Technical Note

Meijers, Arturs; Seller Oria, Carmen; Free, Jeffrey; Langendijk, Johannes A; Knopf, Antje C; Both, Stefan

Published in:
Medical Physics

DOI:
10.1002/mp.14713

Technical Note: First report on an in vivo range probing quality control procedure for scanned proton beam therapy in head and neck cancer patients

Arturs Meijers[a] Carmen Seller Oria, Jeffrey Free and Johannes A Langendijk
Department of Radiation Oncology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

Antje C Knopf
Department of Radiation Oncology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
Division for Medical Radiation Physics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Stefan Both
Department of Radiation Oncology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

(Received 6 October 2020; revised 23 December 2020; accepted for publication 5 January 2021; published xx xxxx xxxx)

Purpose: The capability of proton therapy to provide highly conformal dose distributions is impaired by range uncertainties. The aim of this work is to apply range probing (RP), a form of a proton radiography-based quality control (QC) procedure for range accuracy assessment in head and neck cancer (HNC) patients in a clinical setting.
Methods and Materials: This study included seven HNC patients. RP acquisition was performed using a multi-layer ionization chamber (MLIC). Per patient, two RP frames were acquired within the first two weeks of treatment, on days when a repeated CT scan was obtained. Per RP frame, integral depth dose (IDD) curves of 81 spots around the treatment isocenter were acquired. Range errors are determined as a discrepancy between calculated IDDs in the treatment planning system and measured residual ranges by the MLIC. Range errors are presented relative to the water equivalent path length of individual proton spots. In addition to reporting results for complete measurement frames, an analysis, excluding range error contributions due to anatomical changes, is presented.
Results: Discrepancies between measured and calculated ranges are smaller when performing RP calculations on the day-specific patient anatomy rather than the planning CT. The patient-specific range evaluation shows an agreement between calculated and measured ranges for spots in anatomically consistent areas within 3% (1.5 standard deviation).
Conclusions: The results of an RP-based QC procedure implemented in the clinical practice for HNC patients have been demonstrated. The agreement of measured and simulated proton ranges confirms the 3% uncertainty margin for robust optimization. Anatomical variations show a predominant effect on range accuracy, motivating efforts towards the implementation of adaptive radiotherapy. [https://doi.org/10.1002/mp.14713]

Key words: in vivo dosimetry, proton radiography, proton therapy, quality control

1. INTRODUCTION

Since the early investigations of proton therapy, the physical characteristics of protons have been regarded as promising for the reduction of integral dose to healthy tissues. Proton therapy can, therefore, offer more conformal treatments than conventional photon therapy.[1,2] Nevertheless, since the early adoption of proton therapy in clinical practice, its application has been hampered due to numerous sources of uncertainty, which can potentially severely degrade planned treatment dose distributions.[3–6]

In practice, a discrepancy between the actual range of a proton beam in the patient and the planned one may occur. In literature, this phenomenon is commonly referred to as range uncertainty. Computed tomography (CT) calibration, conversion of CT numbers to proton stopping power ratios (SPR), handling of lateral and longitudinal heterogeneities in the beam path, etc.[3] are referred to as major contributors to range uncertainty.

However, in clinical practice there are more factors that may impact proton range accuracy. Overall, these are (a) machine related, such as, reproducibility and stability of the equipment, (b) physics related, such as, transformation of CT numbers to mass density to SPR, (c) patient related, such as, anatomical and physiological variations, and (dv) biology related, linked to the end-of-range effect and relative biological effectiveness (RBE) uncertainty.[5] Nonetheless, only (a) and (b) are addressed by range uncertainty margin recipes proposed to account for range uncertainty.[3]

Technologically driven developments, such as, the clinical introduction of dual energy computed tomography (DECT)[7] or proton CT,[8,9] aim at eliminating or reducing the effect of some of the physics contributors to range uncertainty. The use of DECT promises to reduce the range uncertainty to about 2%,[10] as opposed to 3%–3.5%, which is often applied in proton clinics, when single energy computed tomography (SECT) is used.[3,11]

Proposed range uncertainty recipes are based on values found in literature (individually quantifying the extent of different possible sources of errors) and theoretical estimates.[3] Furthermore, attempts have been made to develop experimental techniques, which would allow to gain insight into range accuracy predictions in a near-clinical (for example, commissioning phase) or clinical setting. Techniques as proton radiography,[12–16] prompt gamma imaging[17] or positron emission tomography[18] have been investigated and applied for this purpose.

In our clinic we used proton radiography, more specifically range probing (RP),[13] to investigate range accuracy predictions of the treatment planning system (TPS) in near-clinical conditions (during the commissioning phase). A set of experiments was conducted to validate and optimize the CT calibration curve on animal tissue samples (bone and soft tissue).[19] Furthermore, range uncertainties in lung-like tissues were assessed using a porcine lung phantom.[20] As shown in these studies, RP allowed to support the choice of an applied range uncertainty recipe for robust plan optimization in clinical practice.

The RP acquisition method has been introduced into clinical practice and made available for patient-specific range accuracy checks as a part of an in vivo quality control (QC) procedure. This is the first report on the results of Pencil Beam Scanning RP QC after the clinical implementation for head and neck cancer (HNC) patients.

