A New Anthropomorphic Pediatric Spine Phantom for Proton Therapy Clinical Trial Credentialing

Dana J. Lewis, MS1,2,3, Paige A. Taylor, MS1,2,3, David S. Followill, PhD1,2,3, Narayan Sahoo, PhD2,3, Anita Mahajan, MD3,4, Francesco C. Stingo, PhD3,5, and Stephen F. Kry, PhD1,2,3

1Imaging and Radiation Oncology Core Quality Assurance Office, Houston, TX, USA
2Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
3The University of Texas Health Science Center Houston, Graduate School of Biomedical Sciences, Houston TX, USA
4Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
5Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

Purpose: To design and evaluate an anthropomorphic spine phantom for use in credentialing proton therapy facilities for clinical trial participation by the Imaging and Radiation Oncology Core Houston QA Center.

Materials and Methods: A phantom was designed to perform an end-to-end audit of the proton spine treatment process, including simulation, dose calculation, and proton treatment delivery. Because plastics that simulate bone in proton beams are unknown, 11 potential materials were tested to identify suitable phantom materials. Once built, preliminary testing using passive scattering and spot scanning treatment plans (including a field junction) were created in-house and delivered 3 times to test reproducibility. The following measured attributes were compared with the calculated values: absolute dose agreement using thermoluminescent dosimeters, planar gamma agreement, distal range, junction match, and right and left profile alignment using radiochromic film. Finally, credentialing results from 10 institutions were also assessed.

Results: A suitable bone substitute was identified (Techtron HPV Bearing Grade), which had a measured relative stopping power that agreed within 1.1% of its value calculated by Eclipse. In-house passive scatter testing of the phantom demonstrated that the phantom was suitable for assessing craniospinal irradiation dose delivery. However, the in-house scanning beam results were more mixed, highlighting challenges in treatment delivery. Seven of ten institutions passed the
proposed criteria for this phantom, a pass rate consistent with other Imaging and Radiation Oncology phantoms.

**Conclusions:** An anthropomorphic proton spine phantom was developed to evaluate proton therapy delivery. This phantom provides a realistic challenge for centers wishing to participate in proton clinical trials and highlights the need for caution in applying advanced treatments.

**Keywords**
IROC; proton; quality assurance; clinical trials; QA

**Introduction**
Proton therapy is gaining acceptance as a cancer treatment modality, particularly for conducting craniospinal irradiation (CSI), for which it produces far superior dose distributions [1].

The Imaging and Radiation Oncology Core Quality Assurance Center at Houston (IROC Houston), formerly known as the Radiological Physics Center, is funded by the National Cancer Institute to audit radiotherapy institutions for clinical trial participation. This responsibility includes assessment of institutional radiation therapy programs to ensure that dose uncertainty is minimized so results from clinical trials can be reliably interpreted. Before patient enrollment in clinical trials, institutions must complete the National Cancer Institute–mandated IROC Houston proton approval process [2], which includes using anthropomorphic phantoms that verify dose delivery for special treatment techniques. The phantom results ensure accurate dose delivery or help institutions identify discrepancies in the treatment process and implement solutions.

IROC Houston had previously developed an anthropomorphic spine phantom for evaluating proton therapy. However, the physical bone used in this phantom quickly degraded and was soon largely air. In addition, the film used in the phantom was in a curved plane, making extraction and comparison with the treatment planning system dose untenable. The purpose of this study was therefore to design, evaluate, and test a new anthropomorphic spine phantom for use by IROC Houston in credentialing proton therapy facilities for clinical trial participation.

**Methods**
**Phantom Design**
A pediatric thoracic spine phantom was designed to use plastics that are biologically equivalent (radiologically) in a proton beam, lightweight for easy transportation, and discernible on computed tomography (CT) to allow for treatment planning. The phantom was designed to include bone, cartilage, and soft tissue substitutes to simulate heterogeneous anatomy and make the treatment delivery more realistically challenging.

