



Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2017 August 01; 98(5): 1197–1203. doi:10.1016/j.ijrobp.2017.03.049.

Treatment Planning System Calculation Errors Are Present in the Majority of IROC-Houston Phantom Failures

James R. Kerns^{1,2,3}, Francesco Stingo⁴, David Followill^{1,2,3}, Rebecca Howell^{1,2,3}, Adam Melancon¹, and Stephen F. Kry^{1,2,3}

¹Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030

²Imaging and Radiation Oncology Core-Houston, The University of Texas MD Anderson Cancer Center, Houston, TX 77030

³Graduate School of Biomedical Sciences, The University of Texas Health Science Center-Houston, Houston, TX 77030

⁴Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030

Abstract

Purpose—The anthropomorphic phantom program at the Houston branch of the Imaging and Radiation Oncology Core (IROC-Houston) is an end-to-end test that can be used to determine whether an institution can accurately model, calculate, and deliver an intensity-modulated radiation therapy (IMRT) dose distribution. Currently, institutions that do not meet IROC-Houston’s criteria have no specific information with which to identify and correct problems. In this work an independent recalculation system is developed that can identify treatment planning system (TPS) calculation errors.

Methods—A recalculation system was commissioned and customized using IROC-Houston measurement reference dosimetry data for common linear accelerator classes. Using this system, 259 head and neck phantom irradiations were recalculated. Both the recalculation and the institution’s TPS calculation were compared with the delivered dose that was measured. In cases where the recalculation was statistically more accurate by 2% on average or 3% at a single measurement location than was the institution’s TPS, the irradiation was flagged as having a “considerable” institutional calculation error. Error rates were also examined according to the linac vendor and delivery technique.

Results—Surprisingly, on average, the reference recalculation system had better accuracy than the institution’s TPS. Considerable TPS errors were found in 17% (45) of head and neck

Corresponding author: Stephen 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 citable 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

irradiations. 68% (13) of irradiations that failed to meet IROC-Houston criteria were found to have calculation errors.

Conclusion—Nearly 1 in 5 institutions were found to have TPS errors in their IMRT calculations, highlighting the need for careful beam modeling and calculation in the TPS. An independent recalculation system can help identify the presence of TPS errors and pass on knowledge to the institution.

Introduction

Accuracy in treatment planning of radiation therapy is extremely important,¹ and differences between what dose is planned and what dose is actually delivered to the patient must be minimized. Modern technologies such as intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) allow highly conformal dose distributions, making it imperative that the delivered dose distribution matches the planned distribution.

The Houston branch of the Imaging and Radiation Oncology Core (IROC-Houston) offers end-to-end testing through its anthropomorphic phantom program to ensure that planned and delivered doses agree. IROC-Houston has been charged with ensuring that institutions participating in clinical trials calculate and deliver radiation doses consistently and accurately. In the anthropomorphic phantom program, an institution irradiates an IROC-Houston phantom containing thermoluminescent dosimeters (TLDs) and radiochromic film.^{2–4} The institution-calculated dose distribution is compared to the measured dose distribution and, based on the extent of agreement, either passes or fails to meet the irradiation credentialing requirements.

Despite advances in delivery, localization, and imaging, credentialing pass rates have risen only modestly, reaching 85–90% in recent years depending on the phantom type.⁵ This rate is concerning because IROC-Houston's current criteria are less stringent than most institutional criteria (7% TLD agreement and 7%/4 mm film gamma criteria). An institution may fail the phantom test for many reasons: setup or positioning errors, linac or MLC-delivery performance, linac calibration, or TPS calculation errors, including: dose calculation grid size and placement, HU-to-density curve errors, input beam data, or beam modeling errors. One limitation of the phantom program is that because it's an end-to-end test the underlying causes of disagreement between measured and calculated doses are difficult to identify. In case of an irradiation failure, the institutional physicist has relatively little information with which to determine the cause of the discrepancy. Although setup errors are easy to spot, they are relatively rare; typically, the dose is systematically different from the calculation.⁷

