IOP Publishing Phys. Med. Biol. 65 (2020) 025006 (10pp) https://doi.org/10.1088/1361-6560/ab5e97

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
30 September 2019

REVISED
20 November 2019

ACCEPTED FOR PUBLICATION
4 December 2019

PUBLISHED
16 January 2020

PAPER

Validation of linear energy transfer computed in a Monte Carlo dose engine of a commercial treatment planning system

Dirk Wagenaar[1,6], Linh T Tran[2], Arturs Meijers[1], Gabriel Guterres Marmitt[1], Kevin Souris[3], David Bolst[2], Benjamin James[2], Giordano Biasi[2], Marco Povoli[4], Angela Kok[4], Erik Traneus[5], Marc-Jan van Goethem[1], Johannes A Langendijk[1], Anatoly B Rosenfeld[2] and Stefan Both[1]

[1] Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[2] Centre for Medical Radiation Physics, University of Wollongong, Wollongong, Australia
[3] Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[4] Sintef, Trondheim, Norway
[5] Raysearch Laboratories, Stockholm, Sweden
[6] Author to whom any correspondence should be addressed.

E-mail: ir.d.wagenaar@gmail.com

Keywords: microdosimetry, proton therapy, linear energy transfer (LET), relative biological effectiveness (RBE)

Supplementary material for this article is available online

Abstract

The relative biological effectiveness (RBE) of protons is highly variable and difficult to quantify. However, RBE is related to the local ionization density, which can be related to the physical measurable dose weighted linear energy transfer (LET_D). The aim of this study was to validate the LET_D calculations for proton therapy beams implemented in a commercially available treatment planning system (TPS) using microdosimetry measurements and independent LET_D calculations (Open-MCsquare (MCS)).

The TPS (RayStation v6R) was used to generate treatment plans on the CIRS-731-HN anthropomorphic phantom for three anatomical sites (brain, nasopharynx, neck) for a spherical target ($Ø = 5$ cm) with uniform target dose to calculate the LET_D distribution. Measurements were performed at the University Medical Center Groningen proton therapy center (Proteus Plus, IBA) using a μ^+-probe utilizing silicon on insulator microdosimeters capable of detecting lineal energies as low as 0.15 keV μm^{-1} in tissue. Dose averaged mean lineal energy $\overline{\gamma_D}$ depth-profiles were measured for 70 and 130 MeV spots in water and for the three treatment plans in water and an anthropomorphic phantom. The $\overline{\gamma_D}$ measurements were compared to the LET_D calculated in the TPS and MCS independent dose calculation engine. $D \cdot \overline{\gamma_D}$ was compared to $D \cdot LET_D$ in terms of a gamma-index with a distance-to-agreement criteria of 2 mm and increasing dose difference criteria to determine the criteria for which a 90% pass rate was accomplished.

Measurements of $D \cdot \overline{\gamma_D}$ were in good agreement with the $D \cdot LET_D$ calculated in the TPS and MCS. The 90% passing rate threshold was reached at different $D \cdot LET_D$ difference criteria for single spots (TPS: 1% MCS: 1%), treatment plans in water (TPS: 3% MCS: 6%) and treatment plans in an anthropomorphic phantom (TPS: 6% MCS: 1%).

We conclude that $D \cdot LET_D$ calculations accuracy in the RayStation TPS and open MCSquare are within 6%, and sufficient for clinical $D \cdot LET_D$ evaluation and optimization. These findings remove an important obstacle in the road towards clinical implementation of $D \cdot LET_D$ evaluation and optimization of proton therapy treatment plans.

Novelty and significance

The dose weighed linear energy transfer (LET_D) distribution can be calculated for proton therapy treatment plans by Monte Carlo dose engines. The relative biological effectiveness (RBE) of protons is known to vary with the LET_D distribution. Therefore, there exists a need for accurate calculation of

clinical LET_D distributions. Previous LET_D validations have focused on general purpose Monte Carlo dose engines which are typically not used clinically. We present the first validation of mean lineal energy $\overline{\gamma_D}$ measurements of the LET_D against calculations by the Monte Carlo dose engines of the Raystation treatment planning system and open MCSquare.

Introduction

Treatment planning systems for proton therapy typically optimize the physical dose distribution given to the patient. It is generally agreed that protons have a higher biological effectiveness than x-rays which is expressed in the use of a relative biological effectiveness (RBE) factor of 1.1 (Paganetti *et al* 2002) in clinical practice independent on their energy. However, the RBE of protons is actually highly variable and difficult to quantify. Various attempts have been made to effectively model the varying RBE of clinical proton beams (Wedenberg *et al* 2013, Paganetti 2014, McNamara *et al* 2015). Many of these models use the relation between RBE and the local ionization density, which can be related to the physical measurable dose weighted average linear energy transfer (LET_D) (Paganetti 2014, Paganetti *et al* 2019). Comparison studies between different RBE models showed that large differences remain among these models but all models deviate considerably from a constant RBE of 1.1 used in clinical practice (Giovannini *et al* 2016, McMahon *et al* 2018, Rørvik *et al* 2018). The simplest implementation to reflect RBE variability is to use an RBE which increases linearly with the LET_D (Paganetti 2014, Unkelbach *et al* 2016).

Monte Carlo dose calculations have been introduced into some clinical treatment planning systems(TPS) to improve dose calculation accuracy (Paganetti *et al* 2008, Langner *et al* 2018). Monte Carlo dose calculations can in principle be used to simultaneously score dose and LET_D distributions (Giovannini *et al* 2016). Recently, such an implementation has been made in the RayStation (RaySearch Laboratories AB, Stockholm, Sweden) commercial TPS. However, the accuracy of these calculations is currently unknown and therefore ultimately may hinder their clinical use for RBE calculations. Different implementations of integrating LET_D optimization into treatment planning have been published, making it even more important to validate the accuracy of LET_D calculations (Unkelbach *et al* 2016, Cao *et al* 2017, Bertolet *et al* 2019, Traneus and Ödén 2019).

