Published in final edited form as:

Phys Med Biol. ; 63(4): 045003. doi:10.1088/1361-6560/aaa845.

Comparison of Monte Carlo and Analytical Dose Computations for Intensity Modulated Proton Therapy

Pablo Yepes^{1,2,*}, Antony Adair^{1,2}, David Grosshans², Dragan Mirkovic², Falk Poenisch², Uwe Titt², Qianxia Wang^{1,2}, and Radhe Mohan²

¹Physics and Astronomy Department, Rice University, MS 315, 6100 Main Street, Houston, Texas 77005

²Department of Radiation Physics, Unit 1202, The University of Texas M. D. Anderson Cancer, 1515 Holcombe Blvd., Houston, TX 77030, USA

Abstract

Purpose—To evaluate the effect of approximations in clinical analytical calculations performed by a Treatment Planning System (TPS) on dosimetric indices in Intensity Modulated Proton Therapy.

Methods—TPS calculated dose distributions were compared with dose distributions as estimated by Monte Carlo (MC) simulations, calculated with the Fast Dose Calculator (FDC) a system previously benchmarked to full MC. This study analyzed a total of 525 patients for four treatment sites (brain, head-and-neck, thorax and prostate). Dosimetric indices (D02, D05, D20, D50, D95, D98, EUD and Mean Dose) and a gamma-index analysis were utilized to evaluate the differences.

Results—The gamma-index passing rates for a 3%/3mm criterion for voxels with a dose larger than 10% of the maximum dose had a most probable value larger than 98% for all sites. The most probable difference for all dosimetric indices for target volumes was less than 2% for all cases. However, differences for target volumes as large as 10% were found for 2% of the thoracic patients. For organs at risk (OARs), the most probable absolute dose difference was smaller than 2 Gy for all indices and cohorts. However, absolute dose differences as large as 10 Gy were found for some small volume organs in brain and head-and-neck patients.

Conclusion—This analysis concludes that for a fraction of the patients studied, TPS may overestimate the dose in the target by as much as 10%, while for some OARs the dose could be underestimated by as much as 10 Gy. Monte Carlo dose calculations may be needed to ensure more accurate dose computations to improve target coverage and sparing of OARs in proton therapy.

Keywords

proton; therapy; particle; IMPT; Monte Carlo; comparison; analytical; TPS; dosimetric indices

*Correspondence: yepes@rice.edu; Tel.: +1-713-348-5942.

1. Introduction

The use of protons rather than photons for treatment of a variety of cancer types has the potential to reduce the integral dose. The physical characteristics of proton therapy, with its rapid distal falloff, offers potential advantages over traditional photon therapy, and in theory should translate into reducing the risk of toxicities and secondary cancers. Such potential physical and dosimetric benefits along with promising early clinical outcomes, has led to a significant increase in the number of proton therapy centers around the world. As of 2017, there are more than 70 operational hadron-therapy centers worldwide (PTCOG 2017). However, the theoretical advantages of proton therapy do not always translate into better clinical outcomes, which may be attributed, at least partially, to the relative small amount of clinical proton therapy experience compared to photons. While photon therapy has been used for decades, proton therapy is still to reach its full potential. Moreover, the sharp distal dose gradients, as well as uncertainties due to energy straggling, patient motion, intra- and inter-fractional changes in patient position, etc., may render protons more sensitive to changes in patient geometry. Such factors may lead to potentially under dosing target volumes and overexposing OARs in close proximity to the beam, with the corresponding risk of toxicity. Utilizing the most accurate calculation techniques available will reduce some of those uncertainties and will improve the quality of treatment.

An indispensable element necessary to achieve optimal clinical treatment plans is accurate dose calculations. Hitherto, commercial treatment planning systems typically calculate proton dose distributions with analytical pencil-beam algorithms (Russell *et al.*, 2000; Hong *et al.*, 1996; Deasy, 1998; Schneider *et al.*, 1998; Schaffner *et al.*, 1999; Soukup *et al.* 2005; Szymanowski and Oelfke, 2002). However, it is generally accepted that Monte Carlo techniques provide superior precision (Schaffner *et al.*, 1999; Titt *et al.*, 2008; Koch *et al.*, 2008; Paganetti *et al.*, 2008) than analytical algorithms. While commercial calculation engines based on Monte Carlo techniques are starting to become more widely available, there are few studies comparing dose distributions from MC and pencil-beam algorithms in a large number of proton therapy plans. Moreover, few studies have focused on IMPT, the newest and most advanced proton delivery modality.

