IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **62** (2017) 246–257 doi:10.1088/1361-6560/62/1/246

Limitations of analytical dose calculations for small field proton radiosurgery

Changran Geng, Juliane Daartz, Kimberley Lam-Tin-Cheung, Marc Bussiere, Helen A Shih, Harald Paganetti and Jan Schuemann

Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, 30 Fruit Street, Boston, MA 02114, USA

E-mail: cgeng@mgh.harvard.edu and jschuemann@mgh.harvard.edu

Received 24 August 2016, revised 20 October 2016
Accepted for publication 8 November 2016
Published 16 December 2016

Abstract

The purpose of the work was to evaluate the dosimetric uncertainties of an analytical dose calculation engine and the impact on treatment plans using small fields in intracranial proton stereotactic radiosurgery (PSRS) for a gantry based double scattering system. 50 patients were evaluated including 10 patients for each of 5 diagnostic indications of: arteriovenous malformation (AVM), acoustic neuroma (AN), meningioma (MGM), metastasis (METS), and pituitary adenoma (PIT). Treatment plans followed standard prescription and optimization procedures for PSRS. We performed comparisons between delivered dose distributions, determined by Monte Carlo (MC) simulations, and those calculated with the analytical dose calculation algorithm (ADC) used in our current treatment planning system in terms of dose volume histogram parameters and beam range distributions. Results show that the difference in the dose to 95% of the target (D95) is within 6% when applying measured field size output corrections for AN, MGM, and PIT. However, for AVM and METS, the differences can be as great as 10% and 12%, respectively. Normalizing the MC dose to the ADC dose based on the dose of voxels in a central area of the target reduces the difference of the D95 to within 6% for all sites. The generally applied margin to cover uncertainties in range (3.5% of the prescribed range + 1 mm) is not sufficient to cover the range uncertainty for ADC in all cases, especially for patients with high tissue heterogeneity. The root mean square of the *R*90 difference, the difference in the position of distal falloff to 90% of the prescribed dose, is affected by several factors, especially the patient geometry heterogeneity, modulation and field diameter. In conclusion, implementation of Monte Carlo dose calculation techniques into the clinic can reduce the uncertainty of the target dose for proton stereotactic radiosurgery. If MC is not available for treatment planning, using MC dose

distributions to adjust the delivered doses level can also reduce uncertainties below 3% for mean target dose and 6% for the D95.

Keywords: proton stereotactic radiosurgery, analytical dose calculation, Monte Carlo, small field dosimetry

(Some figures may appear in colour only in the online journal)

1. Introduction

Accurate planning and delivery in proton SRS (PSRS) remains challenging due to small target volumes and typically high doses delivered in just one or two fractions. A critical factor is the accuracy of the dose calculation method. The primary goal of analytical dose calculation algorithms (ADC) used for treatment planning is to provide fast computation with clinically acceptable uncertainty. However, approximations are required in order to enable fast calculations. This generally means the effects of multiple Coulomb scattering (MCS) are not taken into account correctly (Grassberger *et al* 2014, Schuemann *et al* 2015). Furthermore, scattering at the aperture and aperture thickness are typically not taken into account as they are generally modeled as binary 2D objects. The impact of these effects is likely most significant when treating with small fields. With decreasing aperture size and increasing beam range in patients, the scattering in small fields results in a decrease in dose along the central axis of the Bragg peak due to lack of lateral equilibrium. This effect has been experimentally quantified for various aperture sizes and SOBP (spread-out Bragg peak) configurations leading to a clinically applied output factor (OF) adjustment to compensate for the loss of lateral equilibrium (Daartz *et al* 2009). The procedure is used to determine field specific calibrations of the monitor unit (MU) to the prescribed dose in the clinic.

Monte Carlo (MC) simulations are the gold standard for proton dose calculations and can accurately describe such effects (Perl *et al* 2012, Paganetti 2014, Guan *et al* 2015). Proton MC simulations have been widely used to assess the uncertainty for dose calculations with analytical methods (Bednarz *et al* 2010, Grassberger *et al* 2014, Schuemann *et al* 2014, 2015). MC simulations do not account for uncertainties in patient setup or conversion from CT Hounsfield Units to materials (Paganetti 2012). While the dominating factors for differences in dose distributions between MC and ADC for typical (larger) treatment fields arise from MCS and loss of secondary particle equilibrium, for small fields the scattering effects will be magnified due to the loss of primary charged particle equilibrium.

There are two consequences of the differences between ADC and MC, i.e. the delivered target dose and the uncertainty in the predicted range (Schuemann *et al* 2014, 2015). This work aims to understand and quantify these limitations for small fields in more detail and assess whether MC can be used to improve on the clinically used output factor and the accuracy of the dose estimation. We investigated correlations between differences in range and absolute target dose as a function of the prescribed range, thickness of the range compensator, field diameter, and heterogeneity for intracranial PSRS patients for a gantry-based double scattering beamline.