Recently developed solid-state microdosimeters are capable of measuring the microdosimetrical equivalent of LET_D, the mean lineal energy ($\overline{\gamma_D}$) with high spatial resolution (Tran *et al* 2015, Rosenfeld 2016). These detectors have been used to evaluate dedicated LET_D Monte Carlo calculations in different particle beams (Anderson *et al* 2017, Chartier *et al* 2017, Tran *et al* 2017). However, dedicated Monte Carlo engines often do not have a clinical beam model and are often incapable of rapid treatment plan evaluation on a clinical CT, limiting their routine use in predicting clinical LET_D distributions. A recent Monte Carlo dose calculation engine called OpenMCsquare (MCS), used for independent dose calculations, can also score LET_D distributions from treatment plans and clinical CTs (Souris *et al* 2016, Sorriaux *et al* 2017, Huang *et al* 2018).

We aimed to validate the LET_D calculations for proton therapy beams implemented in a commercially available TPS (Raysearch v6R) using microdosimetry measurements and MCS independent LET_D calculations.

Methods

Treatment planning

Various treatment plans were generated in a research version of RayStation v6R TPS (RaySearch Laboratories AB, Stockholm, Sweden) using pencil beam scanning (PBS). First, two plans were created consisting of a single pristine Bragg peak of 70 MeV and of 130 MeV in a homogeneous water phantom. Second, three PBS treatment plans were created to treat a spherical target with a 5 cm diameter with uniform dose coverage, in the CIRS 731-HN anthropomorphic head and neck phantom (CIRS, Norfolk, United States of America)

RayStation (v6R) was used to generate treatment plans in the CIRS-731-HN anthropomorphic phantom for three treatment sites (brain, nasopharynx, neck) with a spherical target ($\varnothing = 5$ cm) with uniform physical target dose. One lateral half of the anthropomorphic phantom was used, the other half being replaced with solid water (figure 1). The number of energy layers was reduced to speed up beam delivery so higher measurement accuracy could be achieved. The brain, nasopharynx and neck plans consisted of 1373, 1388 and 1475 spots with a monitor weighted average energy of 122.8 MeV, 118.6 MeV and 108.5 MeV and minimum–maximum energy of 99.5–138.0 MeV, 83.6–138.5 MeV and 70.0–127.6 MeV, respectively.

Dose and LET_D calculations

The dose and LET_D distributions of both the two single spots and of the three treatment plans were calculated in a water phantom, using the TPS which is able to score the LET_D distribution using its Monte Carlo dose engine. In

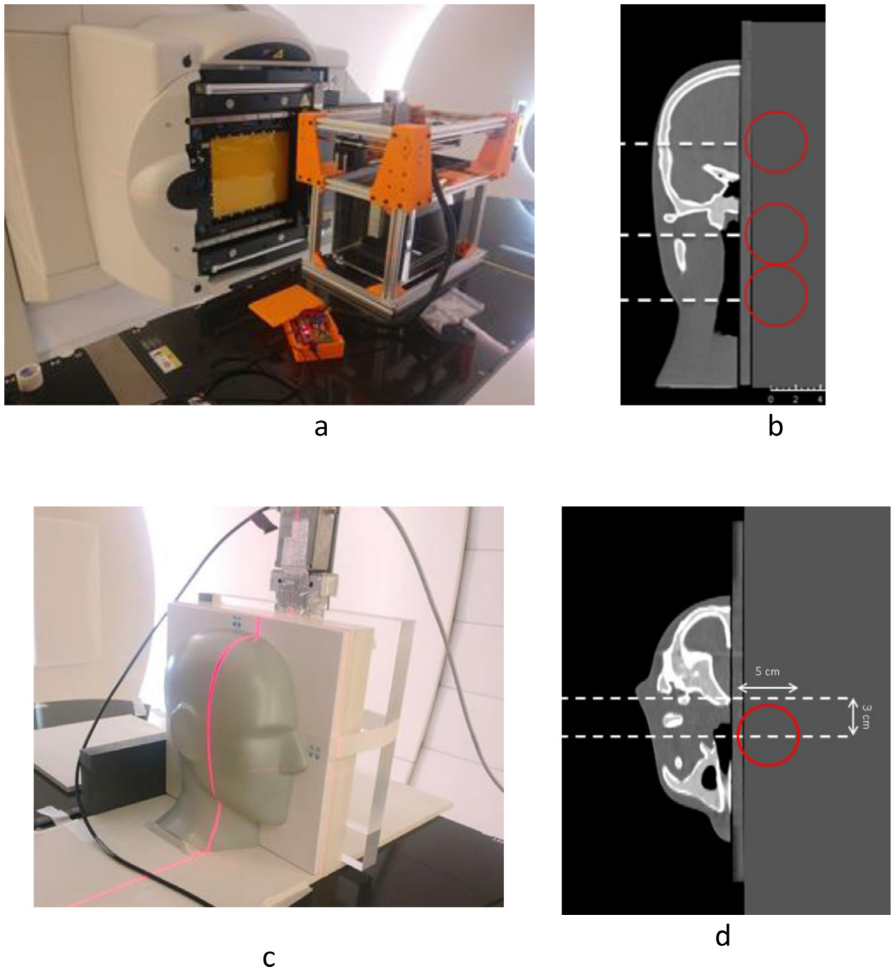

Figure 1. Measurement setup. (a) Experimental setup for the water tank measurements. The gantry was rotated to 270° so the detector could most easily be moved in the beam direction. (b) A coronal projection of CT scan of the CIRS 731-HN phantom. The shown red circles indicate the target areas used in treatment planning to create three treatment plans (i.e. brain, nasopharynx, neck). The left half of phantom was replaced with solid water. (c) Experimental setup for the measurements in the CIRS 731-HN phantom. The detector was placed behind the right half of the phantom. Solid water was added between the phantom and the detector in between measurements to acquire a depth profile. Positioning was verified with cone beam CT scans. (d) Profiles measured for each plan. To reduce the event rate and consequently detector pile-up, measurements were done at a 3 cm offset from the central axis for measurements in and before the spread out Bragg peak.

addition, the dose and LET_D distributions of the three treatment plans were calculated in a high resolution, high exposure (i.e. high mAs) CT scan of the anthropomorphic phantom with 0.5 mm isotropic voxels. The Monte Carlo calculation in the TPS was set to achieve 0.2% statistical accuracy for the physical dose. The statistical accuracy was applied to each beam and defined as the mean standard error of all voxels at least 50% of the maximum dose. The dose voxels were cubic with an edge of 0.1 mm for the single spots, 0.2 mm for the treatment plans in water and 0.5 mm for the treatment plans in the anthropomorphic phantom.