2. MATERIALS AND METHODS

The RP[13] technique, which has been adopted in our clinical practice is based on the use of a multi-layer ionization chamber (MLIC) to measure residual integral depth dose curves (IDDs) distally from an object-of-interest or patient. While there are several groups investigating MLIC-based range probing measurements,[13] the method applied in this work has been proposed and investigated by Farace et al.[21] and makes use of the commercially available MLIC Giraffe (IBA Dosimetry, Schwarzenbruck, DE) detector. A MLIC Giraffe has 180 parallel plane ionization chambers. The electrode diameter of each chamber is 12 cm. The electrodes have 2 mm spacing and the detector provides submillimeter range measurement accuracy for pristine peaks according to the manufacturer's documentation. This allows to measure high energy (relatively small size) spots with a deflection of up to ± 2 cm from the isocenter. The MLIC Giraffe is used in "movie" acquisition mode with a sampling time of 10 ms. The impact of measurement conditions (such as, field size, fluence, detector positioning, etc.) on the measurement accuracy has been assessed in a previous study.[21] For the RP procedure measurement conditions are set such that the accuracy of the detector compared to a baseline as provided by the manufacturer is not deteriorated.

The introduction of the QC procedure in the operational protocol as part of routine clinical practice has been approved by the board of department. On patient-specific basis the procedure is prescribed by the decision of attending MD. All devices, used to perform the procedure, are medical devices and are used as per intent of the device.

The implementation of an in vivo RP procedure for use in clinical routine imposed several implications on the clinical workflow, as described in the subsections below.

2.A. RP in treatment planning

A dedicated treatment field with a gantry angle of 90° is incorporated in the clinical TPS treatment plan. Currently the choice of gantry angle is limited to the lateral orientation (90° or 270°) due to constraints linked to MLIC Giraffe positioning. Positioning the gantry at 90 degrees allows easier access to the patient with the measurement equipment in our site-specific conditions. The field consists of 81 spots, covering a 4×4 cm^2 area around the treatment isocenter. The lowest allowed monitor units (MU) are assigned per spot in order to maintain the delivered dose during the Quality Control (QC) procedure as low as possible. An RP field (81 spots) delivers approximately 1 cGy$_{\mathrm{RBE}}$ of dose per QC procedure. All spots are assigned an energy of 210 MeV, which results in the full width at half maximum in air at the isocenter of 8.2 mm at our facility. In our clinical practice, the treatment isocenter for HNC patients is in the proximity of C3 or C4 vertebrae. Since RP spots are centered around the isocenter, this allows to intersect a broad mixture of tissues (bones, various muscles, fat tissue, nodes and, in some cases, tumor) during the QC procedure. As an example, Fig. 1 shows the setup of the RP field for one of the patients. In addition, Krah *et al* found that the proton radiography accuracy does not vary with its location relative to the treatment volume,[22] although, in the context of adaptive therapy, it might be beneficial to perform RP check through the regions traversed by treatment beams.

2.B. RP in treatment scheduling

After transferring the treatment plan to the Oncology Information System (OIS), the RP field is scheduled in the treatment calendar only on selected days. In the current implementation, the RP QC procedure is performed twice during the treatment course, or more if deemed necessary due to observed previous results or changes as shown by imaging data.

To ensure the availability of an up-to-date CT image of the patient and to support the interpretation of the RP data, the RP procedure is performed on the day of a repeat CT acquisition. The repeat CTs are acquired within 20 min before or after the treatment fraction, with the patient immobilized as in the treatment position.

2.C. RP setup and acquisition

On the day of the scheduled RP QC procedure, prior to the patient entering the treatment room (TR), a gain calibration of the MLIC is performed. The calibration procedure requires delivery of a high-energy proton spot in air. Therefore, the patient should not be present in the room during the

