HIROSAKI UNIVERSITY

Radiation Environment and Medicine 2023 Vol.12, No.1 53–64

REM
Radiat Environ Med

Regular Article

Comparison of the Dose Calculation Accuracy between the Commercial Pencil Beam Algorithm and Various Statistical Uncertainties with Monte Carlo Algorithm in New Proton Pencil Beam Scanning System

Yuki Tominaga[1, 2], Masataka Oita[3]*, Takashi Akagi[4], Junya Miyata[2, 5], Shuichi Harada[4], Tetsunori Matsuda[6] and Masahiro Kuroda[7]

[1]*Department of Radiotherapy, Medical Co. Hakuhokai, Osaka Proton Therapy Clinic, 27-9 Kasugadenaka, Konohana-ku, Osaka, Osaka 550-0022, Japan*
[2]*Graduate School of Interdisciplinary Science and Engineering in Health Systems, Okayama University, 5-1 Shikata-cho, 2-chome, Kita-ku, Okayama, Okayama 700-8558, Japan*
[3]*Faculty of Interdisciplinary Science and Engineering in Health Systems, Okayama University, 3-1 Tsushimanaka, 3-chome, Kita-ku, Okayama, Okayama 700-8530, Japan*
[4]*Hyogo Ion Beam Medical Support, 2-1, 1-chome, Kohto, Shingu, Tatsuno, Hyogo 679-5165, Japan*
[5]*Department of Radiological Technology, Kurashiki Central Hospital, 1-1-1, Miwa, Kurashiki, Okayama 710-8602, Japan*
[6]*Department of Radiological Technology, Tsuyama Chuo Hospital, 1756, Kawasaki, Tsuyama, Okayama 708-0841, Japan*
[7]*Graduate School of Health Sciences, Okayama University, 5-1 Shikata-cho, 2-chome, Kita-ku, Okayama, Okayama 700-8558, Japan*

Received 7 June 2022; revised 29 October 2022; accepted 7 November 2022

We validated the calculation accuracies and times between the pencil beam algorithm (PBA) and the Monte Carlo (MC) algorithm in the new proton pencil beam scanning system. Thirty-three (homogeneous phantom) and three (heterogeneous phantom) uniform-dose plans were verified for cubic targets. These plans were calculated using the PBA and five MC statistical uncertainties of 0.3%, 0.5%, 1.0%, 1.5%, and 2.0%. We also evaluated the required dose calculation times per beam with the statistical analysis. Then, eight clinically realistic beams were validated as the end-to-end test. The dose differences of 0.3% and 0.5% uncertainty MC plans were satisfied our tolerance ($< \pm 3.0\%$). All five uncertainty MC plans improved at least the average/minimum gamma score from 93.8%/33.5% to 98.5%/88.9% for homogeneous (2%/2 mm) and 96.7%/83.9% to 98.8%/94.1% for heterogeneous (3%/3 mm) than PBA plans. However, the average calculation times of 0.3% and 0.5% uncertainties in MC plans were 15.5 and 5.8 times longer than in PBA plans ($p < 0.001$). The end-to-end tests satisfied the acceptable with a statistical uncertainty of MC below 0.5%. Although the dose calculation times would be significantly increased, we suggest that the statistical uncertainty of MC below 0.5% is appropriate for clinical use with PBS plans.

Key words: proton beam therapy, pencil beam scanning, commissioning, Monte Carlo, pencil beam algorithm

1. Introduction

Pencil beam scanning (PBS) proton therapy has been developed as one of the most advanced proton therapies and has been widely introduced globally[1)]. The PBS system can deliver at several individual spots

*Masataka Oita: Faculty of Interdisciplinary Science and Engineering in Health Systems, Okayama University, 5-1 Shikata-cho, 2-chome, Kita-ku, Okayama, Okayama 700-8558, Japan
E-mail: oita-m@cc.okayama-u.ac.jp
https://doi.org/10.51083/radiatenvironmed.12.1_53

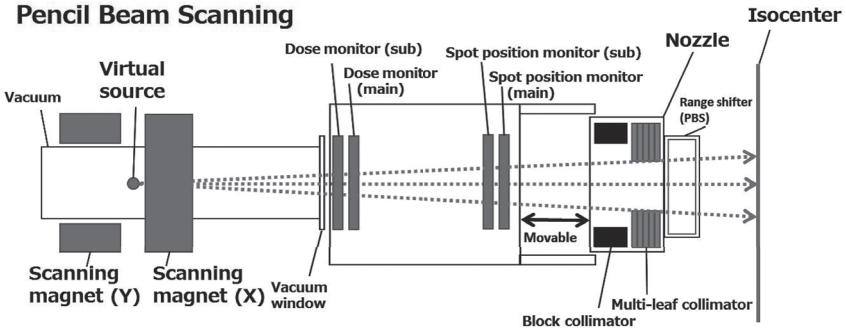

Fig. 1. Schematic of our pencil beam scanning treatment nozzle.

by continuously varying energies and magnetically irradiating spots with unique intensities and positions in lateral planes[2]. This new system also can provide better dose conformity than conventional passive scattering system with compensators, modulation devices, or collimators[3, 4].

Currently, there are two main PBS dose calculation algorithms in the clinical treatment planning system (TPS), namely the pencil beam algorithm (PBA) and the Monte Carlo (MC) algorithm[5, 6]. The PBA is based on an analytical algorithm and can calculate the dose distribution effectively and quickly[7]. Several studies have suggested that the dose agreement between measurements and the PBA calculations deteriorates for plans including a range shifter (RS)[8, 9]. These discrepancies are due to the lack of proper modeling of the secondary protons, which may be induced by the beam passing through the air between the RS and the patient surface. Several reports have also concluded that the dose accuracy of the PBA deteriorates in heterogeneous regions such as skull base, lungs, and implant sites[10-12]. Meanwhile, the MC dose calculation requires the tracking of individual particle trajectories and random sampling of the interaction cross-section data. The statistical nature of the MC dose scoring results in variations in the output dose distribution based on the number of particle histories in the simulations[13]. Although the MC algorithm can calculate doses in heterogeneous regions with higher accuracy, it requires a longer calculation time than that of the PBA[14]. The MC algorithm requires the statistical uncertainty to be set, which is one of the parameters involved in calculating the transport of particles stochastically. Varying the statistical uncertainty causes differences in dose distribution and calculation time[15]. The PBA and MC used in this study were calculated by the RayStation TPS (RaySearch, Stockholm, Sweden), which can calculate dose distribution using both the PBA (version 6.2) and the MC (version 7.0) algorithms and details of the PBS beam calculation model can be found in the reference manual[16].

Our new PBS proton therapy system, MELTHEA V (Hitachi, Ltd., Tokyo, Japan), was installed at Tsuyama Chuo Hospital in January 2016. Although various PBS machine vendors, such as IBA, Varian, Sumitomo, and Mevion, no report is observed on the validation of Hitachi's PBS system installed in our hospital[17-20]. The current proton therapy system which the Mitsubishi Electric Corporation originally developed, was used in this study. And the only research on Monte Carlo calculations for Mitsubishi proton systems was reported by Yamashita *et al.* using conventional passively scattered proton beams[21]. Further, many studies have used a statistical uncertainty of 0.5% for MC plans in the TPS, but the reference manual and other reports do not state which statistical uncertainty is preferable[16, 22, 23].

In this study, we introduced our new proton therapy system firstly. Then, we compared beam commissioning results for PBS using the calculations from the commercial PBA and MC algorithms under various statistical uncertainties for homogeneous and heterogeneous phantoms. To enable the clinical application of the MC algorithm, we evaluated the appropriate statistical uncertainty in terms of dose calculation accuracy and dose calculation time.