The dose and LET_D distributions were calculated independently on the same dose grid in the MCS Monte Carlo dose engine resulting in two complete sets of calculations (Souris *et al* 2016). The MCS dose engine was independently developed from the RayStation Monte Carlo code and was shown to give good agreement in terms of dose 3%/3 mm γ pass rates in a previous comparison study with pass rates of 94.7% and 97.5% for MCS and the TPS respectively, although this was for older versions of the dose engines (Sorriaux *et al* 2017). Nevertheless, such a comparison has not been made in terms of a $D \cdot LET_D$ calculation prior to this work. The MCS system was compiled in single precision and ran on a system with 24 CPU cores. Calculations were performed with $1 \cdot 10^9$ particle simulations in 10 batches.

In the MCS LET_D calculation, the contribution of secondary photons and neutrons are neglected. Secondary electrons are taken into account and assumed to be locally absorbed. For secondary protons, deuterons and alpha's the trajectory is calculated and their contribution to the LET_D is scored. Even heavier particles (i.e. heavy nuclear fragments) are considered in the recoil energy and thus locally absorbed, but not taken into account in the LET_D calculation.In our initial analysis, we observed a 10%–15% difference between the LET_D calculated in the TPS and MCS for a single 70 MeV spot. After investigating the difference an error in the MCS LET_D scoring

of secondary particles was found. MCS presents two methods (i.e. 'stopping power' and 'deposited energy') of scoring the LET_D which can be specified in a settings file. The 'stopping power' method (previously described by Cortés-Giraldo as method 'C') scores a dose weighted average of the stopping power of particles through each voxel (Cortés-Giraldo and Carabe 2015). The 'deposited energy' method (previously described by Cortés-Giraldo as method 'A') scores a dose weighted average of the energy loss divided by length traveled of particles through each voxel (Cortés-Giraldo and Carabe 2015). For the calculations presented in this study the 'deposited energy' method was used as the 'stopping power' method was found to score no LET_D of secondary particles.

The ion transport in the TPS applies a special 'track end stepper' to score response functions that vary rapidly over a voxel which typically occurs near end of range where ionization and kinetic energy varies significantly over single voxel. In the case of LET_D, the scoring proceeds as follows. If an ion track enters a voxel at an energy above a certain energy threshold the stopping power is sampled halfway between the entrance and exit points as part of the regular ion transport. If the energy is below this threshold the dedicated 'track end stepper' mode is used to sample the stopping power. The ion is then transported in a sequence of maximum 90 logarithmic energy loss steps from the entrance energy down to 20 KeV/u while sampling the stopping power at each step. During this transport the ion may leave the voxel and continue in the neighboring voxel. The transport neglects nuclear absorption and multiple scattering and takes place in water at the local voxel mass density. The track end stepper requires the voxel size to be 3 mm or less and the threshold energy is set to 16 MeV/u. Note that the track end stepper is only used for scoring response functions and do not modify the regular dose scoring in any way.

The TPS transports primary protons and secondary protons, deuterons and alphas. Energy transferred to other charged secondary particles (e.g. tritons) is estimated and considered to be absorbed locally. The trajectory of secondary protons is calculated identically to primary protons. Deuterons and alphas are transported without considering nuclear absorption.

Measurement setup and beam delivery

All treatment plans were delivered at the University Medical Center Groningen proton therapy center (Proteus Plus, IBA). Measurements were performed using a μ^+probe utilizing silicon on insulator microdosimeters capable of detecting lineal energies as low as 0.15 keV μm^{-1} (Tran *et al* 2015, 2017, Rosenfeld 2016). The detection rate was managed to avoid detector pile-up while also getting high statistics within a reasonable time period by changing the cyclotron current between 1–50 nA and changing between microdosimeters with different sensitive volume sizes. A single sensitive volume with 30 μm diameter was used in the entrance region and along the Bragg peak positions and three sensitive volumes were used in the distal region of the Bragg peak (figure 2).

The silicon on insulator (SOI) microdosimeter with low noise readout electronics (μ^+-probe) was specifically design for microdosimetry where a strict requirement is to measure energy deposited by charged particles event-by-event in well-defined sensitive volumes (SV) similar to biological cells. The cylindrical geometry of the 3D SVs provide a chord distribution close to that of a sphere which is typically used for cell modelling. Due to the customized requirements for 3D SV of silicon microdosimetry, other pixelated detectors known to authors are not applicable for microdosimetry (Bolst *et al* 2017). The LET_D is a theoretical concept and in reality not measurable. However in protons of not very high energy (less than 100 MeV) secondary electrons can be assumed to deposit their energy within the 3D sensitive volume (SV) of the microdosimeter. Additionally, the mean chord length can be replaced by the mean path length which is approximately the thickness of the 3D SV(Anderson *et al* 2017). Under these conditions the $\overline{\gamma_D}$ and LET_D are similar and their difference decreases for protons with lower energies as is the case near and at the distal edge of the Bragg peak.

The single spots and treatment plans were measured in a water phantom (figure 1(a)). Although we were using the microdosimeter with the smallest available sensitive volume and cyclotron current at our disposal as indicated above, pile up effects were still observed in the spread out Bragg peak for the treatment plans. Therefore, part of the treatment plans were measured at a 3 cm lateral offset (figure 1(d)). The measurement positions were chosen along the beam central axis and were spaced far apart where small dose and LET_D differences are expected but closer in high LET_D gradient areas, most notably at the distal edge of the Bragg peak.

In addition, the brain, nasopharynx and neck treatment plans were measured for a beam shot through the anthropomorphic phantom (figure 1(c)). In this setup, the amount of solid water plates placed between the anthropomorphic phantom and the detector was increased between measurements. The anthropomorphic phantom and solid water plates were fixed to the table and their positions with respect to the beam were verified at three different time points using on board CBCT imaging.