In the present study, we compare the proton dose distributions calculated with the Treatment Planning System (TPS), Eclipse (Eclipse Treatment Planning System, Varian Medical Systems, Palo Alto, CA), with dose distributions obtained with the Fast Dose Calculator (FDC), a fast track-repeating-based Monte Carlo (Yepes *et al.*, 2009). FDC has been validated versus GEANT4 (Agostinelli *et al.*, 2003) for a sample of 25 patients (Yepes *et al.*, 2016b) and versus measurements in water (Yepes *et al.*, 2016a). This study focuses on a large group of patients treated with Intensity Modulated Proton Therapy (IMPT). Recent comparisons of dose calculations and measurements in anthropomorphic phantoms seem to indicate that pencil beam algorithms for IMPT have significant inaccuracies in heterogeneous geometries in thorax (Taylor *et al.*, 2017).

2. Materials and Methods

2.1. Patient Data

Proton treatment plans previously delivered at the University of Texas MD Anderson Cancer Center (MDACC) were selected from the patient database and classified in four clinical sites: brain, head-and-neck, thorax, and prostate, according to the location of the target. Following our previous work (Yepes *et al.*, 2016b), the irradiated volume was represented as a voxelized phantom based on the CT images of the corresponding patients.

For IMPT patients the number of energy layers conveys information about the length of the Spread-Out-Bragg-Peaks (SOBPs), while the minimum and maximum energy are related to the beam range. For some patients, more than one plan was considered when adaptive plans were utilized in the treatment. Table 1 shows for each cohort the following clinical data: the numbers of patients, the number of plans, the average target volume, the number of beams used, the minimum and maximum beam ranges, the number of energy layers per beam, the number of voxels of the CT used for the calculations, and size of the voxels in the x (lateral), y (anterior-posterior), and z (inferior-superior) axes. The number of energy layers per beam is defined as the total number of layers for a plan divided by the number of beams.

For each plan in the study, the Monte Carlo calculation utilizes as input the Radiation Therapy (RT) plan, CT images, and the structure or organ contours all in DICOM (Digital Imaging and Communications in Medicine) format, from the clinical treatment planning system. The patient anatomy was represented as a voxelized phantom based on the CT image data set. Each voxel was assigned a material composition and density that corresponded to the Hounsfield unit value in the CT scan for that voxel, as described by (Schneider *et al.*, 1996). In the transverse plane a reduction in resolution by a factor of two in each axis was implemented to avoid very large total number of voxels. In addition, the spatial grid from CTs, obtained from the clinical database, were reduced in size by an automatic procedure, which crops CT voxels on the edges with an assigned density smaller than 0.05 g cm^{-3} . The procedure eliminates voxels assigned to air on the edges of the image included in the CT images. After the procedure, the number of voxels in the phantoms ranged from approximately 5×10^5 for some brain patients to up to 2.6×10^7 for one of the HN patients, as can be seen in Table 1.

The voxel size in the axes on the transverse plane was in the range 1.37-2.34 mm, while in the inferior-superior axis, the range was 1-2.5 mm. The smallest voxels were used in brain patients with relative small targets, while the largest voxels were used for thoracic patients for whom target volumes can be large. The mean target volume ranged from 87.8 cm^3 for prostate to 366 cm^3 for thorax patients. Most of the treatment plans utilized three beams (or fields) except for prostate, where the standard of care in proton therapy is two, even though occasionally a third posterior/anterior beam is also utilized. For some more complex cases the number of beams was increased to four (brain, thorax) or five (head-and-neck). The plan-averaged minimum beam range was 2 mm for all cohorts, except for prostate where it was 47.8 mm for a patient with three beams. Similarly, the plan-averaged maximum beam range was smallest for brain (147.3 mm) and maximum for prostate (249 mm).

A fraction of the beams used for patient treatment include an ABS resin Range Shifter (RS) with a thickness of 67 mm water-equivalent, used to achieve ranges below the range corresponding to the minimum energy of the synchrotron source. The fraction of the beams using the RS depends on the anatomical site and is: 92.3% for head-and-neck, 50% for brain, 34.8 for thorax, and none for prostate patients. The position of the RS ranges from 95 mm to 475 mm from the isocenter. Results were found not to depend on the presence of a RS.

2.2. Structure Selection

A selection of structures was performed from among the structures contoured by physicians and stored in the DICOM structure file. The selected structures were those most frequently contoured by the physicians at our institution. Since the naming of structures is not standardized, a whitelist of structures of interest was developed. The stored structure label was simplified by deleting non-alphabetic characters. Structures were selected if their simplified-label was found in the whitelist. Among the selected structures for each patient were the various target volumes (TV), and the OARs typical for each site.