Although many plastics have been evaluated for their radiological properties in protons [3], none have been identified that radiologically simulate bone. Therefore, 11 possible bone
substitutes were investigated. Using the Moyers et al methodology [4], the relative stopping power (RSP) and Hounsfield unit (HU) of each material were determined. Each material was scanned via CT at 120kVp to determine the mean HU. The RSP was obtained from depth ionization scans using a Zebra multilayer ionization chamber (IBA, Schwarzenbruck, Germany); this was done at 2 energies: 160 MeV and 250 MeV, to verify consistency across energies. In order to be considered proton equivalent, the material should be very close to the clinical HURSP calibration curve for biological tissues.

The phantom also included dosimeters to measure the delivered dose distribution. Two thermoluminescent dosimeter (TLD-100) capsules were placed in the right superior and left inferior positions of the spinal canal for absolute point dosimetry. Each TLD was analyzed at IROC Houston following its established procedure, correcting for system sensitivity, fading, linearity, and energy dependence [5]. Two sheets of radiochromic film (GAFChromic EBT2 film, Ashland Inc, Covington, Kentucky) were placed in the phantom in the coronal and midsagittal planes. Three localization pinprick marks were used on each film for dose registration. Optical density was converted to dose using a third-degree polynomial calibration curve. The film dose was then normalized to the TLD dose. The films were scanned with the same wait-time interval used to create the calibration curve, avoiding error due to self-darkening of the film [6]. All measured doses were scaled by the proton relative biological effectiveness of 1.1 for comparison with the calculated dose values.

Preliminary Testing

Initial testing of the phantom was conducted in-house using clinical infrastructure. Because it used a clinical system, this initial testing was subject to the same limitations as any irradiation conducted by an institution in terms of possible disagreement between the treatment planning system (TPS) calculation and the dose actually delivered to the phantom. While this testing therefore does not provide true validation of the phantom, it nevertheless allows us to evaluate the visualizability of the phantom and target, feasibility of planning/meeting proposed plan constraints, suitability of detector placement (ie, avoiding high gradient regions for the TLD), appropriateness of proposed analysis, and reproducibility of the dosimeters to phantom setup and repeated irradiations, as well as to create reasonable passing criteria for the phantom.

Preliminary testing used a passively scattered beam delivery technique because that is our institution’s clinical approach. The phantom was simulated on a GE LightSpeed RT16 scanner (GE Healthcare, Waukesha, Wisconsin) using a pediatric spine imaging protocol. Following instructions developed for this phantom, the Eclipse treatment planning system (Varian Medical Systems, Inc, Palo Alto, California) was used to create a clinically acceptable passive scattering treatment plan delivering 600 cGy-CGE to ≥95% of the clinical target volume (CTV) (this dose was selected as it is optimal for the dosimetry protocol used). The CTV was delineated as the spinal column, as typically defined in CSI patients. Phantom-specific apertures and compensators were created to shape the dose distribution. A superior and an inferior field were matched approximately in the middle of the phantom. Two junction plans separated by 1 cm were summed for creation of the final plan to create feathering similar to a clinical treatment.
The phantom setup mimicked patient setup during treatment delivery, relying on external localization marks and kV radiographs. Three independent irradiations of the phantom were completed to evaluate the reproducibility of the measurements. The following measured attributes were evaluated against the TPS calculations: absolute dose, planar dose (gamma analysis) in the sagittal and coronal planes, distal range, and right-left profile alignment. The junction match was also evaluated.

The point dose was evaluated as the ratio of the measured TLD dose to that calculated by the TPS over a region of interest corresponding to the TLD powder. Gamma analysis was performed on the coronal and sagittal films using IROC Houston’s phantom software and a 5%/5 mm criteria (where the percent was taken of the prescription dose) based on IROC Houston phantom experience. Anterior-posterior and right-left profiles were also extracted from the film planes to evaluate the high dose gradient. The spatial displacement between the TPS and measurement at 25%, 50% and 75% of the maximum dose were averaged to determine the distance to agreement (DTA). Anterior-posterior profiles were extracted from the sagittal film (for each of two junctioned treatment fields) to test the range of the treatment. Right-left profiles were extracted from the coronal film (also for each of the 2 junctioned treatment fields) to test the lateral conformality of the treatment.