2. Materials and methods

2.1. Patient cohort and treatment planning

The patient cohort consisted of 50 patients (10 patients from 5 indications) who were treated using a gantry based double scattering system. Patients were sorted into groups

Table 1. Median value and range of values for various treatment properties and analysis results for each diagnostic indication.

Diagnostic indication	AVM	AN	MGM	PIT	METS
No. of patients	10	10	10	10	10
No. of beams	45	31	35	33	55
Range of prescribed ranges (mm)	96 (75–127)	111 (88–170)	108 (80–162)	111 (88–170)	84 (50–158)
Range of prescribed modulation width (mm)	20 (15–58)	20 (15–44)	23 (15–55)	21 (16–35)	15 (13–50)
Range of the maximum thickness of the compensator (mm)	16.3 (6.6–47.3)	12.2 (5.4–32.5)	23.9 (10.6–55.7)	15.9 (7.5–26.8)	26.8 (7.5–64.9)

with respect to indication (i.e. arteriovenous malformation (AVM), acoustic neuroma (AN), meningioma (MGM), metastasis (METS), and pituitary adenoma (PIT)). The significance of clinical diagnosis in our study is not based on target histology but rather physical properties of the target (i.e. location and size) as well as indication-specific treatment planning strategy (i.e. range and lateral margins). All clinical treatment plans were prepared using CMS XiO (Elekta, Stockholm, Sweden). The analytical dose calculation (ADC) algorithm in XiO is based on a pencil-beam approach (Hong *et al* 1996). It should be noted that treatments for AVMs and METs are planned without (AVM) or with small (METS) lateral margins for the target.

Prescribed doses are delivered using an output factor describing the relation between observed monitor units and delivered dose (Kooy *et al* 2005). This output factor (OF) is determined experimentally or theoretically considering a large open field without a patient specific aperture impinging on a homogeneous water phantom. While this is common practice for large fields, for clinical practice with small fields the prescribed monitor unit (MU) is modified to compensate for the decrease in dose due to loss of lateral equilibrium by an OF correction based on experimentally determined parameters (Daartz *et al* 2009). The OF correction is applied to the MUs to match the delivered dose to the (prescribed) ADC dose. Table 1 summarizes the prescription and the beam information of the patient cohort.

2.2. Monte Carlo dose calculation and data analysis

Monte Carlo dose calculations were performed with the validated MC simulation toolkit TOPAS (Perl *et al* 2012, Testa *et al* 2013), which is based on the MC toolkit Geant4 (Agostinelli *et al* 2003). The OF adjustment was applied to the MC computed dose in the same way as done clinically with the prescribed MU to simulate clinical practice.

A measure of the range differences was determined comparing renormalized dose distributions. Normalized MC doses were derived by matching the MC dose distribution at the center volume of the treatment field to the dose predicted by the ADC. Averaging the dose in a well-defined high-dose (target) region was done to determine the normalization factor (NF). The center of the target volume was defined as half the lateral width of the 90% dose in beams eye view (BEV) and the central part of the SOBP defined by the prescribed range (R) and the modulation width (M) as ($R - 0.75M$, $R - 0.25M$). The ratio of the mean values (ADC divided by MC) defines the normalization factor applied to the MC dose distributions. In the results section we refer to this normalized distribution as NORM. Comparisons between ADC and MC were categorized using dose and range as follows.

2.2.1. Dose. Differences were calculated using dose-volume-histogram (DVH) parameters for target volumes and selected organs at risk (OARs). The DVH parameters were compared by extracting the mean (Mean) and the maximum dose that covers 2%, 50%, 90% and 95% of the target volume (D2, D50, D90 and D95, respectively).

2.2.2. Range. Voxel-by-voxel root-mean-square differences (RMS) were determined for 2D range-profiles (Schuemann *et al* 2014). *R*90, the distal point where the dose falls off to 90% of the prescribed dose. RMS of the range difference is calculated by estimating the root-mean-square of the range difference in each voxel of the 2D range difference map from the beam's eye view of each beam.

Differences in dose and range were analyzed with respect to various parameters as follows:

- Normalization factor: The normalization was applied after adjustment for the OF in order to have the central part of the treatment field agree between MC and ADC ('NORM'; see above).
- Field diameter: The diameter of the treatment field defined as the diameter that would be necessary for a circular target area of the same size as the target in beam's eye view. As most fields in SRS have a nearly circular aperture, this is a reasonable approximation of the treatment field size.
- Range compensator maximum thickness (RCmthick): Due to the small size and regular shape of most treatment volumes, this value is a good approximation for the overall thickness of the range compensator. In some cases, the compensator thickness was increased to either reach a minimum range threshold or achieve a clinically acceptable dose rate.
- Prescribed range (*R*): The prescribed range is the *R*90 that the ADC expects if the treatment fields were to be delivered in a water phantom without beam shaping devices (such as apertures and range compensators).
- Modulation width (*M*): The modulation width of the SOBP (from distal *R*90 to the proximal 98% dose falloff) that the ADC expects if the treatment field were to be delivered in a water phantom without beam shaping devices.
- Heterogeneity index (HI): The HI (Pflugfelder *et al* 2007) has been applied to small fields previously (Bueno *et al* 2013). The HI is a measure of density heterogeneities in the patient geometry. High-density gradients lateral to the beam direction cause high HI values. We follow a definition of the HI similar to the one used in Bueno *et al* (2013). In this study, we used a HI normalized to be between 0 and 1 considering heterogeneities at each step (i.e. per voxel) in beam direction. The HI for one beam is defined as the median of the indices for all pencil kernels *i*:

$$\mathrm{HI} = [\widetilde{\mathrm{HI}}_i(x_i, y_i)]^{1/2}_{i\in A} \tag{1}$$

The $\mathrm{HI}_i(x_i, y_i)$ for a pencil kernel at the location of (x_i, y_i) is defined using the pencil kernel i sampled in the beam effective area A, in which voxels receive a dose higher than 90% of the prescribed dose:

$$\mathrm{HI}_i(x_i, y_i) = \sum_{k=0}^{k=R_i} \mathrm{HI}_{ik}(x_i, y_i) = \sum_{k=0}^{k=R_i} \frac{\sum_{j\in T_i(z_k)} \phi_i(x_j, y_j, z_k)\cdot \left[S_{\mathrm{rel}}(x_j, y_j, z_k) - S_{\mathrm{rel}}(x_i, y_i, z_k)\right]^2}{\sum_{j\in T_i(z_k)} \phi_i(x_j, y_j, z_k)} \tag{2}$$

Here, R_i is the water equivalent range ($R90_i$) beyond the range compensator. The HI_{ik} is computed as the sum of the square differences between the relative stopping power ($S_{\mathrm{rel}}(x_i, y_j, z_k)$)

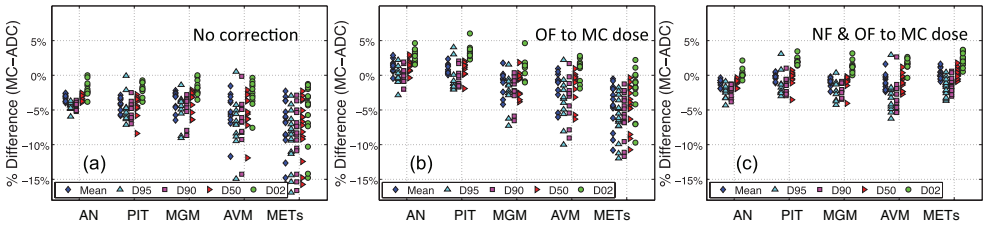

Figure 1. Difference between ADC and MC for target DVH-related parameters (i.e. mean, D95, D90, D50, D02) (a) without corrections, (b) after the OF adjustment to the MC dose and (c) also including the normalization factor (NF) correction to the MC dose.

of the surrounding points on the x-y plane and $S_{\mathrm{rel}}(x_i, y_j, z_k)$ of the central axis of the kernel, weighted by the lateral fluence distribution $\phi_i(x_j, y_j, z_k)$. $T_i(z_k)$ is the set of sampling points in the x-y plane at depth z_k.

3. Results and discussions

3.1. Dose to the target

The ADC can be expected to always overestimate the target dose by 1–2% because of insufficient consideration of the loss of primary and secondary transverse equilibrium in the patient even for large fields (Schuemann *et al* 2015). Similar differences would be expected for small fields if the OF adjustment is sufficient. Figure 1(a) shows the differences (MC-ADC) of DVH-based indices, i.e. the mean dose, D95, D50, and D02, of the target for each treatment group. The DVH indices are nearly all negative, i.e. the delivered (MC) doses are less than predicted by the ADC. The differences in D95 vary by treatment site and are as high as 6% (AN)–18% (METS), reflecting the different planning strategy of no or only small margins for AVM and METS. Figure 1(b) illustrates the effect of applying the small fields OF correction as used in clinical practice. Now the DVH indices generally agree within 5% for the treatment groups of MGM, AN and PIT. However, for AVM and METS, the mean dose loss can still be as high as 11% while the decrease in D95 can be as high as 12%. Applying the treatment field-specific normalization factor (NORM) as determined by MC reduces the overall difference of D95 to less than 6% and the mean dose to less than 3% for all sites, see figure 1(c). This demonstrates that the OF adjustment, a correction for lateral disequilibrium due to small apertures, is not always sufficient to correct for general shortcomings of ADC in a complex patient geometry. The difference in dose is more significant for small fields because of the general lack of equilibrium, i.e. the disequilibrium at the lateral ends of the field can even affect the center of the target volume.