2.3 Treatment Planning System (TPS)

The Treatment Planning System employed at the Proton Therapy Center (PTC) at The University of Texas MD Anderson Cancer center is Eclipse (Varian Medical Systems). For treatment plans in this study, the Eclipse versions comprised the range from 8.9 to 13.6. Quality assurance (QA) measurements of dose profiles in water routinely performed at PTC did not reveal any significant differences between dose distributions computed with the different software versions included in this study.

2.4 Monte Carlo

Monte Carlo calculations were performed by the Fast Dose Calculator (FDC). A description of the basic FDC algorithm can be found elsewhere (Yepes *et al.*, 2009). It can be briefly summarized as follows: a database of pre-generated proton tracks in water is calculated with GEANT4. This step is done only once and it does not need to be repeated each time a new dose calculation is performed. That database of tracks is re-utilized to calculate dose distributions deposited by protons in any material by scaling the ranges and scattering angles of the trajectories of all particles produced by the primary protons, and stored in the database.

FDC was validated versus measurements of mono-energetic beams in water (Yepes *et al.*, 2016a), which proves the accuracy of the code to describe the beam at the facility where patients were treated. Moreover, FDC computations were also validated versus full Monte Carlo simulations (GEANT4) for a set of 25 patients (Yepes *et al.*, 2016b). The latter probes the validity of FDC to simulate proton dose in heterogeneous media, since GEANT4 has been extensively validated (Carrier *et al.*, 2004).

The modeling of the particle source was done in two steps. The first step was performed by matching experimental dose distributions in water with calculation from a full Monte Carlo, MCNPX (Walters *et al.*, 2005). Normal distributions were utilized to model the position of

the particles on the transverse plane at the source. The particle energy distribution was modeled with a normal function, while the particle divergence was modelled with a combination of one square and two normal distributions. In the second step, the functions tuned for MCNPX were further adjusted for FDC, as explained in a previous publication (Yepes *et al.*, 2016a).

For each voxel i , FDC calculates the energy deposited per source proton and per volume, S_i . The delivered dose in voxel i (D_i) by a proton beam of energy E_p , is, then, calculated with the expression: $D_i = S_i \Phi C_p(E_p)$, where Φ is the proton fluence, measured in Monitor Units (MU), and C_p is an energy dependent function, which provides the number of source protons per MU. C_p is implemented as a lookup table for each of the discrete energies used in the treatment plans. Its initial values were obtained from comparisons between FDC and measurements of depth dose profiles in water for all the available energies (Yepes *et al* 2016b). Moreover, C_p values were further scaled by a factor 1.33%, obtained from comparisons of FDC calculations and more precise measurements for a reference treatment plan, intended to fix the overall delivered dose scale. Systematic uncertainties in the total dose was estimated to be less than 2%. The estimate was obtained from comparing integrated dose per beam in measurements and calculations in patient plans delivered in a water phantom, also referred to as Patient Quality Assurance plans.

The physical dose calculated with MC and the TPS was multiplied by a constant factor of 1.1 in order to take into account the relative biological effectiveness of protons.

As in any MC calculation, it is essential to make sure that statistical uncertainties are smaller than an estimated systematic uncertainty. FDC calculates the statistical uncertainty of the dose in each voxel, σ_D . The σ_D voxel average, $\langle \sigma_D \rangle$, for voxels with a dose larger than 10% of the maximum dose, D_{max} , was used to determine whether the number of simulated histories was large enough to satisfy the uncertainty requirements. In this study, dose calculations were performed for each individual beam or field with a default of 100 million histories. However, the simulations were stopped if $\langle \sigma_D \rangle < 0.3\%$.

2.5 Computer Resources

FDC calculations were executed on a server from Exxact Corporation 2016 with 48 Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60GHz processors, with 252 GBytes of RAM, and 1 TBytes of disk space.

2.5 Analysis

Dose to water calculations by MC were performed using the same CT sets as those used for the TPS calculations. For thorax patients, a CT averaged over ten phases of a 4D-CT, representing a breathing cycle, was utilized for both the TPS and FDC calculations. The typical FDC calculation time for one patient was around 5-10 minutes.

The MC dose distributions were then compared to those obtained by TPS. 3D dose distributions, Dose-Volume-Histograms (DVHs) and gamma indices (Low *et al.*, 1998; Ju *et al.*, 2008) of FDC relative to TPS, were calculated.

