogeneity corrections

Plan assessment tools

Delivery software and devices

Equipment

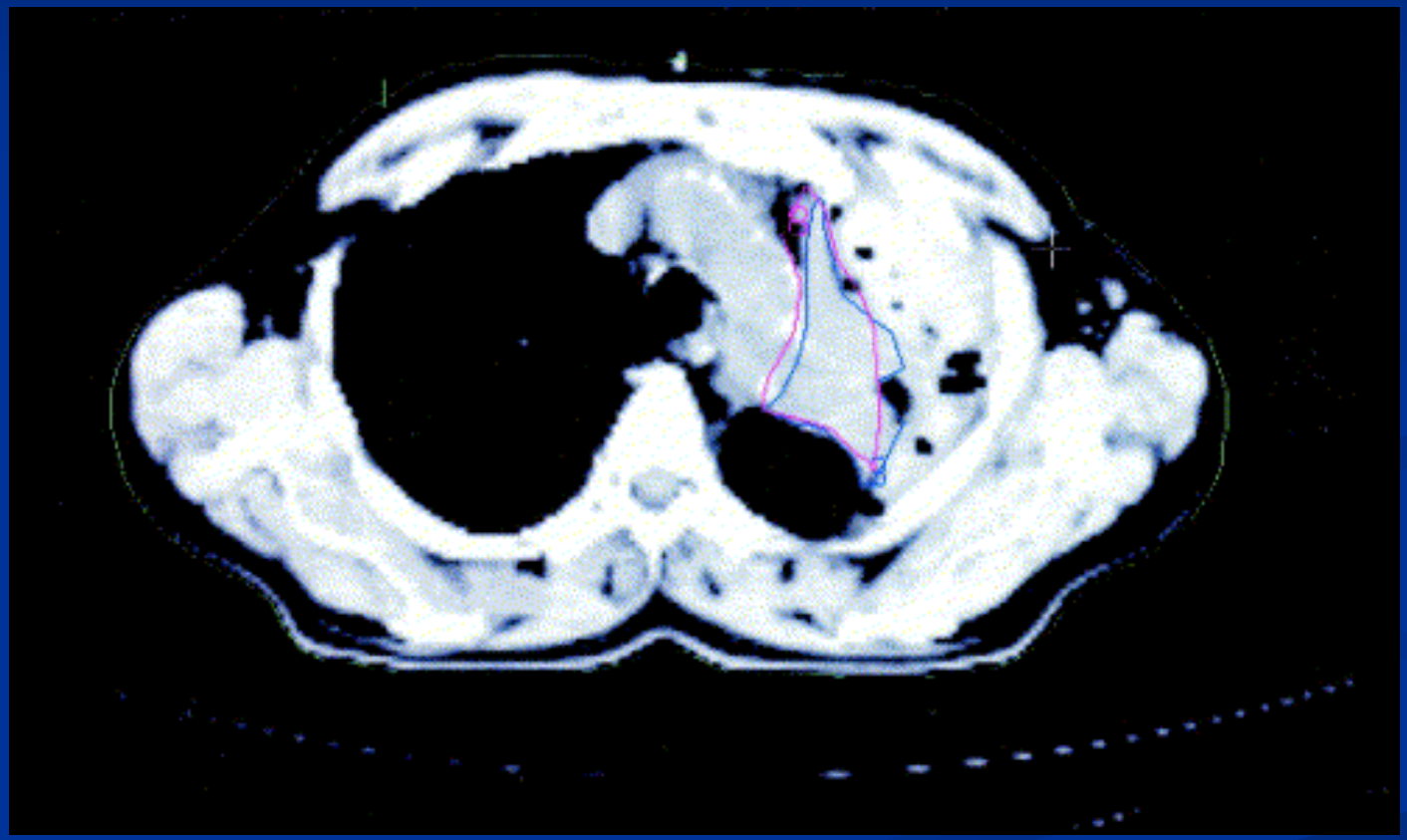
Factors Impacting Data for Clinical Trials

- Patient immobilization
- Imaging for Volume definition
- Patient localization: Setup accuracy
- Treatment planning
- Treatment guidance
- Treatment delivery

- What is the clinical trial designed to study?
- What are the clinical endpoints?
- What data are collected?
- What level of accuracy is required for the treatment planning and delivery chain?
- Will we want to know more later? e.g. modeling of lung response with normal tissue complication probabilities

Importance of Explicit Instructions Regarding Target Definition

Contouring Targets: Effect of Window on Target Volume



Bowden et al. Measurement of lung tumor volumes using three-dimensional computer planning software, IJROBP 53: 566-573, 2002.

Contouring Targets: Effect of Window on Target Volume

Bowden et al. Measurement of lung tumor volumes using three-dimensional computer planning software, IJROBP 53: 566-573, 2002.

	Range (Individual)	Over all patients
Mean Mediastinal Window	23.8-188.7	113.0
Mean Lung Window	47.7-226.7	169.2
Ratio	1.07-2.01	1.5

To improve consistency in trials

Protocols can provide guidance for window/gray levels in the trial design

SAMs Question 1

Bowden et al. compared mean tumor volume values for volumes drawn with a lung window compared to a mediastinal window for 6 physicians on 6 datasets.

The average ratio of the volumes using the lung window compared to the mediastinal window (LW/MW) over all patients was:

a) 0.85
b) 1.0 (no difference)
c) 1.2
d) 1.5
e) 2.2

SAMs Answer

Bowden et al. compared mean tumor volume values for volumes drawn with a lung window compared to a mediastinal window for 6 physicians on 6 datasets.

