Physica Medica 64 (2019) 123–131

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Dosimetric validation of Monte Carlo and analytical dose engines with raster-scanning ^{1}H, ^{4}He, ^{12}C, and ^{16}O ion-beams using an anthropomorphic phantom

Stewart Mein[a,b,c,d,e], Benedikt Kopp[b,c,d,e,f], Thomas Tessonnier[g], Benjamin Ackermann[b,c], Swantje Ecker[b,c], Julia Bauer[b,c], Kyungdon Choi[h,i], Giulia Aricò[j], Alfredo Ferrari[j], Thomas Haberer[b,c], Jürgen Debus[b,c,d,e], Amir Abdollahi[a,b,c,d], Andrea Mairani[b,c,h,*]

[a] Division of Molecular and Translational Radiation Oncology, Heidelberg University Medical School, Heidelberg Institute of Radiation Oncology (HIRO), National Center for Radiation Research in Oncology (NCRO), Heidelberg, Germany
[b] Heidelberg Ion-Beam Therapy Center (HIT), Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany
[c] Radiation Oncology, Department of Radiology, Heidelberg University Hospital, Germany
[d] German Cancer Consortium (DKTK) Core Center, National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany
[e] Heidelberg University, Faculty of Physics, Heidelberg, Germany
[f] Clinical Cooperation Unit Radiooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[g] Centre François Baclesse, Radiation Oncology, Medical Physics Department, Caen, France
[h] National Centre of Oncological Hadrontherapy (CNAO), Medical Physics, Pavia, Italy
[i] University of Pavia, Department of Physics, Pavia, Italy
[j] European Organization for Nuclear Research (CERN), Geneva, Switzerland

ARTICLE INFO

Keywords:
Particle therapy
Pencil beam algorithm
Monte Carlo simulation
GPU
Dosimetry
Anthropomorphic phantom

ABSTRACT

With high-precision radiotherapy on the rise towards mainstream healthcare, comprehensive validation procedures are essential, especially as more sophisticated technologies emerge. In preparation for the upcoming translation of novel ions, case-/disease-specific ion-beam selection and advanced multi-particle treatment modalities at the Heidelberg Ion-beam Therapy Center (HIT), we quantify the accuracy limits in particle therapy treatment planning under complex heterogeneous conditions for the four ions (^{1}H, ^{4}He, ^{12}C, ^{16}O) using a Monte Carlo Treatment Planning platform (MCTP), an independent GPU-accelerated analytical dose engine developed in-house (FRoG) and the clinical treatment planning system (Syngo RT Planning). Attaching an anthropomorphic half-head Alderson RANDO phantom to entrance window of a dosimetric verification water tank, a cubic target spread-out Bragg peak (SOBP) was optimized using the MCTP to best resolve effects of anatomic heterogeneities on dose homogeneity. Subsequent forward calculations were executed in FRoG and Syngo. Absolute and relative dosimetry was performed in the experimental beam room using 1D and 2D array ionization chamber detectors. Mean absolute percent deviation in dose ($|\%\Delta|$) between predictions and PinPoint ionization chamber measurements were within ~2% for all investigated ions for both MCTP and FRoG. For protons and carbon ions, $|\%\Delta|$ values were ~4% for Syngo. For the four ions, 3D-γ analysis (3%/3mm criteria) of FLUKA and FRoG presented mean passing rates of 97.0(± 2.4)% and 93.6(± 4.2)%. FRoG demonstrated satisfactory agreement with gold standard Monte Carlo simulation and measurement, superior to the commercial system. Our preclinical trial landmarks the first measurements taken in anthropomorphic settings for helium, carbon and oxygen ion-beam therapy.

1. Introduction

Particle therapy in cancer treatment affords superior dose conformity to solid tumors compared to conventional modalities [1–3]. To assure optimal treatment delivery, clinicians tailor patient-specific plans via a treatment planning system (TPS). With the rise of high-precision pencil beam scanning (raster-scanning) particle beams, the balance of uncertainty in dose prediction with target conformality between planning and delivery remains a critical issue. In delicate cases, clinical physics must rely on sophisticated measurement tools and time-

* Corresponding author at: Physics Department, Heidelberg Ion-beam Therapy Center (HIT), In Neuenheimer Feld (INF) 450, Heidelberg, DE 69120, Germany.
E-mail address: Andrea.Mairani@med.uni-heidelberg (A. Mairani).

https://doi.org/10.1016/j.ejmp.2019.07.001
Received 12 March 2019; Received in revised form 30 June 2019; Accepted 1 July 2019
Available online 11 July 2019

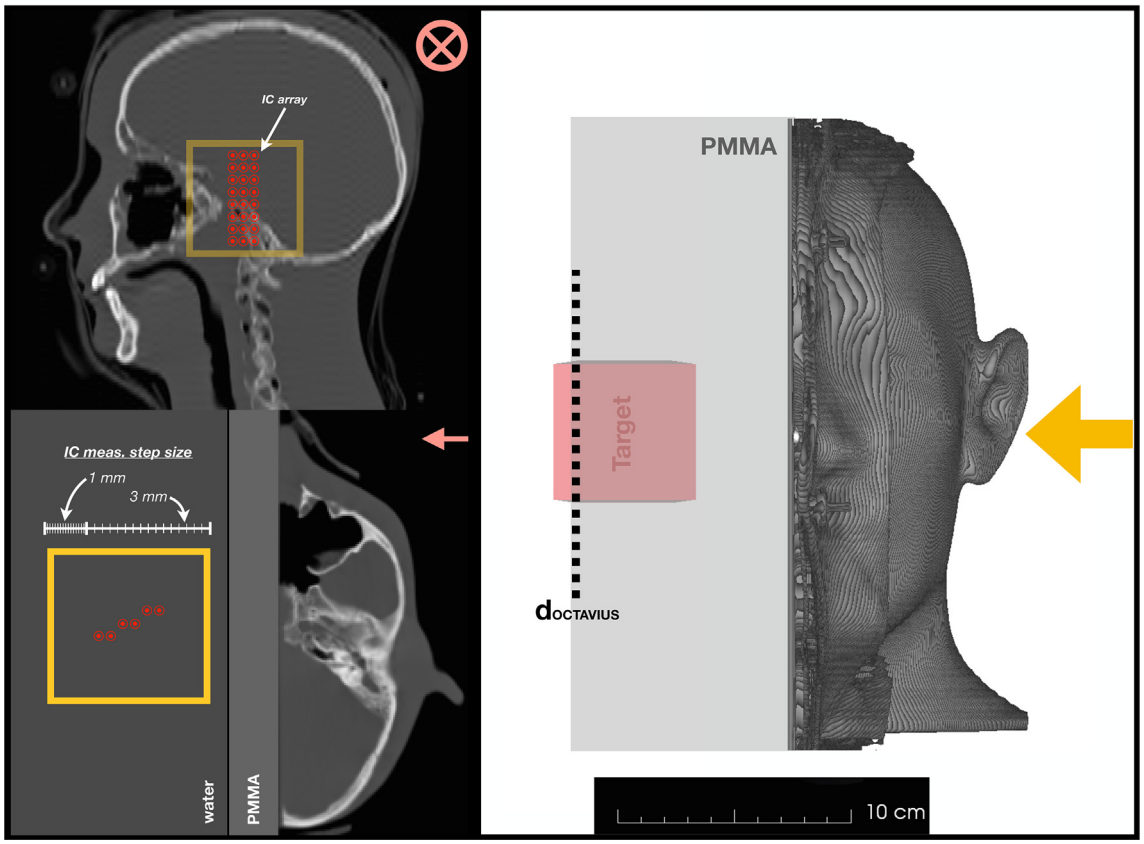

Fig. 1. Planning CT is displayed (left) with the red symbols indicating beam direction and yellow box outlining the target. IC measurement configuration and step size is provided. The schematic on the right illustrates the experimental set-up using the OCTAVIUS® system mounted behind PMMA blocks to reach the desired WED. The yellow arrow denotes beam direction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