2.5.1 Gamma-index Analysis—3D gamma-index calculations were performed using the distance-to-simplex approximation as described in Ju *et al.* (Ju *et al.*, 2008) for the complete treatment plan for all patients under study. The TPS dose distributions were used as the reference and MC as the sample distribution. Three criteria were applied: 1%/1mm, 2%/2mm and 3%/3mm. The passing rate was defined, for voxels with a dose larger than 10% of the maximum dose, as the fraction of voxels with a gamma-index value smaller than unity.

2.5.2 Dosimetric Analysis—DVHs were calculated for selected structures for both, the TPS- and MC dose distributions with the same algorithm. In addition, the following dosimetric indices were obtained to provide a quantitative comparison of the DVH calculated with both approaches:

1. The mean structure dose averaged over all voxels in the structure.
2. D02, D05, D20, D50, D95, and D98: the maximum dose covering 2%, 5%, 20%, 50%, 95% and 98% of the considered structure respectively. For the target volume analysis, those indices were only considered when their value was larger than 10 Gy. Not all indices have the same importance for all structures. For example, D98 or D95 are mostly relevant for target volumes, while in general D02 or D05 are more significant for parallel organs, like the spine or gastrointestinal tract.
3. The Equivalent Uniform dose was also calculated for target volumes with the formula:

$$EUD = \left(\sum_i n_i D_i^a \right)^{\left(\frac{1}{a} \right)}$$

Where n_i is the fraction of the target volume with dose D_i , and a is parameter specific for each structure. For target volumes, a value of $a=10$ was utilized. Such a value of the parameter tends to make EUD close to the D95 and D98.

Under the label of target volumes, TV, “Gross Target Volume” (GTV), “Clinical Target Volume” (CTV), “Plan Target Volume” (PTV), Internal Target Volume (ITV), and Scanning Target Volume (STV) (Zhu 2010) were included.

The differences in the dosimetric indices between the Monte Carlo calculations and TPS dose distributions were analyzed. For the target volumes, the relative differences between the indices were evaluated. This choice was motivated by the fact that some of the plans are so called boost or adaptive plans with a dose significantly smaller than the prescribed total dose.

For each of the four considered treatment sites, a set of OARs was selected. They were: bladder, optic chiasm, cochlea, esophagus, femur, heart, larynx, lung, optic nerve, parotid, rectal wall, rectum, and spinal cord. For those OARs, the difference between the Monte Carlo and TPS results are presented in absolute value, since toxicity is typically estimated

from the absolute dose. Such difference is presented for each plan, even in the case when boost or adaptive plans are utilized. This approach underestimates the differences in the dose received by those OARs during the overall treatment. However, adding the dose from different plans would require an amount of work beyond the scope of this study.

2.5.3. Statistics—We are presenting the results of this study with Descriptive Statistics, in the form of boxplots. Each box plots represents the distribution of a variable, like the difference in a dosimetric index. Boxes have three horizontal lines: the bottom one represents the first quartile, the middle one corresponds to the median, while the top one depicts the third quartile. The top whisker corresponds to the position for which 98% of the entries in the distribution are below, while the bottom whisker depicts the point for which 98% are above. Each point above and below the whiskers represent one entry higher than the 98 or lower than the 2%.

3. Results

3.1. Gamma-index Analysis

Figure 1 depicts the passing rate for the gamma-index with TPS doses as the reference and MC doses as the sample for 1mm/1% (PR11), 2mm/2% (PR22), and 3mm/3% (PR33) passing criteria. As can be seen, PR33 is close to unity for all cohorts, while PR22 is best for brain patients followed by the prostate cohort. This is to be expected because head-and-neck and thorax are anatomical regions with more complex heterogeneities, where the differences between the analytical algorithm (TPS) and the Monte Carlo (FDC) are expected to be larger. PR11 for prostate is of the same order as that for thorax but with a smaller dispersion. Comparison between MC and TPS doses naively is expected to be closer for prostate than for thorax, because the heterogeneities are smaller in the former. However, the results can be explained by the fact that beams in prostate patients traverse a substantial amount of tissue, and, thus, are more sensitive to systematic differences in the proton range between MC and TPS in those tissues.

Figure 2 presents one minus PR11, one minus PR22 and one minus PR33 in log scale in order to better appreciate the proximity of the various passing rates to unity. As can be seen, one minus PR33 is smaller than 1% for all cohorts except for HN for which it is 0.02. The best average agreement is found for brain patients for which it is slightly above 0.1%.