The average ratio of the volumes using the lung window compared to the mediastinal window (LW/MW) over all patients was:

a) 0.85

b) 1.0 (no difference)

c) 1.2

d) 1.5

e) 2.2

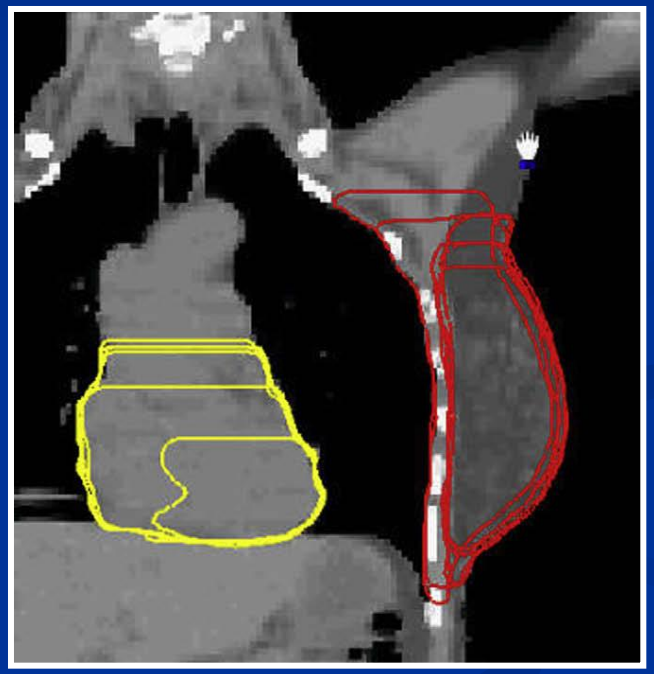
Correct answer: (c) – variation up to 1.5.

Reference: Bowden et al IJROBP 53: 566-573, 2002.

Importance of Atlases for Contouring Target and Organs at Risk Contouring

Radiation Therapy Oncology Group: Breast Group

Contours by 9 physicians from 8 institutions. Structure overlaps as small as 10%. Volumes with standard deviations as high as 60%.



Li et al. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study; IJROBP, 2009.

Impact of Atlas on Consistency of Contours

No Contrast

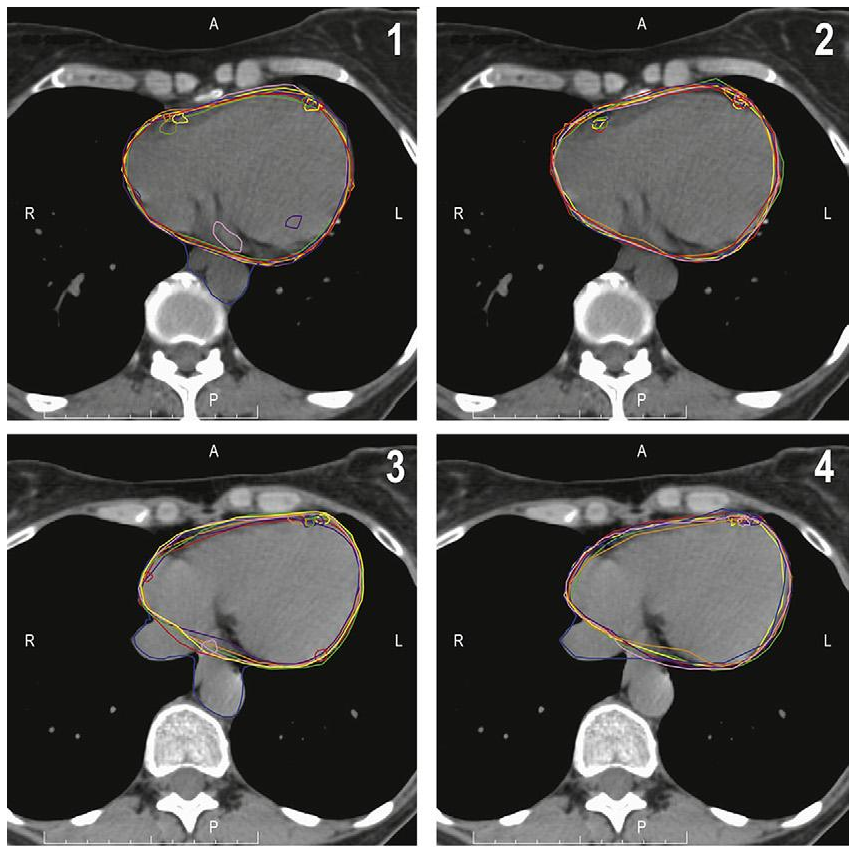
Pre-atlas

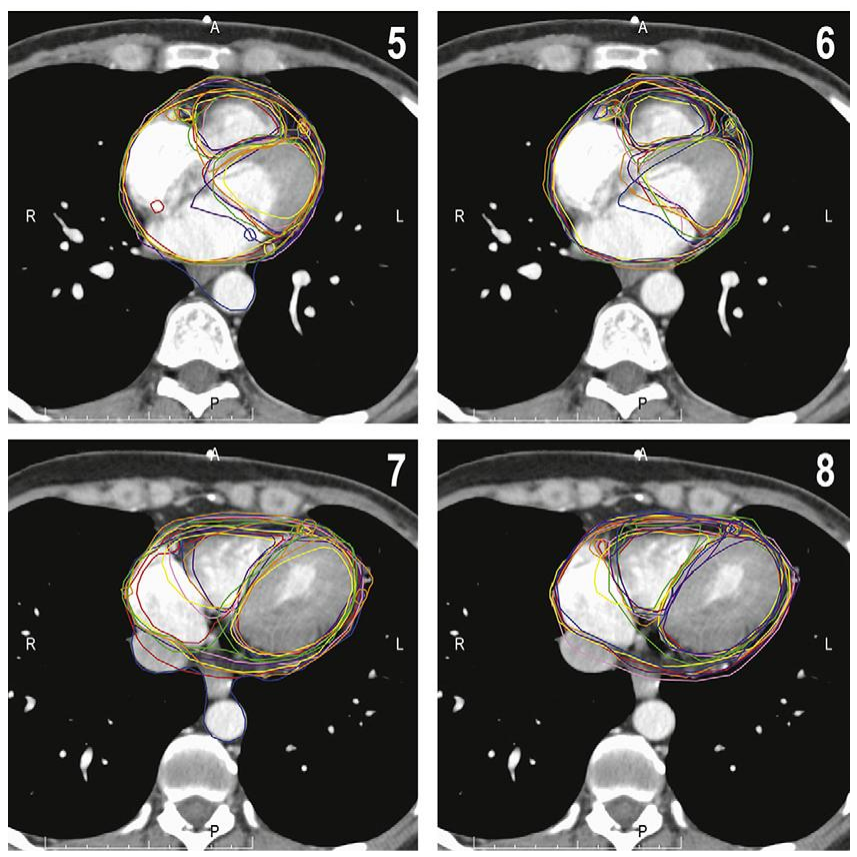
Post-atlas

Contrast

Pre-atlas

Post-atlas





Feng et al, Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer, IJROBP 79: 10-18, 2011.