Similar preliminary testing was also conducted using the spot scanning beam. A treatment plan was developed on the same phantom CT images with the same plan objectives. This plan was also delivered 3 separate times to the phantom. Superior and inferior fields were matched approximately in the middle of the phantom using a multifield optimization process (for which the junction is largely made uniform through field overlap optimization [7]). This delivery technique was evaluated with the same methods as for the scattered beam deliveries.

**Initial Credentialing Results**

Since development, the phantom has been irradiated independently at 10 institutions as part of the baseline clinical trial approval process [2]. As with other IROC phantoms, each of these institutions was provided with the plan objectives but were otherwise instructed to treat the phantom as they would a patient. Two phantoms were irradiated with a scattered beam technique, while the remainder were irradiated with a scanning beam technique. Analysis was conducted according to the methods described previously for the preliminary testing.

**Results**

**Proton Equivalency**

The HU and RSP data for each phantom material tested are shown in Figure 1, along with the clinical calibration curve. Horizontal error bars are included for each material based on the standard deviation of the HU, while vertical error bars were determined based on the RSP uncertainty [4]. Because most craniospinal treatments use a 160 MeV beam, the stopping powers corresponding to this energy were used for determining proton equivalency and are shown in the figure. No tested material showed more than a 1.3% difference in RSP between 160 MeV and 250 MeV. Based on its proximity to the clinical calibration curve (ie, being biologically equivalent) and its suitable HU, the material selected as bone was...
Techtron HPV Bearing Grade (Boedeker Plastics, Shiner, Texas). Based on the work by Grant et al [3], solid water (Gammex, Inc., Middleton, Wisconsin) and blue water (Standard Imaging, Middleton, Wisconsin) were selected as cartilage and soft tissue substitutes, respectively. For the 3 materials used in the phantom, the measured RSP agreed within 1.2% with the value calculated by Eclipse for the corresponding HU. This difference in the theoretical and measured RSP translates to sub-millimeter error in the proton range calculation through the selected phantom materials.

Several of the tested materials, such as Gammex Inner Bone and Gammex Cortical Bone, are designed to be bone equivalent in an x-ray beam. In x-ray radiotherapy, plastics are evaluated in terms of their HU and electron density. However, in proton radiotherapy, the relevant consideration is HU and RSP; therefore, materials that simulate tissue when placed in x-ray beams may not simulate tissue when placed in a proton beam. This was seen for the Gammex Bone materials and most of the other materials tested. The locations of the (HU, RSP) points for most tested materials were far from the proton calibration curve. Such materials are therefore not bone equivalent in a proton beam and could result in up to 35% error in the range. This highlights the importance and challenge in selecting proper phantom materials.

The final phantom, shown in Figure 2, was constructed in-house and contained the following features: a pediatric-sized thoracic spinal column with 12 vertebral bodies, intervertebral cartilage disks, and spinal cord along with surrounding soft tissue. To simulate spine curvature, small wedges of soft tissue substitute were added posterior to the spine. Finally, the length of the phantom was extended with high-impact polystyrene blocks (which did not include any dosimeters) so the phantom could accommodate realistic field sizes that included a field junction region.

Preliminary Testing

The results from the passive scattering delivery analysis (mean and standard deviation) are listed in the Table. Low standard deviations were observed across the measurements, indicating suitable positioning of the dosimeters in the phantom, as well as good reproducibility of the irradiation and read-out procedures. The average measured TLD dose agreed with the calculated dose within 2%, indicating that the TPS was able to calculate the dose accurately, and the treatment unit could accurately deliver it. The gamma analysis performed on both film planes also showed reasonable reproducibility over repeated irradiations and good agreement between measurement and calculation, with average pixel pass rates >90% for a 5%/5 mm passing criteria. The percent dose criteria for the gamma analysis was based on the prescription dose, and the gamma analysis was done on a rectangular region of interest that was cropped just beyond the field edge. The planar gamma results, shown in Figure 3, display good agreement in the CTV except for slight disagreement at the distal edge visible on the sagittal film. Gamma analysis was also performed using the 5%/3 mm criteria, however, the average passing rate decreased to 79% for sagittal film, making this an unreasonable criteria.