5.2. Dosimetric Analysis

The relative difference of the dosimetric indices for target volumes is presented in Figure 3 for the four cohorts in the study. The numbers of plans analyzed is indicated under the title of the cohort, and they range from 408 for the prostate cohort to 1033 for the head-and-neck group.

The most probable difference between FDC and TPS predicted mean dose and D50 is within 1.5% for all cohorts. The brain and head-and-neck cohorts have values for FDC slightly higher than for TPS, while the opposite is the case for the other two cohorts. The maximum differences for the mean dose are found for brain and head-and-neck patients. For all cohorts except for prostate, for 2% of the patients FDC predicts a mean dose and D50 about 4-5%

larger than TPS, while for prostate patients it is around 1%. For D02 and D05, the difference is around 1% larger than for D50 and the mean dose. This is to be expected, and we attribute such a behavior to the fact that the plans were optimized with the TPS. When a plan is optimized with a dose algorithm, and then the dose calculated with an alternative algorithm, the DVHs from the algorithm used for the optimization tend to be sharper than DVHs from the alternative algorithm. This is also illustrated in Figure 4 showing the DVH graph for a typical thorax patient. As can be seen, FDC fall off is less sharp for FDC than for TPS.

On the contrary, for D98 and D95, the dose predicted by FDC is slightly lower than the value for TPS for all cohorts except for brain patients. We attribute that to the fact that brain does not have large heterogeneities and relatively low proton energies are used, thus low range. Head-and-neck and thorax patients have substantial heterogeneities, and prostate has long ranges. The longer the range the more affected is the beam by heterogeneities or range uncertainties. Concerning EUD, for target volumes, it tends to be close to D98, even though it is more sensitive to cold spots in the target, which tend to reduce its value.

Concerning the dispersion of the distributions, it can be observed that the cohort with the least dispersion in all dosimetric indices is the prostate cohort. That can to a great extent be attributed to the homogeneity of that anatomical region. On the other hand, the largest differences are observed for head-and-neck, and, especially, thorax patients. Among the dosimetric indices, D98 shows the largest dispersion, because it is the most sensitive to small deviations. For 2% of thoracic patients, FDC predicts a dose at least 4% larger than the TPS, and another 2% of plans showing FDC 10% less than TPS predictions. The latter indicates the possibility for compromised target coverage (Taylor *et al.*, 2017). An example of plan for which MC dose prediction is significantly lower than TPS is shown in Figure 5.

For the EUD index, one treatment plan for head-and-neck and two plans for thorax exhibited differences below -20% and fall outside of the limits of the plot. However, the differences in D98 for the three plans was within 4%. The large differences in EUD were due to a small target area with lower dose in FDC compared to TPS, what caused such a large difference in the EUD value.

As an example of a rare large discrepancy, Figure 6 presents an axial-plane view of the PTV for a thorax patient for which the difference in D98 between MC and TPS was larger than 10%. As can be seen, there is an area in the right-lower area of the PTV (outlined in blue) for which the dose predicted by MC is 10 Gy lower than that predicted by the TPS. It has to be noted that the target is at the center of the thorax and the beams needed to traverse highly heterogeneous regions to reach the target.

Figure 7 presents the differences in dosimetric indices for the OARs for the brain cohort. As can be seen, the most probable difference, represented by the horizontal line at the center of the boxes, for all OARs is of the order of 1 Gy. Nevertheless, individual plans present significant differences for some of the indices. The top 2% difference is above 6 Gy for all the indices of cochlea and D05 and D02 for chiasm. There is two cochlea plans for which the difference is above 10 Gy for D02, D05, and D20. The cochlea is a very small organ

with volumes of the order of 1 cm^3 , causing increased sensitivity to relatively small dose fluctuations.

Similar results for the head-and-neck cohort are presented in Figure 8. The differences for the most probable values are within 1% for all OARs. The bottom 2% for all indices is above -5 Gy, while the top 2% is smaller than 5 Gy, except for cochlea and optic chiasm. For the former, differences as large as 10 Gy were found in D02 and D05 for one plan.

Figure 9 presents the results in the thorax plans for lung, heart, esophagus and spinal cord. The most probable differences in the indices for all OARs are within 1%, while the top 2% is always below 5 Gy. For D20, there are one lung structure and one heart structure where the difference is close to 10 Gy, falling outside of the limits of the plot.

The differences in the dosimetric indices for prostate patients are presented in Figure 10 for the bladder, femur, rectum and rectal wall. The most likely values are also within 1%, except for D50 for the rectum and bladder. As can be observed, the dispersion of the differences for the prostate case are smaller than for other anatomic areas. The largest differences for the dosimetric indices are around 3%. Such small dispersion of the differences can be understood due to the relative homogeneity of the pelvis area, which minimizes scattering, and thus the differences between TPS doses and MC doses.

