



Published in final edited form as:

*Pract Radiat Oncol.* 2020 ; 10(5): 372–381. doi:10.1016/j.prro.2020.04.004.

## Differences in the patterns of failure between IROC lung and spine phantom irradiations

Sharbacha S. Edward, BS<sup>1,2,3</sup>, Paola E. Alvarez, MS<sup>2,3</sup>, Paige A. Taylor, MS<sup>1,2,3</sup>, H. Andrea Molineu, MS<sup>2,3</sup>, Christine B. Peterson, PhD<sup>1,4</sup>, David S. Followill, PhD<sup>1,2,3</sup>, Stephen F. Kry, PhD<sup>1,2,3</sup>

<sup>1</sup>The University of Texas MD Anderson Cancer Center UT Health Graduate School of Biomedical Sciences, Houston TX 77030

<sup>2</sup>IROC Houston Quality Assurance Center, The University of Texas MD Anderson Cancer Center, Houston TX 77030

<sup>3</sup>Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston TX 77030

<sup>4</sup>Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston TX 77030

### Abstract

**Purpose:** To investigate and classify the reasons why institutions fail the Imaging and Radiation Oncology Core (IROC) SBRT spine and moving lung phantoms, which are used to credential institutions for clinical trial participation.

**Methods:** All IROC moving lung and SBRT spine phantom irradiation failures recorded from January 2012 to December 2018 were evaluated in this study. A failure was a case where the institution did not meet the established IROC criteria for agreement between planned and delivered dose. We analyzed the reports for all failing irradiations, including point dose disagreement, dose profiles, and gamma analyses. Classes of failure patterns were created and used to categorize each instance.

**Results:** There were 158 failing cases analyzed: 116 of 1052 total lung irradiations and 42 of 263 total spine irradiations. Seven categories were required to describe the lung phantom failures, while four were required for the spine. Types of errors present in both phantom groups included: “systematic dose” and “localization” errors. Fifty percent of lung failures were due to a superior-inferior (SI) localization error, i.e., error in the direction of major motion. Systematic dose errors however, contributed to only 22% of lung failures. In contrast, the majority (60%) of spine

---

**Corresponding author:** Stephen F. Kry, sfkry@mdanderson.org, 713-745-8983, 8060 El Rio St. Houston, TX, 77054.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Conflict of interest statement:** none

phantom failures were due to systematic dose errors, with localization errors (in any direction) accounting for only 14% of failures.

**Conclusion:** There were two distinct patterns of failure between the IROC moving lung and SBRT spine phantoms. The majority of the lung phantom failures were due to localization errors, whereas the spine phantom failures were largely attributed to systematic dose errors. Both of these errors are clinically relevant and could manifest as errors in patient cases. These findings highlight the value of independent end-to-end dosimetry audits, and can help guide the community in improving the quality of radiation therapy by focusing attention on where errors manifest in the community.

---

## INTRODUCTION

Radiation therapy is used to treat approximately 50% of cancer cases in the United States[1]. However, it has been repeatedly shown that in order for this treatment to be effective, the correct dose must be delivered to the correct treatment site, otherwise, overall patient survival decreases dramatically[2, 3]. To maintain a high standard of quality and consistency in the radiotherapy community, and particularly among the National Cancer Institute's (NCI's) radiotherapy clinical trials, participating centers across the US and abroad irradiate patient surrogates (phantoms) from the Imaging and Radiation Oncology Core (IROC)[4]. These phantoms comprise tissue equivalent material, along with thermoluminescent dosimeters (TLDs) and radiochromic film, to measure dose delivered to targets and critical structures. These phantom irradiations evaluate an institution's ability to deliver the planned dose correctly.

The IROC head and neck phantom, which assesses intensity modulated radiation therapy (IMRT) performance, was previously evaluated on several criteria including the patterns of failure of the 10% of irradiations that did not meet IROC's acceptability criteria[5]. It was found that the majority of failing phantom irradiations were due to an incorrect dose (i.e., correct shape of the dose distribution in the correct location, but of the wrong magnitude)[5], and that these were associated with inaccuracies in the institutional dose calculation[6]. Modern radiotherapy goes well beyond IMRT, and IROC phantoms have been established to test these other elements, including the moving lung and SBRT spine phantoms (fig. 1). These two phantoms have recorded failure rates of 13% and 17% respectively between 2012 and 2018. These failures translate to a substantial number of institutions, and correspondingly patients, who may be receiving clinically suboptimal treatments. Also, institutions that successfully undergo the credentialing process, are generally better prepared for compliance with clinical trial protocol requirements[7]. To better understand the nature of these failing cases, and thereby to begin any rectification that may be appropriate at the corresponding institutions, we first need to better understand the nature of the failures. The purpose of this study was to evaluate the moving lung and SBRT spine phantom irradiation failures and classify them according to the nature of the failures. Results of this study are critical for the entire radiation oncology team in order to understand the risks and challenges of delivering high quality radiation therapy.