To better inform institutions of where problems may lie, IROC-Houston would benefit from new tools to diagnose specific issues. Through an independent recalculation, IROC-Houston can identify the presence of errors in an institution's TPS model or calculation process. Although other causes or multiple causes may exist, this is the first step toward evidence-based error diagnosis. While detailed analysis of individual factors that could affect the passing rate is the ultimate goal, the goal of this study was a move toward understanding the

gross sources of error with evidence. Not all errors will be identified with the tools developed here, but useful information will always be generated by these tools.

Methods and Materials

We studied the head and neck (H&N) anthropomorphic phantom because it is the most frequently irradiated phantom. Irradiations performed between January 2012 and April 2016 were included in the analysis, so results were up-to-date and reflected current equipment and methodology. The phantom, previously described in detail,² is made of a hollow plastic shell (filled with water during irradiation) with a solid insert containing 6 TLDs in the targets (4 within a primary target, 2 within a secondary target). Two radiochromic films are also positioned axially and sagittally through the center of the primary target volume. Upon receiving a phantom, the institution treats it as a patient, designing and calculating a complete therapy plan. After irradiation, the phantom and its associated DICOM data, containing CT scan images, treatment plan, and TPS calculated dose, are sent back to IROC-Houston. The dose delivered to the TLDs is read out using a well-documented method.⁷ The film's optical density is converted to dose and then normalized to the dose of the adjacent TLDs. The measured TLD dose is compared with the TPS calculated dose, which was taken as the mean dose to the TLD contour. The film dose is compared with the TPS calculation for a rectangular region encompassing the target. If all TLD doses are within $\pm 7\%$ of the calculations and the pixels passing gamma analysis for each of the two films at 7%/4 mm is above 85% the phantom passes the credentialing requirements.

In this study, we independently recalculated institutional treatment plans for 259 H&N phantom irradiations. To independently recalculate the institution's dose distribution, a dose recalculation system (DRS) was used. Mobius3D (v1.5.3, Mobius Medical Systems, Houston, TX) was chosen as the DRS due to minimal additional workflow requirements, IROC-Houston's high volume of irradiations, 3D dose recalculation, and customizable beam models.^{8,9} The beam model customization feature was of particular interest for this project because IROC-Houston has acquired basic dosimetric data from hundreds of linacs and grouped them into representative classes, creating "average" linac models.¹⁰ We created three 6 MV recalculation beam models intended to match the three most common classes of linear accelerator: the Varian base class (including iX, EX, and trilogy machines), the Varian TrueBeam class, and the Elekta Agility class. These three classes of accelerator accounted for ~70% (259) of all the phantom irradiations performed in the time frame of this study. Of the 259 irradiations, 247 were unique (i.e., using a different institution, machine, TPS, or delivery technique). Six cases were irradiations repeated by the institution.

Each of the three beam models started with the default, out-of-the-box DRS model as a starting point and was then iteratively tuned to match its respective IROC-Houston reference beam dataset using Mobius3D's built-in modeling tools. A fitness metric, the absolute sum of local differences between the reference beam data point dose values and the model's calculation of the point dose values under the same geometric conditions, was calculated and minimized through the iterative tuning process. PDD, jaw output factors, MLC output factors, and off-axis factors¹⁰ were each evaluated.