The right-left profiles extracted from the coronal film (1 for each of the superior and inferior fields) were used to evaluate dose conformity lateral to the target as shown in Figure 4. The
anterior-posterior profiles extracted from the sagittal film (1 for each of the 2 treatment fields) were used to verify the range, as shown in Figure 4. The DTA results averaged over all profiles are listed in the Table. The measured and calculated dose profiles typically agreed very well and were reproducible.

Junction region analysis was performed but could not be pursued more broadly. A robust analysis of the junction region was not achievable as it was not possible to consistently assess the hot and cold spot. Because of this, and because most clinical treatments are done with scanning beam delivery (for which there is no clear junction region [7]), junction analysis was not pursued.

The results from the in-house spot scanning delivery analysis are also listed in the Table, including the average value and standard deviation across the 3 repeated irradiations. Agreement between measured and calculated values were notably poorer for this irradiation condition. Both the left inferior and right superior TLD dose agreed with the calculated dose within 5%, but gamma analysis performed on the coronal and sagittal planes (Figure 5) showed poor agreement. The coronal plane showed a pass rate of only 72%, with all of the failing pixels spread uniformly throughout the CTV and a relatively large standard deviation on the result. The sagittal plane had an average pass rate of 87%, with disagreements noted at the distal edge but also in the CTV. The DTA results from the right-left profiles showed good alignment of the treatment field, but failing results were observed in the anterior-posterior profiles for 1 of the 3 irradiations in the distal fall-off region.

Overall, the preliminary testing highlighted that the phantom could be employed for testing CSI delivery. That is, the targets could be visualized, clinically realistic treatment plans could be developed for the phantom, and the detectors were suitably positioned in that they provided consistent results over repeated irradiations. While most proposed analysis metrics for this phantom were readily implemented, the junction analysis was found to be not suitable. Additionally, the ultimate acceptability criteria for passing the phantom could be proposed based on these results, but based on the deviations seen during the scanning beam testing, they could not yet be finalized.

Initial Credentialing Results

The results of the irradiations from the 10 institutions participating in baseline proton approval are shown in the Table, including the median and range for each parameter. Based on these results, the preliminary testing, and criteria for other IROC phantoms, acceptability criteria were established for the spine phantom (Table). Based on these criteria, 7 of the 10 external phantom irradiations were acceptable. This pass rate is comparable to that observed for other IROC Houston phantoms [8]. Of the 3 institutions that failed to meet criteria, 1 showed too low a dose in the TLD (6% low), and the other 2 showed range problems in the anterior-posterior direction (by 8 mm and 9 mm).

Discussion

Initial acceptance criteria of TLD point dose agreement within 5% and planar agreement with >85% of pixels passing a 5%/5 mm gamma criteria were selected to be consistent with
criteria used for the other IROC Houston proton phantoms. These criteria, and the distance to agreement criteria for the profile analysis, were found to be reasonable based on the results of this study and were therefore adopted for IROC Houston credentialing. It is important to note that these criteria are established for, and based on, clinical trial credentialing. While a failing result should be interpreted as a serious problem, simply passing the phantom is not a guarantee of optimal performance. A marginal pass should still likely prompt physicists to internally evaluate their proton radiotherapy program.

Some differences were seen between dosimetric tests within the phantom. The coronal plane had very high gamma passing rates; the sagittal plane showed reasonable, but lower, agreement. The source of this disagreement (Figures 3b and 5b) is most notable in the distal dose fall off region. Compared with the coronal film, the sagittal film measures the range beyond heterogeneities, which is a more challenging TPS calculation compared with the relatively uniform coronal plane. This increased challenge is not only observed for the primary proton calculation but also for the secondary particles that travel beyond the proton range [9]. The sagittal film is also subject to a wide range of proton linear energy transfer, making it a less certain measurement [10] than the comparatively uniform coronal film: with the increase in the linear energy transfer at the end of the range, darkening of the film is quenched due to signal saturation and an increase in dose may not cause an increase in dosimeter response [11]. Based on the more complicated dosimetry as well as the more challenging TPS calculation, it is unsurprising that the sagittal plane shows poorer agreement.