## METHODS

### Phantom Design

The IROC anthropomorphic thorax phantom shell (fig.1) is used to perform both lung and spine irradiations. This phantom is heterogeneous in order to simulate actual patient anatomy[8], and specific lung and spine target and dosimetry inserts are used based on the anatomical site being irradiated (either a lung or spine insert). The target in the lung insert is an ovoid structure, measuring 5 cm in length and 3 cm in diameter, located in the center of low density tissue equivalent material. The target in the spine insert mimics the shape of the vertebral body. Abutting the target is a structure representing the vertebral foramen, within which the organ at risk, the spinal cord, is contained, only 0.8 cm posterior to the edge of the target. The materials that make up this phantom include compressed cork for the lungs, nylon for the heart, polybutylene terephthalate-polyester for the spine, and polystyrene for the tumors [9]. The shell is also filled with water in order to represent soft tissue.

The lung phantom treatments are either static or include motion. Static lung treatments, despite not being representative of a typical patient lung treatment in terms of motion, test other aspects of the treatment process such as heterogeneity corrections and dose delivery. To simulate motion for gated and free-breathing/ internal target volume (ITV) treatments, the lung phantom is placed on a moving platform [10]. In the superior-inferior (SI) direction, the platform (and phantom) moves with a 2 cm amplitude. The breathing cycle contains 2 distinct breaths based on clinical patient breathing patterns, and a cycle of both breaths takes 11 seconds to complete. The phantom also moves a total of 0.5 cm in the anterior-posterior (AP) direction during motion. The AP direction was considered as the direction of minor motion for the purposes of this study, since the extent of this motion was within the distance-to-agreement gamma criterion ( $\pm 5$  mm).

Thermoluminescent dosimeter (TLD) capsules with double-loads of powder are placed inside the targets and organs at risk in order to record the dose to these structures[11]. The lung phantom contains 2 centrally located TLD capsules within the 35 cm<sup>3</sup> target. The spine phantom contains 4 TLD capsules within the 22 cm<sup>3</sup> target. Radiochromic film is also placed in the phantom in orthogonal planes in order to measure dose distributions[11]. Institutions are directed to deliver 6 Gy to the target and treat it as they would any radiotherapy patient in terms of imaging, treatment planning, setup, and delivery.

### Irradiation criteria

Successful irradiation of a lung phantom is achieved by producing TLD measurements in the target that are each within  $\pm 7\%$  of the planned dose (over the TLD contours). Additionally, a film pass rate of 80% is required for each of the axial, coronal and sagittal film planes, and a combined average of 85% for the three planes, with a film gamma index of 7%/5 mm. The criteria for successful irradiation of a spine phantom is that each measured TLD dose agrees within  $\pm 7\%$  of the planned dose, and a film pass rate of 85% each for the axial and sagittal film planes is achieved, with a film gamma index of 5%/3 mm.

## Data collection

Failing phantoms were identified through the IROC phantom records database. For this study, 116 failing lung and 42 failing spine phantom reports were analyzed individually. This includes all attainable data for failing irradiations for these two phantoms from January 2012 to December 2018. Reports were abstracted for irradiation result as well as demographic information (treatment delivery unit, planning system, etc.)

## Phantom classification

We reviewed all phantom reports, including dose disagreements, profiles, and gamma analysis, in order to define categories and then categorize the patterns of failure. While most phantoms had a single clear cause of failure, some cases were found to contain multiple causes of failure (e.g., the dose was systematically low and the dose profile was shifted from target center). In such an instance, where the case would have failed given either one of the failures on its own, the case was counted twice: once in each of the failure categories represented. For a few phantoms, the failure was as the result of a culmination of causes, no single one of which on its own would have caused a failure. In these cases the failure mode was described as a “combination” and placed into the combination category.

## Data Analysis

Failure mode totals were calculated for each category. Due to the double counting of some phantoms, the failure-mode total was greater than the number of individual phantoms evaluated. A 95% confidence interval was calculated for the rate of failures due to each main failure category using the Wilson interval method. This method was used to assess the likelihood of phantom failures falling under each category, given the total number of phantoms in the study. Other criteria that were analyzed for patterns of failure include beam energy, machine model, treatment planning system (TPS) algorithm and treatment technique, with the addition of motion management technique for the lung phantom. The association between failure and this demographic data was analyzed using the Chi-square and Fisher’s exact tests.

