University of Groningen

Comprehensive quality control process for high precision intensity modulated adaptive proton therapy

Meijers, Artürs

DOI:
10.33612/diss.170751488

Publication date:
2021

Link to publication in University of Groningen/UMCG research database

Comprehensive quality control process for high precision intensity modulated adaptive proton therapy

Arturs Meijers

Cover & layout: Lovebird design.
www.lovebird-design.com

Printing: Eikon+

Comprehensive quality control process for high precision intensity modulated adaptive proton therapy

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Wednesday 9 June 2021 at 9:00

by

Arturs Meijers

born on 12 August 1986
in Riga, Latvia

Contents

Introduction

Proton therapy is a form of radiation therapy that uses high energy proton beams for irradiation of cancerous cells. Proton therapy is associated with superior dosimetric treatment plan characteristics compared to photon therapy, due to the physical properties of proton beams [1]. However, these physical properties, specifically, the Bragg peak, are also associated with increased uncertainty — particularly range uncertainty — affecting the accuracy of any treatment plan delivery [2]. Proton range uncertainty is widely considered to be one of the main issue preventing the usage of proton therapy from reaching its full potential [3].

Proton therapy has been established as the preferred treatment modality for pediatric indications, base of skull treatments, ocular melanomas, and for the irradiation of the craniospinal axis. Lately, growing evidence has emerged supporting the use of proton therapy for tumors in the liver. In addition, numerous institutions have recently explored the applicability of proton therapy for other indications, such as, head and neck, lung, breast, neurological, gastrointestinal, and upper and lower abdomen [4-23].

Over the last decade, actively scanned proton beams, also referred to as pencil beam scanning, have become the most commonly used treatment modality of proton therapy, replacing passively scattered proton beams. Active scanning of a narrow proton beam across the target volume introduces an additional dimension to the uncertainty. It specifically affects indications which are subject to intra-fractional motion, such as, thoracic indications. In addition to range uncertainty, which is considered particularly critical for lung-like tissue due to the microstructures present in the lung, delivered dose distributions for thoracic indications are also affected by the so-called interplay effect. Due to these concerns the proton therapy treatment of targets affected by breathing motion has been adopted very slowly [24].

Although various sources of uncertainty, such as range uncertainty and organ motion, have been identified and investigated relatively thoroughly, potentially the greatest contributor to the dose uncertainties is intrinsically linked to patient's anatomical variations. These variations compromise the model of the patient itself, as it is defined during the patient's simulation. Furthermore, anatomical variations are highly

patient specific and therefore difficult to predict and account for in the treatment planning process, while maintaining reasonable safety margins. Robust optimization helps to some extent in coping with uncertainties. Nevertheless, a balance between acceptable dose to healthy tissues and the extent of the scenarios accounted for must be found to avoid creating overly robust treatment plans, which account for a large number of scenarios that will never occur in practice. A promising solution for coping with anatomical variations is adaptive therapy, which would allow the "renewal" of a patient's model as anatomical variations occur [25]. Nevertheless, while the concept has been discussed extensively, the adoption of adaptive radiotherapy in the mainstream clinical practice has been fairly slow. The approach is time consuming, adequate and efficient data handling environments have been lacking and compatible quality control procedures are still to be developed.

In order to support the introduction of adaptive proton therapy, comprehensive quality control procedures must be put in place. These procedures should cover various steps in the workflow, from a patient's virtual model to patient-specific quality assurance and follow up throughout the delivery of the treatment course itself.

Chapters I, II and IV of this thesis focus on investigating proton range uncertainty. Chapter I "Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing" proposes and validates a method to verify the CT calibration curve in near-clinical conditions during the commissioning phase. This method introduces the use of a range probing technique for the purpose of CT calibration curve validation and optimization, if found to be necessary. Chapter II "First report on an in vivo proton radiography quality control procedure for scanned proton beam therapy in head and neck cancer patients" demonstrates the incorporation of range-probing into an in-patient clinical quality control procedure. Chapter II reports on the first experience of an online in vivo range probing quality control measurements performed for head and neck cancer patients. For the first time, this allows the assessment of treatment beam range prediction accuracy on patient- and fraction-specific basis.

For indications affected by respiration, motion adds an additional level of complexity for accurate assessment of delivered treatment dose.

Chapter III "Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: Implementation and proof-of-concept" introduces the use of treatment delivery log files into a dose reconstruction workflow. The method is suitable for use in commercial treatment planning systems and allows considering treatment fraction-specific spot delivery sequence timing, as well as breathing signal.