After the DRS beam models were customized to best match the reference data values, H&N phantom DICOM datasets were imported into the DRS for recalculation. The same dose calculation information, i.e. calculated TLD dose, was extracted from the recalculations as the original TPS calculation. Dose grid size was 1mm; the default Mobius3D HU-to-density curve was used. Because the vast majority of irradiation failures result from TLD disagreement rather than film, we focused on recalculated TLD results to flag potential calculation errors.^{5, 11}

The difference in accuracy between IROC-Houston's recalculation and the institution's calculation was defined as follows:

$$D_n = \left(\left| 1 - \frac{TPS_n}{TLD_n} \right| - \left| 1 - \frac{DRS_{n,class}}{TLD_n} \right| \right) * 100$$

where D_n represents the difference in accuracy between IROC-Houston's recalculation using the same class model ($DRS_{n,class}$) as the institution's machine and the institution's original calculation (TPS_n) for the measured dose (TLD_n) of a given TLD (n). Thus, each irradiation had 6 D_n values for the 6 TLD. Positive D_n values indicate that IROC-H's recalculation was closer to the measured dose (i.e., more accurate) than was the institution's calculation, whereas negative values indicate that the institution's calculation was more accurate. Of particular interest were irradiations in which the DRS recalculation was "considerably better" than the institution's TPS, which would indicate TPS modeling or calculation errors (as they differ only in the calculation). The threshold of being "considerably better" used a clinical and a statistical criterion. The clinical criterion specified that the DRS must have an average positive value of $D > 2\%$ over the 6 PTV TLD locations or a single TLD D value $> 3\%$. In other words, the recalculation must be more accurate by an average of 2%, or 3% at a single location. The statistical criterion required that the mean value of the distribution of the 6 D_n values must be statistically greater than zero ($\alpha=0.05$). This was done by using a 2-sided t-test with a failure detection rate correction applied to the p-values. "Significant" in the results will always imply statistical significance. The 2% average value was chosen due to its common threshold of acceptability in TPS/measurement agreement in ICRU 24, TG-65, and MPPG5a.¹³⁻¹⁵ The 3% single value difference was chosen because it is approximately 2σ of the TLD measurement uncertainty.⁶

The improvement of the recalculation of each phantom plan was examined for the entire cohort of phantoms as well as the subset of irradiations that failed to meet the IROC-Houston credentialing criteria. The entire cohort was also split linac class and delivery technique (segmental IMRT, dynamic IMRT, and VMAT). These subsets were tested for statistical significance using a t-test ($\alpha=0.05$).

Results

Results for tuning the Varian base class DRS beam model are shown in Table 1. The values describe the local difference between the reference data point from IROC-Houston's standard dataset and the model's calculation of the same point.¹⁰ The default recalculation model had a fitness metric value of 11.8. After tuning the model using the built-in tools, the

fitness value was lowered to 5.1 The TrueBeam and Agility models started at 11.8 and 16.3 while the tuned model values were 4.8 and 5.4 respectively. The tuned DRS beam models had overall mean dose differences from the corresponding reference data of 0.27%, 0.27%, and 0.36% for the base, TrueBeam, and Agility models, respectively. These differences are the same or smaller than the average institutional measurement-to-TPS difference of 0.36% as determined by IROC-Houston,¹² meaning the tuned DRS models described reference dosimetric data better than the average institution's TPS described their own linac, and thus the models were considered accurate enough for recalculation use.

Two representative case studies are given here to detail irradiation and recalculation results and to better understand general results. Agreement values of the original TPS calculation and the IROC-Houston recalculation are given in Table 2. The first institution, using a Varian base class linac, Pinnacle, and a VMAT technique failed the credentialing requirements with two TLD dose/calculation discrepancies >7%, and had an overall average 6% dose discrepancy. Upon recalculation by the DRS, the maximum dose/calculation discrepancy was only 4%, and the total (average) improvement for this phantom (D_n) was +3.9%. Notably, the institution would have easily passed the irradiation requirements with the reference DRS model. The second irradiation, using a TrueBeam, Eclipse, and a dynamic IMRT technique had very good accuracy from the original institution's calculation, with the measured TLD dose and TPS calculation having no more than 1% discrepancy. The DRS calculation had poorer accuracy, however, with the average discrepancy being 7% and an overall (average) improvement (D_n) of -5.9%. This institution would have failed credentialing requirements using the reference DRS model.