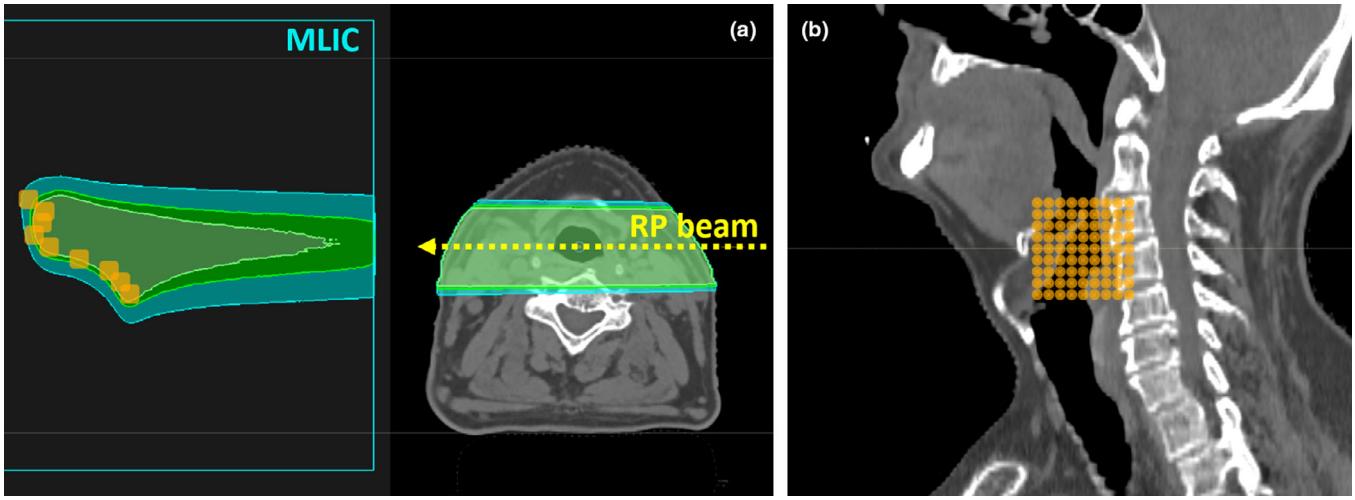

FIG. 1. Visualization of a range probing (RP) field for an example patient geometry (Patient 2). (a) The dose distribution of the RP field is shown from a transversal view of the patient, where the RP field is directed from the patient's left to the right (as from a gantry angle of 90°). The MLIC is represented by a blue box contour at the right side of the patient, and the range of penetration of each RP spot is indicated by orange rectangle. The distance between the patient and the MLIC is not shown at a scale. The integral dose of the whole RP beam as introduced in the planning system is shown in the image, while measurement analysis is performed on spot-by-spot basis. (b) Sagittal plane of the patient, in which the orange circles represent the RP spots.

calibration. The chamber gain is determined by normalizing the calibration measurement per chamber to the reference IDD measurement of the same energy spot acquired in a water tank with a large diameter parallel plane chamber. For the given application, no energy or beam intensity-dependent calibration is performed. The procedure requires approximately 5 min. After the calibration is performed, the patient may enter the TR. The patient is positioned at the TR isocenter using cone-beam computed tomography (CBCT). Afterwards, the gantry is moved to the 90 degrees position and the MLIC is positioned next to the patient on the gantry rolling floor opposite to the nozzle, using a dedicated trolley. The MLIC is aligned to the isocenter along the beam axis using in-room lasers. The distance between isocenter and entrance window of the MLIC is measured and recorded, as it is required for RP simulations in the TPS. Positioning of the MLIC does not require high precision due to the RP field size versus size of the MLIC electrode. Positioning errors along the beam axis will only have minimal impact due to the low density of air. When the alignment of the device is complete, the RP field delivery and acquisition can be performed. Delivery of the RP field does not require any nonstandard modifications to the beam delivery system. After the measurement, the trolley with the MLIC is removed from the gantry rolling floor and is left in the TR until the treatment fraction is complete. The setup of the equipment in the treatment room during RP acquisition is demonstrated in Fig. 2.

2.D. RP simulations in the TPS

In the TPS, a dedicated structure with an override to a material, which has the physical density and elemental composition of water, is added distally to the patient in the beam path to simulate the MLIC. Each spot of the RP field is calculated individually using the clinical TPS Monte Carlo engine on an isotropic 1 mm dose grid with an accuracy of 0.5%. The obtained dose distributions per spot are integrated along the beam axis. Automation of calculations and data extraction was realized using the scripting capabilities of the TPS.

Rigid registration between planning and repeat CTs is required to perform RP simulations. In accordance to the standard operational protocol, during the planning phase, a verification mask is defined on the planning CT. It is a rectangular structure encompassing the target volume, which is used as a region-of-interest (ROI) to perform automatic image registration during patient alignment in the treatment room. The same ROI was also used to register planning and repeat CTs.

2.E. RP data analysis

IDD curves simulated in the TPS are compared to the IDDs as measured with the MLIC. For each spot, the range error was obtained as the optimal offset between measured and simulated IDDs along the depth axis, as obtained by means of the least squares method[19] (also see Fig. 3). Negative / positive range errors correspond to a simulated IDD with a shorter / longer range with respect to the range as obtained from the measured IDD. Eventually, range errors are presented relative to the water equivalent path length (WEPL) of the proton spot passing through the patient. The WEPL for all spots is determined based on the measurements (as a difference between the depth of maximum dose for RP IDD and the maximum depth of a measurement in air for the same energy).

The analysis and comparison of the measured and calculated data sets (exemplary curves shown in Fig. 3) was

FIG. 2. Equipment setup in the treatment room for proton range probing acquisition. The gantry angle is set to 90° and the MLIC is positioned on the trolley on the right side of the patient. After removal of the trolley from proximity of the patient, the planned treatment fields are delivered as usual. The RP QC procedure extends the treatment fraction time by about 5 min, of which approximately 15 s is the time of RP delivery.