2. Materials and methods

2.1. Proton beam delivery system

Our PBS machine comprises a synchrotron and gantry room equipped with a dedicated nozzle. The PBS irradiation method is a raster scanning technique, similar to step-and-shoot spot scanning; however, the beam in PBS does not turn off when moving between spots[24, 25]. The PBS system has 92 energies that span from 70.7 to 235.0 MeV with the beam ranges of 40.0 to 340.0 mm in water. To irradiate depths of less than 40 mm, a pre-absorber called the offset RS may be inserted downstream of the nozzle (Fig. 1). These RSs are prepared with a range of 60 to 66 mm water equivalent thicknesses (WET). Six additional RSs with a 1 to 6 mm WET are also available to adjust the high-energy beam range.

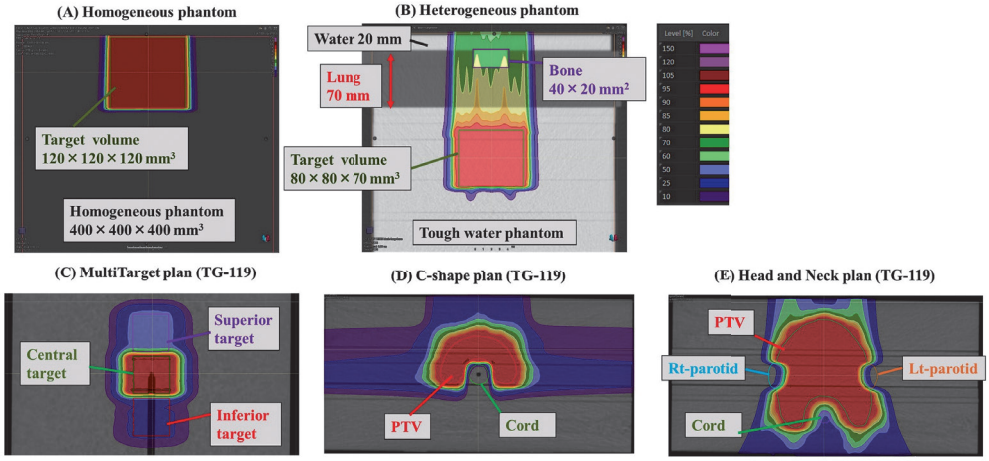

Fig. 2. Screenshots of representative verification plans for (A) homogeneous, (B) heterogeneous, (C) MultiTarget, (D) C-shape, and (E) Head and Neck phantoms.

The snout of the nozzle can move in the beam direction to get as close as possible to the patient's surface. This nozzle includes a multi-leaf collimator (MLC) to create a collimated PBS field area.

2.2. Monte Carlo dose calculations in the TPS

This section describes the MC dose engine for proton PBS in the TPS. The MC calculation needs to model the physical processes involved. The MC code transports primary protons and secondary ions such as protons, deuterons, and alphas. Primary and secondary protons use a class II transport method. The heavier secondaries are transported only considering energy loss in a continuous slowing down approximation (CSDA), while nuclear absorption, multiple coulomb scattering (MCS), and energy straggling are thus neglected for the secondary deuterons and alphas. The modeling code consists of four items: stopping power, energy loss straggling, MCS, and nuclear interactions.

The stopping power model is used based on the well-established Bethe-Bloch formula as follows:

$$S = \frac{1}{u}\cdot\frac{4\pi e^4}{m_e c^2}\sum_i w_i\cdot\frac{Z_i}{A_i}\cdot\frac{\rho}{\beta^2}\left[\ln\frac{2m_e c^2}{I}-\ln\left(\frac{1}{\beta^2}-1\right)-\beta^2\right], \quad (1)$$

$$\beta^2 = 1-\left(1+\frac{E}{m_p c^2}\right)^{-2}. \quad (2)$$

where, E, Z_i, w_i, and A_i are the kinetic energy of the protons, the elemental composition in atomic number, weight, and mass, respectively. The ρ, I, u, c, m_e, m_p, and e are the mass density, mean ionization energy of the medium, anatomic mass unit, speed of light, electron mass, proton mass, and elementary charge, respectively.

The energy loss straggling is calculated by the classic Bohr expression as follows:

$$\sigma_B^2 = 4\pi e^4 n_e\frac{1-\beta^2/2}{1-\beta^2}\Delta_z, \quad (3)$$

$$n_e = \frac{\rho}{u}\sum_i w_i\frac{Z_i}{A_i}. \quad (4)$$

This expression considers a Gaussian energy loss distribution after traversing a slab thickness (Δz). The standard deviation of the straggling is given by Bohr using Eq. (3) and (4). The n_e in Eq. (3) and Eq. (4) indicate the target total electron density.

The modeling of MCS is based on the Goudsmit-Sanderson method[26, 27]. This method includes the angular deflection from an angle probability distribution, Rutherford cross-section, and sampled random hinge step for electron/positron transport.

The MC dose engine considers three nuclear interaction processes: the non-elastic reactions (secondary protons, deuterons, and alpha), elastic proton of proton scattering, and nucleus scattering. The TPS uses a data library with pre-stored tables of the quantities needed to consider non-elastic reactions provided in the ICRU 63 report[28]. The process of the elastic proton (proton scattering) is modeled as a nuclear force (catastrophic event). The elastic proton (nucleus scattering) is modeled as a non-catastrophic event and is embedded in the calculation of MCS. The other details of MC codes are described in the reference manual[16].

2.3. Verification plans

Thirty-three uniform dose plans were created by the TPS for simple cubic targets with various field sizes, beam ranges (target depths), number of energy layers, and

spread-out Bragg peak (SOBP) widths in a homogeneous water phantom without the MLC[29]. The 20 plans did not use any RS (RS-), and the remaining 13 plans inserted RSs (RS+) with a 60 mm WET. In the 13 RS+ plans, the air gaps were set to 100 mm, except for the beam where the nozzle collided with the water surface. For all plans, the field sizes were varied from 30 to 152 mm^2. The SOBP widths were determined to range from 30 to 300 mm. Figure 2A shows a screenshot of the representative dose distribution of cubic plan in the water phantom. The homogeneous plan with the RS was optimized for a target with a field size of 120×120 mm^2 and a SOBP width of 120 mm. Then, to validate heterogeneous regions, three plans were created for a CT with solid water, bone, and lung phantom (Tough Water, Tough Bone, and Tough Lung, Kyoto Kagaku, Kyoto, Japan) as shown in Fig. 2B. There is a 20 mm thick water area from the surface, and a bone phantom (40 mm width and 20 mm thick) is contained in a 70 mm thick lung phantom. The plan was optimized for a target with a field size of 80 × 80 mm^2 and a SOBP width of 70 mm. Then, sufficient thickness solid water phantoms are placed distal to the lung phantom.

In PBS, pencil beams are placed at regular intervals to create a uniform dose distribution. The spot spacing depends on the pencil beam size (spot sigma) of each energy (layer), and it varies with the beam characteristics of the treatment machine[30]. Typically, the default spot sigma is set as 1.0. However, the value can be changed with the complexity of the plan. Therefore, in both our homogeneous and heterogeneous plans, the spot spacing was determined arbitrarily within the range of 0.5–1.0 sigma. The prescribed doses ranged from 1.0 to 2.0 Gy (relative biological effectiveness [RBE]) with 50% of each target volume. The dose calculation grid was 2.0 mm for all plans. All beams consisted of a 0° gantry angle, and the isocenter was set at the center of each SOBP. All plans were optimized using the PBA (version 6.2), and the final doses were calculated using PBA and MC (version 7.0). The reason for optimizing with PBA for all plans is that only PBA could be calculated in the TPS (version 6.2) when the beam verification was performed. As the statistical uncertainties of the MC plans were used 0.5%, 1.0%, and 2.0% according to several papers, we determined the uncertainties in the range of 0.5% to 2.0% at 0.5% intervals, and added 0.3% uncertainty as a set value lower than 0.5%[31, 32]. Thus, the five statistical uncertainties consist of 0.3%, 0.5%, 1.0%, 1.5%, and 2.0%. In this report, these uncertainties are denoted as MC_0.3%, MC_0.5%, MC_1.0%, MC_1.5%, and MC_2.0%, respectively. Then, the required final dose calculation times of 33 cubic plans (from completion of spot optimization to end of dose calculation) were measured using TPS scripting for the six algorithms. The TPS computer used was an Intel® Xeon® with an E5-2667 v3 3.20 GHz CPU.