intensive Monte Carlo transport and interactions codes as a gold standard reference for accuracy in dose prediction. In contrast, the TPS used by clinical oncologists, physicists and dosimetrists provides treatment optimizations in minutes but these capabilities are accompanied by a potential loss in accuracy by using approximations or simplified models to prepare patient-specific dose distributions within the clinical timeframe. For example, pencil beam (PB) algorithms generally assume beam properties associated with dose deposition in water, regardless of patient anatomical structure, applying simplified models for lateral scattering such as Gaussian parameterization, and pencil beam splitting to account for modulated range due to anatomic heterogeneity. In short, analytical algorithms superimpose a field of individual particle beams (varying energy and position), either measured or simulated in a homogenous water block [4–7]. Planning on patient or patient-like anatomical structures introduces heterogeneities, impacting the particle range by "deforming" or "distorting" the PB shape [8]. The method of calculation and accuracy limits of a TPS can vary between vendors and calculation mode. Recently, a surge of graphics processing unit (GPU) accelerated codes is entering the market, making analytical algorithms potentially more accurate or Monte Carlo codes faster [9,10].

Before treating the very first patient, a facility will undergo a series of benchmark tests for TPS accuracy versus delivery in controlled (homogeneous) settings. These processes are well understood and meticulously summarized in various task group (TG) reports by the American Association of Medical Physics (AAPM), e.g. TG053 and TG119 for TPS quality assurance (QA) and commissioning, respectively [11]. Nevertheless, to date, there is no official guideline published by an overseeing body in particle therapy for acceptance testing with anthropomorphic phantoms, a setting more representative of clinical conditions and sensitive to uncertainties. The National Cancer Institute (NCI, Bethesda, MD, USA) in cooperation with the Imaging and Radiation Oncology Core (IROC, Houston, TX, USA) does, however, require a center to adhere to baseline credentialing benchmarks with anthropomorphic phantoms to participate in NCI-funded clinical trials for photon- and proton-based therapies.

Prior works assess plan accuracy and robustness of a clinical TPS for proton therapy in the presence of high-level anatomical heterogeneity using patient and/or patient-like conditions, adopt their own end-to-end quality assurance (QA) procedures and present performance results ranging from outstanding to unacceptable, e.g. concluding that the analytical PB algorithms are unsuitable for lung treatment planning [12–17]. Similar efforts develop phantoms to gauge the effect of maximal geometry heterogeneity on dose calculation performance [18]; however, validations using anthropomorphic phantoms have yet to be performed with ion-beams heavier than protons.

For the four ions (^{1}H, ^{4}He, ^{12}C, ^{16}O) available at the Heidelberg Ion-beam Therapy Center (HIT), we set out to benchmark three dose calculation systems in the context of a challenging scenario for head and neck (H&N) treatments using an anthropomorphic phantom [19]. Recently developed in-house systems, such as the FLUKA-based Monte Carlo Treatment Planning Platform (MCTP) [20] and a GPU-accelerated analytical dose engine (FRoG) [21], and a commercial system for particle therapy, Syngo RT Planning (Siemens AG, Erlangen, Germany), are validated using multi-dimensional dosimetry.

At HIT, clinical activity with protons and carbon ion-beams began in 2009, with over 5000 patients treated to date and plans to start the world's first clinical program using raster-scanning helium ion-beams next year. Currently, there is no commercial TPS available for helium and oxygen ions, and therefore, advanced dosimetry benchmarks with in-house systems are critical. Consequently, FRoG was recently

integrated into the clinical and research pipeline as an independent dose engine at HIT, the National Centre for Oncological Hadrontherapy (CNAO, Pavia, Italy), the Danish Center for Particle Therapy (Aarhus, Denmark) and the Normandy Proton Therapy Center (Caen, France), with other facility partnerships planned or pending. Aside from HIT, Syngo is the primary TPS for various light and heavy ion facilities around the world e.g. CNAO, Shanghai Proton and Heavy Ion Center and Marburg Ion-beam Therapy Center (MIT). Thus, assessing the current standards in particle therapy dose calculation is essential to overcome the known limitations of analytical approaches, making way for both improvement in dose prediction, especially for delicate cases, and the translation of novel ion-beam modalities to the clinic.

2. Methods

2.1. Treatment planning

The RANDO Alderson® phantom (Radiology Support Devices, Long Beach, CA, USA) used in this study consists of a human skull with tissue-equivalent structures and air gaps. CT and dose calculation grid sizes were selected for half of the head phantom following clinical procedure, 0.9766 mm × 0.9766 mm × 3 mm resolution and 2 mm × 2 mm × 3 mm, respectively, with the water phantom entry window positioned at isocenter (Fig. 1). The set-up was developed and applied for clinical commissioning of treatment planning and delivery at HIT [22]. Subsequently, physical dose optimization of spread-out Bragg peak (SOBP) plans (6 cm × 6 cm × 6 cm target, centered at 5 cm depth from the entry window) for 1 Gy in the target was obtained using a FLUKA-based MCTP tool [23,24].

FLUKA inputs and default settings are detailed in the benchmarking of the MCTP in water [20]. More specifically, the HADROTHErapy option was activated to most accurately model multiple Coulomb scattering of charged particles with a particle transport and delta ray production threshold of 100 keV. COALESCE and EVAPORATION cards were also activated for production and evaporation of energetic heavy fragments, respectively. Monte Carlo simulations were performed using a detailed model of the HIT beamline [20,24,25], which incorporates the vacuum window and the beam applications and monitoring systems (BAMS). For carbon and oxygen ions, the ripple filter (RiFi) was implemented following the clinical routine for carbon ions [26]. A forward calculation for the clinical beams, namely protons and carbon ions, was performed using FRoG V2018 [21] and Syngo VC13A [27]. For the experimental beams, namely helium and oxygen ions, forward calculations were performed using FRoG only. All MC simulations were executed with the FLUKA 2018 development version until reaching a level of statistical uncertainty less than 1%, with histories typically ranging from 1% to 6% of the total primary particles for each plan [28].