Results of the average D_n value for each of the 259 phantoms are shown in a waterfall plot in the top and middle panels of Figure 1, which plot the same data but with different color overlays. Recalculations with large negative difference values indicate that the recalculation had poorer accuracy than did the institution's original calculation, like that of the second case study. Values near the center show that the accuracy of the recalculation was comparable to that of the original calculation. Irradiations with high positive difference values are those in which the recalculation system obtained much better accuracy than the original calculation did, like that of the first case study. The color of the difference value in the top panel denotes how accurate the original calculation was. Cyan indicates that the original calculation had less than a 2% maximum discrepancy with the TLD measured dose across the 6 TLDs. Light green indicates between 2% and 5% max discrepancy; orange indicates >5% max discrepancy; red indicates that the irradiation failed to meet IROC-Houston criteria (>7%).

The median DRS recalculation difference value was +0.2%, meaning that on average, the recalculation was closer to the measured TLD dose than the institution's calculation was; this result was not significant ($p=0.9$). Of the 259 recalculations, 45 (17%) had differences above the clinical and statistical thresholds, meaning that the institutional TPS calculation had serious errors compared to the DRS recalculation. These data are shown in pink in the middle panel, which is the same data as the top panel, but color-coded according to whether the irradiation had a considerable calculation error. Irradiations without considerable calculation differences are shown in gray.

The recalculations of the irradiations that failed to meet current IROC-Houston criteria (the red-colored values of the top panel) are shown in the bottom panel of Figure 1. Nineteen phantoms were in this subset, and of those, 13 (68%) had considerable TPS calculation differences, i.e. some sort of calculation or modeling error, with a median DRS difference of +3.1%; the mean was significant ($p \ll 0.01$), emphasizing the sizeable TPS errors seen in these cases. The sensitivity of the tool to detecting IROC failures was 68% and the specificity was 87% using the criterion defined in this study.

All of the failing irradiations were from separate institutions. Of the 19 failures, 6 were from institutions that performed a second irradiation with the same machine, TPS, and technique and subsequently passed the IROC-Houston criteria. The average dose-improvement value of these 6 original failing irradiations was 3.2%; the subsequent passing irradiations had an average improvement of 1.4%. There was no obvious systematic differences of the failing cases versus non-failing cases; the type of linac, TPS, and delivery technique were roughly the same proportions as the entire cohort, and no statistical significance could be determined. The DRS thus showed a large possible improvement of the failing irradiations, while the passing irradiations showed less possible improvement.

Applying the criterion in the opposite direction, where the institution's TPS was considerably more accurate than the DRS, 16% (42) irradiations were identified. In these cases, as well as cases where the DRS was comparable to the TPS, the developed tool does not provide insight into the accuracy of the institution's TPS calculation. As well, the handful of irradiations that were repeats from the same institution were scrutinized for commonalities but due to the low number, no statistically relevant information could be extracted, such as machine type or TPS.

The recalculation cohort was analyzed according to the linac class that delivered the irradiation and the delivery technique. The results are shown in Table 3. For the delivery techniques, VMAT was the most common delivery technique and had a negative mean value of -0.39% ($p=0.01$), indicating that, in general, an individual institution had more accurate results than the recalculation did when using VMAT. In contrast, both the segmental and dynamic IMRT delivery types had positive D values: 1.18% ($p < 0.01$) and 0.32% ($p=0.28$), respectively, meaning that institutional TPSs typically did not calculate the dose as accurately compared to our reference DRS model. Regarding the linac class, the TrueBeam performed the best, with a median D value of -0.51% ($p=0.04$), whereas the Varian base and Elekta Agility class had positive values of 0.32% ($p=0.06$) and 0.68% ($p=0.20$), respectively.

Discussion

Several surprising conclusions can be drawn from the recalculation results. A relatively large percentage of irradiations (17%) were identified as having a considerable TPS calculation errors in the context of a H&N phantom irradiation. Furthermore, of the irradiations that failed to meet IROC-Houston's criteria, two-thirds were shown to have TPS calculation errors; thus, TPS errors are present in the majority of irradiations not meeting the criteria. These values are alarmingly high.