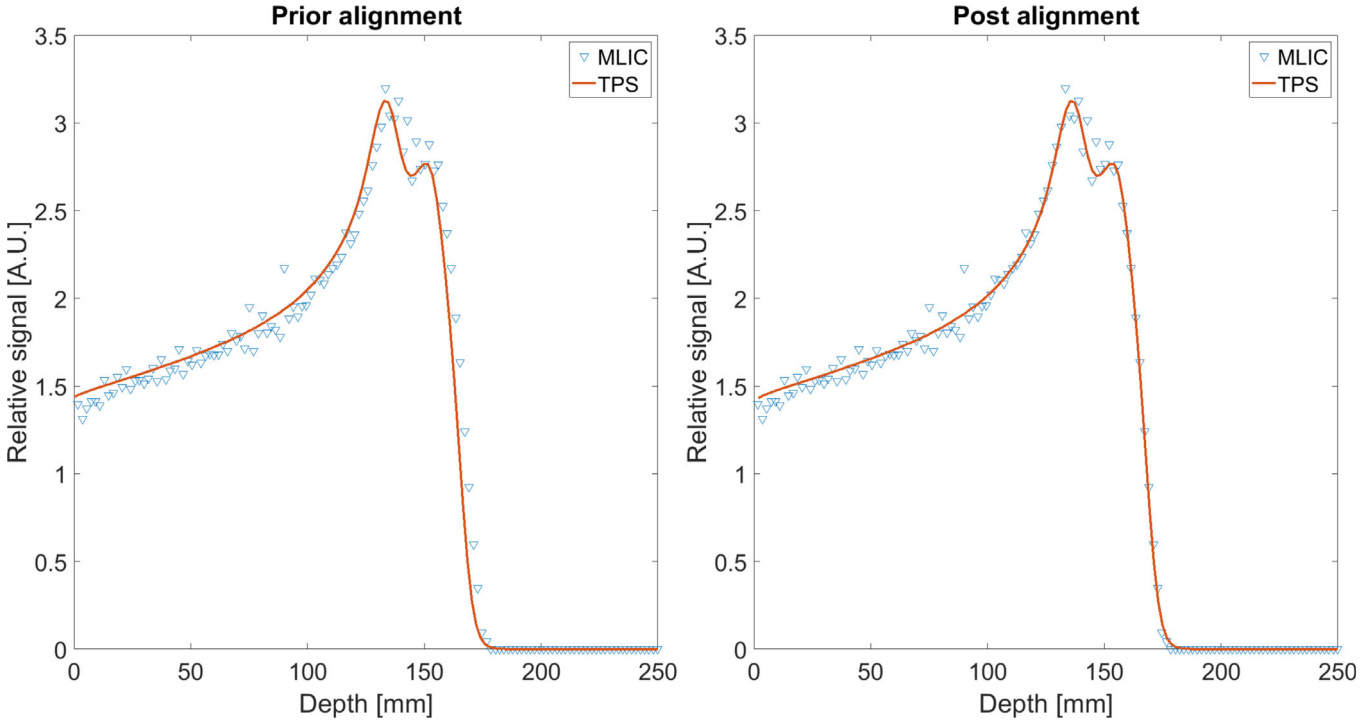

FIG. 3. Exemplary IDD curves as measured by MLIC and calculated by TPS for one of the pencil beams. (Left) plot shows raw MLIC and TPS data sets prior the calculation of an optimal offset, while (right) plot shows both data sets aligned (in this case 2.3 mm offset was calculated using the least squares method). The calculated offset in the context of this work is considered the range error for the given spot.

performed using Matlab tools, which have been developed based on the toolbox provided by the openREGGUI open-source project.[23]

As evaluated by Farace et al, this RP implementation allows to determine range errors with an accuracy of 0.5 mm.[21]

Within the scope of this study, the range accuracy was assessed not only on the basis of the repeat CTs, but also based on the planning CT. This was done to gain insight in how machine- and physics-related sources of range uncertainty (as introduced above) in general compare to patient-related sources of range errors, namely anatomical variations.

2.F. Anatomical inconsistency checks

In addition to reporting results for complete measurement frames (including all 81 spots per measurement, resulting in 1134 analyzed spots in total), all RP measurements have been reviewed by focusing on the agreement between repeat CT and CBCT of the same fraction, in order to identify spots affected by anatomical variations. The most common areas of inconsistency were identified and corresponded to (a) proximity of trapezius muscle and shoulders, (b) swallowing muscles, (c) base of tongue. An additional analysis was performed, where spots intersecting areas of common anatomical variations were excluded in all RP frames. To clarify, spots were excluded based on anatomical location instead of observed range errors. Spots, which were excluded for a specific patient, were excluded consistently in all RP frames for that patient. In total 48.5% of spots were excluded for the reduced data set analysis, resulting in 584 remaining spots.

3. RESULTS

Results of the clinical utilization of the RP QC procedure in the first seven consecutive HNC patients are presented.

As an example, relative range error maps for Patient 3 are shown in Fig. 4.

Relative range errors have been determined based on the planning CT and two repeat CTs (fractions 6 and 11 for this patient). Relative range errors for all 81 spots per RP frame are shown as an overlay on the sagittal plane of the planning CT scan at the treatment isocenter. For fraction 6, the comparison between the measured and simulated RP based on the repeat CT of that fraction shows a mean range error of 0.1% (1.5 SD 1.8%, which means 1.5 standard deviation is 1.8%), while the comparison with the simulated RP based on the planning CT shows a higher mean range error of −1.8% (1.5 SD 2.0%). Correspondingly for fraction 11, a mean range error of 0.7% (1.5 SD 2.4%) is seen for the comparison of measured and simulated RP based on the repeat CT and a mean error of −3.1% (1.5 SD 1.8%) when basing the simulation of the RP on the planning CT.

Mean WEPL for RP spots considering all spots (1134) combined was 145 mm, with values varying from 66 to 278 mm.

Mean relative range errors and 1.5 SD per RP frame (81 spots or less for the reduced data set) are listed in Table I for the QC procedures performed for all seven patients. Additionally, fraction numbers, during which QC procedures were performed, are added. During the same fractions also repeat CTs used for simulations were acquired. Table I shows that the mean range error obtained by comparison of the measured RP to the RP simulated based on the day-specific repeat CT is typically smaller than if the comparison is done to the RP simulated based on the planning CT. Mean absolute relative range errors for both data sets are provided in Supplementary Material S1.