2.4. Clinically realistic plan verifications

To validate clinically realistic PBS beams, we created three plans in the MultiTarget, C-shape, and Head and Neck phantoms of the American Association of Physics in Medicine (AAPM) Task Group Report (TG-119)[33]. These plans consist of two, three, and three beams for the MultiTarget, C-shape, and Head and Neck phantoms, respectively. All plans were created by the multi-field optimization (MFO)[34]. The MFO is used to create clinically advanced PBS plans, which optimizes the spots on all beams simultaneously, producing a very conformal dose distribution in the sum of all MFO beams. Thus, individual MFO beams are non-conformal, resulting in steep dose distributions. The TPS optimization automatically selected the RSs. The prescribed doses were 50 GyRBE in 25 fractions for all plans. The MultiTarget and Head and Neck plans could satisfy all dose constraints of the targets and OARs in Table II and Table IV in Ref. 28, respectively. The C-shape plan was satisfied with the hard constraints (10% volume to receive less than 10 GyRBE) of the core OAR and 95% of PTV to receive 50 GyRBE. The spot spacing (1.0 sigma), optimization algorithm (PBA), final dose calculation (PBA, MC_0.3%, MC_0.5%, MC_1.0%, MC_1.5%, MC_2.0%), and calculation grid (2.0 mm) were the same as for the cubic plans. For the validation of the plans in water, the eight beams of TG-119 plans were copied to the same homogeneous phantom that was used as the cubic plans. The isocenter was determined to be the center of the SOBP for each beam. Figure 2C–E shows the screenshots in the TPS for the three TG-119 plans. Table 1 also shows the beam characteristics of verification plans for the 33 homogeneous, three heterogeneous, and eight TG-119 plans.

2.5. Beam measurements

Verification measurements for homogeneous and TG-119 plans were performed with the depth doses, lateral profiles, and absolute doses at the isocenter. We measured only lateral profiles for the heterogeneous plans. The depth doses were measured using a 3-dimensional (3D) water phantom (MP3-M, PTW, Freiburg, Germany), six ionization chambers (Type 31015, PinPoint, PTW, Freiburg, Germany) with a dedicated attachment, and a multi-channel electrometer (MULTIDOS, PTW, Freiburg, Germany). The calibration factor of six PinPoint chambers was obtained in the 141.2 MeV mono-energetic square fields of 102 mm^2 with a spot spacing of 3.0 mm and at a depth of 25 mm. The measured doses and planned doses were analyzed by the gamma index method[35] (3%/3 mm and a dose threshold of 10% of the maximum planned dose) with an in-house python-based software.

To measure 2-dimensional (2D) dose distributions

Table 1. Field parameters of verification uniform dose plans for 33 homogeneous, three heterogeneous, and eight TG-119 plans.

Phantom type	Number of plans/ Beam name	Field size (mm^2)	Energies (MeV)	SOBP width (mm)	Isocenter (mm)	RS (mm)
Homogeneous	1	152	70.7-235.0	300	187	0.0
	2	150	70.7-235.0	150	112,263	0.0
	3	100	70.7-235.0	100	88,190,289	0.0
	5	50	70.7-235.0	50	65, 124, 184, 246, 309	0.0
	9	30	70.7-235.0	30	52, 85, 117, 149,183, 217, 251, 284, 321	0.0
	1	152	89.8-222.1	250	124	60.0
	2	120	89.8-222.1	120	57, 187	60.0
	3	80	89.8-222.1	80	38, 121, 206	60.0
	7	30	89.8-222.1	30	15, 48, 85, 119, 155, 194, 231	60.0
Heterogeneous	1	80	99.0-157.6	70	155	0.0
	1	80	147.0-194.1	70	230	0.0
	1	80	141.2-174.9	40	200	0.0
TG-119	MultiTarget 1	Planned value	104.9-143.1	Planned value	105	5.0
	MultiTarget 2		100.4-139.3		105	0.0
	C-shape 1		72.1-111.5		50	4.0
	C-shape 2		128.7-170.4		120	5.0
	C-shape 3		128.7-170.4		120	4.0
	Head and Neck I		103.4-166.1		75	60.0
	Head and Neck 2		111.5-182.0		75	62.0
	Head and Neck 3		113.2-184.4		75	64.0

conveniently with absolute doses, the lateral dose profiles were measured using a 2D ion chamber detector array (OCTAVIUS 729 XDR, PTW, Freiburg, Germany) and the Tough Water, Tough Bone, and Tough Lung phantoms. We measured four lateral planes per beam at the plateau (25 or 50 mm for the homogeneous plans, 100 mm for the heterogeneous plans, and 11 or 25 mm for the TG-119 plans), proximal, middle, and distal planes. The total measurements consisted of 128 (80 and 48 planes for the RS- and RS+ plans, excluding plateau planes with the four RS+ plans with no plateau regions),12, and 32 planes for the homogeneous, heterogeneous, and TG-119 plans, respectively. The measured lateral profiles were also analyzed using the gamma index at 2%/2 mm or 3%/3 mm and a dose threshold of 10% of the maximum dose. The gamma pass rate (gamma score) tolerance of lateral profiles was determined based on the AAPM Task Group 185[36]. The tolerances of more than 98% at 2%/2 mm for the homogeneous plans and 95% at 3%/3 mm for the heterogeneous and TG-119 plans were allowed for the lateral profile for each measurement plane.

The absolute doses at the isocenter were measured using the MP3-M and an ionization chamber Type 31013 (Semiflex, PTW, Freiburg, Germany). A factor of 1.1, considering RBE, scaled up the values. The absolute dose differences (DDs) were calculated between the measured and calculated doses using Eq. (5). The D_{meas} and D_{calc} indicate the measured and calculated doses, respectively, and a positive value for DD indicates that the measured dose is higher than the calculated dose. Based on the report of beam commissioning with the MC algorithm for different PBS systems with the same TPS, a DD of less than ± 3% was allowed for the absolute DDs for six algorithm plans[37].

$$\sigma = 100 \times \frac{(D_{meas} - D_{calc})}{D_{calc}} (\%). \quad (5)$$

2.6. Statistical analysis

We analyzed the Steel method to perform the nonparametric multiple comparisons of five MC plans against the PBA plans such as the gamma scores of the lateral profiles with the RS- and RS+ plans, absolute DDs, and calculation times. The statistical analyses were performed by the EZR (Easy R) version 1.54 (Jichi Medical University, Saitama, Japan). A p-value of less than 0.05 was considered as statistically significant.