The cases in which IROC-Houston's independent recalculation was considerably more accurate than a given institution's TPS were of particular concern. There is serious concern that these beam models would similarly miscalculate dose in patient cases given that the phantoms irradiated were treated according to the same patient workflow as normally used by the institution.

The positive overall median value of D for the phantom recalculations is surprising, given that the DRS beam models were meant to represent an "average" machine. If all institutions modeled their TPS to perfectly match their linac, the recalculation error values would always be negative, or at best zero; IROC-Houston's recalculation would never be more accurate. However, in 17% of cases, the reference recalculation was considerably more accurate.

Figure 1 demonstrates several important findings. First, irradiations that had good agreement between the institution's calculation and measured dose (cyan colored) generally had negative values. These institutions accurately modeled their linac characteristics and customized their TPS accordingly. A generic calculation model like the ones developed here are inferior for these cases (as demonstrated by the second case study). Second, recalculations of irradiations that failed to meet IROC-Houston criteria (red colored) almost always had substantially and significantly improved accuracy with the DRS (as demonstrated in the first case study). The single case of a failing irradiation with a negative difference value (bottom panel of Fig 1) was determined to be a setup error, and thus improvement would not be expected.

The 6 institutions that initially failed and then subsequently passed the IROC-Houston criteria also underscore the detection ability and accuracy of the DRS. The average improvement of the DRS of 3.2% for failing irradiations changed to 1.4% for the subsequent passing irradiation. This highlights that these institutions calculated the delivered doses more accurately upon repeating the test, and illustrates the lower potential improvement of the DRS for passing irradiations.

Based on these data, the question of using "stock" data as a model in the TPS can be addressed. The recalculation models used in this study were based on representative beam data from the community. Although there were cases where using averaged data improved overall accuracy, there were an abundance of cases (those with negative difference values) where using such averaged data reduced accuracy, in some cases by a large amount. This result clearly demonstrates that using stock or averaged data to commission a TPS is not reasonable in every case. Rather, these data underscore the need for physicists to validate the match of their TPS calculations to their linac characteristics. The goal is not to match stock data, but to carefully tailor the beam model and calculation to the linac. Given the abundance of TPS errors observed in this study, this clearly remains a serious challenge to the medical physics community, and something that requires broad attention.

The ultimate cause of why an institution failed or performed poorly could have a large number of sources including calculation errors, delivery errors (MLC positioning errors, output errors, etc.), machine calibration errors, setup or positioning errors, or others. In this work we have developed an infrastructure to identify cases where a computational error is

involved, which ends up being a very frequent issue. However, there are a host of reasons why a computational error may arise. Beam modeling is a clearly known problem¹², but poor commissioning data, the dose grid spacing,¹⁶ and the HU density curve¹⁷ could also contribute. Ultimately, to meet the goal of IROC Houston identifying the underlying cause(s) of a phantom failure, more refined information is necessary. To this end, future work includes sensitivity analysis of individual parameters and statistical evaluations of the causes of phantom failures. To broaden the scope of the current infrastructure, future work must also include analysis of less common classes of linac as this study evaluated solely the three dominant classes.

Conclusions

IROC-Houston has utilized an independent dose recalculation tool, tuned to match extensive reference data, to identify institutions that have considerable errors in their TPS beam model or calculation via the anthropomorphic phantom program. We recalculated 259 H&N phantom irradiations, 17% (45) of which were found to have considerable TPS errors. Of the irradiations that failed to meet current IROC-Houston criteria, 68% (13) had TPS errors, making this type of error present in the majority of irradiations not meeting the criteria. IROC-Houston now has the ability to flag when an institution has a TPS error and can pass that information along to the institution. Given the high rate of TPS errors, medical physicists must be extremely vigilant during creation and validation of their TPS beam models and settings to ensure it accurately calculates dose in clinical conditions.