2.2. Irradiation & dosimetric measurements

Irradiation was performed via raster-scanning delivery in the HIT experimental room. With the half-head phantom set-up described in the prior section, point measurements (absolute dose) were acquired with the beam impinging on a water-tank positioned with its entrance window at the isocenter and the half-head Alderson phantom attached to the tank's surface (Fig. 1). An array of twenty-four PinPoint ionization chambers (0.03 cm^3 sensitive volume, 2.9 mm diameter, TM31015, PTW) was submerged in the tank with the configuration details reported in the validation work [29]. A detailed description and schematic of the IC apparatus is also available in a previous report which performed a comprehensive integration and validation of the geometry in FLUKA Monte Carlo [30]. The procedure and experimental set-up used throughout the study has been previously validated and is used for patient plan verification at various hadrontherapy centers (HIT, CNAO, MIT, etc.) [29–31]. The beam quality correction factor (k_{Q,Q_0}) was applied for the four ions following clinical procedure with protons and carbon ions, calculated according to IAEA TRS-398 [32].

MCTP-optimized plans were delivered and ionization chamber readings were recorded at various depths near and along the central axis in the target region and dose fall-off gradient. During the data collection stage, the ion chamber block was translated, depth-wise, to acquire measurements with 3 mm resolution in the proximal/mid-target and 1 mm resolution at the distal end/fall-off. For each ion, 192 individual measurement points were collected in total.

Additionally, lateral measurements were acquired using the OCTAVIUS® 1000SRS P (PTW-Freiburg, Freiburg, Germany), a prototype 2D ionization chamber array detector differing from its predecessor by replacing liquid-filled chambers with air. The prototype's effective depth was measured ($d_{eff} \cong 5$ mm) and detector uniformity (within acceptable tolerances i.e. $\sigma = \pm 0.5\%$) was validated internally for use during beam commissioning, as well as through preliminary works to evaluate the system as a convenient alternative to film [33]. Relative measurements were performed at the distal end of the target, behind the half-head Alderson phantom and 6.0 cm of PMMA, where the largest dose deviations due to heterogeneities are expected (Fig. 1). The measured water equivalent depth (WED) for 1 cm of PMMA insert was 1.168 cm.

2.3. Dosimetric evaluation

Initial dosimetric comparisons are performed for each calculation mode against FLUKA Monte Carlo [34,35] taken as the reference benchmark. Dose-volume histogram (DVH) metrics such as D_{95}, D_{50}, and D_5 quantified the dose at 95%, 50% and 5% of the target volume, respectively. All external dose calculations and ion chamber positioning vectors from the water tank system were loaded into FRoG's QA-mode for dose point extraction to directly compare predictions with measurements [36]. For 1D dosimetry, absolute percent dose deviation $|\%\Delta|$ was determined by the absolute difference between an extracted dose prediction and corresponding physical measurement for each data point collected, with a standard deviation provided for the mean $|\%\Delta|$.

Subsequently, 3D-γ analysis was performed for dose maps from the three dose engines against corresponding 2D OCTAVIUS® measurements as the reference [37]. Global passing rates were computed using a 3%/3-mm dose difference and distance-to-agreement (DTA) criterium for the 5 cm × 5 cm central region of the detector. Range uncertainties can be due to misalignment/positioning of head phantom between planning CT acquisition and treatment delivery, variations in planned to delivered pencil beam spot position/size and inadequate range prediction using conventional Hounsfield unit (HU) to water-equivalent path length (WEPL) conversions. In order to consider these uncertainties, the following procedure was adopted: a standard deviation (σ_γ) for each 3D-γ passing rate is provided in the text corresponding to ± 2 mm range uncertainty in the experimental set-up and computation [38]. These σ_γ values were calculated by performing 3D-γ analysis for dose prediction maps with range shifts of −2 mm and + 2 mm (in addition to the central slice i.e. 0 mm shift) against the reference OCTAVIUS® measurement.

3. Results

Dose maps and DVH plots are depicted in Fig. 2 with corresponding metrics provided in Table 1, detailing the percent deviation of the analytical engines with FLUKA MC as the reference. In general, results are consistent with the validation work for MCTP and FRoG [20,21,36] showing target dose predictions in agreement with measurements by ~1% for protons and ~1–3% for the heavier ions, with FRoG overestimating FLUKA by ~1–2%. Results from the initial dosimetric study are shown in Fig. 3, displaying measurements and line profiles extracted from the three calculation modes after a highly heterogeneous region.