## RESULTS

From January 2012 to December 2018, the lung phantom was irradiated 1052 times and recorded a failure rate of 13%. The SBRT spine phantom was irradiated 263 times and recorded a higher failure rate of 17%. For this study, all available failing phantom records, totaling 116 lung (82% of failures) and 42 spine (91% of failures) were evaluated and categorized based on failure type. The 116 failing lung phantom cases were from 106 different institutions, with 7 institutions repeating the phantom 2 or more times. The 42 failing spine cases were from 33 different institutions, also with 7 institutions repeating the phantom 2 or more times. Four institutions recorded failures in both phantoms.

While some phantom irradiation cases contained multiple error types, all cases were categorized based on their most egregious error type that ultimately caused that irradiation result to fall outside of the established IROC criteria.

## Category descriptions

Seven lung and four spine categories were formed. These categories are described as follows:

### Lung

1. **Systematic dose:** uniform overdosing or underdosing of the PTV
2. **Local dose:** dose error in an isolated area of the plan
3. **Localization – major motion:** dose distribution improperly aligned with target in the superior – inferior direction. The phantom moves 2.0 cm in this direction during the breathing cycle. This error is illustrated in fig 2 (top), where the measured dose profile is shifted almost 2 cm inferiorly to the institution’s planned dose profile.
4. **Localization – minor motion:** dose distribution improperly aligned with target in the anterior-posterior direction. The phantom moves 0.5 cm in this direction during the breathing cycle.
5. **Localization – no motion:** dose distribution improperly aligned with target in the left-right direction, in which the target does not move at all.
6. **Global Error:** grossly irregular dose distributions. This is illustrated in fig. 2 (middle), where the measured profile is of a different shape, contains an inferior shift and has a lower dose in half of the PTV relative to the planned profile.
7. **Combination Category:** contributions from two separate error types, not individually sufficient to each cause a failure, but when combined, caused the irradiation results to fall outside of criteria. Combinations which fell into this category included: SI localization + localization AP, systematic underdose + localization AP, SI localization+ ITV exaggeration (illustrated in fig. 2 (bottom)) and systematic underdose + ITV exaggeration. The combination component “ITV exaggeration” is a side effect (or consequence) of the ITV motion management technique, which affects the dose profile in the shoulder region, causing it to be more susceptible to otherwise negligible dose/localization effects.

### Spine

1. **Systematic dose:** uniform overdosing or underdosing of the PTV. This is illustrated in fig. 3 (top) where the measured dose profile has a lower value relative to the institution’s planned dose profile.
2. **Dose fall-off region:** dose error in the steep dose gradient between the PTV and spinal cord. This is illustrated in fig. 3 (middle) where the measured dose profile is of a higher value in the dose fall-off region relative to the institution’s planned dose profile.

3. **OAR overdose:** overdose of the spinal cord structure. This is illustrated in fig. 3 (bottom), where the measured dose is higher than planned in the spinal cord region, essentially causing an overdosing of that OAR.
4. **Localization:** dose distribution improperly aligned with the target

## Lung

Results of the various category assignments are summarized in table 1. The majority (62%) of the lung phantom failures were due to localization errors; however, almost all of these localization errors (64 out of 79) were in the direction of motion (SI), making up 50% of all lung phantom failures. The average observed error among both superiorly and inferiorly shifted phantoms was  $1.29 \pm 0.54$  cm, with the maximum recorded shift in each direction being 2.5 cm (four cases). Localization error in the anterior-posterior direction of motion was  $0.47 \pm 0.09$  cm on average, which was within the established gamma threshold of 0.5 cm. Systematic dose errors accounted for 28 (22%) of all lung failures, with 19 (16%) of those being due to an underdose and 9 (6%) due to an overdose. The remaining phantom failure modes comprised relatively few cases.

The phantom failures are classified by motion management technique in table 2. In the SI localization category, which was the most prominent error type, the failure rates were markedly elevated for cases irradiated using gated and free-breathing/ITV techniques (table 2). Failure rates in the SI localization category were 13% and 10% for gated and ITV cases respectively, versus 0.3% ( $p < 0.001$ ) for static phantom cases. In the other two major error categories: dose and localization in other directions (non-motion), the rates were fairly evenly distributed and were all at or below 5%.

## Spine

The majority (60%) of the spine phantom failures were due to systematic dose errors, with 9 (21%) of them being due to an overdose and 16 (38%) due to an underdose. The phantoms with a systematic underdose error had an average dose difference of  $-7.4\% \pm 2.4$ , and those with a systematic overdose error had an average dose difference of  $4.8\% \pm 1.8$ . The remaining phantoms were nearly evenly distributed among the remaining categories. Of note, the failure rate due to localization errors of the spine phantom (14%) was nearly identical to the failure rate due to localization errors of the lung phantom in the non-motion direction (12%).