As mentioned, a reduced measurement data set was created in order to exclude proton spots intersecting areas where anatomical inconsistencies are often observed. In this way, the influence of anatomical variations on the analysis is greatly reduced. Results of the analysis in the reduced data set can also be seen in Table I.

Again, considering Patient 3 as an example, for the reduced data set the mean range error for RP acquired during fraction 6 is 0.2% (1.5 SD 1.4%), when considering simulations based on the day-specific repeat CT. The mean range error is again noticeably larger (−1.9% (1.5 SD 1.9%)) when performing the analysis versus RP simulations based on the planning CT.

4. DISCUSSION

An RP QC procedure has been applied in vivo for the first time and results for the first seven HNC patients have been presented.

It can be noticed from Table I, that the analysis performed on the reduced RP data set (in areas of stable anatomy) shows range errors within the ±3% range uncertainty margin, which is applied in our clinical practice for the robust plan optimization. This holds if anatomical variations between the measurement and simulation in the TPS are minimal, meaning that simulations are performed on the same-day repeat CT. If simulations are performed on a planning CT image (reduced data set), which is already 3–4 weeks old at the time of RP acquisition, agreement between simulated and measured residual IDDs deteriorates. This can be explained (also confirmed by comparing images visually) by weight changes of the patients. For HNC patients undergoing concurrent chemotherapy we often observe skin contour increase / decrease caused by weight changes and/or swelling. For example, for Patient 3 and 6 in Table I simulations on the pCT systematically show under- or over-range compared to simulations on the rCT. This is caused by post-chemo treatment swelling of the Patient 3 and weight loss for Patient 6. Such a trend was not observed for Patients 5 and 7. Although standard deviations were smaller for the rCT-based data set compared to the pCT-based data set, mean relative range errors were slightly smaller (~1%) for the pCT-based data set. For these patients no obvious swelling or weight changes were observed, which is an effect that results in pronounced mean error fluctuations.

For the complete data sets (both, pCT and rCT), where all spots of each RP frame have been included in the analysis, similar tendencies as for the reduced data set can be observed: typically, rCT-based data sets show better agreement between measured and simulated residual IDDs than pCT-based data set. This is mostly caused by anatomical

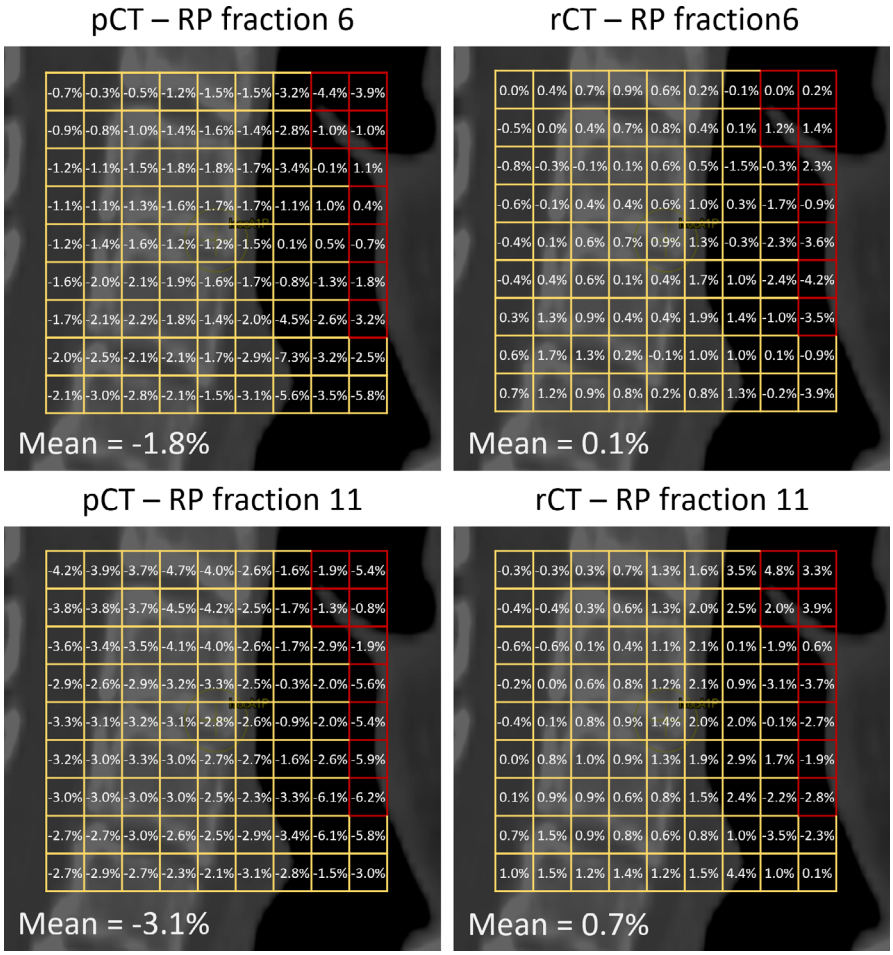

FIG. 4. Grids of relative range errors overlaid with the sagittal planes for an exemplary patient (Patient 3). Relative range errors are shown for a comparison between measured IDDs and individually simulated IDDs on the planning CT (left) and the two repeat CTs (right). For the example patient, proton radiograms were acquired during fractions 6 and 11. For each frame, each cell in the grid corresponds to a RP proton spot. Squares shown in red color were excluded for Patient 3 due to anatomical inconsistencies to obtain the reduced data set. Additionally, mean relative range errors for each frame are shown.