3. Results

Figure 3 shows the representative depth dose results for the PBA and MC algorithms under five statistical uncertainties. The plan was passing through the RS when all the energy is irradiated, result in causing a 60 mm range shift. This plan includes a beam range of 277 mm, SOBP width of 250 mm, and field size of 152×152 mm^2. This plan was the all SOBP regions from the surface to the distal fall-off. Each title represents the name of the selected dose calculation algorithm. The calculated depth dose of the RS+ plan using PBA was overestimated at less than 50 mm depth regions, particularly at shallower depths. Table 2 shows the summary of gamma scores for

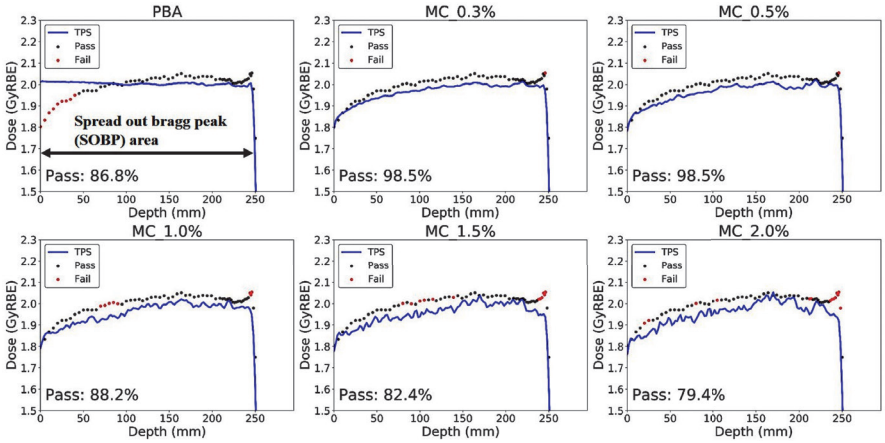

Fig. 3. Representative depth dose comparisons between the measured (scatters) and calculated (plots) doses for six dose calculation algorithms. The bottom left region in each graph shows the gamma score at 3%/3 mm. The gamma index passed and failed points are indicated black and red scatters, respectively. The RBE in the graph means relative biological effectiveness.

Table 2. Validation results for the six dose algorithms, which consist of the average/median (minimum-maximum) gamma scores of depth doses and lateral profiles at 2%/2 mm and 3%/3 mm for the homogeneous and TG-119 plans, and the gamma scores of lateral profiles at 3%/3 mm for the heterogeneous plans.

	Depth doses Average/Median (Min-Max)		Lateral profiles Average/Median (Min-Max)		
	Homogeneous at 3%/3 mm(%)	TG-119 at 3%/3 mm(%)	Homogeneous at 2%/2 mm(%)	Heterogeneous at 3%/3 mm(%)	TG-119 at 3%/3 mm(%)
PBA	89.4/93.1 (11.1-100.0)	87.9/91.3 (51.7-100)	93.8/99.7 (33.5-100.0)	96.7/100.0 (83.9-100)	87.9/91.3 (51.7-100)
MC_0.3%	95.7/100.0 (70.0-100.0)	97.8/100 (91.7-100)	98.5/100.0 (88.9-100.0)	98.8/100.0 (94.1-100)	97.8/100 (91.7-100)
MC_0.5%	95.6/100.0 (58.0-100.0)	97.2/100 (91.3-100)	98.5/100.0 (90.0-100.0)	98.8/99.6 (94.1-100)	97.2/100 (91.3-100)
MC_1.0%	94.5/100.0 (70.0-100.0)	97.8/100 (91.7-100)	98.7/99.7 (91.2-100.0)	99.2/100.0 (95.6-100)	97.8/100 (91.7-100)
MC_1.5%	91.0/95.8 (38.9-100.0)	90.6/93.2 (69.6-100)	99.0/100.0 (94.0-100.0)	99.7/100.0 (98.4-100)	90.6/93.2 (69.6-100)
MC_2.0%	86.9/90.3 (38.9-100.0)	84.9/84.8 (69.6-95.7)	99.4/100.0 (95.6-100.0)	99.9/100.0 (99.3-100)	84.9/84.8 (69.6-95.7)

the depth doses and lateral profiles. For the homogeneous depth dose results, the average/median gamma scores at 3%/3 mm were 89.4%/93.1%, 95.7%/100.0%, 95.6%/100.0%, 94.5%/100.0%, 91.0%/95.8%, and 86.9%/90.3% for PBA, MC_0.3%, MC_0.5%, MC_1.0%, MC_1.5%, and MC_2.0%, respectively. For the PBA plans, the gamma score of the depth doses at 3%/3 mm averaged less than 90%. At uncertainties below MC_0.5%, the average gamma scores exceeded 95% at 3%/3 mm. The average gamma scores of depth doses were decreased with increasing statistical uncertainty.

For the homogeneous plans, the average/median gamma scores of 128 lateral profiles at 2%/2 mm were 93.8%/99.7%, 98.5%/100.0%, 98.5%/100.0%, 98.7%/99.7%, 99.0%/100.0%, and 99.4%/100.0% for PBA, MC_0.3%, MC_0.5%, MC_1.0%, MC_1.5%, and MC_2.0%, respectively. The average gamma score of MC plans was improved by at least 4.8% to that of the PBA plans. All the MC plans exceeded 90.0% gamma scores except for MC_0.3%. Table 3 shows the summary of average/median (score ranges) gamma scores of 80 and 48 lateral profiles for the RS- and RS+ plans, and multiple comparison results between the PBA and five MC plans. In the RS-plan, the gamma scores of all MC plans were not significantly different from those of PBA. Meanwhile, the MC plans in the RS+ plans had significantly better gamma pass rates than the PBA plans except for MC_1.0% ($p < 0.05$). Contrary to the results of the depth doses, the gamma scores of lateral profiles in homogeneous plans improved with larger statistical uncertainty.

Figure 4 shows the representative lateral profile comparisons at the isocenter between the measurements and calculations in the heterogeneous phantom. The top two rows represent 2D dose distributions for six

Table 3. Summary of the average/median (minimum-maximum) gamma scores of 80 and 48 lateral profiles for the RS- and RS+ plans, and multiple comparison results between the PBA and five MC plans. The * and ** marks indicate statistically significant differences of $p < 0.05$ and $p < 0.01$, respectively.

	80 lateral profiles of RS- (%)		48 lateral profiles of RS+ (%)	
	Average/Median (Min-Max)	p value	Average/Median(Min-Max)	p value
PBA	95.8/100.0 (38.3-100.0)	-	90.6/99.1 (33.5-100.0)	-
MC_0.3%	98.4/100.0 (89.7-100.0)	0.872	98.7/100.0 (88.9-100.0)	0.006 **
MC_0.5%	98.4/99.8 (90.0-100.0)	0.995	98.8/100.0 (90.9-100.0)	0.034 *
MC_1.0%	98.7/99.7 (92.2-100.0)	0.945	98.7/99.7 (91.2-100.0)	0.189
MC_1.5%	99.0/100.0 (95.6-100.0)	0.539	99.0/100.0 (94.0-100.0)	0.047 *
MC_2.0%	99.3/100.0 (96.0-100.0)	0.253	99.5/100.0 (95.6-100.0)	0.001 **