The phantom and dosimeters were shown to be accurate and robust under established passive scatter irradiation techniques during the preliminary testing. However, the preliminary test results using the pencil-beam scanning technique showed poorer agreement; it should be noted though that this technique was not used clinically in-house at the time of this experiment and the dose calculation and beam delivery would therefore be subject to increased uncertainty. Overall, most parameters evaluated still passed the initial acceptance criteria, although both sets of TLD results were approximately 4% lower than the TPS predicted dose. Previous studies have found an overestimation of the predicted dose by analytical dose calculations (compared with Monte Carlo calculations) [12], particularly near heterogeneities [13,14]. The coronal plane gamma analysis did not meet the proposed criteria, and the sagittal plane gamma analysis, along with the DTA in the anterior-posterior profiles, were only marginally within the proposed passing criteria. This preliminary test result contrasted with the larger community results, where the majority of institutions were able to successfully irradiate this phantom using a scanning beam technique. Of note, these other institutions had clinically implemented scanning proton therapy for this disease site. More thorough clinical testing of calculation and delivery likely explain the difference in the results.

Conclusion

An anthropomorphic proton spine phantom was designed to evaluate craniospinal radiotherapy. The inclusion of multiple tissue substitutes introduced heterogeneity and provided a realistic challenge for plan development and delivery. Consistent with previous
IROC Houston proton phantoms, a 5% absolute dose, 5%/5 mm gamma analysis criteria with 85% of pixels passing, and 5-mm profile analysis criteria were found to be reasonable. While most institutions were able to meet the established criteria, some dose disagreements were observed between the calculated and measured dose that could warrant follow up by the institution. This phantom was found to be suitable for clinical trial credentialing.

Acknowledgments

Acknowledgments: This work was supported by NIH (DHHS) CA180803, NIH (DHHS) CA214526 and federal share income earned by Massachusetts General Hospital Proton Therapy Research and Treatment Center, contract C06 CA059267.

References

1. Howell RM, Giebeler A, Koontz-Raisig W, Mahajan A, Etzel CJ, D’Amelio AM, Homann KL, Newhauser WD. Comparison of therapeutic dosimetric data from passively scattered proton and photon craniospinal irradiations for medulloblastoma. Radiat Oncol. 2012;7:116. [PubMed: 22820873]

2. National Cancer Institute. Guidelines for the Use of Proton Radiation Therapy in NCI-Sponsored Cooperative Group Clinical Trials. Bethesda, MD 2012.

3. Grant RL, Summers PA, Neihart JL, Blatnica AP, Sahoo N, Gillin MT, Followill DS, Ibott GS. Relative stopping power measurements to aid in the design of anthropomorphic phantoms for proton radiotherapy. J Appl Clin Med Phys. 2014;15: 4523. [PubMed: 24710437]

4. Moyers MF, Sardesai M, Sun S, Miller DW. Ion stopping powers and Ct numbers. Med Dosim. 2010;35:179–94. [PubMed: 19931030]

5. Kirby TH, Hanson WF, Johnston DA. Uncertainty analysis of absorbed dose calculations from thermoluminescence dosimeters. Med Phys. 1992;19:1427–33. [PubMed: 1461205]

6. Arjomandy B, Tailor R, Anand A, Sahoo N, Gillin M, Prado K, Vicic M. Energy dependence and dose response of Gafchromic EBT2 film over a wide range of photon, electron, and proton beam energies. Med Phys. 2010;37:1942–47. [PubMed: 20527528]

7. Liao L, Jiang S, Wang X, Li H, Zhu X, Sahoo N, Gillin M, Mahajan A, Grosshans D, Lim G, Zhang X. TH-C-BRD-12: robust intensity modulated proton therapy plan can eliminate junction shifts for craniospinal irradiation. Med Phys. 2014;41:553.