TABLE I. Overview of relative range errors as determined by proton range probing-based quality control checks. Results are shown for the analysis performed on the complete ("compl.") and reduced data sets ("red.").

	Mean relative range error (1.5SD) (%)									
	QC session 1					QC session 2				
		pCT		rCT			pCT		rCT	
Pat #	Fx #	compl.	red.	compl.	red.	Fx #	compl.	red.	compl.	red.
1	3	0.7 (4.6)	2.0 (1.3)	0.5 (3.1)	0.8 (1.3)	8	3.1 (4.2)	2.7 (2.7)	0.2 (2.6)	−0.2 (1.4)
2	6	2.8 (4.4)	2.1 (2.0)	1.8 (3.9)	0.2 (1.6)	11	0.5 (3.9)	1.6 (2.5)	1.8 (2.6)	1.1 (1.2)
3	6	−1.8 (2.0)	−1.9 (1.9)	0.1 (1.8)	0.2 (1.4)	11	−3.1 (1.8)	−3.0 (1.5)	0.7 (2.4)	0.7 (1.9)
4	2	4.5 (7.4)	3.5 (2.5)	2.2 (6.0)	0.1 (0.7)	9	−2.6 (8.9)	−2.5 (3.3)	1.9 (6.3)	−0.1 (2.4)
5	2	1.2 (2.3)	0.5 (1.9)	2.3 (2.2)	1.3 (0.9)	7	1.2 (3.2)	0.4 (2.8)	−0.6 (1.7)	−0.7 (1.0)
6	3	3.9 (2.4)	4.1 (1.7)	0.2 (2.2)	0.6 (1.3)	8	2.5 (7.7)	3.2 (2.9)	−0.1 (8.0)	−0.1 (1.3)
7	9	0.9 (4.9)	−0.1 (2.0)	4.4 (5.5)	1.1 (1.4)	14	−9.0 (8.7)	−4.3 (1.7)	−0.8 (7.4)	−1.2 (1.3)

variations between CT image acquisition and patient's anatomy at the time of the RP measurement, an effect that range uncertainty is not intended to account for.

Farace et al.[25] have indicated that range errors, as determined by RP, are sensitive to spatial misalignment errors between the simulated and measurement data sets. RP

acquisitions, as per proposed methodology, are acquired after alignment of the patient to the treatment isocenter. Although the patient's position is representative of the treatment position, residual setup errors may affect the RP measurements. As demonstrated by Farace *et al,* residual setup errors would cause increased range errors along areas of high heterogeneity index Hi values. Effects of residual misalignment and density interface have also been investigated by Hammi et al.[14,24] Such patterns were not observed in our data set along, for example, the spine (see Fig. 4), however to some extent could have been a contributing factor towards mean relative range error variation between 1st and 2nd QC based on the reduced data set of Patient 7.

Overall RP QC results are affected by multiple sources of uncertainty: (a) overall accuracy of range error determination according to the RP QC method (~0.5 mm),[19–21,25] which includes spot position accuracy, (b) energy fluctuations of the treatment delivery machine (~0.1 mm short-term to ~0.5 mm long-term), (c) residual setup errors and intrafraction motion, including, for instance, muscle relaxation, in the treatment room (~1 mm), (d) anatomical inconsistencies of the patient between the treatment room and CT imaging room, (e) rigid registration of the planning and repeat CTs. Sources (d) and (e) may be highly patient- and day-specific. For the current data set we estimate the uncertainty of the relative range error in the order of 1% (relative to the observed WEPL values in the obtained data sets). This is also indirectly supported by assessing the reproducibility of data between both QC sessions for the same patient, when reduced data set based on rCT (least affected by anatomical inconsistencies) is considered. Additionally, in this work the WEPL of a spot has been defined as the difference between the depth of the range probe peak and the depth of the peak for an in-air measurement. In case of a degraded peak shape due to heterogeneities (for example a double peak as shown in Fig. 3), this approach might not result in self-evident definition of the WEPL. For the purpose of this study, any bias in WEPL definition introduced by the use of maximum dose approach due to double peaks (as shown in Fig. 3) was considered negligible, since in practice such double peaks were rare. In fact, only about 3% of the measured spots per frame showed double peak pattern. In addition, in some cases dose maximum was coinciding with the more proximal peak, while in other cases, with the more distal peak. Furthermore, in some cases of highly degraded IDDs, the use of the least square method may result in good alignment at the distal fall-off region, while the alignment (or rather shape of the curves) at the peak region is suboptimal. While the used openreggui tools allow to visually review the alignment of the curves on spot-by-spot basis and no anomalies where observed in this data set, for the purpose of further automation, the introduction of a more robust metrics is desirable. Similarly, chamber gain fluctuations (noise) may influence the fitting process. In practice, noise on the proximal side of the peak has limited impact on the fitting since gradients on the proximal side are much lower than on the distal side of the peak. Additional metrics for quantifying range probing measurements have been investigated and proposed in literature, such as the weighted mean range and range dilution.[24] Future improvements of the RP QC procedure could be warranted by investigating applicability of additional metrics and introducing them into result reports.