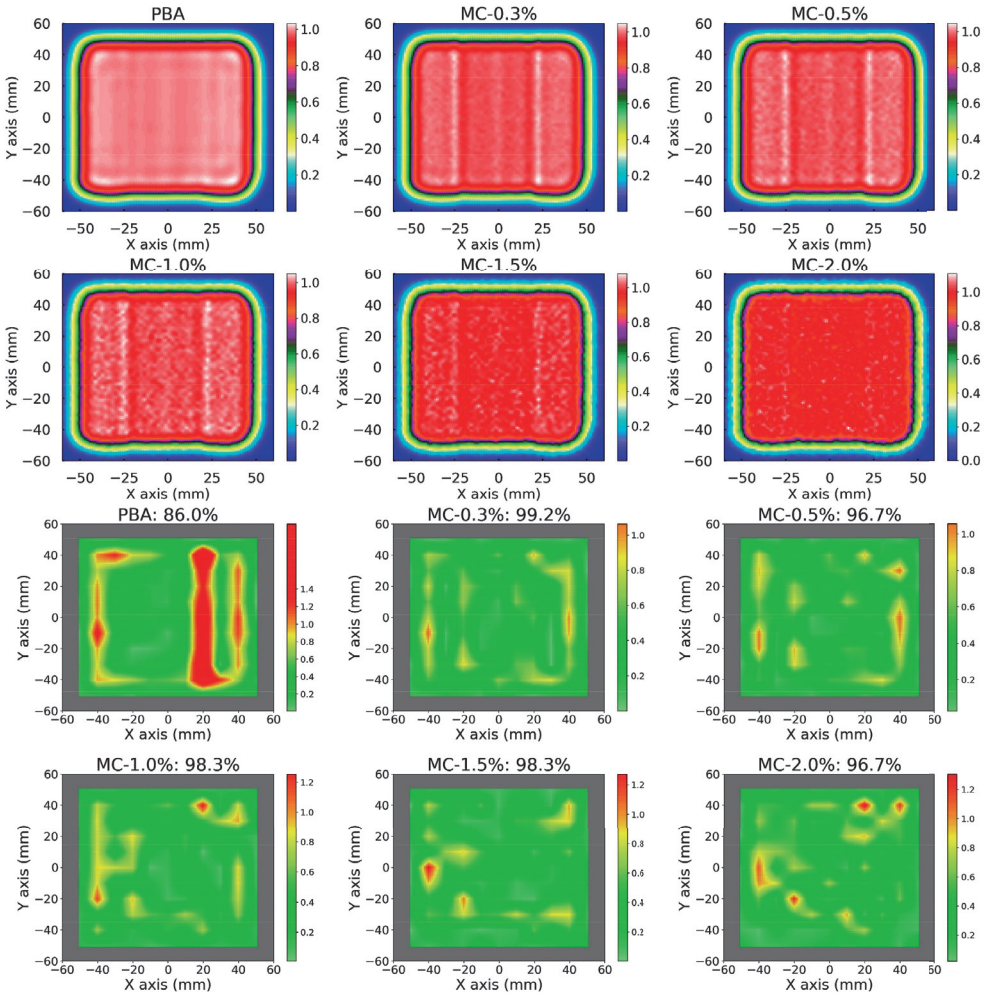

Fig. 4. Representative lateral profile comparisons at the isocenter between the measurements and calculations in the heterogeneous phantom. The top two rows represent 2D dose distributions for six calculation algorithms. The bottom two rows indicate the gamma distributions at 2%/2 mm. Each title represents the name of the selected dose calculation algorithms and the gamma score between the measurements and calculations. The red regions indicate gamma failed points as a gamma index of more than 1.0.

calculation algorithms. The bottom two rows indicate the gamma distributions at 2%/2 mm. The calculated doses for all MC plans have more significant dose variations as the statistical uncertainty increases. The minimum/average gamma scores of 12 lateral profiles at 3%/3 mm were improved at least from 83.9%/96.7% for the PBA plans to 94.1%/98.8% for MC_0.3% plans. All the MC plans with other statistical uncertainties also improved their gamma scores over the PBA plans, especially for the RS+ and heterogeneous plans.

Figure 5 shows the representative verification results of three TG-119 beams between the measurements and calculations with MC_0.3% for (A) MultiTarget, (B) C-shape, and (C) Head and Neck plans with the

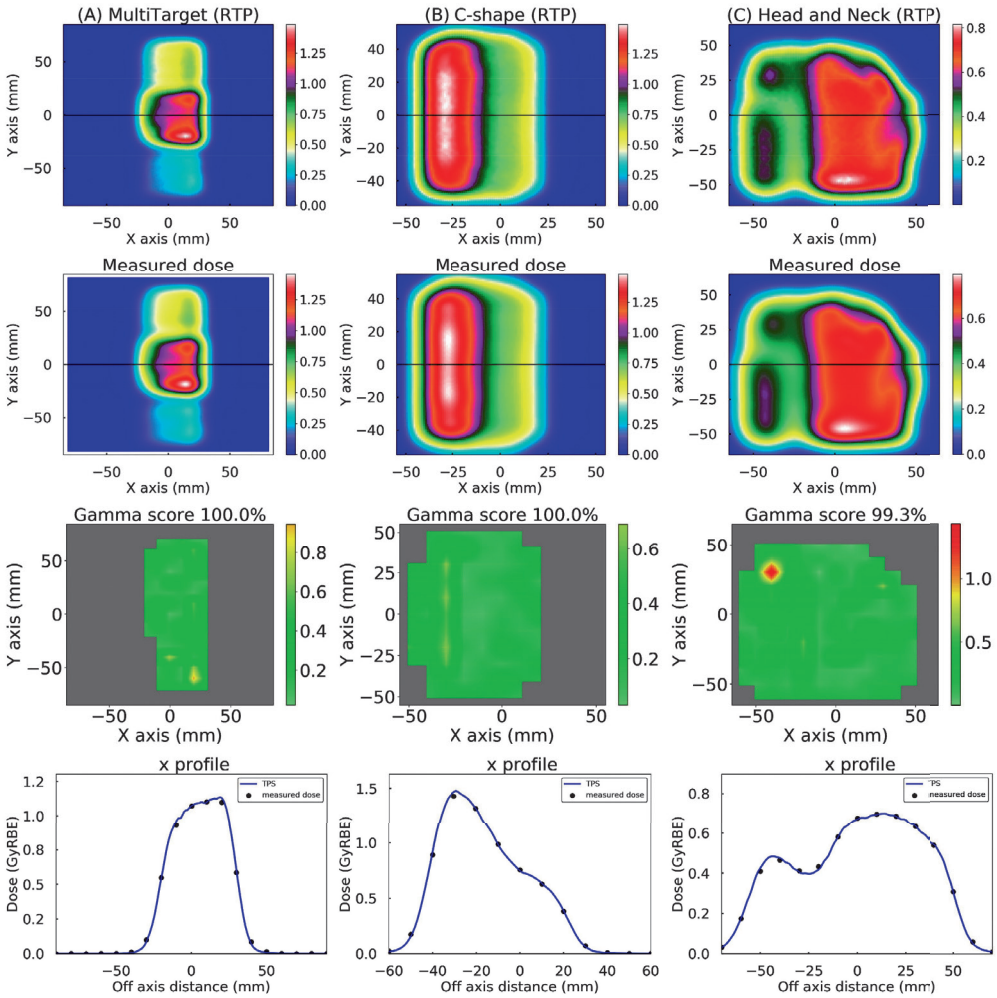

Fig. 5. Representative TG-119 beam's verifications of between the measurements and calculations for (A) MultiTarget, (B) C-shape, and (C) Head and Neck plans with MC_0.3%. The rows 1, 2, 3, and 4 show the 2D calculated doses, 2D measured doses, gamma distributions at 3%/3 mm (gamma scores were shown in each title), and lateral profiles comparisons between the measured (scatter) and calculated (plots) doses at the black horizontal line in the 2D dose distribution.

2D calculated doses, 2D measured doses, gamma distributions at 3%/3 mm, and lateral profiles comparisons between the measured and calculated doses at the black horizontal lines in the 2D dose distributions. The average/median gamma scores of 32 lateral profiles at 3%/3 mm were improved at least from 87.9%/91.3% for the PBA plans to 97.8%/100.0% for the best one, the MC_0.3% plans.