For protons, global $|\%\Delta|$ values in the target structure were

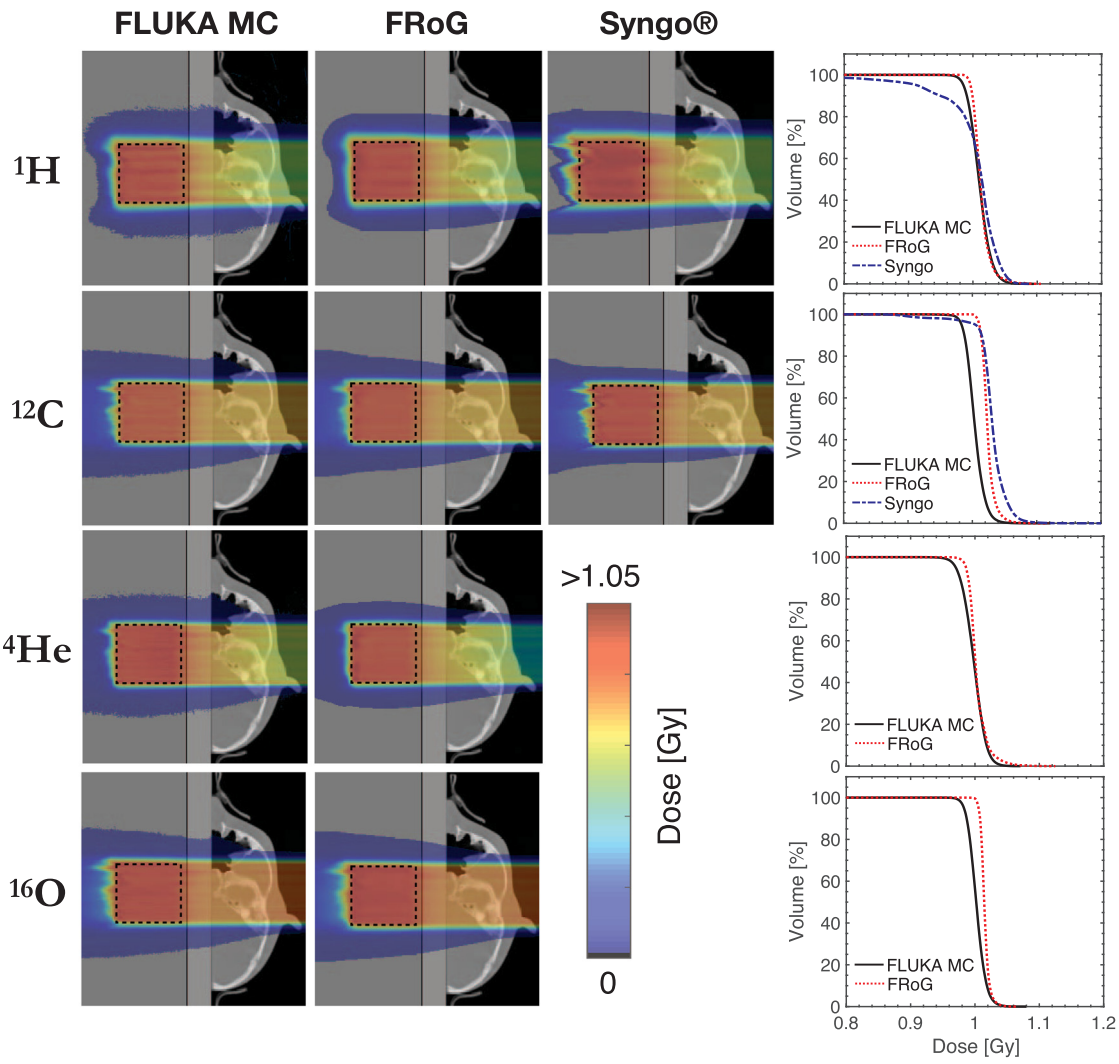

Fig. 2. Dose maps and DVHs of the forward calculated MCTP-optimized plans (^{1}H, ^{4}He, ^{12}C, and ^{16}O) are shown for the following dose engines: FLUKA, FRoG and, Syngo (when applicable).

Table 1
DVH metrics.

DVH Metrics—FRoG & Syngo vs. FLUKA							
Particle	Engine	D95%	Δ%	D50%	Δ%	D5%	Δ%
^{1}H	FLUKA	0.985	–	1.009	–	1.037	–
	FRoG	0.996	1.1	1.012	0.23	1.036	−0.07
	Syngo	0.913	−7.2	1.014	0.46	1.049	1.20
^{4}He	FLUKA	0.970	–	0.998	–	1.024	–
	FRoG	0.999	2.9	1.014	1.64	1.044	2.03
^{12}C	FLUKA	0.983	–	1.002	–	1.027	–
	FRoG	1.011	2.7	1.023	2.05	1.043	1.59
	Syngo	1.009	2.6	1.035	3.23	1.068	4.07
^{16}O	FLUKA	0.982	–	1.002	–	1.024	–
	FRoG	1.005	2.3	1.014	1.20	1.027	0.32

1.16(± 0.84)%, 2.23(± 1.48)% and 3.76(± 1.73)% for FLUKA, FRoG and Syngo. Measurements acquired in and beyond the high dose gradient (distal fall-off) were not considered in calculating global metrics. In the studied profiles and measurements, however, the range where distal dose fall-off reaches 80% of the 1 Gy planned dose (R_{80}) for the Syngo calculation varied by up to ± 5 mm compared to FLUKA.

For carbon ions, |%Δ| were 1.51(± 0.92)%, 1.47(± 1.51)%, 3.23(± 2.24)% for FLUKA, FRoG and Syngo, respectively. For the experimental beams, |%Δ| values for FLUKA and FRoG, respectively, were 1.50(± 0.49)% and 1.21(± 1.14)% for helium ions, and 1.91(± 0.70)% and 2.06(± 1.49)% for oxygen ions. In addition, region-specific |%Δ| are provided in Table 2, listing variations at mid-SOBP (first 50 mm in the target) and the distal (final 10 mm in the target) for the four ions, with percent deviation in dose ($\%\Delta_{Gy}$) between the three dose calculation modes and reference dosimetric measurements depicted in Fig. 4. In general, the highest dose deviations occurred at the distal end of the target as opposed to proximal- and mid-SOBP.

Regarding global |%Δ|, FLUKA obtained superior agreement with measurement for both clinical beams. FRoG demonstrated excellent agreement, especially in mid-SOBP positions. The commercial system Syngo yields net values within the 5% tolerance window of clinical acceptability for TPS benchmarking [29]. Despite these findings, further inspection reveals a spatial dependency along the target. Greater levels of disagreement occur near the distal edge compared to the entrance and mid-SOBP positions.

Analysis of data collected with the OCTAVIUS® detector positioned at the distal end of the target are presented in Fig. 5, depicting a 2D surface map of the raw measurements and corresponding slices extracted from the various dose engine calculations. Considering the four ion-beams of this study, FLUKA presented the highest mean 3D-γ passing rates against the measurements acquired by the OCTAVIUS® system, with 97.0(± 2.4)% pixels passing, on average. Structure in

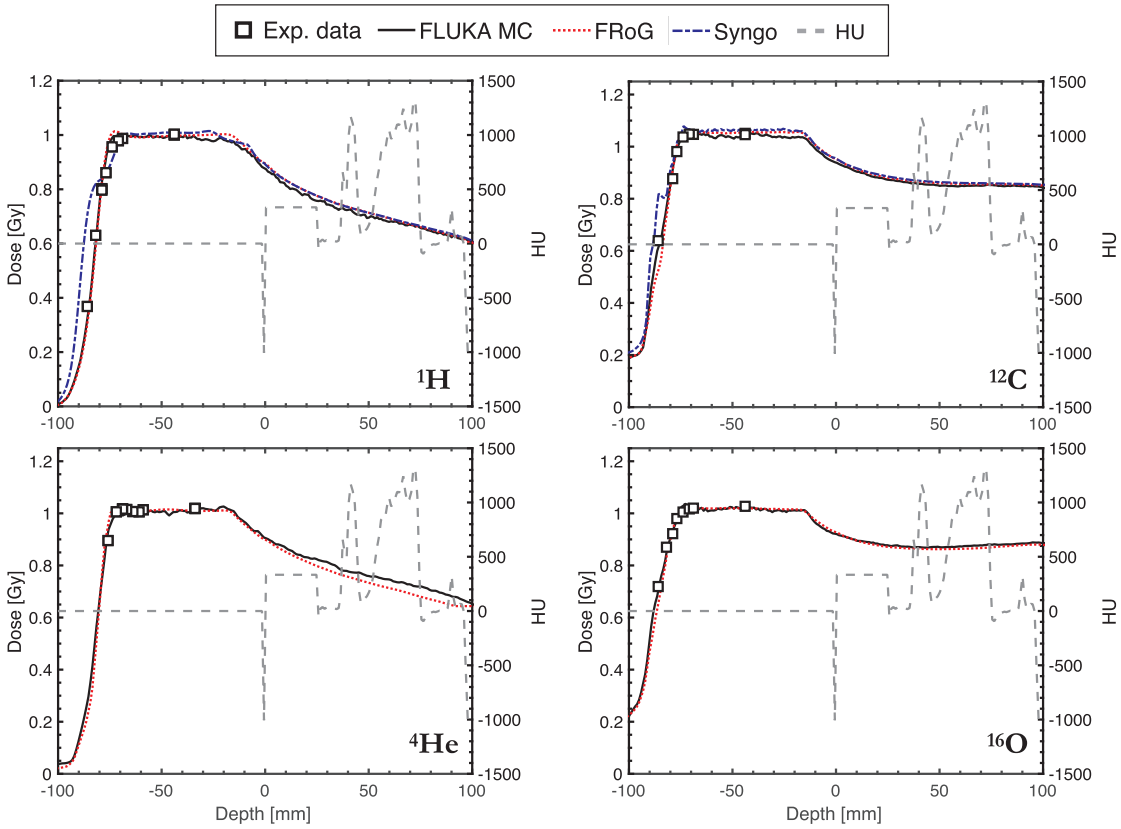

Fig. 3. Depth-dose measurements using the twenty-four PinPoint ionization chamber apparatus submerged in the water tank versus dose calculation with various engines, when applicable, for the four ions (^{1}H, ^{4}He, ^{12}C, ^{16}O): FLUKA, FRoG and Syngo (when applicable). Corresponding HU profiles are provided, demonstrating great anatomic variability.