It is important to note that large range errors in specific areas experiencing anatomical changes (as observed experimentally) do not necessarily translate into dosimetrically unacceptable plans. In fact, for none of the seven presented patients dosimetric evaluation of the treatment plan based on the standard weekly repeat CT triggered a plan adaptation. Multiple factors determine if robust target and organ-at-risk (OAR) dose is maintained during the treatment course. The number of treatment fields, their orientation, spot size, weight and placement, the dose modulation within a field and the robustness margins play a role in preserving an appropriate dose distribution.

The method applied in this work focuses on the evaluation of range errors in a clinical context by comparing measured residual ranges to simulated ones in the TPS. This method is not suitable to establish ground truth SPR per tissue type. In case major deviations between measured and simulated IDDs are observed, further investigation to establish the root cause would be necessary.

Furthermore, the method looks at a mixture of tissues in an integral manner. Theoretically it is possible to observe good range agreement, while this could be a result of balancing over- and under-estimation of SPRs for various tissue types in the beam path. Nevertheless, this is unlikely to systematically occur in practice since the geometry and anatomy of HNC patients is diverse. For anatomically less complex cases, for instance, intracranial indications, such compensation effects have been reported in literature.[26]

The purpose of the applied RP QC procedure is to provide data on the range calculation accuracy in the TPS on patient-specific bases. RP spots intersect a broad mixture of tissues and therefore provide data on range calculation accuracy also relevant for the tissue types in the beam path of treatment fields. However, RP fields do not exactly overlap with the treatment fields. This could be considered as a limitation of the method.

The results presented in this paper allow to bring the range uncertainty, as defined in literature,[3] and actual range errors, which regularly occur in clinical practice and towards which anatomical variations contribute most, into perspective. Anatomical variations in the beam path may have a more severe degrading effect on the range accuracy than other sources of uncertainty that are accounted for by range uncertainty recipes, as also reported in literature.[27] To account for anatomical variations, the ability to perform plan adaptations at a higher frequency (or online) would be required.

The described in vivo RP QC procedure is currently applied to assess range calculation accuracy on a patient-specific basis. It provides an insight on range prediction accuracy in a workflow based on single energy computed

tomography (SECT) imaging, but it could also be applied as a QC tool for workflows based on DECT imaging.

In our current practice, repeat CTs are systematically performed on weekly basis. It is not intended that RP QC procedure could replace a need for repeat CTs in future, as the information obtained is complimentary.

Furthermore, in the future, such a procedure could have a major role in the validation of synthetic CTs intended for proton dose calculations. CBCTs suffer from the large uncertainty of the Hounsfield Units (HU), which makes them unsuitable for proton dose calculation. Promising results have been shown on performance of neural networks (NN) in generating synthetic CTs based on daily CBCTs,[28] rendering synthetic images suitable for dose calculation. Nevertheless, to some extent NN may be considered as a "black box". Therefore, extensive QC procedures should be introduced to validate the output of NNs. In this context, in vivo RP QC may provide means to validate HU accuracy in synthetic CT images and may confirm their usability for dose calculation on a patient- and/or a fraction-specific basis. For such use case, more frequent (potentially, daily) RP QC acquisitions might be necessary. In such case 1 cGy RP dose might not be considered acceptable by some clinicians. Although not ideal, 1 cGy dose level is comparable to an imaging dose required by earlier generation MV portal imagers. To reduce RP dose further, modifications to the clinical proton delivery system would be required, as this dose level is currently determined by minimum monitor unit per spot limit. Nevertheless, to determine if 1 cGy dose is clinically acceptable for a QC procedure, it should be weighed against the possible dosimetric gains from using synthetic CTs in adaptive workflows.

5. CONCLUSIONS

A proton range probing-based quality control procedure has been deployed in clinical practice for HNC patients. It allows to evaluate range calculation accuracy on a patient-specific basis. Initial results show that anatomical inconsistencies that occur during the HNC treatment course often have a predominant effect on range errors. However, there is an agreement between calculated and measured ranges for spots in anatomically stable areas within 3%, which is the currently used range uncertainty margin for robust Monte Carlo-based optimization in our clinic.

ACKNOWLEDGMENTS

The Authors would like to acknowledge the Dutch Cancer Society (KWF research project 11518) for providing grant support towards the "INCONTROL- Clinical Control Infrastructure for Proton Therapy Treatments" project. The analysis for this study was partially performed by personnel funded by the INCONTROL project.

The authors would like to acknowledge the openREGGUI community for developing open-source tools, which were helpful in conducting this study.

CONFLICT OF INTEREST

Langendijk JA is a consultant for proton therapy equipment provider IBA.

FUNDING INFORMATION

No direct funding was made available for the study as such. The position of Carmen Seller Oria is funded by KWF (Dutch cancer society) research project 11518.

DISCLOSURES

University of Groningen, University Medical Centre Groningen, Department of Radiation Oncology has active research agreements with RaySearch, Philips, IBA, Mirada, Orfit.

Meijers A at the time of submission is full-time employee of Varian Medical Systems, USA. This study was conducted prior to that and without any involvement or support of Varian Medical Systems.