Analysis

The analysis was performed by comparing the $D \cdot LET_D$ to the $D \cdot \overline{\gamma_D}$ rather than comparing the LET_D and $\overline{\gamma_D}$ directly. The goal of validating the LET_D calculations is that they may be used to effectively calculate the RBE weighed dose defined as $D \cdot RBE$ where D is the physical dose. In our analysis we aim to assess the uncertainty of the additional biological dose defined as $(1 - RBE) \cdot D$. Similarly to Unkelbach *et al*. we used an approximation

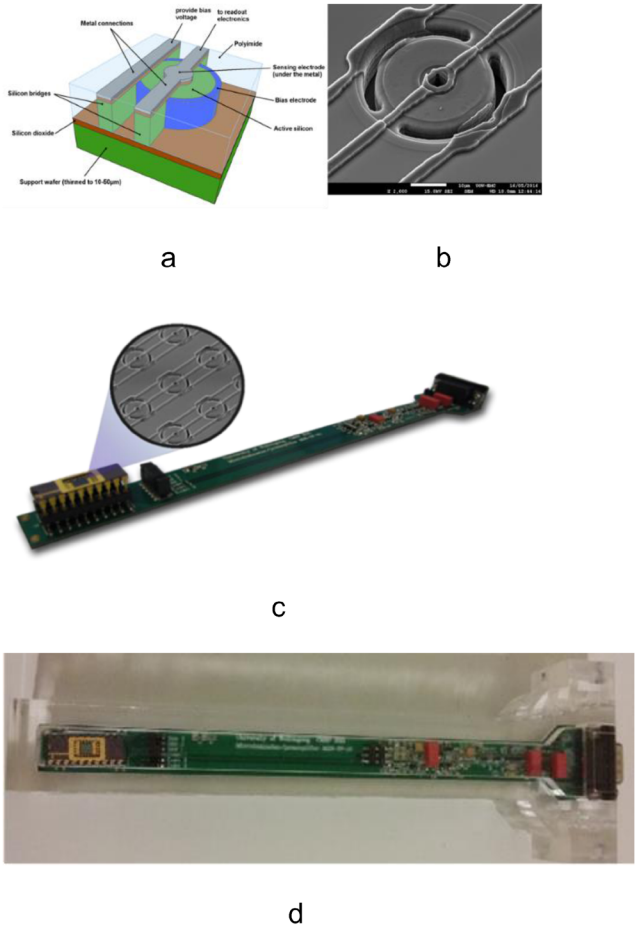

Figure 2. Schematics of used detectors. (a) Schematic of a single sensitive volume with diameter of 30 μm and (b) scanning electron microscope image of a single volume of the mushroom microdosimeter (c) schematic of a μ^+ probe connected to the microdosimeter chip. (d) A waterproof Polymethyl Methacrylate (PMMA) sheath containing the μ^+ microdosimetric probe. The readout electronics (right) are placed 10 cm away from the detector (left) to avoid radiation damage.

for the RBE with a linear dependency on LET_D, namely $RBE = 1.0 + 0.04 \cdot LET_D$ (Unkelbach *et al* 2016). Using this equation, the quantity $0.04 \cdot D \cdot LET_D$ can be used as a surrogate for the additional biological dose (Unkelbach *et al* 2016). Analyzing the additional biological dose instead of comparing LET_D and $\overline{\gamma_D}$ directly ensures that high values occur in clinically relevant areas of the Bragg peak and shifts focus away from regions with negligible dose where an increase in RBE is clinically irrelevant.

The $\overline{\gamma_D}$ was calculated from the measured microdosimetric spectrum with the μ^+-probe and converted from silicon to tissue at all positions using a conversion method as described in a previous Monte Carlo simulation study (Bolst *et al* 2017). Next, the $\overline{\gamma_D}$ was multiplied with the dose calculated by the TPS and the dose calculated with MCS resulting in two sets of $D \cdot \overline{\gamma_D}$ for all measurement points which can be compared to the $D \cdot LET_D$ of the TPS and MCS respectively.

A 3D gamma analysis with interpolation was performed to find the lowest dose uncertainty which would result in a 90% passing rate (Low *et al* 1998). The μ^+-probe positioning in the water phantom was within 1 mm accuracy as in an earlier experiment (Tran *et al* 2017). During the measurements in the anthropomorphic phantom, the on board CBCT imaging found an isocentric position uncertainty of 1 mm. We chose a distance to agreement setting of 2 mm for all measurements. This value was chosen to be slightly larger than the found uncertainties as it is the combination of positional and computational inaccuracies. The gamma passing rates were calculated for each measurement set using increasing global $D \cdot LET_D$ difference setting to find the lowest global $D \cdot LET_D$ difference setting with a 90% passing rate.

The criterion for acceptable $D \cdot LET_D$ differences was determined based on what would be clinically acceptable. No pre-existing criterion for the accuracy of LET_D or $D \cdot LET_D$ is currently available. Historically, the International Commission on Radiation Units and Measurements suggested a total treatment uncertainty of up to 5% on the prescribed dose can be deemed acceptable (ICRU 1978). A common criteria for the accuracy of the dose calculation is currently within 3%/3 mm (Both *et al* 2007, Ezzell *et al* 2009, Mans *et al* 2015). Before the LET_D can be used for RBE weighed dose calculations, its uncertainty has to be small enough to not negatively impact the total dose calculation uncertainty. Since LET_D has a much smaller impact on the RBE weighed dose distribution

than physical dose, its accuracy constraints should be less stringent. We chose the conservative criterion that the LET_D calculation inaccuracy should not impact the average RBE weighed dose to the target by more than 1%. For the planned cases in the anthropomorphic head and neck phantom, the additional biological dose to the targets was 14% of the physical dose distribution. An LET_D calculation uncertainty of 7% or less is required for the uncertainty in LET_D calculation to have result in an additional uncertainty of 1% or less on the RBE weighed dose distribution.