Acknowledgments

This work was funded in part by Public Health Service Grant CA10953 (NCI) and Cancer Center Support Grant P30CA016672.

References

1. Nelms BE, Chan M, Jarry G, Lemire M, Lowden J, Hampton C, Feygelman V. Evaluating IMRT and VMAT dose accuracy: Practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. *Medical Physics*. 2013; 40(11) 111722-1-15.
2. Molineu A, Followill DS, Balter PA, Hanson WF, Gillin MT, Huq MS, Eisbruch A, Ibbott GS. 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, biology, physics*. 2005; 63(2):577–583.
3. Followill DS, Evans DR, Cherry C, Molineu A, Fisher G, Hanson WF, Ibbott GS. Design, development, and implementation of the radiological physics center's pelvis and thorax anthropomorphic quality assurance phantoms. *Medical physics*. 2007; 34(6):2070–2076. [PubMed: 17654910]
4. Ibbott GS, Molineu A, Followill DS. Independent evaluations of IMRT through the use of an anthropomorphic phantom. *Technology in cancer research & treatment*. 2006; 5(5):481–487. [PubMed: 16981790]
5. Molineu A, Hernandez N, Nguyen T, Ibbott G, Followill D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. *Medical physics*. 2013; 40(2):022101. [PubMed: 23387762]
6. Kirby TH, Hanson WF, Johnston DA. Uncertainty analysis of absorbed dose calculations from thermoluminescent dosimeters. *Med Phys*. 1992; 19(6):1427–1433. [PubMed: 1461205]

7. Carson ME, Molineu A, Taylor P, Followill D, Stingo F, Kry S. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. *Medical Physics*. 2016; 43(12):6491–6496. [PubMed: 27908168]
8. Nelson CL, Mason BE, Robinson RC, Kisling KD, Kirsner SM. Commissioning results of an automated treatment planning verification system. *Journal of applied clinical medical physics/American College of Medical Physics*. 2014; 15(5):4838.
9. Fontenot JD. Evaluation of a novel secondary check tool for intensity-modulated radiotherapy treatment planning. *Journal of applied clinical medical physics/American College of Medical Physics* 15 (5). 2014:4990.
10. Kerns JR, Followill D, Lowenstein J, Molineu A, Alvarez P, Taylor PA, Stingo F, Kry SF. Technical Report: Reference photon dosimetry data for Varian accelerators based on IROC-Houston site visit data. *Medical physics*. 2016; 43
11. Kry SF, Molineu A, Kerns JR, Faught AM, Huang JY, Pulliam KB, Tonigan J, Alvarez P, Stingo F, Followill DS. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. *International journal of radiation oncology, biology, physics*. 2014; 90(5):1195–1201.
12. Kerns JR, Followill D, Lowenstein J, Molineu A, Alvarez P, Taylor PA, Kry SF. Agreement between institutional measurements and treatment planning system calculations for basic dosimetric parameters as measured by IROC-Houston. *International journal of radiation oncology, biology, physics*. 2016; 95(5):1527–1534.
13. Determination of absorbed dose in a patient irradiated by beams of x- or gamma-rays in radiotherapy procedures. *International Committee on Radiation Units, Report 24*. 1976
14. Papanikolaou N, Battista J, Boyer A, Kappas C, Klein E, Mackie TR, Sharpe M, Dyk JV. Tissue inhomogeneity corrections for megavoltage beams. *Medical Physics*. 2004
15. Smilowitz J, Das I, Feygelman V, Fraass B, Kry S, Marshall I, Mihailidis D, Ouhib Z, Ritter T, Snyder M, Fairbent L. AAPM Medical Physics Practice Guideline 5.a.: Commissioning and QA of Treatment Planning Dose Calculations–Megavoltage Photon and Electron Beams. *Journal of applied clinical medical physics*. 2015; 16(5):14. [PubMed: 28296110]
16. Chung H, Jin H, Palta J, Suh T-S, Kim S. Dose variations with varying calculation grid size in head and neck IMRT. *Physics in Medicine Biology*. 2006; 51:4841–4856. [PubMed: 16985274]
17. Kilby W, Sage J, Rabett V. Tolerance levels for quality assurance of electron density values generated from CT in radiotherapy treatment planning. *Physics in Medicine Biology*. 2002; 47:1485–1492. [PubMed: 12043814]