There are several reasons for the general overestimation of dose by the ADC even after the OF adjustment if no normalization is used (see figure 1), in particular for the group of AVM and METS. First, the difference in the plan design as (lower margins) results in more of the low dose region that is affected by the loss of lateral equilibrium being inside the target. Second, some AVM and METS are small and shallow targets, requiring extra range compensator thickness in order to reach the minimum deliverable range and the minimum acceptable dose rate. Figure 2 shows the relation between the field diameter, prescribed range, the thickness of the range compensator and the normalization factor, and the difference of the mean target dose between ADC and MC. ADC increasingly underestimates the loss of target dose

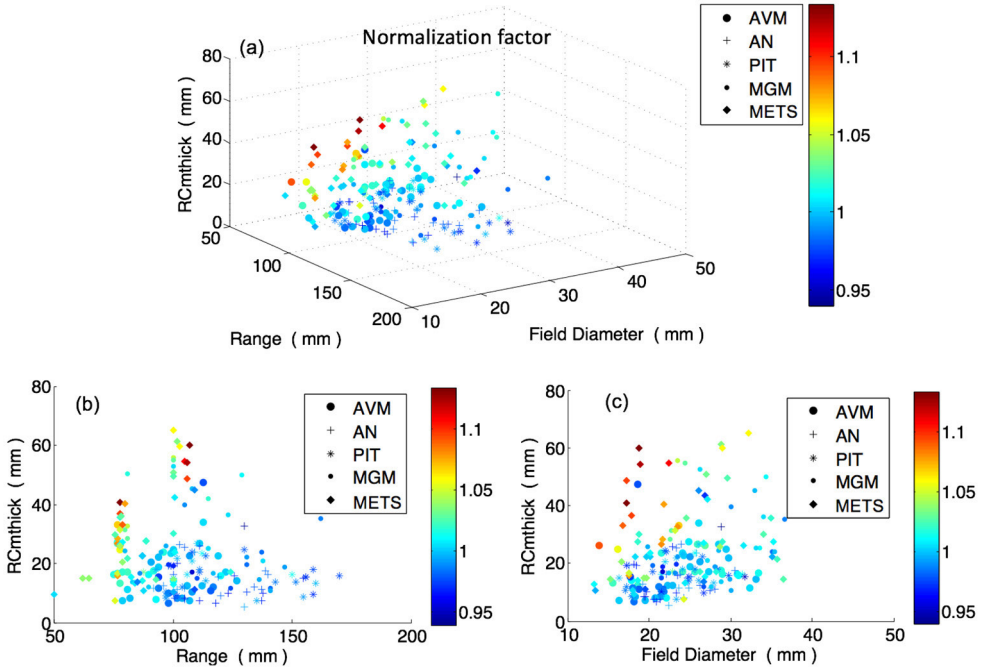

Figure 2. The relation between the normalization factor and three beam parameters. In (a), each data point with a color represents the value of the normalization factor for each beam with respect to range, field diameter and the thickness of range compensator in a 4D manner. (b) and (c) 3D relationship between the normalization factor and the beam parameters.

for shallow targets with thick range compensators and small field sizes. This could be attributed to the insufficient consideration in range compensator scattering in the ADC, especially with smaller field diameter.

Figure 3 shows an example beam of the METS patient with the largest normalization factor, i.e. 1.134, which means the total dose calculated by MC was 13.4% lower than that predicted by ADC for the central voxels even after applying a 4.6% OF adjustment. The ADC generally underestimates scattering effects resulting in the pattern observed in the dose difference plot in figure 3(c) and also figures 4(c) and (f) (the dose distribution for a PIT case). The increased scattering causes more protons to be scattered out from the high dose region to the penumbra. The resulting lateral disequilibrium causes a dose lower than predicted in the central region, surrounded by a ring of higher than predicted dose. The high-dose ring is the sum of an MCS dose ring and the secondary dose halo. The size of this effect depends on the field size and beam parameters.

Figures 3(d) and (e) illustrate differences in MC and ADC dose profiles as well as the two normalization regiments used in this analysis. First we apply the clinically used OF (dotted line) to the MC dose to describe the clinically delivered dose. In the second step, we normalize the OF-corrected dose distribution in the central region (indicated by vertical lines) to the dose distribution expected by ADC. This latter factor is the field specific normalization factor (NORM) and results in the average dose in the central target region to be the same for MC and ADC. The normalized distribution represents the best dose that can be achieved with a simple scaling factor. Since this does not address the differences in the shape of the dose distributions