Concordance of Structure Contours: Pre and post-atlas

Table 1. Percent overlap of observer and GS contours

Structure	Pre-atlas mean% ± SD% (% range)	Post-atlas mean% ± SD% (% range)	*p* value
Heart	79 ± 13 (50–99)	91 ± 4 (73–99)	<0.001
Left main coronary artery	10 ± 22 (0–72)	22 ± 20 (0–67)	<0.001
LAD artery	35 ± 21 (0–77)	62 ± 16 (28–89)	<0.001
Right coronary artery	11 ± 14 (0–49)	24 ± 18 (0.2–59)	0.002
Left ventricle	87 ± 11 (62–99)	92 ± 6 (79–99)	0.06
Right ventricle	65 ± 10 (55–80)	74 ± 8 (61–88)	0.003

Abbreviation: SD = standard deviation.

Structure	Post-atlas/Pre-atlas
Heart	1.15
Left main coronary artery	2.2
LAD artery	1.77
Left ventricle	2.18
Right ventricle	1.06

Feng et al, Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer, IJROBP 79: 10-18, 2011.

SAMS Question 2

The RTOG and independent investigators have created a number of atlases to better support consistency in clinical trials. Li et al evaluated the accuracy of multiple targets and organs-at-risk.

What was the smallest amount of overlap between volumes defined by different investigators in %?

a) 5%
b) 10%
c) 15%
d) 20%
e) 25%

SAMS Answer

The RTOG and independent investigators have created a number of atlases to better support consistency in clinical trials. Li et al evaluated the accuracy of multiple targets and organs-at-risk.

What was the smallest amount of overlap between volumes defined by different investigators in %?

a) 5%

b) 10%

c) 15%

d) 20%

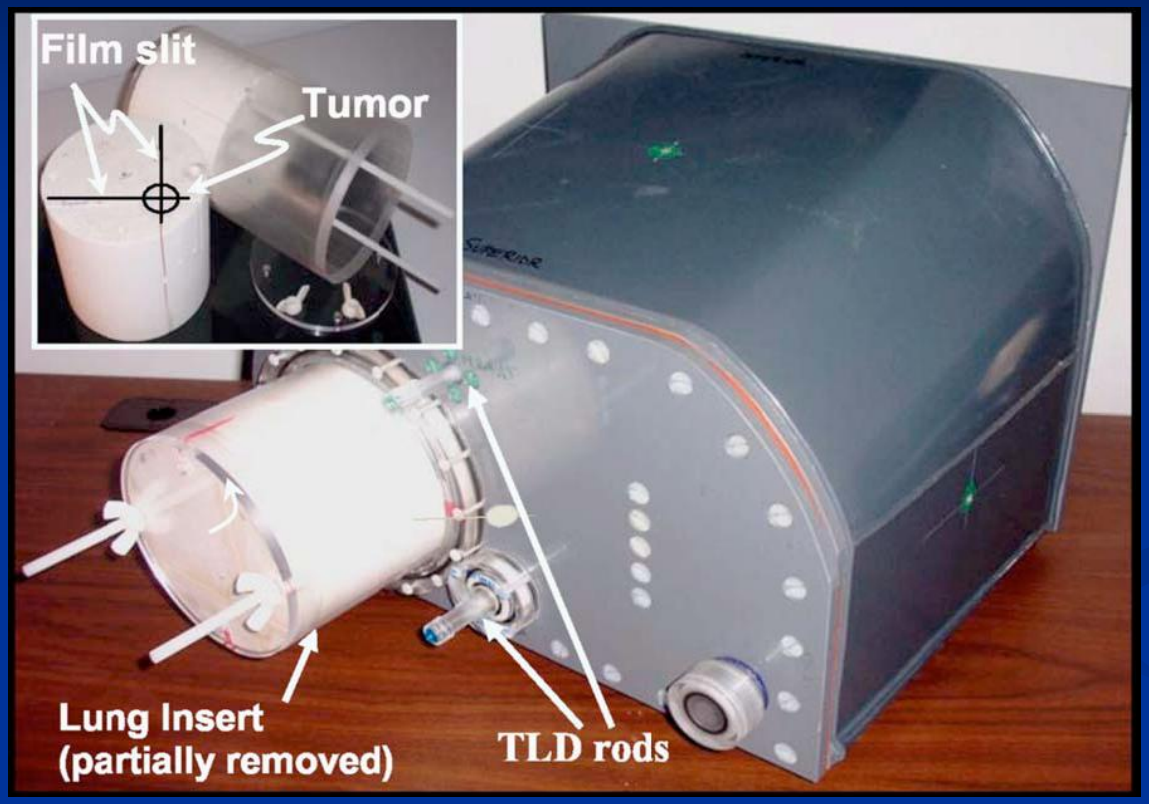
e) 25%

Correct answer: b) 10% There was significant variability among investigators.

Reference: Feng et al, Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer, IJROBP 79: 10-18, 2011.

Accuracy of Delivery

Thorax Phantom - Benchmark



- Designed for verification of stereotactic body irradiation
- Can be placed on a motion stage
- Supports reproducible positioning of TLDs and film