[a)]Author to whom correspondence should be addressed. Electronic mail: a.meijers@umcg.nl

REFERENCES

1. Wilson RR. Radiological use of fast protons. *Radiology.* 1946;47:487–491.
2. Kjellberg RN, Koehler AM, Preston WM, Sweet WH. Stereotaxic instrument for use with the Bragg peak of a proton beam. *Confin Neurol.* 1962;22:183–189.
3. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol.* 2012;57:R99–117.
4. Urie M, Goitein M, Holley WR, Chen GT. Degradation of the Bragg peak due to inhomogeneities. *Phys Med Biol.* 1986;31:1–15.
5. Maeda K, Yasui H, Matsuura T, et al. Evaluation of the relative biological effectiveness of spot-scanning proton irradiation in vitro. *J Radiat Res.* 2016;57:307–311.
6. Knopf AC, Hong TS, Lomax A. Scanned proton radiotherapy for mobile targets-the effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. *Phys Med Biol.* 2011;56:7257–7271.
7. Bär E, Lalonde A, Zhang R, et al. Experimental validation of two dual-energy CT methods for proton therapy using heterogeneous tissue samples. *Med Phys.* 2018;45:48–59.
8. Dedes G, Dickmann J, Niepel K, et al. Experimental comparison of proton CT and dual energy x-ray CT for relative stopping power estimation in proton therapy. *Phys Med Biol.* 2019;64:165002.
9. Meyer S, Kamp F, Tessonnier T, et al. Dosimetric accuracy and radiobiological implications of ion computed tomography for proton therapy treatment planning. *Phys Med Biol.* 2019;64:125008.
10. Li B, Lee HC, Duan X, et al. Comprehensive analysis of proton range uncertainties related to stopping-power-ratio estimation using dual-energy CT imaging. *Phys Med Biol.* 2017;62:7056–7074.
11. Yang M, Zhu XR, Park PC, et al. Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration. *Phys Med Biol.* 2012;57:4095–4115.
12. Johnson RP. Review of medical radiography and tomography with proton beams. *Rep Prog Phys.* 2018;81:16701.
13. Mumot M, Algranati C, Hartmann M, Schippers JM, Hug E, Lomax AJ. Proton range verification using a range probe: definition of concept and initial analysis. *Phys Med Biol.* 2010;55:4771–4782.

14. Hammi A, Koenig S, Weber DC, Poppe B, Lomax AJ. Patient positioning verification for proton therapy using proton radiography. *Phys Med Biol*. 2018;63:245009.
15. Krah N, De Marzi L, Patriarca A, Pittá G, Rinaldi I. Proton radiography with a commercial range telescope detector using dedicated post processing methods. *Phys Med Biol*. 2018;63:205016.
16. Rinaldi I, Brons S, Jäkel O, Voss B, Parodi K. A method to increase the nominal range resolution of a stack of parallel-plate ionization chambers. *Phys Med Biol*. 2014;59:5501–5515.
17. Richter C, Pausch G, Barczyk S, et al. First clinical application of a prompt gamma based in vivo proton range verification system. *Radiother Oncol*. 2016;118:232–237.
18. Knopf A, Parodi K, Paganetti H, Cascio E, Bonab A, Bortfeld T. Quantitative assessment of the physical potential of proton beam range verification with PET/CT. *Phys Med Biol*. 2008;53:4137–4151.
19. Meijers A, Free J, Wagenaar D, et al. Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing. *Phys Med Biol*. 2020;65:3NT02.
20. Meijers A, Seller Oria C, Free J, et al. Assessment of range uncertainty in lung-like tissue using a porcine lung phantom and proton radiography. *Phys Med Biol*. 2020;65:155014.
21. Farace P, Righetto R, Meijers A. Pencil beam proton radiography using a multilayer ionization chamber. *Phys Med Biol*. 2016;61:4078–4087.
22. Krah N, Patera V, Rit S, Schiavi A, Rinaldi I. Regularised patient-specific stopping power calibration for proton therapy planning based on proton radiographic images. *Phys Med Biol*. 2019;64:65008.
23. https://openreggui.org/ Last visit: 29. Mar. 2020.
24. Hammi A, Placidi L, Weber DC, Lomax AJ. Positioning of head and neck patients for proton therapy using proton range probes: a proof of concept study. *Phys Med Biol*. 2017;63:15025.
25. Farace P, Righetto R, Deffet S, Meijers A, Vander SF. Technical Note: a direct ray-tracing method to compute integral depth dose in pencil beam proton radiography with a multilayer ionization chamber. *Med Phys*. 2016;43:6405.
26. Wohlfahrt P, Möhler C, Troost EGC, Greilich S, Richter C. Dual-energy computed tomography to assess intra- and inter-patient tissue variability for proton treatment planning of patients with brain tumor. *Int J Radiat Oncol Biol Phys*. 2019;105:504–513.
27. Wang P, Yin L, Zhang Y, et al. Quantitative assessment of anatomical change using a virtual proton depth radiograph for adaptive head and neck proton therapy. *J Appl Clin Med Phys*. 2016;17:427–440.
28. Thummerer A, Zaffino P, Meijers A, et al. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. *Phys Med Biol*. 2020;65:95002.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Overview of relative range errors as determined through proton range probing-based quality control checks. Results are shown for the analysis performed on the complete ("compl.") and reduced data sets ("red.").