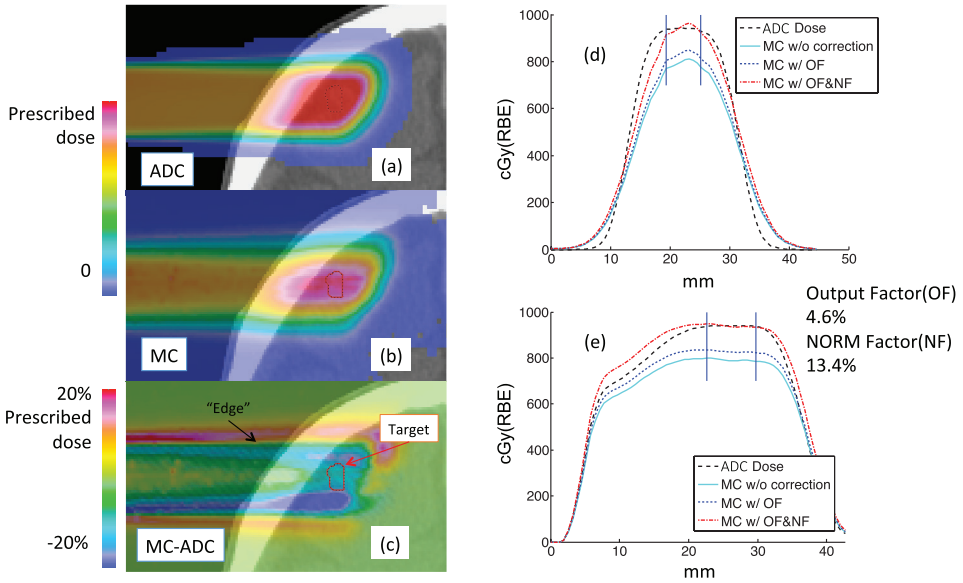

Figure 3. Example of the dose distribution calculated by ADC (a), MC with only the OF adjustment factor applied (b), and the difference (MC-ADC) (c). The illustrated treatment field enters from the left side after passing through aperture and compensator. The lateral (d) and depth (e) dose profiles show the two normalization steps. The vertical blue lines in d and e represent the area being used for normalization.

using such an MC derived scaling factor can result in a slight hotspot at the field center and does not fully resolve the cold regions in the lateral penumbra.

3.2. Dose to normal tissues

Organs at risk often see increased dose levels compared to the ADC predicted doses due to scattering effects. Figure 4 shows the dose distribution and the DVH of a typical PIT case. The target doses as predicted by ADC and MC for this case agree well but the doses to the organs at risk show an underestimation by ADC for the chiasm, optic nerves and the brainstem. It is important that these uncertainties be factored into the treatment planning approach, for instance by reducing planned doses to OARs. We did not find any case where the OAR dose exceeded the clinical constraints due to our standard precautions given the potential dosimetric uncertainties. Our clinical practice is to use conservative dose constraints, including dose limits of optic pathway structures $\leqslant$8 Gy(RBE) and brainstem surface $\leqslant$12 Gy(RBE).

3.3. Range

Figure 5 shows box plots of the RMS of the *R*90 range difference (*R*90 RMS) for each site. The analyses for negative and positive values focus on shorter (underestimated) and longer (overestimated) ranges predicted by the ADC. For the box plots, the center mark is the median value of the *R*90 RMS, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points not considering outliers. Outliers are defined as those points that are greater than 75th percentiles + 1.5 $\times$ (75th percentiles $-$ 25th percentiles) or less than 25th percentiles $-$ 1.5 $\times$ (75th percentiles $-$ 25th percentiles), and are shown as red

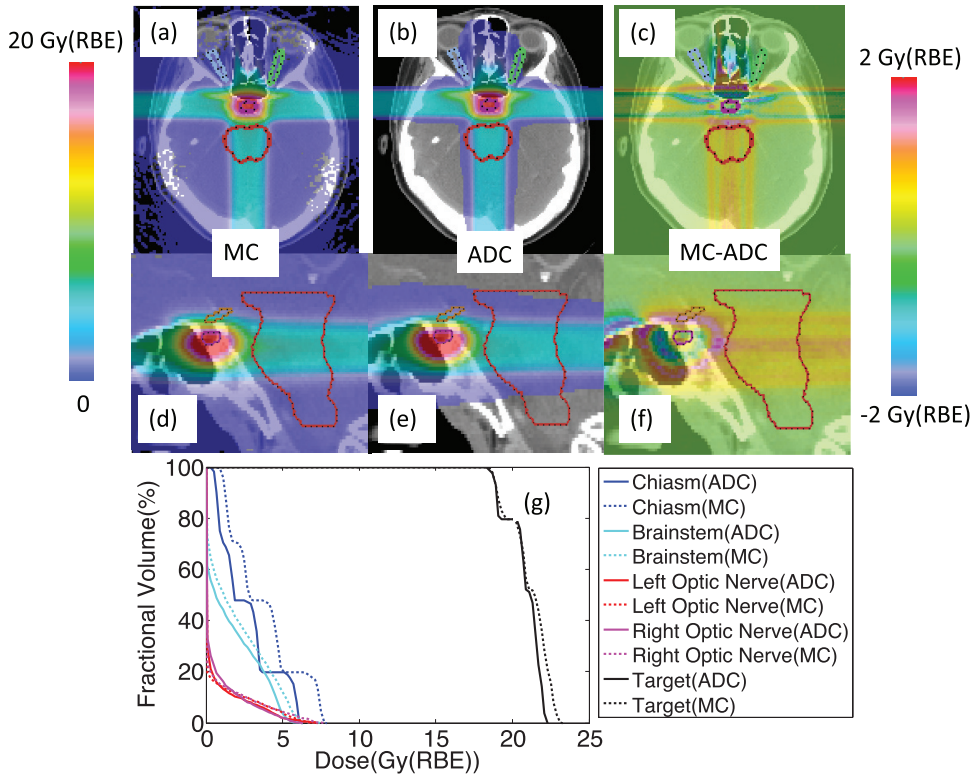

Figure 4. Example of the underestimated dose to the organ at risk for a PIT patient. (a)–(c) Transverse and (d)–(f) sagittal plane of the dose distribution and the difference calculated by ADC and MC. (g) Differences between the two methods in the DVH after OF correction.