Table 2
Mean absolute percent deviation in dose (|%Δ|) and standard deviation (± σ) for mid-SOBP, distal end, and whole target (total) for PinPoint measurements against various calculation modes.

PinPoint Measurements vs. Monte Carlo vs. Analytical Algorithms							
Particle	Type	Mid-SOBP		Distal end		Total	
		|%Δ|	± σ	|%Δ|	± σ	|%Δ|	± σ
^{1}H	FLUKA	1.01	0.94	1.58	0.43	1.16	0.84
	FRoG	1.73	1.07	3.33	0.73	2.23	1.48
	Syngo	3.32	1.12	4.70	2.31	3.76	1.73
^{4}He	FLUKA	1.58	0.97	1.96	1.02	1.50	0.49
	FRoG	0.70	1.45	2.02	1.25	1.21	1.14
^{12}C	FLUKA	1.34	0.56	1.76	1.33	1.51	0.92
	FRoG	1.25	1.31	1.64	1.58	1.47	1.51
	Syngo	2.74	1.27	4.60	2.89	3.23	2.24
^{16}O	FLUKA	1.95	0.82	1.33	0.88	1.91	0.70
	FRoG	1.79	1.32	2.48	1.46	2.06	1.49

dose homogeneity (i.e. slight fluctuations in target uniformity observed at the distal edge of the target) are visible in the FLUKA dose map. The mean 3D-γ passing rate for FRoG was 93.6(± 4.2)% averaged over the four ions.

As for 3D-γ analysis using 3%/3-mm dose-difference/DTA criterium, individual passing rates are presented with ± σ, considering ± 2 mm range uncertainties. For proton irradiations, passing rates were 98.7(± 0.2)%, 93.6(± 0.4)% and 70.7(± 7.0)% for FLUKA, FRoG and Syngo. For irradiation with carbon ions, γ-passing rates were 98.9(± 0.6)%, 99.1(± 0.1)% and 91.8(± 2.1)% for FLUKA, FRoG and Syngo, respectively. The experimental beams presented the following 3D-γ passing rates for FLUKA and FRoG, respectively: 96.3(± 0.4)% and 92.5(± 1.9)% for helium ions, and 93.9(± 0.4)% and 89.1(± 0.6)% for oxygen ions.

4. Discussion

Using multi-dimensional measurement strategies, we evaluated particle therapy dose calculation performance for the four ion-beams available at HIT in heterogenous setting. In this study, accuracy of clinical and research-based systems, both analytical algorithms and Monte Carlo codes, are addressed in a challenging clinical scenario (i.e. deep-seated field after severe anatomic heterogeneity without beam angle optimization). Aside from providing benchmarks for the clinically applied particle beams (i.e. protons and carbon ions) in patient-like geometries, the study provides preliminary data using novel ion-beams for future clinical programs: (1) the upcoming translation of raster-scanning helium ion-beams and (2) investigating advanced treatment planning and delivery techniques such as intensity modulated composite particle therapy (IMPACT) [39]. Both motivations first require systematic characterization of full Monte Carlo codes and fast analytical algorithms in clinical-like settings for each ion.

Visual inspection of the DVH curves in Fig. 2 reveals noticeable discrepancy between the reference FLUKA Monte Carlo and analytical codes. Generally, analytical engines for the heavier ions have difficulty reproducing Monte Carlo results due to the complexity of nuclear interactions and imperfect characterization of the particle beam as a superposition of Gaussians. In summary, the analytical algorithms overestimate, with FRoG exhibiting agreement in D_{50} within ~1.28(± 0.78)% averaged over the four ions, in line with prior findings [21,36]. In regards to quality of the proton plan calculation using Syngo, the mean deviation in D_{50} from FLUKA was ~1.85%. For carbon ions, deviations between FLUKA and Syngo of ~3% for D_{50} are

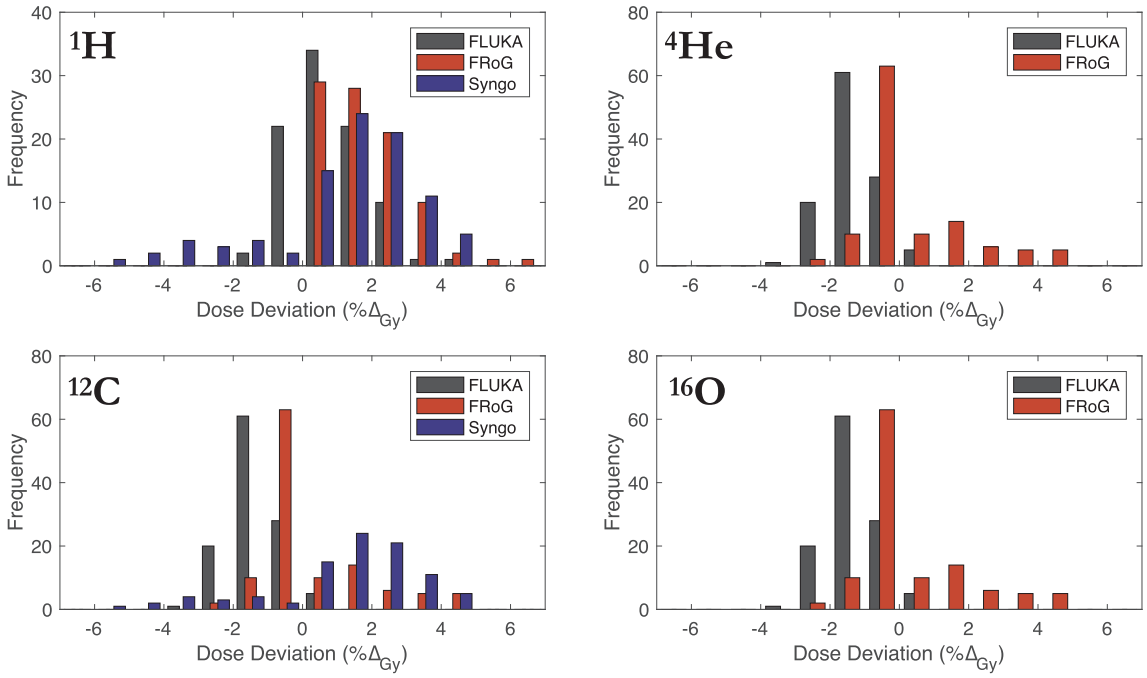

Fig. 4. Percent dose deviations for the four ions (^{1}H, ^{4}He, ^{12}C, and ^{16}O) compared to reference measurements for the following dose engines: FLUKA, FRoG and, Syngo (when applicable).

consistent with works which assessed the clinical TPS against the internal Monte Carlo framework [28]. For all four ions, agreement between FLUKA and measurements are comparable with the MCTP validation study, showing roughly $\lesssim$2% deviations on average [20]. Due to recently acquired cross-section measurements for helium ion-beams within the clinical energy range 70–220 MeV/u) and their incorporation into the FLUKA 2018 development version, improvements in reproducing physical measurements were observed, decreasing from nearly 5% to $\lesssim$1% for individual helium ion Bragg peaks [40].