## Additional analysis

Table 3 is a demographic display of the phantoms, grouping them by machine, energy, TPS algorithm, treatment technique and for the lung, respiratory motion management. The number of failing phantoms, and its value as a percentage of the total number of irradiated phantoms within the 2012–2018 time frame are included. LINACs were categorized by machine classes as defined previously [12, 13]. For both phantoms, the majority of irradiations were performed on Varian and Elekta machines; Siemens and Accuray machines were used for the remaining phantoms. The energies used were mostly 6 MV regular, with a few cases of 6 SRS, 6 FFF, 10 MV and 10 FFF beams. Five different TPS algorithms were

recorded with the most popular ones being Eclipse AAA and Superposition Convolution. Treatment techniques varied between the two phantoms with Dynamic MLC, Segmental (step & shoot) MLC and VMAT being the most popular techniques among the lung irradiations, while Segmental (step & shoot) MLC and VMAT were most common among the spine irradiations. The lung phantom respiration motion was managed using either the gating, free breathing (ITV) or tracking techniques, or the phantom was irradiated with no motion i.e. static. Tracking is a method specific to Cyberknife users and saw a 100% pass rate. The failure rates for the gating and ITV techniques were much greater than for static cases at 18% each vs 2% respectively, while there were no recorded failures for tracking. Statistical analysis using Fisher's exact test showed that lung phantoms were overall more likely to fail when treated using gating or ITV respiratory motion management techniques over the static technique ( $p < 0.001$ ).

## DISCUSSION

There were 116 moving lung phantom failures placed into 7 categories. The majority of the failing lung phantoms had an SI localization error, which is the direction of major target motion. As shown in table 2, phantom motion contributed substantially to SI localization errors, as 10% of all free-breathing/ITV and 13% of all gated treatments failed in this way, versus 0.3% of all static treatments. Poor respiratory motion management is likely responsible for these failures, including components like image-guided radiation therapy (IGRT) setup and tumor localization, given that only 1 of 64 phantom failures in this category was stationary during irradiation.

Gating did not improve the phantom SI localization results when compared with ITV, as a similar fraction of phantoms (13% vs 10% respectively) failed in this manner (table 2). Gating and ITV techniques were previously shown to also perform similarly when compared dosimetrically in lung treatments [14, 15].

Additionally, statistical analysis of the demographic data (table 3) showed that lung phantoms were more likely to fail overall when treated using gating or free-breathing motion management techniques. No statistical significance was found among failure rates due to TPS algorithm, despite previous indications that some algorithms perform better in heterogeneity corrections[16]. This is likely due to low power because of the limited sample size.

It is worthy to note that the failure rates under the minor (AP) and no (LR) motion categories were equal, which makes sense, because the motion in the AP direction is within the bounds of the gamma criteria ( $\pm 0.5$  mm). Treatment error in this direction was therefore more representative of treatment setup errors.

Applications of these clinically relevant findings can be made to similar treatment sites containing motion caused by patient respiration, such as the liver [17, 18]. Tumor motion caused by gastrointestinal activity[19] is of even greater concern as this motion is irregular, unlike the very regular motion pattern of the lung phantom. Gastrointestinal motion would

require more careful planning and monitoring, for accurate delivery and to avoid irradiating organs at risk.

There were 42 SBRT spine phantom failures placed into 4 categories. These failures, in contrast to those of the lung phantom, were mostly dosimetric in nature, displaying underdosing and overdosing of the PTV by as much as  $-11\%$  and  $+10\%$  respectively, and overdosing of OARs and the dose fall-off region. Dose calculation errors have been shown to indicate inaccuracies in institutions' dose calculation software[5], which is even more crucial for cases such as these highly modulated, high dose plans. In addition, the criteria used in IROC's phantom credentialing processes are loose compared to clinical standards, biological needs, and dosimeter precision[5]. As such, these failures should be considered dramatic underperformances in terms of dose delivery. The results of the spine phantom irradiation are also consistent with those of the IMRT head neck phantom, where 62% of failures were classified as systematic dose errors [5]. Further study is warranted to understand if the same TPS errors are manifesting in both phantoms as head and neck IMRT (comprised of highly modulated but larger fields) tends to be different from spine SBRT (comprised of less modulated but smaller fields)[20].

The IROC phantom serves as a patient surrogate, and allows us to record post treatment doses received by the target, through TLDs and radiographic film. The phantom is therefore the most accurate representation of a patient that we have, and provides us with vital information about the performance and accuracy of our current procedures in actual patient cases among the radiation oncology community. The various categories of failure for each phantom highlight the nature of failures which are most likely to occur for a given treatment type. The majority of the failures occurring in each phantom were errors in the treatment process, and did not appear to be the result of random mistakes or human error. Human error is a likely explanation for localization errors (in the non-motion direction), as this is often explained by a failure to setup to the correct isocenter, and such errors would often (although not always) be caught by clinical image review procedures. For all of the spine, lung, and head and neck phantoms, the rate of this error is very consistent: 12–14% of failures result from this cause. This is a small minority of the causes of failure. The dominant failure modes are ones that are testing the challenge of the irradiation technique being performed, namely SI localization for the moving lung target and systematic dose for the highly modulated spine target. The systematic dose calculation errors, in particular, are hard to imagine as arising from human error. The existence of problems within an institution's treatment process is highlighted by the fact that of the 7 institutions that failed the lung phantom twice, in 4 of them the error type was the same. One institution failed the lung phantom 4 times, and had a similar SI localization error 3 times in a row. Among institutions that recorded multiple failures of the spine phantom, every institution recorded the same error for each phantom failure. The results therefore serve to support that the majority of these errors are not random, but highlight a problem with the radiotherapy process. As such, these problems would likely show up in actual patient treatments. This information can guide quality assurance practices, and alert clinicians and physicists to the components in the radiation therapy treatment process that are most error prone, to properly guide future rectification efforts.