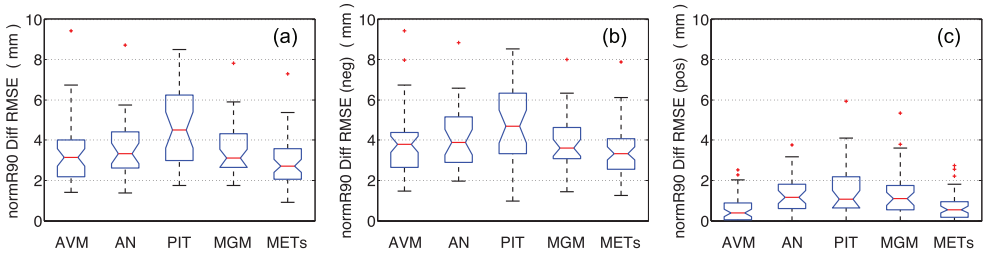

Figure 5. Box plot of the *R*90 (MC-ADC) RMS for each site, (a) for *R*90 RMS, (b) for *R*90 RMS considering only positive values (range overestimated by ADC), and (c) for negative values (range underestimated by ADC).

dots. The notch represents the 95% confidence interval of the median value. The RMS of positive *R*90 is generally smaller than the negative RMS, which indicates that ADC mostly overestimated the ranges of the beams, i.e. the actual delivered dose falls off earlier than predicted. The RMS *R*90 for PIT has the largest median. This is due to the target being located close to the nasal cavity resulting in the beam stopping in a highly heterogeneous patient geometry. The other four groups have similar behavior in range differences. The median value of the

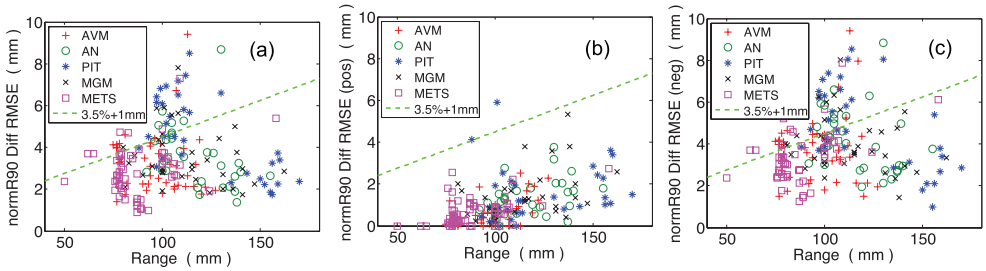

Figure 6. Relationship between the *R*90 RMS and the prescribed range, (a) for *R*90 RMS, (b) for *R*90 RMS considering only positive values (range overestimated by ADC), and (c) for negative values (range underestimated by ADC). The dashed green line indicates the reference line of the clinically used range margin (3.5% + 1 mm).

*R*90 RMS is 4.1 mm, and the maximum RMS is in this cohort of 50 patients was found in the AVM group with 9.4 mm.

Figure 6 shows the relationship between the prescribed range and the *R*90 RMS. Each point represents the *R*90 RMS for a single treatment field. The dashed green line indicates the reference line of the clinically used range margin (3.5% + 1 mm). This margin definition assumes a linear correlation of *R*90 RMS with the prescribed range. However, we found no significant correlation for small fields. *R*90 RMS seems to be correlated stronger to other parameters of the treatment plan. For instance, for lesions in the PIT class, treatment plans are typically designed with three fields, i.e. two lateral fields and a posterior–superior field. Although the lateral beams have a smaller prescribed range, they can encounter higher density heterogeneities along the interface of the bone and the nasal cavity, e.g. see figure 4(a). Also for the posterior–superior field, smaller modulation widths were observed, which contribute to a smaller *R*90 RMS; the correlation of modulation width with *R*90 RMS will be further discussed in the following section. For METS, the spread in *R*90 RMS for the same range is much larger than the correlation with range. For small fields, patient geometry plays a bigger role in range uncertainties than the overall prescribed range, which is used to determine range uncertainty margins.

Figure 7 shows the correlation between the HI and *R*90 RMS relative to the prescribed range. Table 2 summarizes the Spearman's correlation coefficient and *p*-value of the coefficient for each beam parameter correlated with *R*90 RMS (relative to the prescribed range). The positive and negative values of the correlation coefficient in the table represent that *R*90 RMS increased/decreased with the beam parameter. Significant correlation (Spearman's p-value < 0.05) was found between HI and R90 RMS (relative to the prescribed range) for all sites. While a correlation of *R*90 RMS and HI can be seen independent of the treatment group (see figure 7 and table 2), the spread in the data points suggests that other factors have an impact on the accuracy of range prediction as well, e.g. field diameter, range compensator and modulation. Smaller field diameters tend to cause larger *R*90 RMS because the loss of lateral equilibrium is enhanced at the field edges. Smaller fields have a larger surface (edge) to volume ratio (thus the root mean square is larger).