Figure 6 shows the absolute DDs at the isocenter between the measurements and calculations for the six dose algorithms in the homogeneous and TG-119 plans. The black, red, and blue scatter indicate the RS-, RS+, and TG-119 plans, respectively. The blue lines in each figure indicate our absolute dose error tolerance of ±3%. No trends in absolute DDs were observed for five MC algorithms due to the difference in measurement depths. Most RS+ plans of PBA do not appear to have a measured depth-dependent trend of DD compared to the RS- plans. However, the DD was worse at -6.7% for the plan measured at the shallowest depth of 15 mm (Fig. 6 top left). Table 4 shows the summary of absolute DDs and multiple comparisons. The minimum/maximum absolute DDs were -6.7%/2.8%, -2.0%/2.7%, -2.4%/2.9%, -4.1%/4.3%, -3.1%/5.3%, and -6.5%/8.1% for PBA, MC_0.3%, MC_0.5%, MC_1.0%, MC_1.5%, and MC_2.0%, respectively. Only the MC_0.3% and MC_0.5% plans had a satisfactory DD less than ±3.0%. No significantly average DDs were observed between the PBA and five MC plans ($p > 0.05$), but the standard deviation of the DDs gradually expanded as the uncertainty increased.

Table 4 also shows the summary of average dose calculation times, the ratio to PBA plans, and multiple comparison results. The average dose calculation times per beam were 0.6, 9.3, 3.4, 1.0, 0.5, and 0.3 minutes

Table 4. Summary of the average ± standard deviation (Av. ± SD) absolute DDs, mean dose calculation times, mean time ratios of MC plans compared to that of PBA plans, and the multiple comparisons for DDs and dose calculation times between the PBA and five MC plans. The values in parentheses indicate the minimum and maximum ranges. The * and ** marks indicate statistically significant differences of $p < 0.05$ and $p < 0.01$, respectively.

	Absolute dose differences (%)		Mean calculation time (min)		
	Av. ± SD (Min-Max)	p value	Av. ± SD (Min-Max)	Ratio to PBA time (Min-Max)	p value
PBA	-0.1 ± 1.5 (-6.7-2.8)	-	0.6 ± 0.8 (0.1-3.6)	1.0	-
MC_0.3%	0.3 ± 1.1 (-2.0-2.7)	0.797	9.3 ± 11.1 (1.2-42.1)	15.5 (6.7-37.0)	<0.001 **
MC_0.5%	0.3 ± 1.2 (-2.4-2.9)	0.811	3.4 ± 4.1 (0.5-15.3)	5.8 (2.6-13.6)	<0.001 **
MC_1.0%	0.3 ± 1.7 (-4.1-4.3)	0.998	1.0 ± 1.1 (0.2-4.2)	1.7 (0.8-3.7)	0.020 *
MC_1.5%	0.5 ± 2.0 (-3.1-5.3)	0.927	0.5 ± 0.5 (0.1-2.0)	0.8 (0.4-2.0)	0.774
MC_2.0%	0.0 ± 2.9 (-6.5-8.1)	0.921	0.3 ± 0.3 (0.1-1.3)	0.5 (0.2-2.0)	0.992

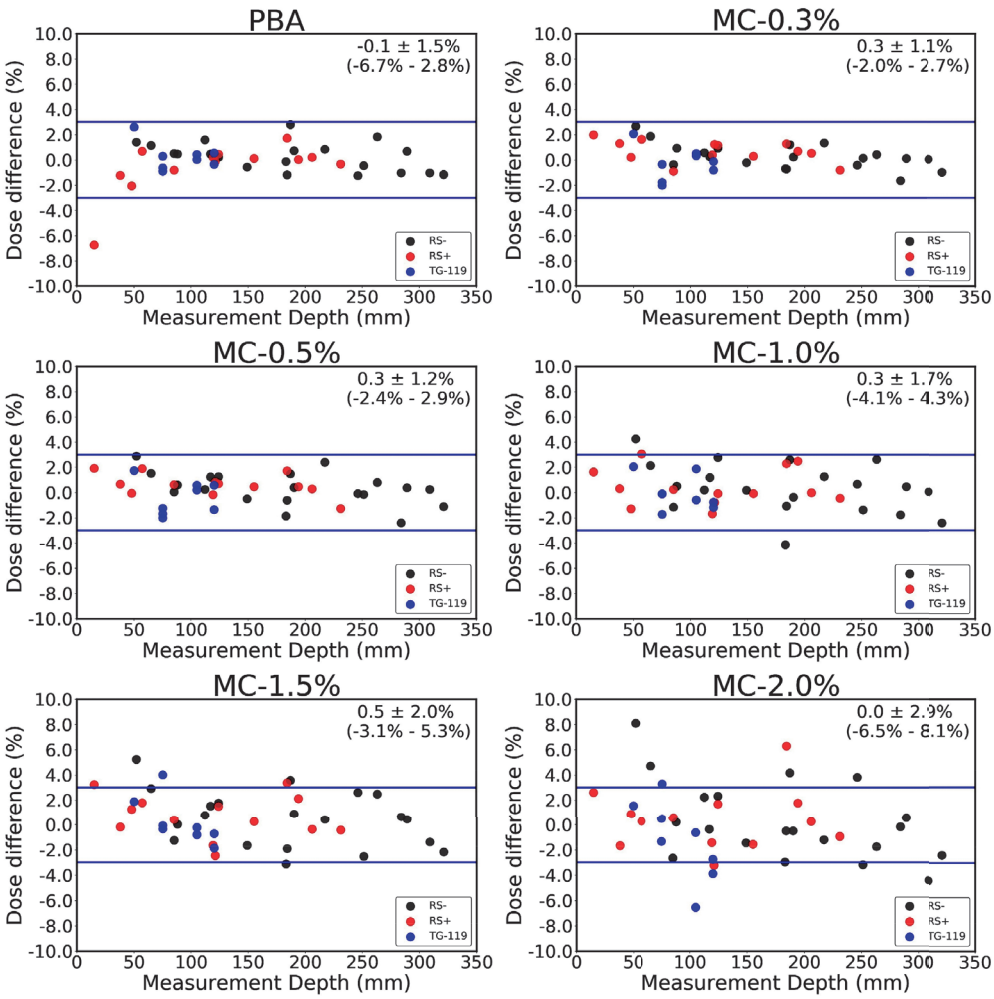

Fig. 6. Absolute DDs at the isocenter for 33 cubic and eight TG-119 plans as a function of measurement depth for six algorithms. Each title represents the name of the dose calculation algorithm used. The black, red, and blue scatters indicate the RS-, RS+, and TG-119 plans, respectively. The blue lines in each figure indicate our absolute DD tolerance of ±3%. The average and standard deviation (minimum-maximum) of DDs is shown in the upper right corner of each graph.

for PBA, MC_0.3%, MC_0.5%, MC_1.0%, MC_1.5%, and MC_2.0%, respectively. The average dose calculation times were 15.5, 5.8, 1.7, 0.8 and 0.5 times for MC_0.3%, MC_0.5%, MC_1.0%, MC_1.5%, and MC_2.0% longer than that of PBA. The average dose calculation times were not significant at MC_1.5% ($p = 0.774$) and MC_2.0% ($p = 0.992$), and increased significantly at MC_0.3% ($p < 0.001$), MC_0.5% ($p < 0.001$), and MC_1.0% ($p = 0.020$) than that of PBA.

4. Discussion

In this study, we validated the clinical beam commissioning of our new PBS system that comprises both PBA and MC dose calculation engines. We also compared the calculation accuracies of MC, which can be used as an alternative to PBA. Saini *et al.* reported that a beam calculated by PBA with a range of less than 30 mm causes dose overestimation (up to 8.0%) in the plateau region and within the SOBP[37]. Similar to the report, our RS plan with PBA less than 30 mm beam ranges led to a large dose discrepancy of -6.7% in the absolute DD (Fig. 6 and Table 4). Our MC plans with five uncertainties showed that all absolute dose overestimations were improved in the ranges of 2.0%-3.3% for the RS+ plan less than 30 mm ranges. Thus, more significant dose overestimation will occur in our PBA plans, especially for the RS+ plans (Table 3). These PBA dose overestimations are caused by the nuclear halo, which occurs when air passes between the RS and patient surface[31].