## CONCLUSION

The patterns of failure among IROC lung and spine phantoms were investigated for phantoms irradiated from 2012 to 2018. The majority of lung phantom failures were due to a localization error in the direction of major target motion and described the situation of missing the moving target. In contrast, spine phantoms failed mostly because of underdosing of the PTV (target) or overdosing of the organs at risk. These errors are clinically relevant and have high potential to manifest as errors in patient cases. This study can be used as a guide when treating sites with similar parameters, i.e. affected by motion or high modulation. Knowing what is most likely to go wrong for a particular case affords physicists and clinicians the opportunity to exercise precaution in these stages of the radiotherapy process.

## Acknowledgments

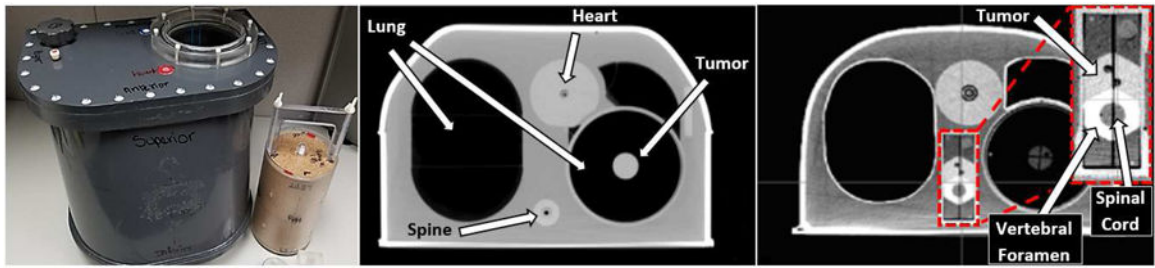
**Funding:** Work supported by IROC NCI grants CA180803 and CA214526

Christine B. Peterson is partially supported by NIH/NCI CCSG grant P30CA016672 (Biostatistics shared resource)