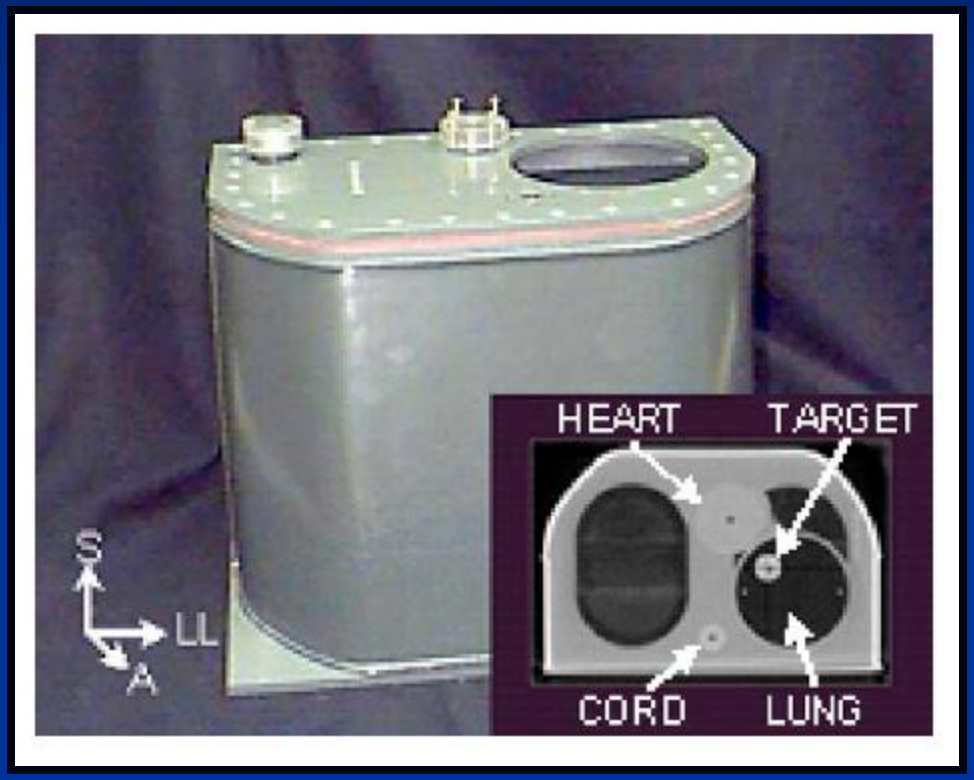
Figure 6, Followill et al Medical Physics 34: 2070-2076, 2007.

Thorax Phantom - Benchmark

- Designed for verification of stereotactic body irradiation
- Can be placed on a motion stage
- Supports reproducible positioning of TLDs and film

See Figure 6, Followill et al Medical Physics 34: 2070-2076, 2007.

RPC Thorax Phantom – Dose Calculation Accuracy

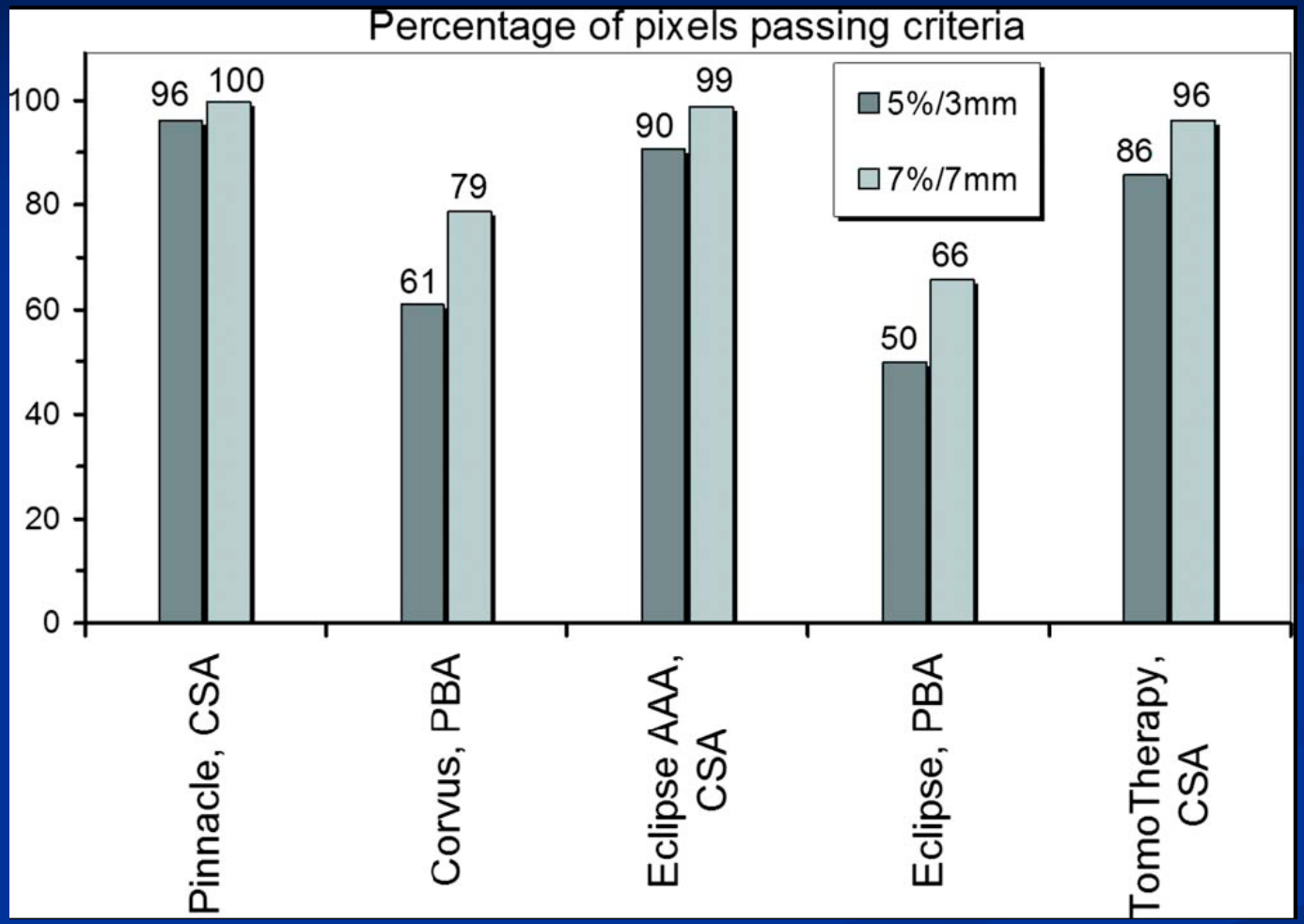


Commercial Systems tested
-3 convolution/superposition or an analytical method
-2 pencil beam algorithms

TLDs and radiochromic film used to measure the doses

Davidson et al. Technical note: Heterogeneity dose calculation accuracy in IMRT: study of five commercial treatment planning systems using an anthropomorphic thorax phantom. Medical Physics 35: 5434-5439, 2008.

Thorax Phantom – Dose Calculation Analysis



Davidson et al. Technical note: Heterogeneity dose calculation accuracy in IMRT: study of five commercial treatment planning systems using an anthropomorphic thorax phantom. Medical Physics 35: 5434-5439, 2008.