The previous report showed that the gamma scores of lateral profiles at 3%/3 mm were improved from 70.7-85.3% for the PBA plan to 92.0-99.1% for the MC plans in a realistic lung (heterogeneous) phantom[38]. Widesott *et al.* also showed that the MC-calculated beams had all more than 95% gamma scores of lateral profiles at 3%/3 mm for 0 degree and oblique gantry angle beams in a realistic head phantom[39]. Our results showed that all MC algorithms as an alternative to the PBA satisfied the average gamma scores for the lateral profiles of 98% at 2%/2 mm for homogeneous plans and 95% at 3%/3 mm for heterogeneous and TG-119 plans. As the heterogeneous plans were created assuming treatment through the chest wall, we showed that the MC plan calculations are useful even in the border between the lung and bone. The gamma analysis of the 2D dose distribution resulted in an improved gamma score for larger statistical uncertainty for the homogeneous and heterogeneous plans (Table 2). The reason why that the gamma analysis of Monte Carlo dose distributions may underestimate the gamma scores is because both the Distant-To-Agreement (DTA) and the gamma functions are sensitive to noise (variations due to different statistical uncertainties)[40]. However, the gamma scores of TG-119 plans deteriorated when the statistical uncertainty of more than 1.0%. Thus, the statistical uncertainty of MC at least below 1% is desirable for the use of patient fields in terms of the lateral profile agreements between the calculations and measurements.

Our absolute DDs showed that the MC_0.3% and the MC_0.5% plans were satisfied both ICRU 24 (less than ± 5.0%) and our tolerance (less than ± 3.0%) for all measurements depths[41]. According to our validation results, similar to the statistical uncertainty often used in previous reports, the dose calculations were in good agreement with the measurements if the statistical uncertainty was used below 0.5% for our PBS system[38, 42]. The statistical analysis showed no significant difference between the PBA and five MC plans. However, a more significant statistical uncertainty will result in a larger maximum and standard deviation of the absolute dose, which is an expected tendency of the MC computational nature[43]. Therefore, an increased statistical uncertainty is expected to deteriorate the target uniformity and dose-volume histograms of the PBS plans by contributing to the target uniformity.

The dose calculation times were compared between the PBA and the MC_0.5% plans in different PBS machines with the same TPS; the average calculation time of the MC plans increased 5.7 times longer than that of PBA plans[44]. Our average time ratio of the MC_0.5% plans to the PBA plans was similar to their results (5.8 times). Although our MC calculations were longer than those of the PBA, the MC calculation time of the TPS is faster than other in-house MC codes, which takes more than eight hours[45]. As shown in our results for various uncertainties, the calculation time to get better dose calculation accuracy (lower statistical uncertainty) is a trade-off. Considering the results of our validations in the PBS system, the statistical uncertainty of below 0.5% was desirable to perform clinical use even if the time is significantly extended. Recently, a new version of the TPS with a GPU-based dose calculation engine has been developed[46]. We expect that the abovementioned trade-off between the calculation time and calculation accuracy will be reduced.

Two limitations were identified in the present study. First, we performed dose verification using a 2D ionization chamber that has a more widely spaced ionization chamber of 10 mm. The verification results may differ when measured using a radiochromic film with a higher resolution. Second, no verification was performed on the dose distributions owing to the differences between the PBA and the MC plans, as all plans were optimized by the PBA. The TPS could select both algorithms for the final dose calculation and plan optimization, thereby optimizing the beam spot position and MU. Several reports have concluded that optimization by the MC algorithm is more effective than that of the PBA for the lung and breast treatment sites[47, 48]. Detailed plan evaluations will be required for the PBS plan that was created for the MC optimization and the MC final calculation.

5. Conclusion

We introduced the newly installed PBS proton therapy system and performed beam commissioning before clinical use. The MC plans with less than 0.5% statistical uncertainties were in good agreements between

the calculations and measurement and satisfied our tolerances of absolute dose differences and lateral profiles. However, the calculation times of below 1.0% statistical uncertainty were significantly longer than that of the PBA. We suggest that the MC plans below 0.5% statistical uncertainty is appropriate for clinical use with PBS plans even if requiring a longer calculation time.

Acknowledgements

We would like to thank the staff of Tsuyama Chuo Hospital for their cooperation in obtaining the experimental data for this study.

Conflict of interest

The authors have no conflicts of interest.

Author contribution

Yuki Tominaga, Masataka Oita and Junya Miyata involved in designing this study. The manuscript was drafted by Yuki Tominaga, Masataka Oita, Takashi Akagi, Junya Miyata, Shuichi Harada, Tetsunori Matsuda, and Masahiro Kuroda.

References

1. Pedroni E, Bacher R, Blattmann H, Böhringer T, Coray A, Lomax A, *et al*. The 200-MeV proton therapy project at the Paul Scherrer Institute: conceptual design and practical realization. Med Phys. 1995;22(1):37–53.
2. Zhu XR, Poenisch F, Lii M, Sawakuchi GO, Titt U, Bues M, *et al*. Commissioning dose computation models for spot scanning proton beams in water for a commercially available treatment planning system. Med Phys. 2013;40(4):041723.
3. Zhang X, Li Y, Pan X, Xiaoqiang L, Mohan R, Komaki R, *et al*. Intensity-modulated proton therapy reduces the dose to normal tissue compared with intensity-modulated radiation therapy or passive scattering proton therapy and enables individualized radical radiotherapy for extensive stage IIIB non-small-cell lung cancer: a virtual clinical study. Int J Radiat Oncol Biol Phys. 2010;77(2):357–66.
4. Ares C, Khan S, Macartain AM, Heuberger J, Goitein G, Gruber G, *et al*. Postoperative proton radiotherapy for localized and locoregional breast cancer: potential for clinically relevant improvements? Int J Radiat Oncol Biol Phys. 2010;76(3):685–97.
5. Hong L, Goitein M, Bucciolini M, Comiskey R, Gottschalk B, Rosenthal S, *et al*. A pencil beam algorithm for proton dose calculations. Phys Med Biol. 1996;41(8):1305–30.
6. Tourovsky A, Lomax AJ, Schneider U, Pedroni E. Monte Carlo dose calculations for spot scanned proton therapy. Phys Med Biol. 2005;50(5):971–81.
7. 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(1):27–41.
8. Winterhalter C, Fura E, Tian Y, Aitkenhead A, Bolsi A, Dieterle M, *et al*. Validating a Monte Carlo approach to absolute dose quality assurance for proton pencil beam scanning. Phys Med Biol. 2018;63(17):175001.
9. Huang S, Kang M, Souris K, Ainsley C, Solberg TD, McDonough JE, *et al*. Validation and clinical implementation of an accurate Monte Carlo code for pencil beam scanning proton therapy. J Appl Clin Med Phys. 2018;19(5):558–72.
10. Molinelli S, Russo S, Magro G, Maestri D, Mairani A, Mastella E, *et al*. Impact of TPS calculation algorithms on dose delivered to the patient in proton therapy treatments. Phys Med Biol. 2019;64(7):075016.
11. 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):750–6.
12. Righetto R, Clemens LP, Lorentini S, Fracchiolla F, Algranati C, Tommasino F, *et al*. Accurate proton treatment planning for pencil beam crossing titanium fixation implants. Phys Med. 2020;70:28–38.
13. Grassberger C, Daartz J, Dowdell S, Ruggieri T, Sharp G, Paganetti H. Quantification of proton dose calculation accuracy in the lung. Int J Radiat Oncol Biol Phys. 2014;89(2):424–30.
14. Yepes P, Adair A, Grosshans D, Mirkovic D, Poenisch F, Titt U, *et al*. Comparison of Monte Carlo and analytical dose computations for intensity modulated proton therapy. Phys Med Biol. 2018;63(4):045003.
15. Liang X, Li Z, Zheng D, Bradley JA, Rutenberg M, Mendenhall N. A comprehensive dosimetric study of Monte Carlo and pencil-beam algorithms on intensity-modulated proton therapy for breast cancer. J Appl Clin Med Phys. 2019;20(1):128–36.
16. RaySearch Laboratory: RayStation 7 reference manual. 2017.
17. Pidikiti R, Patel BC, Maynard MR, Dugas JP, Syh J, Sahoo N, *et al*. Commissioning of the world's first compact pencil-beam scanning proton therapy system. J Appl Clin Med Phys. 2018;19(1):94–105.
18. Chang CW, Huang S, Harms J, Zhou J, Zhang R, Dhabaan A, *et al*. A standardized commissioning framework of Monte Carlo dose calculation algorithms for proton pencil beam scanning treatment planning systems. Med Phys. 2020;47(4):1545–57.
19. Kim DH, Park S, Jo K, Cho S, Shin EH, Lim DH, *et al*. Investigations of line scanning proton therapy with dynamic multi-leaf collimator. Phys Med. 2018;55:47–55.
20. Kang M, Pang D. Commissioning and beam characterization of the first gantry-mounted accelerator pencil beam scanning proton system. Med Phys. 2020;47(8):3496–510.
21. 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(22):7673–88.
22. Maes D, Regmi R, Taddei P, Bloch C, Bowen S. Parametric characterization of penumbra reduction for aperture-collimated pencil beam scanning (PBS) proton therapy. Biomed Phys Eng Express. 2019;5(3):035002.
23. Grewal HS, Ahmad S, Jin H. Characterization of penumbra sharpening and scattering by adaptive aperture for a compact pencil beam scanning proton therapy system. Med Phys. 2021;48(4):1508–19.
24. Furukawa T, Inaniwa T, Sato S, Tomitani T, Minohara S, Noda K, *et al*. Design study of a raster scanning system for moving target irradiation in heavy-ion radiotherapy. Med Phys. 2007;34(3):1085–97.
25. Fukumitsu N, Yamashita T, Mima M, Demizu Y, Suzuki T, Soejima T. Dose distribution effects of spot-scanning proton beam therapy equipped with a multi-leaf collimator for pediatric brain tumors. Oncol Lett. 2021;22(2):635.
26. Goudsmit SA, Saunderson JL. Multiple scattering of electrons.