Summary

IROC-Houston recalculated over 250 head and neck phantom irradiations performed between 2012 and 2016 using a customized independent calculation system that was modeled after IROC-Houston linear accelerator measurement reference data. Dose recalculations that were substantially more accurate than the institution's calculations suggested that the institution had treatment planning system calculation errors. Among all recalculations, 17% revealed such errors; among irradiations that failed IROC-Houston criteria, 68% had such errors.

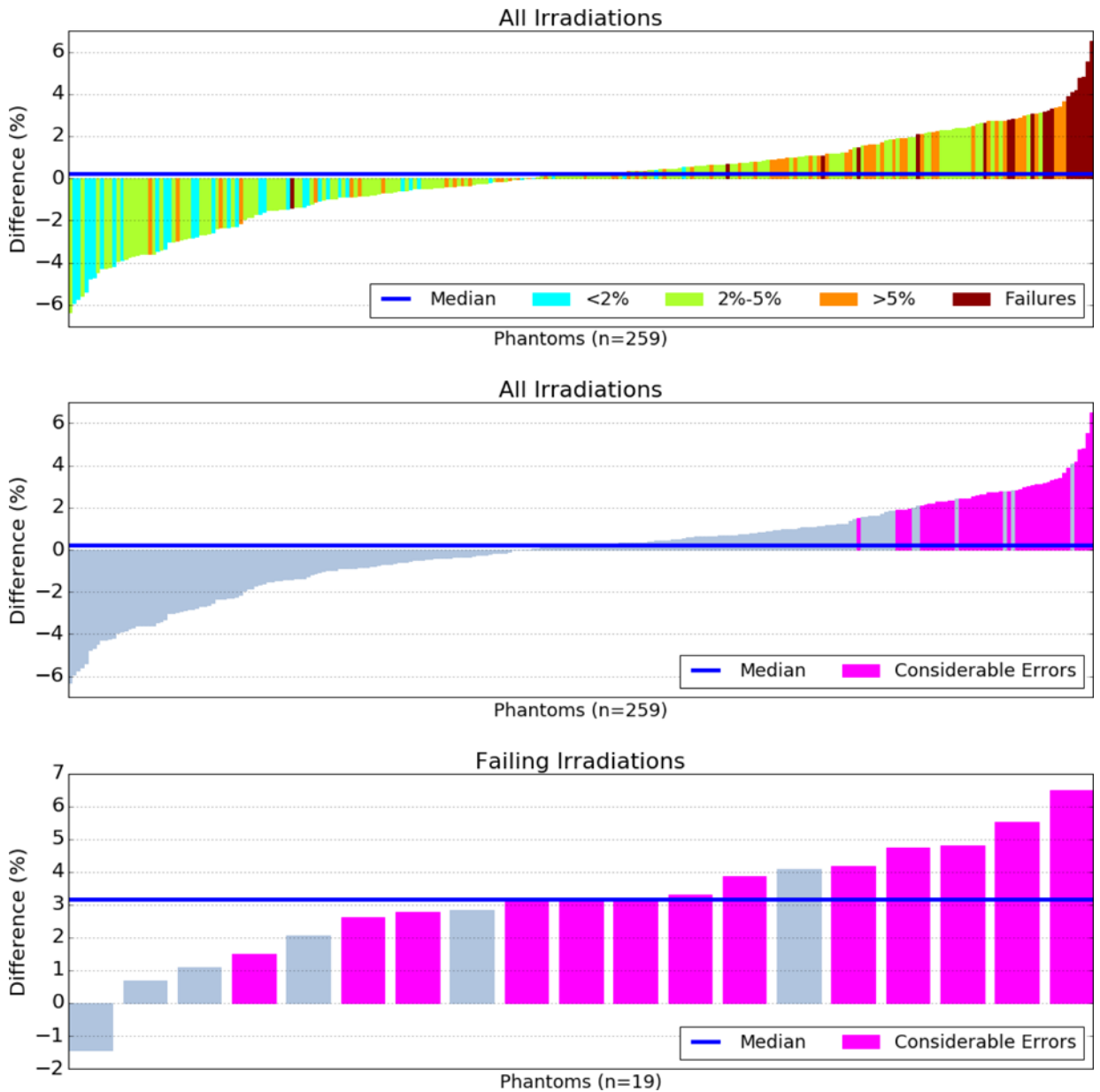


Figure 1.

Difference values in accuracy between the institution's treatment planning system (TPS) calculation and IROC-Houston's reference recalculation. Positive values indicate the recalculation was more accurate. The top and middle panels show the same data with different color overlays. The top overlay indicates the institution's original agreement with the phantom's thermoluminescent dosimeters. Cyan indicates the institution calculated the delivered dose very accurately, whereas dark red indicates the institution calculated the delivered dose very poorly. Pink values in the middle and bottom panels indicate a considerable TPS calculation error for that irradiation.

DRS model discrepancies between the reference data and calculation for the default beam model and the final customized model for the Varian base class. Reference dosimetric parameters were determined previously.⁸