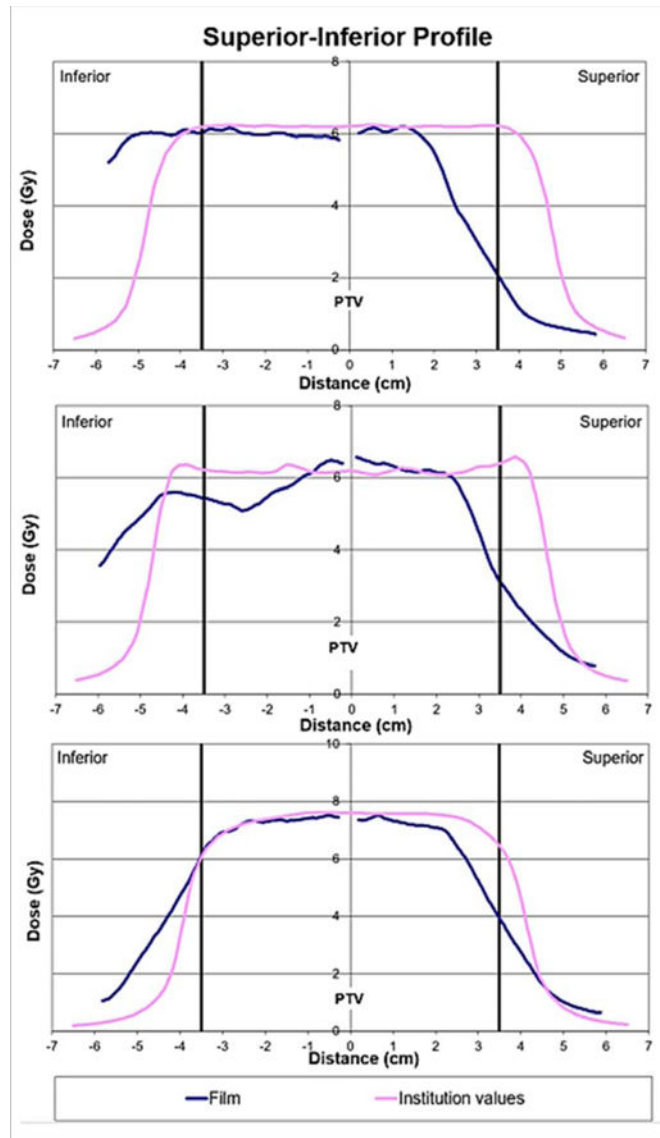
## References

1. Baskar R, et al., Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences*, 2012 9(3): p. 193–199. [PubMed: 22408567]
2. Ohri N, et al., Radiotherapy Protocol Deviations and Clinical Outcomes: A Meta-analysis of Cooperative Group Clinical Trials. *Jnci-Journal Of The National Cancer Institute*, 2013 105(6): p. 387–393.
3. Peters L, et al., Critical Impact of Radiotherapy Protocol Compliance and Quality in the Treatment of Advanced Head and Neck Cancer: Results From TROG 02.02. *Journal Of Clinical Oncology*, 2010 28(18): p. 2996–3001. [PubMed: 20479390]
4. Ibbott GS, QA in Radiation Therapy: The RPC Perspective. *Journal of Physics: Conference Series*, 2010 250: p. 012001.
5. Carson ME, et al., Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. *Medical Physics*, 2016 43(12): p. 6491–6496. [PubMed: 27908168]
6. Kerns JR, et al., Treatment Planning System Calculation Errors Are Present in Most Imaging and Radiation Oncology Core-Houston Phantom Failures. *International Journal of Radiation Oncology\*Biography\*Physics*, 2017 98(5): p. 1197–1203.
7. Ibbott GS, et al., Challenges in credentialing institutions and participants in advanced technology multi-institutional clinical trials. *International journal of radiation oncology, biology, physics*, 2008 71(1 Suppl): p. S71–S75.
8. Followill DS, et al., Design, development, and implementation of the Radiological Physics Center's pelvis and thorax anthropomorphic quality assurance phantoms. *Medical Physics*, 2007 34(6Part1): p. 2070–2076. [PubMed: 17654910]
9. Kry S, et al., Algorithms Used in Heterogeneous Dose Calculations Show Systematic Error as Measured With the Radiological Physics Center's Anthropomorphic Thorax Phantom Used for RTOG Credentialing. *International Journal of Radiation Oncology • Biology • Physics*, 2012 84(3): p. S128–S129.
10. Shoales J, et al., WE-D-T-617–07: Development of An Independent Audit Device for Remote Verification of 4D Radiotherapy. *Medical Physics*, 2005 32(6Part19): p. 2138–2138.
11. Molineu A, et al., Design and implementation of an anthropomorphic quality assurance phantom for intensity-modulated radiation therapy for the Radiation Therapy Oncology Group. *International Journal of Radiation Oncology\*Biography\*Physics*, 2005 63(2): p. 577–583.