Impact of voxel size on LET_D calculation

TPS calculations for a single 70 MeV spot in water were done multiple times using a 0.1, 0.5 and 2.0 mm isotropic dose voxel sizes to investigate a potential voxel size dependency on the calculated $0.04 \cdot D \cdot LET_D$ (Cortés-Giraldo and Carabe 2015). These three $D \cdot LET_D$ calculations were compared to ascertain whether or not the LET_D calculations were independent of dose voxel size. Comparison was done in terms of the mean gamma analysis with a $D \cdot LET_D$ difference of 3% and distance to agreement of 3 mm (i.e. 3%/3 mm). The MCS algorithm was not tested for dose voxel size dependency as the MCS code scores dose and LET_D on the same grid as the CT image and resampled the result to match the dose grid of the TPS. In the TPS Monte Carlo transport code, dose and LET_D are scored directly in the dose grid defined be the user.

Results

In total, 28, 47 and 38 $\overline{\gamma_D}$ measurements were performed for single spots in water, treatment plans in water and treatment plans in an anthropomorphic phantom respectively. The comparison between the TPS and MCS calculations and measured microdosimetric parameters for a 70 MeV spot and the brain plan is shown in figure 3. The comparisons of all measured spots and treatment plans is shown in the online supplementary materials (figure S1(stacks.iop.org/PMB/65/025006/mmedia)).

The gamma passing rates for the TPS $D \cdot LET_D$ calculations compared to the $D \cdot \overline{\gamma_D}$ reached 90% for a $D \cdot LET_D$ difference criteria of 1%, 3% and 6% for single spots in water, treatment plans in water and treatment plans in the anthropomorphic phantom respectively (figure 4(a)). The TPS LET_D calculations for 2.0, 0.5 and 0.1 mm dose voxel sizes are shown in figure 5. The mean 3%/3 mm gamma values were 0.08 and 0.07 when comparing the calculations with 2.0 and 0.5 mm and 0.5 and 0.1 mm dose voxels respectively.

The gamma pass rates for the MCS $D \cdot LET_D$ calculations compared to the $D \cdot \overline{\gamma_D}$ reached 90% for a $D \cdot LET_D$ difference criteria of 1%, 6% and 1% for single spots in water, treatment plans in water and treatment plans in the anthropomorphic phantom respectively (figure 4(b)).

When comparing the TPS and MCSquare calculations in terms of a 3%/3 mm gamma analysis the mean gamma in the area receiving at least 10% of the maximum $D \cdot LET_D$ was 0.10 for the spots in water, 0.49 for treatment plans in water and 0.62 or treatment plans in the anthropomorphic phantom.

Discussion

The $D \cdot LET_D$ calculations agreed to the measured $D \cdot \overline{\gamma_D}$ within 6% for both the TPS and MCS. The differences between the algorithms were relatively small and there was no observed dependence of LET_D on dose voxel size for the TPS. The found differences should lead to an error of the RBE weighed target dose of less than 1%.

The found accuracies of within 6% are well within the postulated clinically acceptable level of 7%. The uncertainty could be considered large compared to the uncertainty of dose calculations which are typically within 2%. However, one should take into consideration that the LET_D calculation will be used to calculate only the additional biological dose which is less than the physical dose. Using the RBE approximation made by Unkelbach *et al*, the found uncertainty will lead to an uncertainty of the RBE weighed target dose of less than 1% (Unkelbach *et al* 2016). However this dependency will vary depending on the used RBE model (Carabe-Fernandez *et al* 2007, Wedenberg *et al* 2013, Paganetti 2014, McNamara *et al* 2015, Rørvik *et al* 2018). Furthermore, when considering a new modality larger uncertainties are typically accepted, for example for the introduction of intensity modulated radiation therapy (IMRT) in 2003, where a 7%/4 mm passing criteria was considered acceptable (Dong *et al* 2003, Both *et al* 2007).

In the measurements we included single spots of two energies. Including a larger variety of spot energies was our initial intention; however we were unable to acquire these measurements within the allotted time. This was in part due to technical challenges in acquiring a low enough dose rate for measurements points before the Bragg peak and due to the long measurement time for measurement points beyond the Bragg peak as a result of a low event rate. The same holds true for acquiring only depth profiles. For lateral profiles, the event rate varies greatly for each position which requires constant changing of detectors in the μ^+probe composed of varying numbers of sensitive volumes and the cyclotron current to avoid both detector pile-up and too long measurement times.