Table 1

PDD		Jaw Output Factor	
Default	Tuned	Default	Tuned
5 cm	-0.1%	6×6 ² cm	0.9%
10 cm	-0.2%	15×15 cm ²	-0.3%
15 cm	0.6%	20×20 cm ²	-0.2%
20 cm	0.3%	30×30 cm ²	0.3%

IMRT - style Output Factors				SBRT - style Output Factors			
Default		Tuned		Default		Tuned	
6×6 cm ²	0.4%	0.1%	6×6 cm ²	1.0%	0.0%		
4×4 cm ²	-0.3%	-0.8%	4×4 cm ²	1.3%	-0.1%		
3×3 cm ²	-0.2%	-0.8%	3×3 cm ²	1.7%	0.0%		
2×2 cm ²	-0.7%	-1.2%	2×2 cm ²	2.1%	-0.4%		

Off - Axis Factors	
Default	Tuned
5 cm	-0.6%
10 cm	-0.2%
15 cm	-0.4%

Table 2

Two irradiations comparing the institution's original dose agreement (TPS/TLD) and IROC-Houston's recalculation agreement (DRS/TLD).

TLD #	Inst #1		Inst #2	
	TPS/TLD	DRS/TLD	TPS/TLD	DRS/TLD
1	0.96	0.98	1.00	1.04
2	0.94	0.97	0.99	1.04
3	0.92	0.97	0.99	1.06
4	0.92	0.98	1.00	1.07
5	0.96	0.99	1.00	1.10
6	0.94	0.96	1.01	1.11
Avg Ratio	0.94	0.97	1.00	1.07
Avg <i>D</i>		+3.9 %		-5.9%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Recalculation data broken down by delivery technique and linac class. *N* is the number of recalculations. *D* is the difference value of the recalculation and the percent of recalculations with a considerable TPS error is also shown. An asterisk indicates statistical significance.

	<i>N</i>	Mean <i>D</i> (%)	Considerable Error (%)
Technique			
Segmental IMRT	32	+1.18*	30
Dynamic IMRT	72	+0.32	26
VMAT	153	-0.39*	11
Linac Class			
Varian Base	140	+0.32	21
Varian TrueBeam	74	-0.51*	12
Elekta Agility	23	+0.73	26