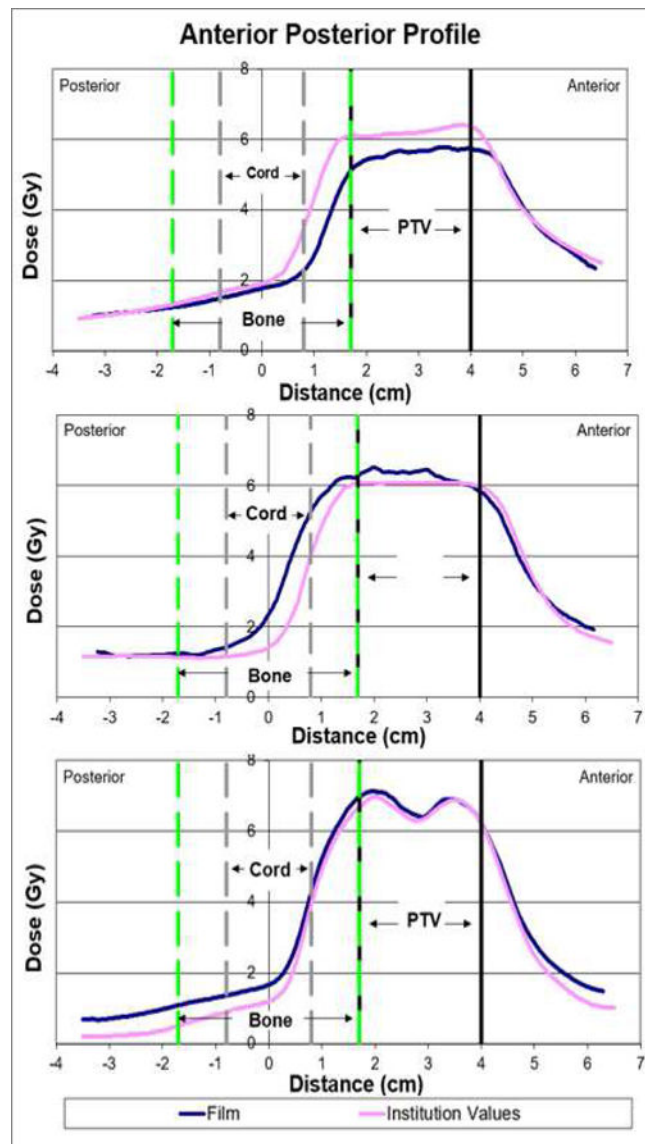
12. Kerns JR, et al., Technical Report: Reference photon dosimetry data for Varian accelerators based on IROC-Houston site visit data. *Medical Physics*, 2016 43(5): p. 2374–2386. [PubMed: 27147349]
13. Kerns JR, et al., Reference dosimetry data and modeling challenges for Elekta accelerators based on IROC-Houston site visit data. *Medical Physics*, 2018 45(5): p. 2337–2344. [PubMed: 29537634]
14. Kim J, et al., To gate or not to gate - dosimetric evaluation comparing Gated vs. ITV-based methodologies in stereotactic ablative body radiotherapy (SABR) treatment of lung cancer. (Report). *Radiation Oncology*, 2016 11(1).
15. Kim J, et al., SU-C-210-01: Are Clinically Relevant Dosimetric Endpoints Significantly Better with Gating of Lung SBRT Vs. ITV-Based Treatment?: Results of a Large Cohort Investigation Analyzing Predictive Dosimetric Indicators as a Function of Tumor Volume and Motion Amplitude. *Medical Physics*, 2015 42(6): p. 3204–3204.
16. Rana S, et al., Evaluation of Acuros XB algorithm based on RTOG 0813 dosimetric criteria for SBRT lung treatment with RapidArc. *Journal of Applied Clinical Medical Physics*, 2014 15(1): p. 118–129.
17. Wu QJ, et al., The impact of respiratory motion and treatment technique on stereotactic body radiation therapy for liver cancer. *Medical Physics*, 2008 35(4): p. 1440–1451. [PubMed: 18491539]
18. Osborn VW, Lee A, and Yamada Y, Stereotactic Body Radiation Therapy for Spinal Malignancies. *Technol Cancer Res Treat*, 2018 17: p. 1533033818802304. [PubMed: 30343661]
19. Abbas H, Chang B, and Chen ZJ, Motion management in gastrointestinal cancers. *Journal of gastrointestinal oncology*, 2014 5(3): p. 223–235. [PubMed: 24982771]
20. Du W, et al., Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. *Medical Physics*, 2014 41(2): p. 021716. [PubMed: 24506607]



**Figure 1.** IROC thorax phantom with cork lung insert beside (left). Transverse axial CT image of IROC thorax phantom with lung target insert showing the lung tumor (middle), and with spine target insert showing the spine tumor, vertebral foramen, and the cord as the avoidance structure (right).



**Figure 2.** Lung phantom dose profiles showing an SI localization error (top), a global error (middle), and a combination error (bottom). This combination error comprises an SI localization + ITV exaggeration effect. The consequences of the ITV technique can be seen in the shoulder region of the profiles, where the planned dose (pink) is broader than the measured (blue) film profile.



**Figure 3.** Spine phantom dose profiles showing a systematic underdose (top), a dose fall-off region error (middle) and an OAR overdose error (bottom). The OAR in this case is the spinal cord.

**Table 1.**

Categories of failure for IROC lung and SBRT spine phantom irradiations from Jan 2012 – Dec 2018, along with head and neck phantom failure data from a previous IROC study[5] (Nov 2014 – Oct 2015).

<b>Error Type</b>	<b>Number of phantoms (%)</b>	<b>95% C.I. (%)</b>
<b>Lung</b>		
Systematic dose	<b>28 (22%)</b>	(17,33)
➤ overdose	9 (6%)	
➤ underdose	19 (16%)	
Local dose	<b>7 (6%)</b>	(3, 12)
Localization: <i>major</i> motion	<b>64 (50%)</b>	(46,64)
➤ superior	26 (20%)	
➤ inferior	38 (30%)	
Localization: <i>minor</i> motion (ant-post)	<b>8 (6%)</b>	(4,13)
Localization: <i>no</i> motion (left-right)	<b>7 (6%)</b>	(3,12)
Global	<b>3 (2%)</b>	(1,7)
Combination category	<b>10 (8%)</b>	(5,15)
<b>Spine</b>		
Systematic dose	<b>25 (60%)</b>	(44,73)
➤ overdose	9 (22%)	
➤ underdose	16 (38%)	
Dose fall-off region	<b>5 (12%)</b>	(5,25)
OAR overdose	<b>6 (14%)</b>	(7,28)
Localization	<b>6 (14%)</b>	(7,28)
<b>Head and Neck [5]</b>		
Systematic dose	<b>32 (62%)</b>	(48,74)
➤ overdose	6 (12%)	
➤ underdose	26 (50%)	
Local dose	<b>8 (15%)</b>	(8,28)
Localization	<b>6 (12%)</b>	(5,23)
Global	<b>6 (12%)</b>	(5,23)

**Table 2.**

Main lung phantom error types grouped by respiratory motion management technique. Values are represented as a fraction of all irradiated IROC lung phantoms irradiated with that type of motion management within the time period (2012–2018). Combination and global error categories were excluded due to the specific and unique nature of these failures.