The modulation width (*M*) shows significant correlation to *R*90 RMS for all the sites. This can be attributed to two reasons. First, due to the increasing loss of dose in the center of the treatment field for very small field sizes, the SOBP is becoming tilted. When the treatment field was normalized to the center of the ADC SOBP, a larger modulation results in the tilt of the SOBP to have a larger effect on the *R*90. Second, as pointed out in previous studies (Titt *et al* 2008, Daartz *et al* 2009), for small field dosimetry the measurable scattered dose from

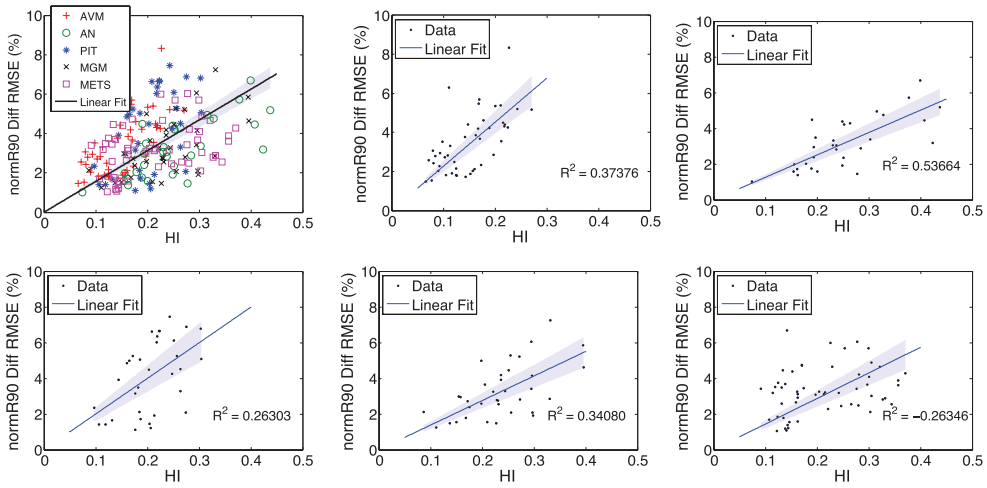

Figure 7. Correlation between the *R*90 RMS and the heterogeneity index (a) for all sites, (b) for AVM, (c) for AN, (d) for PIT, (e) for MGM, and (f) for METS, respectively. The RMSE is shown in % of the prescribed range.

Table 2. The Spearman's correlation coefficient and *p*-value of the coefficient between beam parameters with the *R*90 RMS (relative to the prescribed range). Positive/negative values of the correlation coefficient mean that the *R*90 RMS is increased/decreased with the beam parameter.

Diagnostic indications	AVM	AN	PIT	MGM	METS
Heterogeneity index	**+ 0.630 (0.000)**	**+ 0.700 (0.000)**	**+ 0.531 (0.002)**	**+ 0.542 (0.002)**	**+ 0.292 (0.030)**
Field diameter	−0.177 (0.243)	−0.273 (0.137)	**−0.562 (0.001)**	−0.036 (0.839)	−0.217 (0.112)
Maximum thickness of the compensator	**+ 0.448 (0.002)**	**+ 0.397 (0.027)**	+0.339 (0.053)	**+ 0.380 (0.024)**	−0.153 (0.264)
Prescribed modulation width	**+ 0.402 (0.006)**	**+ 0.588 (0.001)**	**+ 0.441 (0.010)**	+0.282 (0.101)	**+ 0.372 (0.005)**

the beam specific shaping device, such as the collimator and the range compensator, is not negligible. With larger modulation, an increased number of scattered protons will contribute dose to the point for normalization, which could affect the range difference as a consequence.

4. Conclusion

Dose calculation for small fields in proton therapy continues to provide challenges. Effects such as aperture scatter, multiple-Coulomb scattering and secondary dose halo have a more drastic impact on dose distribution characteristics the smaller the cross section of the beam. Inaccurate modeling of scatter and primary fluence loss in the range compensator poses an additional problem. For small fields, there is not one dominating effect—for most accurate results, all must be considered. Capturing all contributing physical effects is challenging for

analytical dose engines. Therefore, small field dosimetry lends itself to investigation using Monte Carlo simulations.

In this study, we examined the accuracy of an analytical dose calculation algorithm (ADC) with respect to predicting dose and range for treatment planning of proton stereotactic radiosurgery (PSRS). Our results show that ADC dose calculations with (without) output factor corrections provide an estimate of D95 for PIT, MGM, and AN within 6% (9%) and, due to smaller planning margins, to within 11% (18%) for AVM and METS for treatments using small field diameter and thick range compensators. By applying a simple scaling factor obtained from Monte Carlo simulations the inaccuracies can be reduced to be within 6% for the D95 of the target for all sites.