Thorax Phantom – Dose Calculation Analysis

Without a credentialing mechanism, institutions could prescribe a given dose as per the protocol but the delivered doses to the targets and normal structures could vary in a way that adversely affects the trial results.

Consistency is important. Early SBRT lung trials had no heterogeneity corrections due to the variability and the lack of widespread commercial availability of advanced algorithms.

RTOG Clinical Trials and Heterogeneity Corrections

- Modern algorithms with improved heterogeneity corrections are available in more commercial planning systems
- In 2011, the Medical Physics Committee of the RTOG determined that modern algorithms should be used for trials in the thorax
  - Convolution/superposition and Monte Carlo

SAMS Question 3

The RPC developed a thorax phantom for credentialing for trials involving heterogeneities. Convolution/superposition and Monte Carlo algorithms agree within approximately 5% with measurements for the criteria of 7% dose and 7 mm distance.

What was the approximate percentage range of pixels passing those criteria for pencil beam algorithms?

a) 30-40%
b) 50-60%
c) 70-80%
d) 80-90%
e) 90-95%

SAMS Answer

The RPC developed a thorax phantom for credentialing for trials involving heterogeneities. Convolution/superposition and Monte Carlo algorithms agree within approximately 5% with measurements for the criteria of 7% dose and 7 mm distance.

What was the approximate percentage range of pixels passing those criteria for pencil beam algorithms?

a) 30-40%
b) 50-60%
c) 70-80%
d) 80-90%
e) 90-95%

Correct answer (b).

Reference: S. E. Davidson, R. A. Popple, G. S. Ibbott, and D. S. Followill. Technical note: Heterogeneity dose calculation accuracy in IMRT: Study of five commercial treatment planning systems using an anthropomorphic thorax phantom. Medical Physics 35: 5434-5439, 2008.

SAMS Answer (continued)

- This is why the RTOG Medical Physics Committee requires modern algorithms (such as CV/SP, Monte Carlo, and deterministic) to be used for advanced technology trials funded by the National Cancer Institute.

SAMS Question 4

The National Cancer Institute requires that institutions participate in a remote dosimetry audit that has been conducted by the Radiological Physics Center (RPC).

Participation in the remote dosimetry audit has been required by the NCI because:

a) It limits the variability in dose delivery for institutions participating in external beam trials.

b) It is considered an important safety check for participants.

c) It can be used in place of phantom irradiations for all trials.

d) RPC data have shown little variability in the output so it is not a requirement for participation in NCI-sponsored trials.

SAMS Answer

The National Cancer Institute requires that institutions participate in a remote dosimetry audit that has been conducted by the Radiological Physics Center (RPC).

Participation in the remote dosimetry audit has been required by the NCI because:

a) It limits the variability in dose delivery for institutions participating in external beam trials.

b) It is considered an important safety check for participants.

c) It can be used in place of phantom irradiations for all trials.

d) RPC data have shown little variability in the output so it is not a requirement for participation in NCI-sponsored trials.

Correct answer: (a) The audit is required to limit the variability in dose delivery.

Reference: Ibbott QA in Radiation Therapy: The RPC Perspective; Journal of Physics Conference Series 250 - IC 3D Conf Proceedings, 2010.

Credentialing for RT Clinical Trials:

Credentialing for RT Clinical Trials: Prior to patient enrollment

- Questionnaires – frequently general and tied to equipment, personnel, and delivery techniques for a given institution
- Dry run – tests ability of team members for one or more aspects of a trial
  - Contouring, image registration, treatment planning
- Phantom irradiations
  - Evaluates the team and system performance from simulation through delivery

Credentialing for Advanced Technology Trials

- Credentialing can be used to verify that:
  - The institution's team understands the protocol
  - The team can complete an end-to-end test from simulation through delivery
- Benchmark phantoms can be used to limit factors that can degrade the quality of a clinical trial
  - Failure of a phantom irradiation can be due to a weak link in any part of the chain from the imaging simulation to localization and delivery

Benchmarks

- As new technology becomes commercially available, it may be implemented clinically without a full understanding of the technology
  - It may also occur prior to the development of guidance or consensus documents on the QA of the involved software and hardware
- Therefore, credentialing for trials allowing use of advanced technologies may include performance of an end-to-end phantom test
  - This is to ensure consistency in the treatment for patients treated at multiple institutions

To improve consistency in clinical trials

- Clear role for dosimetric verification during clinical introduction of complex technologies
- RPC phantoms tests have found dosimetric errors that would degrade the quality of the clinical trial results
  - e.g. head and neck phantom – initially had a 30% failure rate