Chapter IV "Assessment of range uncertainty in lung-like tissue using porcine lung phantom and proton radiography" brings the range probing range evaluation technique to 4D space by exploring range accuracy in lung-like tissue. Furthermore, in Chapter IV, the log file-based dose reconstruction technique is combined with the range probing method to allow for range accuracy evaluations for an air-ventilated (in other words, "breathing") lung tissue phantom to validate the magnitude of the range uncertainty margin to be employed in clinical thorax treatment planning. Chapter V "Evaluation of interplay and organ motion effect by means of 4D dose reconstruction and accumulation" validates the clinical IMPT treatment of 4D indications. In Chapter V, the 4D dose reconstruction method based on use of treatment delivery log files is applied in a clinical setting. For a set of 10 IMPT patients treated for thoracic indications, the 4D dose reconstruction is applied on fraction-specific basis, therefore allowing observation and monitoring the interplay and organ motion effects on the delivered dose distribution throughout the treatment course more accurately. Additionally, this approach offers a dose reconstruction and accumulation workflow, which can be used to monitor patient treatment progress and guide plan adaptation decisions.

A significant hurdle, especially online, towards the introduction of adaptive proton therapy is linked to the Patient Specific Quality Assurance (PSQA). Measurement-based PSQA procedures require in-beam time to acquire QA measurements. This has an influence on the timing of offline treatment plan adaptations, while makes online treatment plan adaptations non-feasible. Chapters VI and VII are focusing on a paradigm shift in PSQA processes by proposing alternative methods. Chapter VI "Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy" demonstrates development and deployment of automated PSQA platform, which relies on the use of treatment delivery

log files and independent dose recalculation for the quality assurance of treatment plans. While Chapter VII "Feasibility of patient specific quality assurance for proton therapy based on independent dose calculation and predicted outcomes" brings the revised PSQA procedure a step further and proposes a way to interpret PSQA results in more clinically relevant manner by using predicted outcomes.

In summary, the performed studies look into the range uncertainty problem and propose a method to monitor range accuracy on fraction- and patient-specific basis, which is applied into clinical practice. In addition, a method for more realistic fraction-specific dose reconstruction and course-wise dose accumulation compatible with 4D indications is proposed and applied in clinical practice. Finally, a patient specific quality assurance procedure compatible with offline and online adaptive proton therapy workflows is proposed and implemented in clinical practice. In such a way, a set of comprehensive quality control and assurance procedures were developed and implemented, making a further step towards enabling adaptive proton therapy in clinical practice.

References

[1]   Nyström H, Blomqvist E, Høyer M, Montelius A, Muren LP, Nilsson P, Taheri-Kadkhoda Z, Glimelius B. Particle therapy — a next logical step in the improvement of radiotherapy. Acta Oncol. 2011 Aug;50(6):741-4. doi: 10.3109/0284186X.2011.590150. PMID: 21767169.

[2]   Unkelbach J, Paganetti H. Robust Proton Treatment Planning: Physical and Biological Optimization. Semin Radiat Oncol. 2018 Apr;28(2):88-96. doi: 10.1016/j.semradonc.2017.11.005. PMID: 29735195; PMCID: PMC5942229.

[3]   Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol. 2012 Jun 7;57(11):R99-117. doi: 10.1088/0031-9155/57/11/R99. Epub 2012 May 9. PMID: 22571913; PMCID: PMC3374500.

[4]   Weber DC, Malyapa R, Albertini F, Bolsi A, Kliebsch U, Walser M, Pica A, Combescure C, Lomax AJ, Schneider R. Long term outcomes of patients with skull-base low-grade chondrosarcoma and chordoma patients treated with pencil beam scanning proton therapy. Radiother Oncol. 2016 Jul;120(1):169-174. doi: 10.1016/j.radonc.2016.05.011. Epub 2016 May 28. PMID: 27247057

[5]   Ares C, Albertini F, Frei-Welte M, Bolsi A, Grotzer MA, Goitein G, We-
        ber DC. Pencil beam scanning proton therapy for pediatric intracranial
        ependymoma. J Neurooncol. 2016 May;128(1):137-145. doi: 10.1007/
        s11060-016-2090-4. Epub 2016 Mar 5. PMID: 26945580

[6]   Anderson J, Niska JR, Thorpe CS, Bruso ME, McGee LA, Hartsell WF,
        Larson GL, Rossi CJ, Rosen LR, Chang AL, Vargas CE. Proton Beam Ac-
        celerated Partial-Breast Irradiation: Prospective Multi-Institutional PCG
        Registry Analysis. Int J Rad Oncol Biol Phys. 2019 Sep 1;105(1):E49. doi:
        https://doi.org/10.1016/j.ijrobp.2019.06.2376

[7]   Chao HH, Berman AT, Simone CB 2nd, Ciunci C, Gabriel P, Lin H, Both
        S, Langer C, Lelionis K, Rengan R, Hahn SM, Prabhu K, Fagundes M,
        Hartsell W, Mick R, Plastaras JP. Multi-Institutional Prospective Study of
        Reirradiation with Proton Beam Radiotherapy for Locoregionally Recurrent
        Non-Small Cell Lung Cancer. J Thorac Oncol. 2017 Feb;12(2):281-292. doi:
        10.1016/j.jtho.2016.10.018. Epub 2016 Nov 5. PMID: 27826034