When evaluating against experimental data, the commercial TPS presented larger deviations in dose prediction, particularly for the proton case, as demonstrated in |%Δ| from point measurements (Fig. 4 and Table 2) and 3D-γ passing rates (Fig. 5). To compensate for pencil beam deformation in conditions with severe lateral heterogeneity, Syngo makes use of the water equivalent depth at point of interest (WED-at-POI) approach, which can elicit notable deviations in range, as depicted dose map (Fig. 2) and line profiles (Fig. 3). In WED-at-POI, the radiological depth of each dose voxel is computed via raytracing and WED spread due to particle scattering is neglected. For protons, the variation in $R_{80\%}$, the distance where the SOBP dose at the distal edge drops to 80% of the planned value (1 Gy), between Syngo and FLUKA ranged between ± 5 mm within the target. For carbon ions, the disparity in range prediction was less pronounced, with a ± 2 mm fluctuation, on average, for Syngo against FLUKA, an anticipated result considering the beam characteristics of carbon ions (i.e. reduced spot-size and multiple Coulomb scattering (MCS) in medium compared to protons). These relatively minor uncertainties in range, however, form visible hot-spots in the distal region of the target, observed in the upper right panel of Fig. 3.

To best describe lateral dose evolution, reduce range errors in heterogeneous conditions and properly handle intra-beam spot widening for tangentially-oriented fields, FRoG uses a higher order approximation for the Gaussian model, an increased beam decomposition multiplicity during PB splitting, and the dual-PB model for explicit handling of the secondary beam profile due to large angle scattering in the nozzle [21]. In turn, FRoG exhibited satisfactory agreement with measurements, with |%Δ| ranging between ~1% mid-SOBP to a maximal ~3% at the distal edge, as reported in Table 2.

For the assessment of lateral distributions using the OCTAVIUS®, the target's distal end was of particular interest due to the large variations observed in prediction between the three dose engines. For distal field positions, the selected 3D-γ analysis criteria (3%/3-mm) are relatively stringent considering the associated uncertainties in particle therapy treatment planning in an anthropomorphic phantom. The only passing rate well above 95% for the proton case was for the full Monte Carlo simulation. Additional gamma tests were performed for the proton case to determine the limits using more loose criteria, i.e. 5%/3-mm and 7%/4-mm (IROC criteria), respectively. For protons, passing rates using a 5%/3-mm criteria were 99.1% and 96.4% for FLUKA and FRoG, while using a 7%/4-mm criteria, passing rates were 100% for both platforms. As for the commercial system, passing rates using a 5%/3-mm and 7%/4-mm criteria were 81.9% and 91.0%, respectively. Additionally, a baseline measurement was acquired as a reference for the proton plan irradiation by positioning the detector at mid-target (~3 cm depth), which yielded 3D-γ analysis passing rates under 3%/3-mm criteria of 100.0%, 98.6% and 76.2% for FLUKA, FRoG and Syngo. The commercial software in this study presented less than satisfactory agreement compared to measurements, especially at the distal end. Other works present significantly smaller deviations for commercially available dose engines against measurements using IBA systems (Ion Beam Applications, Louvain-Neuve, Belgium) alongside Monte Carlo based RaySearch TPS solutions [15]. Consequently, commissioning of the RayStation Monte Carlo TPS is currently underway for clinical use at HIT.

From a clinical perspective, the TPS beam model can meaningfully impact target coverage between planning and delivery in patients. As a proof of concept, an additional measurement was acquired with the OCTAVIUS® in a fixed-beam treatment room (as opposed to the experimental cave) by irradiating a proton plan optimized using the Syngo TPS in place of MCTP. 3D-γ analysis between measurements and Syngo in the mid-SOBP region yielded passing rates of 86.1% and 94.0% under a 3%/3-mm and 5%/3-mm criteria, a result within the specified clinical tolerances. Despite these improvements mid-target, hot spots on the order of 10% were measured at the distal end and beyond the target. These findings further demonstrate the impact of anatomic heterogeneities on dose prediction and the importance of