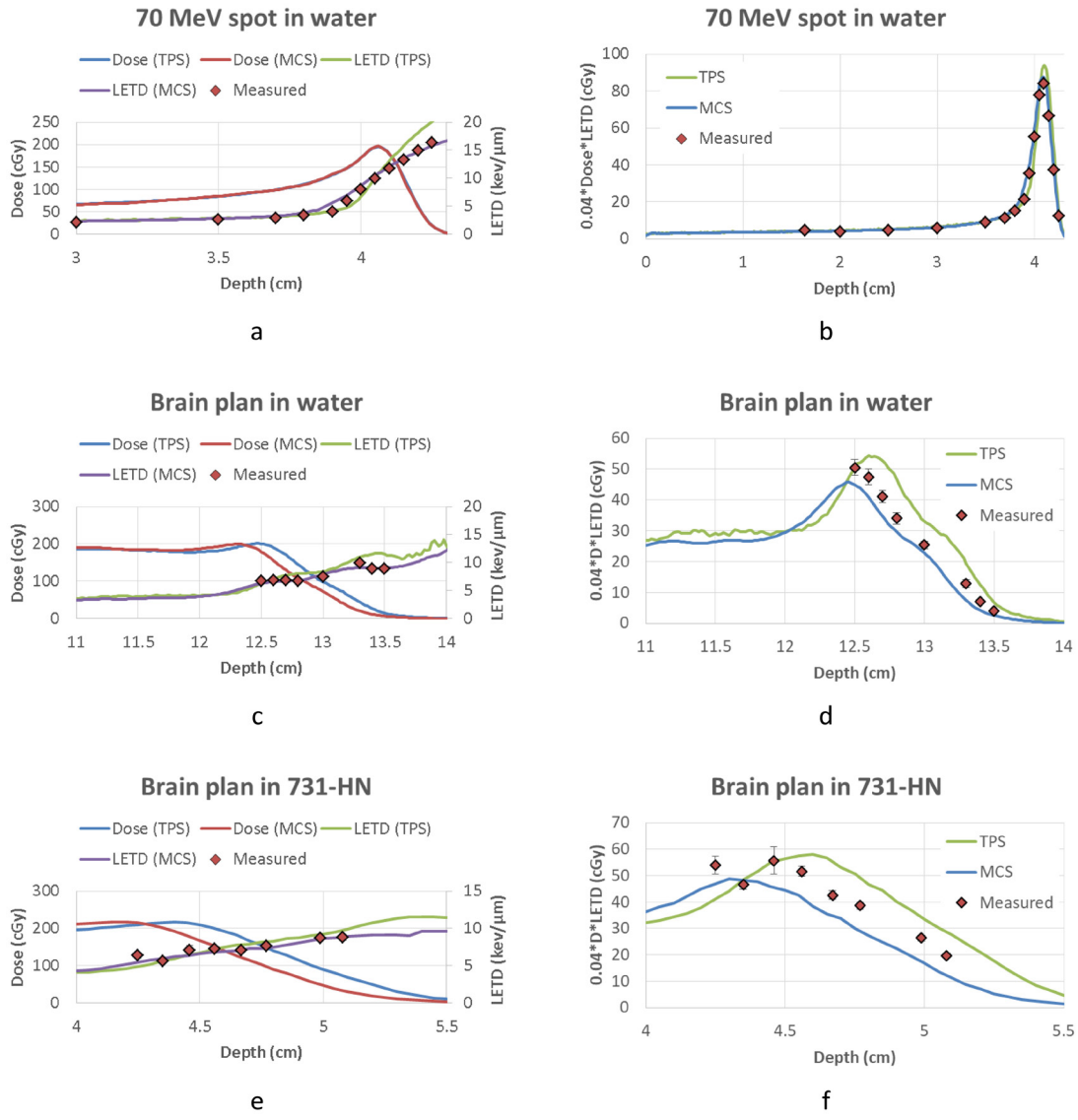

Figure 3. LET_D measurement results for a single spot and a brain treatment plan. RayStation TPS calculations, open MCSquare calculations and measurements for a 70 MeV spot in water (top), the brain treatment plan measured in water (middle) and the brain plan measured in the CIRS 731-HN phantom (bottom). For the treatment plan, only the central axis measurements are shown. Both the calculated dose and LET_D and measured $\overline{\gamma_D}$ (left) and $D \cdot LET_D$ and $D \cdot \overline{\gamma_D}$ (right) are shown. For the $D \cdot \overline{\gamma_D}$ calculation the average dose was taken from the MCS and TPS calculations. Due to a range difference in the dose calculation of Error bars represent the statistical standard error.

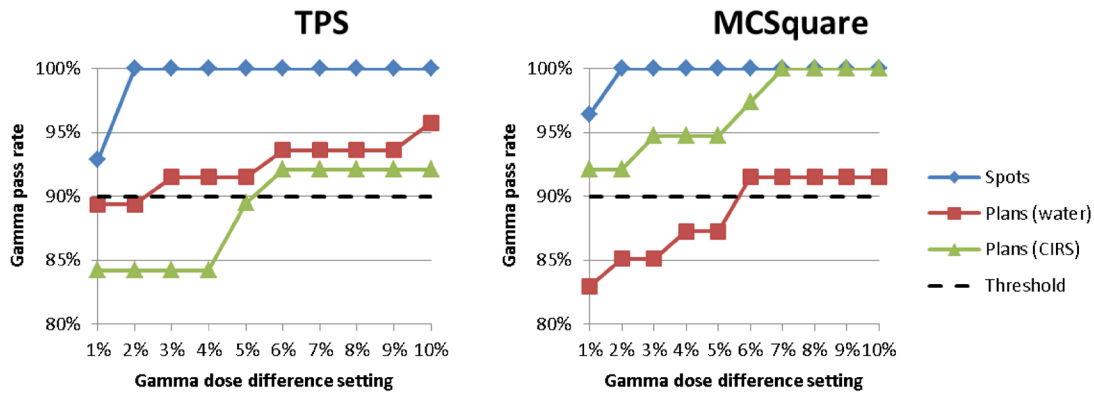

Figure 4. Gamma pass rates for different $LET_D \cdot$ Dose difference criteria.

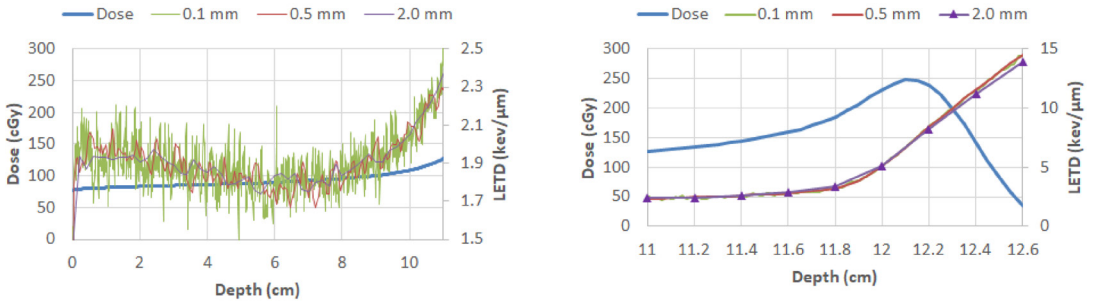

Figure 5. LET_D calculations in the RayStation TPS for different voxel sizes. The LET_D for various dose voxel sizes and dose for the entry region of the proton beam (left) and at the Bragg peak (right). Dose was calculated on a dose grid with 0.5 mm dose voxels.

The analysis in this study was done in terms a gamma analysis comparing $D \cdot LET_D$ to $D \cdot \overline{\gamma_D}$ for a set distance to agreement and varying $D \cdot LET_D$ difference settings. This type of analysis focusses on high dose regions which are most clinically relevant. The use of a gamma index was necessary due to the positional inaccuracy of the measurements and the variability of LET_D. Even for the water phantom measurements, small inaccuracies were likely to occur and the inaccuracy for the anthropomorphic phantom was found to be 1 mm as verified with the CBCTs acquired in between measurements.