3. Discussion

The Monte Carlo method, if properly validated, is generally accepted to be a more accurate method to calculate dose distributions. The main purpose of this study was to evaluate, in a large cohort of patients treated with IMPT, the significance of the deviations between Monte Carlo and analytical dose calculations currently used in the TPS. Schuemann *et al.* published results in a similar study for a smaller cohort treated with Passive Scattering Proton Therapy (Schuemann *et al.*, 2015), as opposed to IMPT, used in this analysis.

The main concern when analyzing the dose differences in target volumes (TV) is whether TPS predicts significantly higher doses than the estimates from MC (Figure 3). That would imply that the targets may be irradiated with a lower than prescribed dose increasing the risk of recurrence. The number of plans for which the MC under-predicts TPS for any dosimetric index by more than 5% is 2 out of 134 (1.5%) for brain, 5 out of 125 (4%) for head-and-neck, 28 out of 159 (18%) for thorax, and none out of 107 for prostate. For the case of an under-prediction by more than 10%, 2 out of 125 head-and-neck and 8 out of 159 thorax patients are found. If EUD is excluded, which is particularly sensitive to dose variations, 4 thorax patients have any index under predicted by MC by more than 10%, and 2 by more than 15% for both of the patients with an under-prediction larger than 15%, the target is located in the center of the thorax. These results are in agreement with the above mentioned PSPT study by Schuemann *et al.*, which has also found thorax as the most challenging site for the analytical dose computation algorithms, followed by head-and-neck.

For the OARs, the concern is the opposite, cases where MC predicts significant higher doses than the TPS. The cases where the largest increases are seen are small organs (optic chiasm and optic nerve, as well as cochlea). This can be explained due to their small volume they

are more sensitive to changes in dose distributions. Increases in absolute dose as large as 10 Gy are observed for D02 and D05.

Our results are in agreement with the phantom study performed by (Taylor *et al.*, 2017). However, discrepancies seem to be less pronounced in our analysis. This might be attributed to the fact that the anthropomorphic phantoms used in that study were implemented with a motion larger than what is seen in most of the proton plans processed in this study. Even though, neither study considered motion implicitly, both used moving targets. In their case, a moving phantom was used, while our study analyzed thoracic patients for whom a CT averaged over the breathing cycle was utilized for the dose calculations, and the treatment was delivered in the free breathing mode.

Future works will aim at achieving a better understanding of the root causes for the differences between Monte Carlo and TPS.

5. Conclusions

Proton treatment plans of cohorts of above 100 patients each for brain, head-and-neck, thorax and prostate were used to compare MC and TPS dose calculations. Agreement within 2% was found for the most probable values of dosimetric indices calculated with both sets of dose prediction algorithms. However, differences in dosimetric indices as large as 10% were found for 2% of thoracic plans for target volumes. The largest differences for target volumes were found for thoracic plans. Conversely, absolute differences in dosimetric indices for OAR as large as 10 Gy were found for some head-and-neck and brain plans, mostly for relatively small organs.

Given the differences found in this analysis, we recommend the use of MC techniques for all dose calculations in proton therapy, or at least for the most challenging sites like thorax and head-and-neck.

Acknowledgments

This study was supported by the support provided by the Cancer Research Institute of Texas grant award RP160232, and the National Cancer Institute at the National Institute of Health award 2U19CA021239-35.

References

- Agostinelli S, Allison J, Amako K, Apostolakis J, Araujo H, Arce P, Asai M, Axen D, Banerjee S, Barrand G, Behner F, Bellagamba L, Boudreau J, Broglia L, Brunengo A, Burkhardt H, Chauvie S, Chuma J, Chytrcek R, Cooperman G, Cosmo G, Degtyarenko P, Dell'Acqua A, Depaola G, Dietrich D, Enami R, Feliciello A, Ferguson C, Fesefeldt H, Folger G, Foppiano F, Forti A, Garelli S, Giani S, Giannitrapani R, Gibin D, Gomez Cadenas JJ, Gonzalez I, Gracia Abril G, Greeniaus LG, Greiner W, Grichine V, Grossheim A, Gumplinger P, Hamatsu R, Hashimoto K, Hasui H, Heikkinen A, Howard A, Ivanchenko V, Johnson A, Jones FW, Kallenbach J, Kanaya N, Kawabata M, Kawabata Y, Kawaguti M, Kelner S, Kent P, Kodama T, Kokoulin R, Kossov M, Kurashige H, Lamanna E, Lampen T, Lara V, Lefebvre V, Lei F, Liendl M, Lockman W, Longo F, Magni S, Maire M, Medernach E, Minamimoto K, Mora de Freitas P, Morita Y, Murakami K, Nagamatsu M, Nartallo R, Nieminen P, Nishimura T, Ohtsubo K, Okamura M, O'Neale S, Oohata Y, Paech K, Perl J, Pfeiffer A, Pia MG, Ranjard F, Rybin A, Sadilov S, Di Salvo E, Santin G, Sasaki T, Savvas N, Sawada Y, Scherer S, Sei S, Sirotenko V, Smith D, Starkov N, Stoecker H, Sulkimo J, Takahata M, Tanaka S, Tcherniaev E, Safai Tehrani F, Tropeano M, Truscott P, Uno H, Urban L, Urban P, Verderi