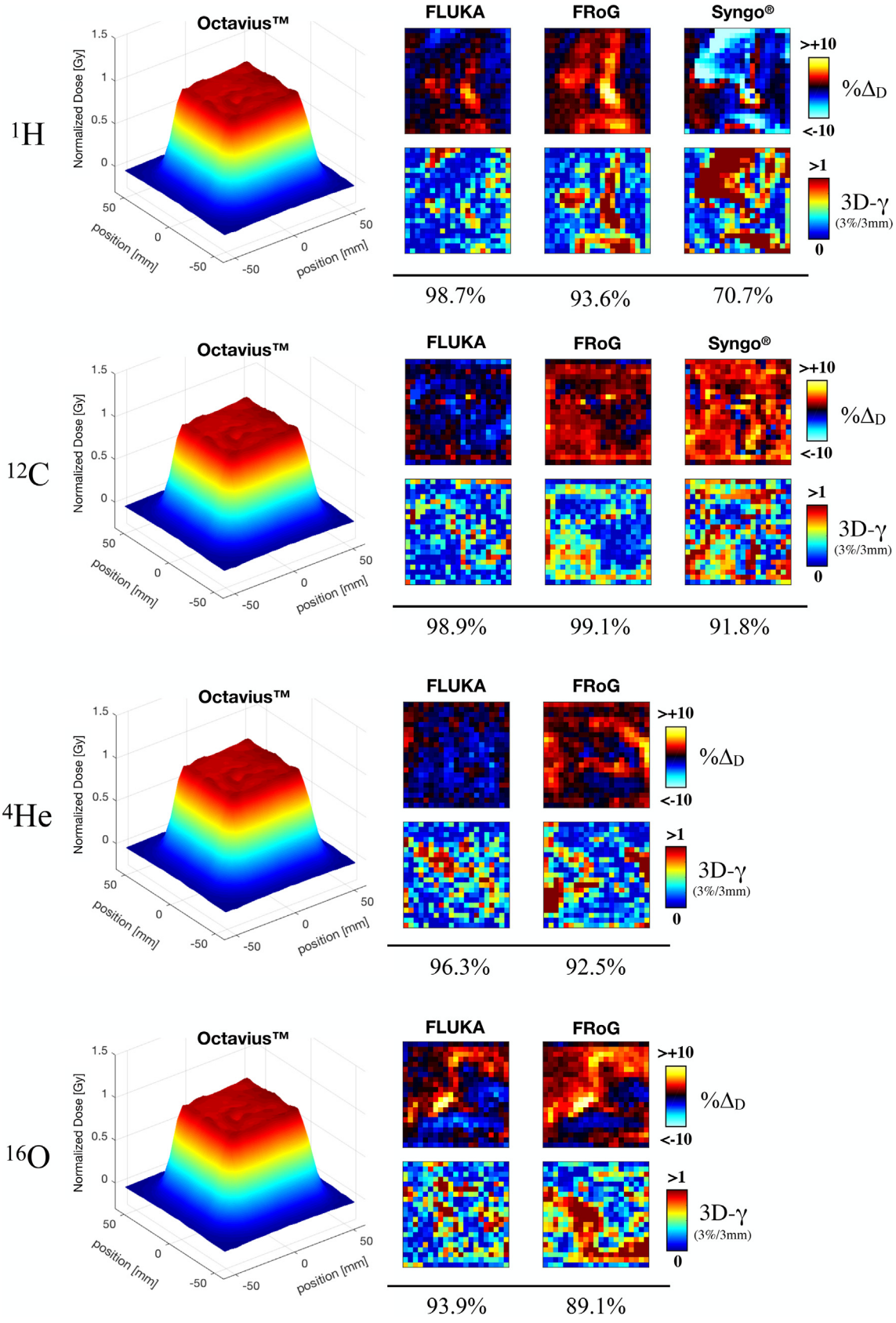

Fig. 5. Lateral dose distribution measurements acquired with the OCTAVIUS® 1000SRS P. Corresponding slices are displayed with percent dose-difference (%Δ_D) and 3D-γ maps, with passing rates listed below each column.

implementing advanced beam models during optimization and robust treatment planning. Albeit demonstrating excellent agreement with measurements, FLUKA and MCTP are non-viable for routine clinical use or large scale patient studies, a circumstance in which fast dose engines present numerous advantages [41]. Despite discrepancies identified in the target's distal end due to various limitations in analytical dose calculation, FRoG achieves superior agreement with measurements overall compared to the commercial TPS.

Regarding dose prediction performance by the analytical algorithms, overestimations were observed. This result was anticipated considering the simplified beam models assumption of lateral scattering effects in water, particularly in airgaps and low density tissue/bone interfaces (i.e. lung, nasal cavity, paranasal sinuses, etc.) where beam spread is not properly considered [42]. Consequently, hot spots in dose calculation for the analytical engines were generally found distally, along trajectories with distinct anatomical complexity. From a practical standpoint, the overestimation made by the commercial system cannot be solved via calibration or scaling since the physical distribution is severely altered by anatomic heterogeneities. Compared to our study, other groups present significantly smaller differences for commercially available analytical algorithms versus measurements in a brain phantom [43]. It is therefore worth noting that disparity in calculation and measurement is strongly dependent on the mode of calculation/optimization (Monte Carlo versus analytical), the extent of heterogeneity, depth/size of the target volume, field configuration and the particle beam-line design.

Depending on the degree of range and dose prediction uncertainty when using a particular TPS, an increased risk of toxicity or adverse effects, especially for skull-based diseases (e.g. hippocampal/limbic system avoidance), is of major concern. Since the analytical codes are considerably more practical in terms of runtimes compared to full Monte Carlo, they are often an attractive means for dose computation in clinical outcome studies involving large patient cohorts but may jeopardize the validity of the study. Further investigation using anthropomorphic set-ups is required to make more conclusive statements regarding the accuracy limits of the various dose engines available for particle therapy; however, considering the findings in this work and the related studies, warranted application of certain algorithms may be disease and/or location-dependent. Nevertheless, based on the level of agreement with Monte Carlo and measurements for the highly demanding tests in this study, dose engines such as FRoG which maximize PB splitting and employ triple Gaussian parametrization to describe lateral dose evolution in highly heterogeneous anatomy should be implemented for the upcoming helium ion-beam therapy programs. Future work will investigate calculation accuracy in various treatment sites (e.g. thorax and pelvic) for light and heavy ions. As for testing performance with beam-modifiers (e.g. range shifter), FRoG has demonstrated excellent agreement with both FLUKA Monte Carlo and physical measurements, well within 2% [36] and comparable to results in clinical studies on the impact of new commercial software, such as RayStation®, which use Monte Carlo codes for proton therapy treatment planning [44]. Similar agreement between FRoG and FLUKA with measurements is anticipated when testing in anthropomorphic H&N phantoms. Further assessment of dose engine performance in more computational demanding anatomical sites, e.g. thorax, is foreseen.

Regarding biological dose optimization, carbon ion-beam therapy requires treatment planning with biophysical/mechanistic models for relative biological effectiveness (RBE), e.g. local effect model (LEM) or microdosimetric kinetic model (MKM). Conversely, a constant RBE of 1.1 is assumed clinically for protons worldwide despite *in vitro* and *in vivo* evidence of enhanced cell-kill end-of-track [45]. For future study of RBE variability in patient treatment, biological dose prediction using FRoG was validated against FLUKA using the LEM-I and MKM and found to be within ~2% [36]. Moreover, fast and accurate tools like FRoG can be used for calculating LET_d and linking clinical outcome to various biophysical parameter for carbon ions. Effective dose calculation is also made possible in FRoG using LET-based phenomenological approaches to model variable RBE for proton therapy [46]. Future studies will investigate the impact of physical dose engines accuracy and RBE-weighted dose optimization on *in vitro* response in homogenous setting and anthropomorphic phantoms.

5. Conclusion

MCTP and FRoG were successfully validated in a complex dosimetric scenario for four ions in preparation for advanced applications in hadrontherapy, e.g. clinically informed particle selection and biologically optimal treatments using multi-ion modalities, and provide for the first time clinical benchmarks for PB scanning helium, carbon and oxygen ion-beams in an anthropomorphic set-up. These platforms were compared against a commercially available system, which presented clinically-relevant discrepancies particularly at end-of-range depths (distal target), demonstrating that limitations of clinical algorithms should be thoroughly examined through site-specific testing. Considering the limitations of WED-at-POI for protons and carbon ions presented in this work, we advise using next-generation analytical dose engines such as FRoG to reduce treatment planning uncertainties for centers introducing protons, heavy and/or novel ions to the clinic.

Declaration of Competing Interest

None declared.

Acknowledgements

The authors thank Dr. Silvia Molinelli at the National Centre of Oncological Hadrontherapy (CNAO, Pavia, Italy) for providing professional feedback during manuscript write-up.