Earlier measurements also showed good agreement of LET_D when it was compared to a GEANT4 Monte Carlo dose engine for single spots and single-energy square PBS fields (Anderson *et al* 2017, Tran *et al* 2017). To the knowledge of the authors, there is no earlier publication of a comparison between the measured $\overline{\gamma_D}$ and calculated LET_D by a TPS for a PBS treatment plan delivered in water and an anthropomorphic phantom. In our study, the agreement with calculations was better for single spots than for treatment plans. The TPS calculation showed better agreement than MCS to measurements for the treatment plans in water. For single spots, the LET_D seemed to rise more steeply with depth for the TPS calculations compared to the MCS calculations. For the treatment plans the $D \cdot LET_D$ was higher for the TPS calculations compared to the MCS calculations (figure S1). The worst results were found in the measurements of the neck case in the anthropomorphic phantom for the TPS and for the off-axis measurements of the brain case in the water phantom for MCS.

Overall, we found a similar $D \cdot LET_D$ difference criterion for the calculation of the TPS than for MCS. It is possible these results could be improved by putting more effort into beam modelling. In our clinical practice, the beam model in the TPS is used by the treatment plan optimization engine and MCS functions as an independent quality assurance tool for the treatment plan. Hence, more effort was placed in the tuning of the TPS beam model to agree with the measured physical dose. We expect that the performance of the open MCS $D \cdot LET_D$ calculation would improve if more time was invested into beam modelling based on physical dose. Ultimately, fitting the beam model to the measured $D \cdot \overline{\gamma_D}$ directly might further improve the results, but this is likely to reduce the agreement with calculated physical dose. Furthermore, for the purpose of treatment plan quality assurance this higher uncertainty may be found acceptable. At the time of writing a new MCS beam model was created aimed at showing better agreement for physical dose in the penumbra and halo region. Potentially this beam model could also improve the $D \cdot LET_D$ results.

In our study no relation was found between 2.0, 0.5 and 0.1 mm dose voxel size and calculated $D \cdot LET_D$ in the TPS. In a study evaluating different LET_D scoring methods in GEANT4, Cortes-Giraldo *et al* found a factor 1.8 difference in the plateau region of a 160 MeV beam when the dose voxel size was reduced from 2.0 to 0.2 mm (Cortés-Giraldo and Carabe 2015). The dependency of LET_D on voxel size was found to be different for different methods of scoring LET_D (Cortés-Giraldo and Carabe 2015). Even though the LET_D algorithm implemented in the Raystation TPS is not published, the results clearly indicate no voxel size dependency. As explained in the Methods section, the LET_D scoring method implemented in the TPS applies a special transport method. The applied LET_D scoring method is, by construction, not sensitive to the voxel boundaries and is thus not expected to be sensitive to voxel size as demonstrated in these results and will in fact also produce accurate results for voxels much smaller than the lower limit of 1 mm allowed in the clinical version of the TPS.

Good results were found for MCS only after adjusting the LET_D scoring method. The 10%–15% difference that was observed before adjusting the MCS was large enough to have a clinical impact but small enough that it might not be immediately noticed. This stresses the importance of quality assurance of Monte Carlo dose and LET_D calculations. While independent calculations can help detect errors in dose and LET_D calculations, measurements can provide a much more rigorous validation as they do not rely on assumptions of CT density conversion or beam models.

Several aspects of these measurements could be improved upon to better characterize the spot $D \cdot LET_D$. Additional measurements may be taken at spots with higher energies and include lateral $D \cdot LET_D$ profiles as

higher energies than 130 MeV are routinely used in clinic. These measurements should focus on regions where the $D \cdot LET_D$ is sufficiently high to have clinical impact which occurs close to the distal part of the spread out Bragg peak. While we did include measurements inside an anthropomorphic phantom in our study, measuring in biological tissue would give an even more realistic measurement setup. Additionally, comparing the spectrum of measured lineal energy against the calculated LET_D spectrum might give more insight about the correctness of the underlying physics assumptions of the dose calculation algorithms. Furthermore, LET_D measurements could be used to validate the effectiveness LET_D optimization.

This is the first report on the comparison of the measured $\overline{\gamma_D}$ to the LET_D calculation algorithms available in a commercially available TPS and MCS. As our understanding of the relation between RBE and LET_D and the evidence on the benefit of LET_D optimization increases, it becomes more relevant that LET_D calculations fit into the clinical workflow. If Monte Carlo dose calculations are already performed, LET_D calculations require no additional simulation. Therefore, integration into a clinical TPS and into an automated independent dose calculation quality assurance system is indispensable and warrants measurement based commissioning.

Conclusion

We conclude that $D \cdot LET_D$ calculations in the RayStation TPS (v6R) and Open-MCsquare are within 6% and are accurate enough for use in clinical RBE weighed dose evaluation and optimization. Healthcare providers should take care that the relation between LET_D and RBE still has important uncertainties, but LET_D can be accurately calculated in commercially available Monte Carlo dose algorithms. These findings remove an important obstacle in the road towards clinical implementation of $D \cdot LET_D$ evaluation and potentially LET_D based optimization of clinical proton therapy treatment plans.