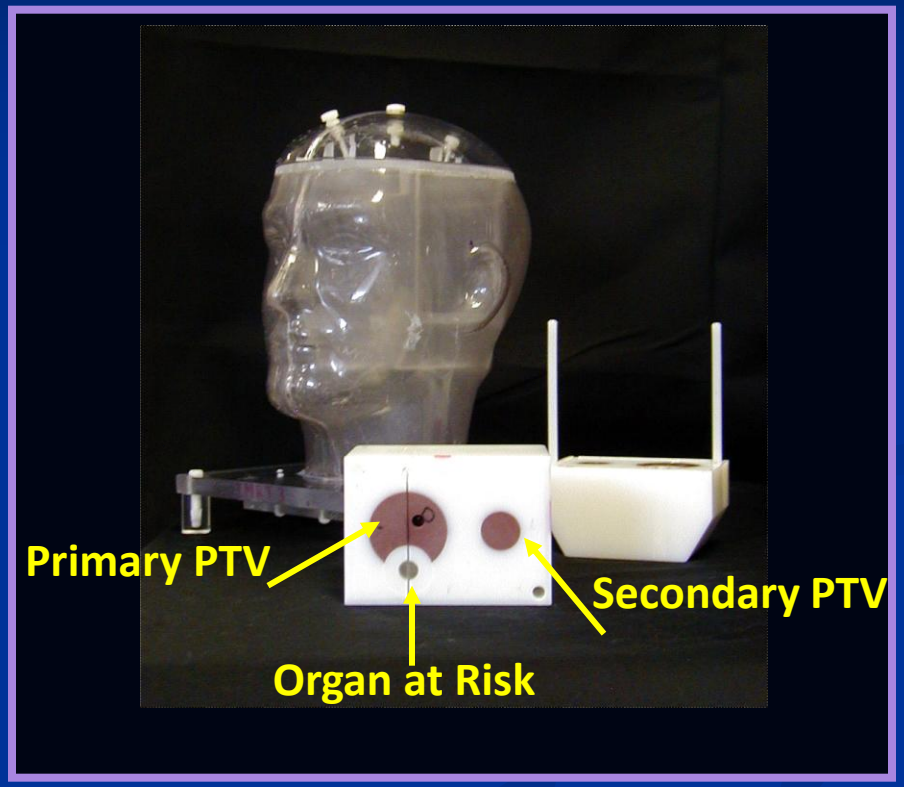


Image courtesy of Andrea Molineu

Self-testing for Clinical Trials

- Goals:
  - Straightforward tests that give the physicist a quick check of the acceptability of their system before full credentialing
  - Accumulate data and publish confidence intervals
- Problems during full credentialing can be difficult to diagnose
  - Format problems, e.g. with headers
  - Missing contours, etc.
  - Identify core behavior such as volume definitions as a quick pre-test by the clinical medical physicist
  - Credentialing often involves the whole team and problems can lead to rework

Implications for Future Trials

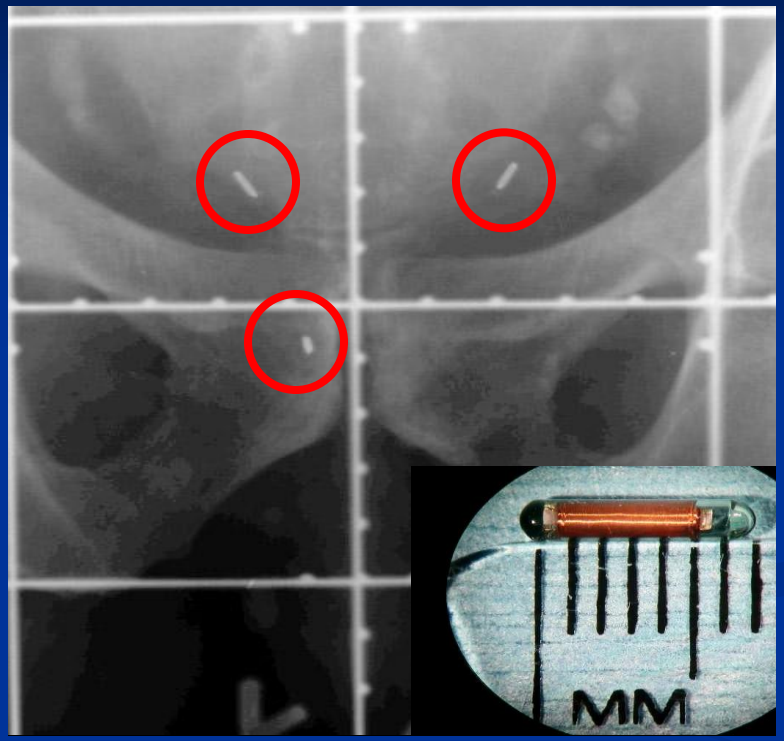
- The TG113 report recommends pre-testing where possible.
  - Specific tests should be done in advance of more complex benchmark tests.
  - Such tests are in use by some of the QA centers
- AAPM Task Group 119 provides example tests and results for IMRT
  - Common combinations of planning, delivery, and measurement systems
  - Results were collated and confidence intervals determined for the results
  - Individuals can conduct the same tests and compare to the results

Example Considerations with respect to Trial Design

Target Localization and Frequency of Treatment Guidance

- There are many methods commercially available for tracking the position of a target or of surrogate anatomy
  - MV or kV portal imaging
  - Cone beam CT
  - kV fluoroscopic imaging
  - Radiofrequency beacons
- Systems in use can vary by institution and by treatment site

Example: Treatment Guidance



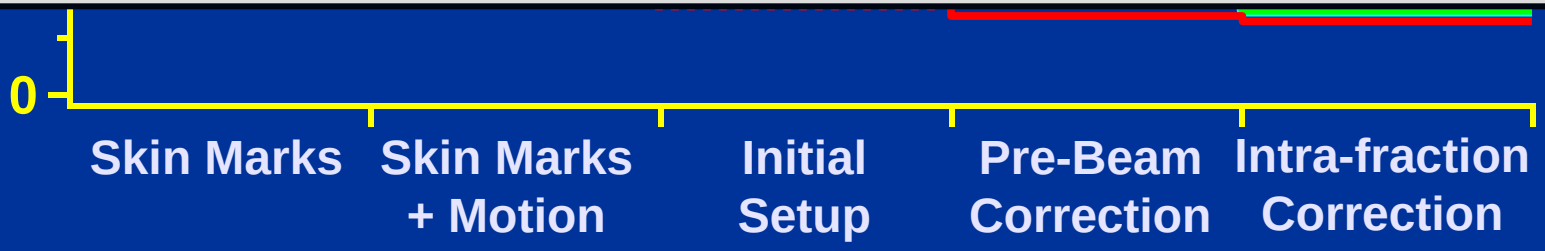
- 3 radiofrequency transponders implanted transrectally in the prostate under ultrasound guidance
- Good positional stability over 8 weeks (σ_{ave} = 0.8 mm)

Image courtesy of Dale Litzenberg

Prostate: Necessary PTV margin

This demonstrates that the PTV margin depends on the localization method, tolerance level, and frequency of intervention.

Note: This does not consider changing anatomy or deformation of organs over time.