- M, Walkden A, Wander W, Weber H, Wellisch JP, Wenaus T, Williams DC, Wright D, Yamada T, Yoshida H, Zschiesche D. GEANT4: A Simulation toolkit. Nucl Instrum Meth. 2003; A506:250–303.
- Carrier JF, Archambault L, Beaulieu L, Roy R. Validation of GEANT4, an object-oriented Monte Carlo toolkit, for simulation in medical physics. Med Phys. 2004; 31:484–92. [PubMed: 15070244]
- Deasy JO. A proton dose calculation algorithm for conformal therapy simulations based on Moliere's theory of lateral deflections. Med Phys. 1998; 25:476–83. [PubMed: 9571613]
- Digital Imaging and Communications in Medicine D. <http://dicom.nema.org>
- Hong L, Goitein M, Bucciolini M, Comiskey R, Gottschalk B, Rosenthal S, Serago C, Urie M. A pencil beam algorithm for proton dose calculations. Phys Med Biol. 1996; 41:1305–30. [PubMed: 8858722]
- Ju T, Simpson T, Deasy JO, Low DA. Geometric interpretation of the gamma dose distribution comparison technique: interpolation-free calculation. Med Phys. 2008; 35:879–87. [PubMed: 18404924]
- Koch N, Newhauser WD, Titt U, Gombos D, Coombes K, Starkschall G. Monte Carlo calculations and measurements of absorbed dose per monitor unit for the treatment of uveal melanoma with proton therapy. Phys Med Biol. 2008; 53:1581–94. [PubMed: 18367789]
- Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys. 1998; 25:656–61. [PubMed: 9608475]
- Paganetti H, Jiang H, Parodi K, Slopsema R, Engelsman M. Clinical implementation of full Monte Carlo dose calculation in proton beam therapy. Phys Med Biol. 2008; 53:4825–53. [PubMed: 18701772]
- PTCOG. 2017. <https://www.ptcog.ch/index.php/facilities-in-operation>
- Russell KR, Isacson U, Saxner M, Ahnesjö A, Montelius A, Grusell E, Dahlgren CV, Lorin S, Glimelius B. Implementation of pencil kernel and depth penetration algorithms for treatment planning of proton beams. Phys Med Biol. 2000; 45:9–27. [PubMed: 10661580]
- Schaffner B, Pedroni E, Lomax A. Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation. Phys Med Biol. 1999; 44:27–41. [PubMed: 10071873]
- Schneider U, Pedroni E, Lomax A. The calibration of CT Hounsfield units for radiotherapy treatment planning. Phys Med Biol. 1996; 41:111–24. [PubMed: 8685250]
- Schneider U, Schaffner B, Lomax T, Pedroni E, Tourovsky A. A technique for calculating range spectra of charged particle beams distal to thick inhomogeneities. Med Phys. 1998; 25:457–63. [PubMed: 9571611]
- Schueman J, Giantsoudi D, Grassberger C, Moteabbed M, Min CH, Paganetti H. Assessing the Clinical Impact of Approximations in Analytical Dose Calculations in Proton Therapy. Int J Radiat Oncol Biol Phys. 2015; 92:1157–64. [PubMed: 26025779]
- Soukup M, Fippel M, Alber M. A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo Simulations. Phys Med Biol. 2005; 50:5089–104. [PubMed: 16237243]
- Szymanowski H, Oelfke U. Two-dimensional pencil beam scaling: an improved proton dose algorithm for heterogeneous media. Phys Med Biol. 2002; 47:3313–30. [PubMed: 12375823]
- Taylor PA, Kry SF, Followill DS. Pencil Beam Algorithms Are Unsuitable for Proton Dose Calculations in Lung. Int J Radiat Oncol Biol Phys. 2017; 99:750–56. [PubMed: 28843371]
- Titt U, Sahoo N, Ding X, Zheng Y, Newhauser WD, Zhu XR, Polf JC, Gillin MT, Mohan R. Assessment of the accuracy of an MCNPX-based Monte Carlo simulation model for predicting three-dimensional absorbed dose distributions. Phys Med Biol. 2008; 53:4455–70. [PubMed: 18670050]
- Waters, LS., Hendricks, J., McKinney, G. MCNPX version 2.5.0. Los Alamos: Los Alamos National Laboratory; 2005.
- Yepes P, Guan F, Kerr M, Randeniya S, Li Y, Bronk L, Liu A, Mirkovic D, Sahoo N, Titt U, Anand A, Mohan R. Validation of a track-repeating algorithm versus measurements in water for proton scanning beams. Biomed Phys Eng Express. 2016a; 2:037002.
- Yepes P, Randeniya S, Taddei PJ, Newhauser WD. A Track-Repeating Algorithm for Fast Monte Carlo Dose Calculations of Proton Radiotherapy. Nucl Technol. 2009; 168:736–40. [PubMed: 20865140]