Phys Rev. 1940;57:24–29.

27. Goudsmit SA, Saunderson JL. Multiple scattering of electrons II. Phys Rev. 1940;58:36–42.
28. ICRU. Nuclear Data for Neutron and Proton Radiotherapy and for Radiation Protection. ICRU Report 63. Bethesda: ICRU;2000.
29. Lomax A. Intensity modulation methods for proton radiotherapy. Phys Med Biol. 1999;44(1):185–205.
30. Bijian A, Paige T, Christopher A, Sairos S, Narayan S, Mark P, *et al*. AAPM task group 224: comprehensive proton therapy machine quality assurance. Med Phys. 2019;46(8):e678–e705.
31. Saini J, Maes D, Egan A, Bowen SR, St James S, Janson M, *et al*. Dosimetric evaluation of a commercial proton spot scanning Monte-Carlo dose algorithm: comparisons against measurements and simulations. Phys Med Biol. 2017;62(19):7659–81.
32. Fracchiolla F, Fellin F, Innocenzi M, Lipparini M, Lorentini S, Widesott L, *et al*. A pre-absorber optimization technique for pencil beam scanning proton therapy treatments. Phys Med. 2019;57:145–52.
33. Ezzell GA, Burmeister JW, Dogan N, Losasso TJ, Mechalakos JG, Mihailidis D, *et al*. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. Med Phys. 2009;36(11):5359–73.
34. Cubillos-Mesías M, Baumann M, Troost EGC, Lohaus F, Löck S, Richter C, *et al*. Impact of robust treatment planning on single- and multi-field optimized plans for proton beam therapy of unilateral head and neck target volumes. Radiat Oncol. 2017;12(1):190.
35. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys. 1998;25(5):656–61.
36. Farr JB, Moyers MF, Allgower CE, Bues M, Hsi WC, Jin H, *et al*. Clinical commissioning of intensity-modulated proton therapy systems: report of AAPM Task Group 185. Med Phys. 2021;48(1):e1–30.
37. Saini J, Cao N, Bowen SR, Herrera M, Nicewonger D, Wong T, *et al*. Clinical commissioning of a pencil beam scanning treatment planning system for proton therapy. Int J Part Ther. 2016;3(1):51–60.
38. Schreuder AN, Bridges DS, Rigsby L, Blakey M, Janson M, Hedrick SG, *et al*. Validation of the RayStation Monte Carlo dose calculation algorithm using a realistic lung phantom. J Appl Clin Med Phys. 2019;20(12):127–37.
39. Widesott L, Lorentini S, Fracchiolla F, Farace P, Schwarz M. Improvements in pencil beam scanning proton therapy dose calculation accuracy in brain tumor cases with a commercial Monte Carlo algorithm. Phys Med Biol. 2018;63(14):145016.
40. Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. Med Phys. 2003;30(9):2455–64.
41. Cunningham J, Cohen M, Dutreix, A. ICRU Report 24: Determination of Absorbed Dose in a Patient Irradiated by Beams of X-or Gamma-rays in Radiotherapy Procedures. Washington: ICRU;1976.
42. Saini J, Traneus E, Maes D, Regmi R, Bowen SR, Bloch C, *et al*. Advanced proton beam dosimetry part I: review and performance evaluation of dose calculation algorithms. Translational Lung Cancer Res. 2018;7(2):171–79.
43. Maes D, Saini J, Zeng J, Rengan R, Wong T, Bowen SR. Advanced proton beam dosimetry part II: Monte Carlo vs. pencil beam-based planning for lung cancer. Transl Lung Cancer Res. 2018;7(2):114–21.
44. Schreuder AN, Bridges DS, Rigsby L, Blakey M, Janson M, Hedrick SG, *et al*. Validation of the RayStation Monte Carlo dose calculation algorithm using realistic animal tissue phantoms. J Appl Clin Med Phys. 2019;20(10):160–71.
45. Molinelli S, Mairani A, Mirandola A, Vilches Freixas G, Tessonnier T, Giordanengo S, *et al*. Dosimetric accuracy assessment of a treatment plan verification system for scanned proton beam radiotherapy: one-year experimental results and Monte Carlo analysis of the involved uncertainties. Phys Med Biol. 2013;58(11):3837–47.
46. Fracchiolla F, Engwall E, Janson M, Tamm F, Lorentini S, Fellin F, *et al*. Clinical validation of a GPU-based Monte Carlo dose engine of a commercial treatment planning system for pencil beam scanning proton therapy. Phys Med. 2021;88:226–34.
47. Rana S, Greco K, Samuel EJJ, Bennouna J. Radiobiological and dosimetric impact of RayStation pencil beam and Monte Carlo algorithms on intensity-modulated proton therapy breast cancer plans. J Appl Clin Med Phys. 2019;20(8):36–46.
48. Teoh S, Fiorini F, George B, Vallis KA, Van den Heuvel F. Is an analytical dose engine sufficient for intensity modulated proton therapy in lung cancer? Br J Radiol. 2020;93(1107):20190583.