Litzenberg Int J Radiat Oncol Biol Phys 2006

Courtesy of Dale Litzenberg

To improve consistency in trials:

- Margins should be based on each institution's image guidance protocols

or

- The trial should specify the acceptable methods and frequency of interventions
- Patient enrollment information should include immobilization method and localization information
  - Frequency
  - Action level
- There may be further changes as more adaptive protocols are conducted

Importance of Credentialing for Multi-institutional Clinical Trials

Credentialing

- Credentialing tasks should be performed by those who will participate in the protocol
  - Therapists for simulation and delivery for tests that include those roles (e.g. experiments involving phantoms)
  - Physician – contouring and guidance in treatment planning and delivery
  - Dosimetrist and physicist for treatment planning, image registration, and final QA
- At the institution, the team should review its procedures and document any changes from their standard practice for a given protocol

Rapid or Interventional Review

- Data review of all volumes and treatment plans for patients enrolled
  - Can be done for the first few patients enrolled on a study or for all patients
  - Frequency is determined at the time of protocol design
- Frequently involves a QA center such as QARC or RTOG; data may be submitted to the ITC in DICOM RT format
  - An expert performs a review of the information from the institution on a patient by patient basis
  - Feedback is given to the institution if not in compliance and data are resubmitted

Rapid Review for QA of Patient Data

- RTOG 0415/NSABP 39 – hypo-fractionated partial breast irradiation trial had a pre-treatment review
  - Contoured volumes (target and normal organs) and dose volume histograms were reviewed by RPC staff
  - Of the 99 patients submitted, 66% had to be resubmitted. 11% of those cases were due to a dose deviation that exceeded the acceptance criteria.
- Use of rapid review resulted in improved integrity of the data submitted for the trial

Leif et al RPC/MDACC, IJROBP S617, 2009.

Impact of Protocol Compliance on Patient Outcome in Prospective Clinical Trials

Clinical Impact of Interventional Review

- TROG (Trans-Tasman Radiation Oncology Group 02.02 trial for Head and Neck
  - Phase III study including radiation therapy with chemotherapy: investigating addition of Tirapazamine (TPZ) to cisplatin-based chemotherapy
- Radiation therapy (2D or 3DCRT) doses were the same for both arms
  - 70 Gy in 35 fractions; sites near gross disease treated to 60 Gy; sites with subclinical disease to 50 Gy
- Diagnostic scans and treatment plans were submitted to QARC and reviewed within 5 days
- Any required modifications were sent back to the investigators; additional reviews were performed if necessary

Peters et al JCO 28: 2996-3001, 2010.

Clinical Impact of Interventional Review

- Final study population: 853 patients
  - 84% of these had an interventional review
  - 820 of these patients had enough data for further blinded review
- Of the 820 evaluated, 208 (25%) were not compliant with the protocol
- A secondary analysis was performed on 206 of those records
  - Blinded review found that 97 of the cases (47%) of the non-compliant volumes and treatment plans would likely have an adverse impact on tumor control
  - Review included redrawing target volumes if necessary and re-evaluating dose distributions
- Further analysis of locoregional and overall survival was performed based on this sub-analysis

Peters et al JCO 28: 2996-3001, 2010.

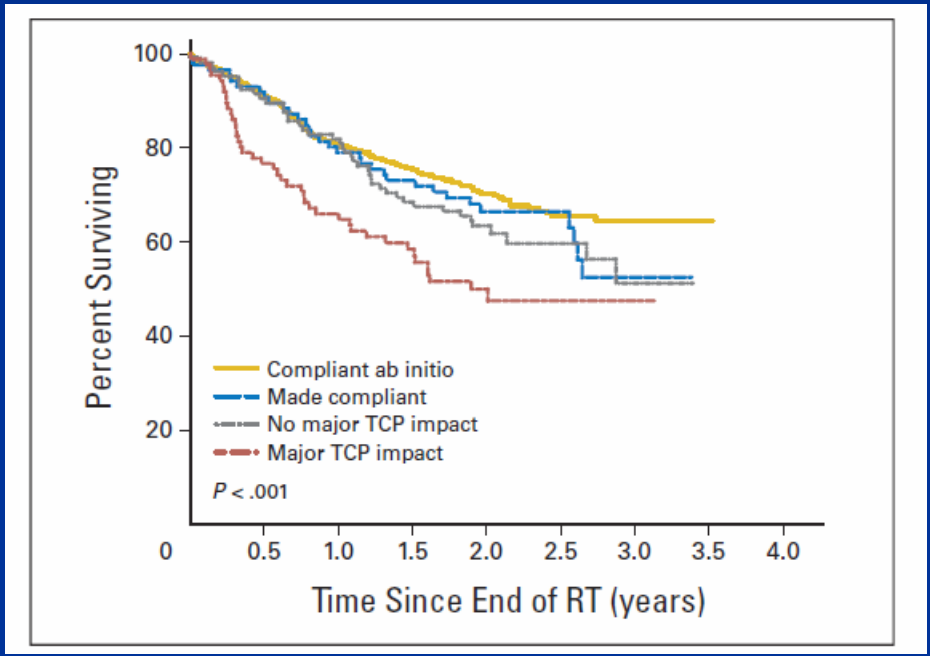
Clinical Impact of Interventional Review

- Expected dose delivered from treatment plans was derived for the revised volumes
  - Factors include minimum and maximum doses to the gross tumor volumes, planning target volumes, and regions adjacent to tumor volumes
- Treatment time ≤ 9 weeks

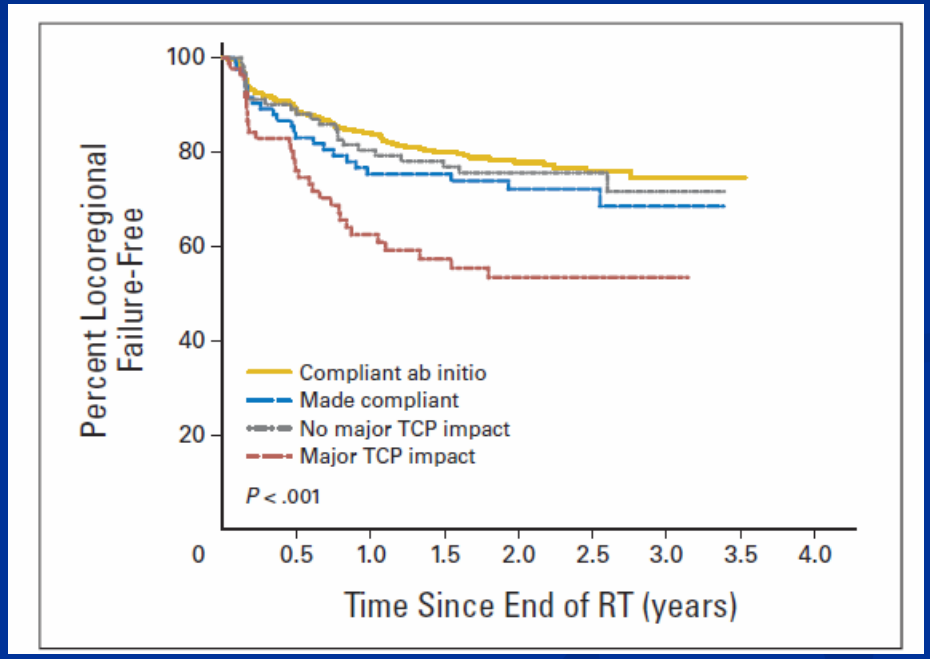
Peters et al JCO 28: 2996-3001, 2010.