- Yepes PP, Eley JG, Liu A, Mirkovic D, Randeniya S, Titt U, Mohan R. Validation of a track repeating algorithm for intensity modulated proton therapy: clinical cases study. *Phys Med Biol.* 2016b; 61:2633–45. [PubMed: 26961764]
- Zhu XR, Sahoo N, Zhang X, Robertson D, Li H. Intensity modulated proton therapy treatment planning using single-field optimization: The impact of monitor unit constraints on plan quality. *Med Phys.* 2010; 37:1210–19. [PubMed: 20384258]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
The passing rate of the gamma-index of FDC relative to the TPS for 1mm/1%, 2mm/2% and 3mm/3% in linear vertical scale for the four cohorts considered in the study. The number of plans considered for each cohort is indicated below the cohort name.

Figure 2. One minus the passing rate of the gamma-index of FDC relative to the TPS for 1mm/1%, 2mm/2% and 3mm/3% in log vertical scale. The number of plans considered for each cohort is indicated below the cohort name.

Figure 3.
Relative difference between TPS and FDC for dosimetric parameters for target volumes for the four cohorts considered in the study.

Figure 4.
Example of a Dose-Volume-Histogram for a thorax patient showing the distributions for TPS and FDC for two target structures and several organs at risk.

Figure 5.
Example of a Dose-Volume-Histogram for a thorax patient with a significant difference between TPS and MC for the target coverage.

Figure 6. View of the dose distributions by MC (left), TPS (center) and their difference (right) for a view on the axial plane of a thorax patient with a large difference of the D98 for PTV.

Figure 7.
Dosimetric indices difference between MC and TPS for organs at risk for the brain cohort

Figure 8.
Dosimetric indices difference between MC and TPS for organs at risk for the head-and-neck cohort.

Figure 9.
Dosimetric indices difference between MC and TPS for organs at risk for the thorax cohort.

Figure 10.
Dosimetric indices difference between MC and TPS for organs at risk in the prostate cohort.

Table 1

Some characteristics of the treatment plans and the calculation CT grids used for the four cohorts of patients.

Site	Brain	Head-and-Neck	Thorax	Prostate
Number of Patients	134	125	159	107
Number of Plans	195	144	185	130
Average Target Volume (cm ³)	98.4	178.6	366.2	87.8
Number Beams	Average Minimum Maximum	2.9 2 4	3.0 2 5	2.9 1 4
Minimum Range (mm)	Average Minimum Maximum	33.3 2.0 88.5	10.4 2.0 132.1	41.7 2.0 114.6
Maximum Range (mm)	Average Minimum Maximum	147.3 91.9 208.3	173.0 64.8 238.1	191.5 37.4 297.7
# Energies/Beam	Average Minimum Maximum	31.3 11 54	44.5 12 62	41.0 15 62
# CT Voxels (10 ⁶)	Average Minimum Maximum	9.7 0.5 18.5	9.0 2.1 25.6	6.7 2.5 13.5
Transverse plane CT voxel size (mm)	Average Minimum Maximum	1.91 1.37 2.34	1.97 1.95 2.34	2.07 1.95 2.34
Inferior-superior CT voxel size (mm)	Average Minimum Maximum	1.61 1 3	2.33 1 3	2.48 1.25 2.5