It should be noted, that different institutes use different dose calculation algorithms, some performing better for small fields than others. Accordingly, output factor correction methods vary, and sometimes are deemed not necessary. Caution is still indicated—our results show that conclusions drawn from single point measurements in water, as they are typically performed to assess small-field computation accuracy, are oftentimes insufficient to characterize target coverage. In-patient heterogeneity and dose-loss at the field edges significantly affect the dose distribution, and may be missed by single-point measurements in water.

The root mean square of the *R*90 difference, the difference in the position of distal falloff to 90% of the prescribed dose, is affected by several factors. A correlation with the prescribed range was not observed as the range differences were dominated by the patient heterogeneity, the modulation width and the field diameter. Consequently, the generally applied margin to cover uncertainties in range (3.5% of the prescribed range + 1 mm) is not sufficient to ensure full target coverage, most notably at the field edges. To guarantee the target coverage, margins should be adjusted accounting for the specifics of each treatment field delivery and the patient heterogeneity along the beam path. It might be beneficial to assign a range margin for PSRS patients not based on range, but instead based on the field diameter, heterogeneity and modulation width. Avoiding the use of thick range compensators, when possible, would further reduce the uncertainties. The discrepancies observed for single fields are also often mitigated by the utilization of multi-field treatment.

Note that the results in this study are generally applicable in all scatter systems considering the common scattering pattern inside the patient and in the beam shaping devices. However, the small size of the targets and delivery fields increase the relative size of the observed effects.

Acknowledgment

This work was funded by the National Cancer Institute under U19 CA-21239.

References

Agostinelli S *et al* 2003 Geant4—a simulation toolkit *Nucl. Instrum. Methods Phys. Res.* A **506** 250–303

Bednarz B, Daartz J and Paganetti H 2010 Dosimetric accuracy of planning and delivering small proton therapy fields *Phys. Med. Biol.* **55** 7425–38

Bueno M, Paganetti H, Duch M A and Schuemann J 2013 An algorithm to assess the need for clinical Monte Carlo dose calculation for small proton therapy fields based on quantification of tissue heterogeneity *Med. Phys.* **40** 081704

Daartz J, Engelsman M, Paganetti H and Bussière M R 2009 Field size dependence of the output factor in passively scattered proton therapy: influence of range, modulation, air gap, and machine settings *Med. Phys.* **36** 3205–10

Grassberger C, Daartz J, Dowdell S, Ruggieri T, Sharp G and Paganetti H 2014 Quantification of proton dose calculation accuracy in the lung *Int. J. Radiat. Oncol. Biol. Phys.* **89** 424–30

Guan F *et al* 2015 Spatial mapping of the biologic effectiveness of scanned particle beams: towards biologically optimized particle therapy *Sci. Rep.* **5** 9850

Hong L, Goitein M, Bucciolini M, Comiskey R, Gottschalk B, Rosenthal S, Serago C and Urie M 1996 A pencil beam algorithm for proton dose calculations *Phys. Med. Biol.* **41** 1305–30

Kooy H M, Rosenthal S J, Engelsman M, Mazal A, Slopsema R L, Paganetti H and Flanz J B 2005 The prediction of output factors for spread-out proton Bragg peak fields in clinical practice *Phys. Med. Biol.* **50** 5847–56

Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Phys. Med. Biol.* **57** R99–117

Paganetti H 2014 Monte Carlo simulations will change the way we treat patients with proton beams today *BJR* **87** 20140293

Perl J, Shin J, Schümann J, Faddegon B and Paganetti H 2012 TOPAS: An innovative proton Monte Carlo platform for research and clinical applications *Med. Phys.* **39** 6818–21

Pflugfelder D, Wilkens J J, Szymanowski H and Oelfke U 2007 Quantifying lateral tissue heterogeneities in hadron therapy *Med. Phys.* **34** 1506–13

Schuemann J, Dowdell S, Grassberger C, Min C H and Paganetti H 2014 Site-specific range uncertainties caused by dose calculation algorithms for proton therapy *Phys. Med. Biol.* **59** 4007–31

Schuemann J, Giantsoudi D, Grassberger C, Moteabbed M, Min C H and Paganetti H 2015 Assessing the clinical impact of approximations in analytical dose calculations for proton therapy *Int. J. Radiat. Oncol. Biol. Phys.* **92** 1157–64

Testa M, Schümann J, Lu H M, Shin J, Faddegon B, Perl J and Paganetti H 2013 Experimental validation of the TOPAS Monte Carlo system for passive scattering proton therapy *Med. Phys.* **40** 121719

Titt U, Zheng Y, Vassiliev O N and Newhauser W D 2008 Monte Carlo investigation of collimator scatter of proton-therapy beams produced using the passive scattering method *Phys. Med. Biol.* **53** 487–504