Error Type	RESPIRATORY MOTION TECHNIQUE							
	Free breathing (ITV)	95% C.I. (%)	Gating	95% C.I. (%)	Static	95% C.I. (%)	Tracking	95% C.I. (%)
<b>Dose (over, under and local)</b>	$\frac{23}{466}$ (5%)	(3,7)	$\frac{5}{125}$ (4%)	(2,9)	$\frac{7}{391}$ (2%)	(1,4)	$\frac{0}{35}$ (0%)	(0,10)
<b>SI localization</b>	$\frac{47}{466}$ (10%)	(8,13)	$\frac{16}{125}$ (13%)	(8,20)	$\frac{1}{391}$ (0.3%)*	(0,1)	$\frac{0}{35}$ (0%)	(0,10)
<b>Other localization</b>	$\frac{10}{466}$ (2%)	(1,4)	$\frac{3}{125}$ (2%)	(1,7)	$\frac{1}{391}$ (0.3%)	(0,1)	$\frac{0}{35}$ (0%)	(0,10)

\*  $p < 0.001$ , Fisher's exact test for pairwise comparisons of ITV, gating and static respiratory motion management techniques for SI localization failures

**Table 3.**

Demographics of the sample set. “% of this irradiation type” refers to the % of all IROC phantoms irradiated under these conditions that failed the irradiation.

Demographic	Lung Phantoms		Spine Phantoms	
Treatment Machine	Number of Failing Phantoms	% of this irradiation type	Number of Failing Phantoms	% of this irradiation type
<b>Varian</b>	<b>94</b>	<b>12%</b>	<b>32</b>	<b>18%</b>
Base class <sup>a</sup>	43	11%	18	27%
Trilogy	1	6%	4	44%
TrueBeam	50	13%	10	10%
<b>Elekta</b>	<b>17</b>	<b>15%</b>	<b>4</b>	<b>25%</b>
Agility	1	13%	0	0%
Infinity	6	16%	1	13%
Synergy	5	12%	0	0%
Versa HD	4	15%	3	38%
Precise	1	50%	0	0%
<b>Siemens</b>	<b>3</b>	<b>7%</b>	<b>1</b>	<b>50%</b>
Artiste	2	15%	1	50%
Oncor	1	17%	0	0%
<b>Accuray</b>	<b>1</b>	<b>4%</b>	<b>5</b>	<b>17%</b>
Cyberknife	0	0%	4	15%
Hi-Art Tomotherapy	1	4%	1	25%
<b>Energy (MV)</b>				
6	91	10%	31	16%
6 SRS	1	6%	4	44%
6 FFF	14	16%	2	50%
10	4	10%	4	20%
10 FFF	6	24%	1	9%
<b>TPS Algorithm</b>				
Eclipse AAA	54	11%	20	17%
Grid-based Boltzmann solvers <sup>b</sup>	19	13%	3	11%
Measured <sup>c</sup>	0	0%	3	43%
Monte Carlo	5	5%	2	6%
Superposition Convolution	38	13%	14	21%
<b>Treatment Technique</b>				
3D CRT	9	6%	0	0%
Dynamic MLC	23	14%	5	19%
Segmental (Step & Shoot) MLC	16	14%	11	26%
VMAT	67	13%	21	14%
Tomotherapy	1	3%	1	14%



Demographic	Lung Phantoms		Spine Phantoms	
	Treatment Machine	Number of Failing Phantoms	% of this irradiation type	Number of Failing Phantoms
Cyberknife	0	0%	4	15%
<b>Respiration Motion Technique</b>				
Gating <sup>*</sup>	23	18%	N/A	N/A
ITV <sup>*</sup>	84	18%	N/A	N/A
Static	9	2%	N/A	N/A
Tracking	0	0%	N/A	N/A

<sup>\*</sup>  $p < 0.001$ , Fisher's exact test for pairwise comparisons of gating, ITV and static respiratory motion management techniques

<sup>a</sup> Clinac 21EX, Clinac 21iX, Clinac 23iX, Clinac iX, Clinac 2300CD[11]

<sup>b</sup> Acuros dose calculation algorithm[20, 21]

<sup>c</sup> Ray tracing was the only measurement based algorithm used[21]