ORCID iDs

Dirk Wagenaar https://orcid.org/0000-0001-8321-6824
Gabriel Guterres Marmitt https://orcid.org/0000-0002-8486-7001

References

Anderson S E *et al* 2017 Microdosimetric measurements of a clinical proton beam with micrometer-sized solid-state detector *Med. Phys.* **44** 6029–37

Bertolet A, Baratto-Roldán A, Cortés-Giraldo M A and Carabe-Fernandez A 2019 Segment-averaged LET concept and analytical calculation from microdosimetric quantities in proton radiation therapy *Med. Phys.* **46** 4204–14

Bolst D, Guatelli S, Tran L T, Chartier L, Lerch M L F, Matsufuji N and Rosenfeld A B 2017 Correction factors to convert microdosimetry measurements in silicon to tissue in ^{12}C ion therapy *Phys. Med. Biol.* **62** 2055–69

Both S, Alecu I M, Stan A R, Alecu M, Ciura A, Hansen J M and Alecu R 2007 A study to establish reasonable action limits for patient-specific quality assurance in intensity-modulated radiation therapy *J. Appl. Clin. Med. Phys.* **8** 1–8

Cao W, Khabazian A, Yepes P P, Lim G J, Poenisch F, Grosshans D R and Mohan R 2017 Linear energy transfer incorporated intensity modulated proton therapy optimization *Phys. Med. Biol.* **63** 015013

Carabe-Fernandez A, Dale R G and Jones B 2007 The incorporation of the concept of minimum RBE (RbEmin) into the linear-quadratic model and the potential for improved radiobiological analysis of high-LET treatments *Int. J. Radiat. Biol.* **83** 27–39

Chartier L *et al* 2017 Microdosimetric applications in proton and heavy ion therapy using silicon microdosimeters *Radiat. Prot. Dosimetry* **180** 365–71

Cortés-Giraldo M A and Carabe A 2015 A critical study of different Monte Carlo scoring methods of dose average linear-energy-transfer maps calculated in voxelized geometries irradiated with clinical proton beams *Phys. Med. Biol.* **60** 2645–69

Dong L, Antolak J, Salehpour M, Forster K, O'Neill L, Kendall R and Rosen I 2003 Patient-specific point dose measurement for IMRT monitor unit verification *Int. J. Radiat. Oncol. Biol. Phys.* **56** 867–77

Ezzell G A *et al* 2009 IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119 *Med. Phys.* **36** 5359–73

Giovannini G, Böhlen T, Cabal G, Bauer J, Tessonnier T, Frey K, Debus J, Mairani A and Parodi K 2016 Variable RBE in proton therapy: comparison of different model predictions and their influence on clinical-like scenarios *Radiat. Oncol.* **11** 68

Huang S *et al* 2018 Validation and application of a fast Monte Carlo algorithm for assessing the clinical impact of approximations in analytical dose calculations for pencil beam scanning proton therapy *Med. Phys.* **45** 5631–42

ICRU 1978 *ICRU 29 Dose Specification for Reporting External Beam Therapy with Photons and Electrons Report* (Bethesda, MD: International Commission on Radiation Units & Measurements)

Langner U W, Mundis M, Strauss D, Zhu M and Mossahebi S 2018 A comparison of two pencil beam scanning treatment planning systems for proton therapy *J. Appl. Clin. Med. Phys.* **19** 156–63

Low D A, Harms W B, Mutic S and Purdy J A 1998 A technique for the quantitative evaluation of dose distributions *Med. Phys.* **25** 656–61

Mans A, Schuring D, Arends M, Vugts L, Wolthaus J, Lotz H, Admiraal M, Louwe R, Öllers M and van de Kamer J 2015 Code of practice for the quality assurance and control for volumetric modulated arc therapy *NCS Report* 24 (Delft: Netherlands Commission on Radiation Dosimetry) (https://doi.org/10.25030/ncs-024)

McMahon S J, Paganetti H and Prise K M 2018 LET-weighted doses effectively reduce biological variability in proton radiotherapy planning *Phys. Med. Biol.* **63** 225009

McNamara A L, Schuemann J and Paganetti H 2015 A phenomenological relative biological effectiveness (RBE) model for proton therapy based on all published *in vitro* cell survival data *Phys. Med. Biol.* **60** 8399–416

Paganetti H 2014 Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer *Phys. Med. Biol.* **59** R419–72

Paganetti H *et al* 2019 Report of the AAPM TG-256 on the relative biological effectiveness of proton beams in radiation therapy *Med. Phys.* **46** e53–78

Paganetti H, Jiang H, Parodi K, Slopsema R and Engelsman M 2008 Clinical implementation of full Monte Carlo dose calculation in proton beam therapy *Phys. Med. Biol.* **53** 4825–53

Paganetti H, Niemierko A, Ancukiewicz M, Gerweck L E, Goitein M, Loeffler J S and Suit H D 2002 Relative biological effectiveness (RBE) values for proton beam therapy *Int. J. Radiat. Oncol. Biol. Phys.* **53** 407–21

Rørvik E, Fjæra L F, Dahle T J, Dale J E, Engeseth G M, Stokkevåg C H, Thörnqvist S and Ytre-Hauge K S 2018 Exploration and application of phenomenological RBE models for proton therapy *Phys. Med. Biol.* **63** 185013

Rosenfeld A B 2016 Novel detectors for silicon based microdosimetry, their concepts and applications *Nucl. Instrum. Methods Phys. Res.* A **809** 156–70

Sorriaux J, Testa M, Paganetti H, Orban de Xivry J, Lee J A, Traneus E, Souris K, Vynckier S and Sterpin E 2017 Experimental assessment of proton dose calculation accuracy in inhomogeneous media *Phys. Med.* **38** 10–5

Souris K, Lee J A and Sterpin E 2016 Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core CPU architectures *Med. Phys.* **43** 1700–12

Tran L T *et al* 2015 3D-Mesa 'bridge' silicon microdosimeter: Charge collection study and application to RBE studies in ^{12}C radiation therapy *IEEE Trans. Nucl. Sci.* **62** 504–11

Tran L T *et al* 2017 Characterization of proton pencil beam scanning and passive beam using a high spatial resolution solid-state microdosimeter *Med. Phys.* **44** 6085–95

Traneus E and Ödén J 2019 Introducing proton track-end objectives in intensity modulated proton therapy optimization to reduce linear energy transfer and relative biological effectiveness in critical structures *Int. J. Radiat. Oncol. Biol. Phys.* **103** 747–57

Unkelbach J, Botas P, Giantsoudi D, Gorissen B L and Paganetti H 2016 Reoptimization of intensity modulated proton therapy plans based on linear energy transfer *Int. J. Radiat. Oncol. Biol. Phys.* **96** 1097–106

Wedenberg M, Lind B K and Hårdemark B 2013 A model for the relative biological effectiveness of protons: the tissue specific parameter α/β of photons is a predictor for the sensitivity to LET changes *Acta Oncol.* **52** 580–8