References

[1] Durante M, Loeffler JS. Charged particles in radiation oncology. Nat Rev Clin Oncol 2010;7:37–43.
[2] Mohan R, Grosshans D. Proton therapy – present and future. Adv Drug Deliv Rev 2017.
[3] Ohno T. Particle radiotherapy with carbon ion beams. EPMA J 2013.
[4] Hong L, Goitein M, Bucciolini M, et al. A pencil beam algorithm for proton dose calculations. Phys Med Biol 1996;41:1305–30.
[5] Kanematsu N, Komori M, Yonai S, et al. Dynamic splitting of Gaussian pencil beams in heterogeneity-correction algorithms for radiotherapy with heavy charged particles. Phys Med Biol 2009;54:2015–27.
[6] Parodi K, Mairani A, Sommerer F. Monte Carlo-based parametrization of the lateral dose spread for clinical treatment planning of scanned proton and carbon ion beams. J Radiat Res 2013.
[7] Russo G, Attili A, Battistoni G, et al. A novel algorithm for the calculation of physical and biological irradiation quantities in scanned ion beam therapy: the beamlet superposition approach. Phys Med Biol 2016;61:183–214.
[8] 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.
[9] Schiavi A, Senzacqua M, Pioli S, et al. Fred: a GPU-accelerated fast-Monte Carlo code for rapid treatment plan recalculation in ion beam therapy. Phys Med Biol 2017;62:7482–504.
[10] Jia X, Schümann J, Paganetti H, et al. GPU-based fast Monte Carlo dose calculation for proton therapy. Phys Med Biol 2012;57:7783–97.
[11] Ezzell GA, Burmeister JWDN. TG-119 IMRT commissioning tests instructions for planning, measurement, and analysis. Med Phys 2009.
[12] 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.
[13] Albertini F, Casiraghi M, Lorentini S, et al. Experimental verification of IMPT treatment plans in an anthropomorphic phantom in the presence of delivery uncertainties. Phys Med Biol 2011;56:4415–31.
[14] Branco D, Taylor P, Zhang X, et al. An anthropomorphic head and neck quality assurance phantom for credentialing of intensity-modulated proton therapy. Int J Part Ther 2017;4:40–7.
[15] Saini J, Maes D, Egan A, et al. Dosimetric evaluation of a commercial proton spot scanning Monte-Carlo dose algorithm: comparisons against measurements and simulations. Phys Med Biol 2017;62:7659–81.
[16] Widesott L, Lorentini S, Fracchiolla F, et al. Improvements in pencil beam scanning

proton therapy dose calculation accuracy in brain tumor cases with a commercial Monte Carlo algorithm. Phys Med Biol 2018.
[17] Winterhalter C, Zepter S, Shim S, et al. Evaluation of the ray-casting analytical algorithm for pencil beam scanning proton therapy. Phys Med Biol 2019.
[18] Sorriaux J, Testa M, Paganetti H, et al. Experimental assessment of proton dose calculation accuracy in inhomogeneous media. Phys Med 2017;38:10–5.
[19] Haberer T, Debus J, Eickhoff H, et al. The heidelberg ion therapy center. Radiother Oncol 2004;73:186–90.
[20] Tessonnier T, Böhlen TT, Ceruti F, et al. Dosimetric verification in water of a Monte Carlo treatment planning tool for proton, helium, carbon and oxygen ion beams at the Heidelberg Ion Beam Therapy Center. Phys Med Biol 2017;62:6579–94.
[21] Mein S, Choi K, Kopp B, et al. Fast robust dose calculation on GPU for high-precision 1H, 4He, 12C and 16O ion therapy: the FRoG platform. Sci Rep 2018.
[22] Jäkel O, Ackermann B, Ecker S, et al. Methodology paper: a novel phantom setup for commissioning of scanned ion beam delivery and TPS. Radiat Oncol 2019.
[23] Böhlen TT, Bauer J, Dosanjh M, et al. A Monte Carlo-based treatment-planning tool for ion beam therapy. J Radiat Res 2013:54.
[24] Mairani A, Böhlen TT, Schiavi A, et al. A Monte Carlo-based treatment planning tool for proton therapy. Phys Med Biol 2013;58:2471–90.
[25] Parodi K, Mairani A, Brons S, et al. Monte Carlo simulations to support start-up and treatment planning of scanned proton and carbon ion therapy at a synchrotron-based facility. Phys Med Biol 2012;57:3759–84.
[26] Weber U, Kraft G. Design and construction of a ripple filter for a smoothed depth dose distribution in conformal particle therapy. Phys Med Biol 1999;44:2765–75.
[27] Wieser HP, Cisternas E, Wahl N, et al. Development of the open-source dose calculation and optimization toolkit matRad. Med Phys 2017;44:2556–68.
[28] Bauer J, Sommerer F, Mairani A, et al. Integration and evaluation of automated Monte Carlo simulations in the clinical practice of scanned proton and carbon ion beam therapy. Phys Med Biol 2014;59:4635–59.
[29] Karger CP, Jäkel O, Hartmann GH, et al. A system for three-dimensional dosimetric verification of treatment plans in intensity-modulated radiotherapy with heavy ions. Med Phys 1999.
[30] Lima TVM, Dosanjh M, Ferrari A, et al. Monte Carlo calculations supporting patient plan verification in proton therapy. Front Oncol 2016.
[31] Molinelli S, Mairani A, Mirandola A, 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;58:3837–47.
[32] IAEA TRS-398. Absorbed dose determination in external beam radiotheraphy: an international code of practice for dosimetry based on standards of absorbed dose to water. Iaea Trs-398 2006.
[33] Bauer D. Study of the Octavius ionization chamber array as a film replacement for clinical ion beam quality assurance (Thesis) Univ. Heidelb; 2018.
[34] Ferrari A, others. FLUKA: A multi-particle transport code (Program version 2005). CERN-2005-010. 2005.
[35] Böhlen TT, Cerutti F, Chin MPW, et al. The FLUKA Code: Developments and challenges for high energy and medical applications. Nucl Data Sheets 2014;120:211–4.
[36] Choi K, Mein S, Kopp B, et al. FRoG—a new calculation engine for clinical investigations with proton and carbon ion beams at CNAO. Cancers (Basel) 2018;10:395.
[37] Low DA, Harms WB, Mutic S, et al. A technique for the quantitative evaluation of dose distributions. Med Phys 1998.
[38] Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol 2012.
[39] Inaniwa T, Kanematsu N, Noda K, et al. Treatment planning of intensity modulated composite particle therapy with dose and linear energy transfer optimization. Phys Med Biol 2017.
[40] Horst F, Aricò G, Brinkmann K-T, et al. Measurement of 4He charge- and mass-changing cross sections on H, C, O, and Si targets in the energy range 70–220 MeV/u for radiation transport calculations in ion-beam therapy. Phys Rev C 2019;99:014603.
[41] Huang S, Kang M, Souris K, et al. Validation and clinical implementation of an accurate Monte Carlo code for pencil beam scanning proton therapy. J Appl Clin Med Phys 2018.
[42] Ciangaru G, Polf JC, Bues M, et al. Benchmarking analytical calculations of proton doses in heterogeneous matter. Med Phys 2005.
[43] Taylor PA, Kry SF, Alvarez P, et al. Results from the imaging and radiation oncology core Houston's anthropomorphic phantoms used for proton therapy clinical trial credentialing. Int J Radiat Oncol Biol Phys 2016;95:242–8.
[44] Molinelli S, Russo S, Maestri D, et al. Impact of TPS calculation algorithms on dose delivered to the patient in proton therapy treatments. Phys Med Biol 2019.
[45] Paganetti H, Blakely E, Carabe-Fernandez A, et al. Report of the AAPM TG-256 on the relative biological effectiveness of proton beams in radiation therapy. Med Phys 2019.
[46] Mairani A, Dokic I, Magro G, et al. A phenomenological relative biological effectiveness approach for proton therapy based on an improved description of the mixed radiation field. Phys Med Biol 2017;62:1378–95.