TROG 02.02: Protocol Compliance

Overall Survival



Time to locoregional failure



Peters et al JCO 28: 2996-3001, 2010.

TROG 02.02: Protocol Compliance

The secondary review found that 47% of the noncompliant plans were determined to have an impact on the tumor control probability.

Correlation with the number of patients enrolled on the study at the center (higher likelihood of non-compliance if less than 5 patients enrolled).

When at least 60 Gy delivered, the 2 year overall survival was 70% for compliant plans vs. 50% for plans with major deviations.

Impact of Protocol Guidelines on Quality

- Definition of dose-volume coverage for targets and normal tissues permits consistency in treatment of patients
- Definition of major and minor deviations has permitted additional evaluation of the outcome results
- For HN cancers, major deviations from the study protocol have resulted in poorer prognosis for those patients
- Quality measures work together to ensure consistent treatment of patients
  - These concepts can be applied in a multi-institutional setting

Importance of Adhering to the Protocol

- RTOG 9704 – Phase III of adjuvant chemotherapy and chemo-radiation therapy for resected adenocarcinoma of the pancreas
- Scored adherence to protocol independent of patient outcome
  - 216 per protocol vs. 200 did not adhere to protocol

Abrams et al IJROBP doi:10.1016/j.ijrobp.2010.11.039; In press

Importance of Adhering to the Protocol RTOG 9704 (Pancreas)

- Patients were randomly assigned to pre- and post-chemotherapy arm vs. gemcitabine
  - Stratified by tumor diameter, nodal status, surgical margins
  - RT: 45 Gy to tumor bed and regional lymph nodes; boost 5.4 Gy
- Adherence to the protocol
  - Resulted in improved survival: 1.74 (protocol) vs. 1.46 years (did not adhere to protocol)
  - Trend towards less toxicity in patients receiving gemcitabine

Abrams et al IJROBP 82: 809-816, 2012.

Importance of Protocol Adherence: RTOG 9704 – Phase III Adjuvant Chemo and ChemoRT – resected adenocarcinoma of the pancreas

Table 1. Criteria for assessing radiotherapy field size parameters

Initial fields (180–4,500 cGy)	Per protocol	Variation acceptable	Variation unacceptable
AP/PA fields			
a) Length	4–5 VB		<4 or >5 VB
b) Distance from primary tumor bed	2–3 cm	1 cm to <2 cm >3 cm to <4 cm	<1 cm or >4 cm
c) Distance from VB edge	2 cm	1 cm to <2 cm >2 cm to <3 cm	<1 cm or >3 cm
Lateral fields			
a) Length	Same as for AP/PA fields		
b) Distance from anterior tumor bed	1.5–2 cm	0.5 cm to <1.5 cm >2–3 cm	<0.5 cm or >3 cm
c) Distance from rest of tumor bed	2–3 cm	1 cm to <2 cm >3–4 cm	<1 cm or >4 cm
d) Posterior edge	Mid VB	Within 1 cm of mid VB	>1 cm away from mid VB
e) Distance from anterior field edge to VB	3.5–4 cm	2.5–<3.5 cm >4–5 cm	<2.5 cm or >5 cm
Boost fields (4,680–5,040 cGy)			
Distance from tumor bed	1.5–2 cm	1 cm to <1.5 cm >2–3 cm	<1 cm or >3 cm
Other parameters	Per protocol	Variation acceptable	Variation unacceptable
Dose delivered	±5%	Not within ±5% Within ±10%	Not within ±10%
Elapsed time	<7 days' break	8–14 days' break	>14 days' break

Abbreviations: AP = anteroposterior; PA = posteroanterior; VB = vertebral bodies.

Abrams et al IJROBP 82: 809-816, 2012.

RTOG 9704 (Pancreas)

All patients were evaluated with respect to compliance with the protocol.

Median survival (p=0.0077):

- **1.75 years per protocol**
- **1.46 years not protocol compliant**

SAMS Question 5

A secondary analysis of the TROG head and neck protocol study investigated protocol compliance and its impact on tumor control probability.

For patients receiving at least 60 Gy, they found 70% overall survival for patients whose treatment plans were compliant with the protocol compared to what % overall survival for those whose plans had major deviations?

a) 70% - No difference between the groups

b) 55-65% - A trend towards difference but not statistically significant

c) 50% (statistically significant)

d) 35% (statistically significant)

e) 20% (statistically significant)

SAMS Answer

A secondary analysis of the TROG head and neck protocol study investigated protocol compliance and its impact on tumor control probability.

For patients receiving at least 60 Gy, they found 70% overall survival for patients whose treatment plans were compliant with the protocol compared to what % overall survival for those whose plans had major deviations?

a) 70% - No difference between the groups

b) 55-65% - A trend towards difference but not statistically significant

c) 50% (statistically significant)

d) 35% (statistically significant)

e) 20% (statistically significant)

Peters et al Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. J Clin Oncol 2010;28:2996–3001.

Summary

- AAPM TG113 report focuses on improving consistency in all parts of the process
  - The principles can be applied for different applications
- During trial design, all parts of the process should be evaluated to determine the level of accuracy and consistency required
- Protocol guidelines should be explicit
  - e.g. extent of regions to be imaged, treatment plan requirements, approved modality, credentialing requirements, and dose-volume criteria with minor and major deviations
- Digital download of imaging and treatment planning information is essential

Summary (continued)

- Rapid or interventional review can have an impact on the study quality
  - It is invaluable to have analysis tools that can be used remotely by study coordinators independent of a given commercial platform
- Trials must be able to be implemented in a wide range of practice settings
  - Supplemental information customized to the protocol
  - In RT, the use of phantoms has been invaluable to establish a level of quality for the planning and delivery process with advanced technologies
- Quality and explicit guidelines are important for the entire treatment planning and delivery process

Acknowledgments

- Members of AAPM TG113
- Andrea Molineu
- Dale Litzenberg