8. Taylor PA, Kry SF, Alvareza P, Keith T, Lujano C, Hernandez N, Followill DS. Results from the Imaging and Radiation Oncology Core Houston’s anthropomorphic phantoms used for proton therapy clinical trial credentialing. Int J Radiat Oncol Biol Phys. 2016;95:242–8. [PubMed: 27084644]

9. Paganetti H Nuclear interactions in proton therapy: dose and relative biological effect distributions originating from primary and secondary particles. Phys Med Biol. 2002;47:747–64. [PubMed: 11931469]

10. Zhao L, Das IJ. Gafchromic EBT film dosimetry in proton beams. Phys Med Biol. 2010;55:N291–301. [PubMed: 20427858]

11. Kirby D, Green S, Palmons H, Hugtenburg R, Wojnecki C, Parker D. LET dependence of GafChromic films and an ion chamber in low-energy proton dosimetry. Phys Med Biol. 2010;55:417–33. [PubMed: 20019399]

12. Schuemann J, Giantsoudi D, Grassberger C, Moteabbed M, Min CH, Paganetti H. Assessing the clinical impact of approximations in analytical dose calculations for proton therapy. Int J Radiat Oncol Biol Phys. 2015;92:1157–64. [PubMed: 26025779]

13. Yamashita T, Akagi T, Aso T, Kimura A, Sasaki T. Effect of inhomogeneity in a patient’s body on the accuracy of the pencil beam algorithm in comparison to Monte Carlo. Phys Med Biol. 2012;57:7673–88. [PubMed: 23123683]

14. Taylor PA, Kry SF, Followill DS. Pencil beam algorithms are unsuitable for proton dose calculations in lung. Int J Radiat Oncol Biol Phys. 2017;99(3):750–756. [PubMed: 28843371]
Figure 1.
Relative stopping power versus Hounsfield unit calibration curve comparing tested materials with the clinical calibration curve (Eclipse). Materials used in the phantom were those that fell onto the calibration curve, including solid water (tissue), blue water (cartilage), and Techron HPV bearing grade (bone).
Figure 2.
Anthropomorphic proton spine phantom. (a) 3-dimensional renderings of the phantom design, including an isolated image of the vertebral bodies. The sagittal and coronal film planes are shown in green in the full 3-dimensional render. (b) Computed tomography–based sagittal view of the phantom. The right superior thermoluminescent dosimeter location is visible in this slice and is circled in red.
Figure 3.
Gamma analysis color maps from scattered beam irradiation in (a) coronal and (b) sagittal planes.
Figure 4.
(a) Right-left and (b) anterior-posterior profiles from scattered beam delivery analysis. Images show film measurement and institution calculated results, with emphasis on the distance-to-agreement analysis that was conducted by averaging the agreement at the 75%, 50%, and 25% isodose lines.
Figure 5.
Gamma analysis color maps from scanning beam irradiation in (a) coronal and (b) sagittal planes.
Table 1.
TLD, gamma analysis, and profile evaluation results from the in-house passive scattered beam preliminary test, the in-house scanning beam preliminary test, and the independent irradiation of this phantom by 10 institutions. Acceptance criteria for each metric are also presented.

| Evaluation parameter                        | Scattered average (SD) | Scanning average (SD) | Independent median (range) | Current acceptance criteria |
|---------------------------------------------|------------------------|-----------------------|-----------------------------|-----------------------------|
| Superior TLD Ratio (Measured/Calculated)    | 1.009 (0.005)          | 0.960 (0.010)         | 0.980 (0.95–1.00)           | 0.95–1.05                   |
| Inferior TLD Ratio (Measured/Calculated)    | 1.020 (0.003)          | 0.957 (0.005)         | 0.975 (0.94–1.01)           | 0.95–1.05                   |
| Gamma result: coronal plane                 | 99% (0%)               | 72% (13%)             | 96% (57%–98%)               | ≤85% pixels passing 5%/5 mm |
| Gamma result: sagittal plane                | 90% (8%)               | 87% (2%)              | 91% (57%–97%)               | ≤85% pixels passing 5%/5 mm |
| Average right-left profile DTA (mm)         | 1.6 (0.3)              | 1.8 (0.8)             | 1.0 (0–5 mm)                | ≤5 mm                      |
| Average distal fall-off DTA (mm)            | 1.2 (0.6)              | 5.0 (1.3)             | 3.0 (0–9 mm)                | ≤5 mm                      |

Abbreviations: DTA, distance to agreement; TLD, thermoluminescent dosimeter;