[8]   Chuong MD, Hartsell W, Larson G, Tsai H, Laramore GE, Rossi CJ,
        Wilkinson JB, Kaiser A, Vargas C. Minimal toxicity after proton beam
        therapy for prostate and pelvic nodal irradiation: results from the proton
        collaborative group REG001-09 trial. Acta Oncol. 2018 Mar;57(3):368-374.
        doi: 10.1080/0284186X.2017.1388539. Epub 2017 Oct 14. PMID: 29034790

[9]   DeCesaris CM, McCarroll R, Mishra MV, Glass E, Greenwald BD, Carr S,
        Burrows W, Mehra R, Regine WF, Simone CB 2nd, Choi JI, Molitoris JK.
        Assessing Outcomes of Patients Treated With Re-Irradiation Utilizing Proton
        Pencil-Beam Scanning for Primary or Recurrent Malignancies of the Esoph-
        agus and Gastroesophageal Junction. J Thorac Oncol. 2020 Jun;15(6):1054-
        1064. doi: 10.1016/j.jtho.2020.01.024. Epub 2020 Mar 4. PMID: 32145427

[10]   Hess CB, Indelicato DJ, Paulino AC, Hartsell WF, Hill-Kayser CE, Perkins
        SM, Mahajan A, Laack NN, Ermoian RP, Chang AL, Wolden SL, Mangona
        VS, Kwok Y, Breneman JC, Perentesis JP, Gallotto SL, Weyman EA, Bajaj
        BVM, Lawell MP, Yeap BY, Yock TI. An Update From the Pediatric Pro-
        ton Consortium Registry. Front Oncol. 2018 May 24;8:165. doi: 10.3389/
        fonc.2018.00165. PMID: 29881715; PMCID: PMC5976731

[11]   Hoppe BS, Nichols RC, Pham DC, Mohindra P, Hartsell WF, Mohammed N,
        Chon BH, Morris CG, Li Z, Flampouri S, Simone CB. UF-Pcg Phase II Study
        of Hypofractionated Proton Therapy with Concurrent Chemotherapy for
        Stage II-III NSCLC. Int J Rad Oncol Biol Phys. 2019 Sep 1;105(1):S144-S145.
        doi: https://doi.org/10.1016/j.ijrobp.2019.06.143.

[12]   Jayakrishnan R, Mehta MP, Tseng YD, Vargas CE, Gondi V, Tsai HK,
        Halasz LM, Rossi CJ, Wang CJ, Badiyan SN, Kotecha R. A Prospective
        Multi-Institutional Study of Clinical Outcomes for Meningioma Patients
        Treated with Proton Beam Radiotherapy. Int J Rad Oncol Biol Phys. 2019
        Sep 1;105(1):E66-E67. doi: https://doi.org/10.1016/j.ijrobp.2019.06.2413

[13]   Jhaveri J, Cheng E, Tian S, Buchwald Z, Chowdhary M, Liu Y, Gillespie TW, Olson JJ, Diaz AZ, Voloschin A, Eaton BR, Crocker IR, McDonald MW, Curran WJ, Patel KR. Proton vs. Photon Radiation Therapy for Primary Gliomas: An Analysis of the National Cancer Data Base. Front Oncol. 2018 Nov 28;8:440. doi: 10.3389/fonc.2018.00440. PMID: 30547008; PMCID: PMC6279888

[14]   Koroulakis A, Molitoris JK, Kaiser A, Hanna N, Jiang Y, Regine WF. Re-Irradiation for Rectal Cancer: A Single Institution Experience Utilizing Photons and Protons. Int J Rad Oncol Biol Phys. 2019 Sep 1;105(1):E167. doi: https://doi.org/10.1016/j.ijrobp.2019.06.2146

[15]   Murray FR, Snider JW, Bolsi A, Lomax AJ, Walser M, Kliebsch U, Schneider RA, Weber DC. Long-Term Clinical Outcomes of Pencil Beam Scanning Proton Therapy for Benign and Non-benign Intracranial Meningiomas. Int J Radiat Oncol Biol Phys. 2017 Dec 1;99(5):1190-1198. doi: 10.1016/j.ijrobp.2017.08.005. Epub 2017 Aug 12. PMID: 28939227

[16]   Niska JR, Thorpe CS, Anderson J, Bruso ME, McGee LA, Hartsell WF, Larson GL, Tsai HK, Rossi CJ, Rosen LR, Vargas CE. Post-Mastectomy Radiotherapy using Proton Beam Therapy: Prospective Multi-Institutional PCG Registry Analysis. Int J Rad Oncol Biol Phys. 2019 Sep 1;105(1):E6. doi: https://doi.org/10.1016/j.ijrobp.2019.06.629

[17]   Smith AW, Gallitto M, Wasserman I, Gupta V, Sharma S, Westra W, Genden E, Haidar Y, Yao M, Teng MS, Miles B, Bakst RL. Redefining Patients at Risk of Contralateral Neck Disease for HPV-related Oropharyngeal Cancer: A Pathologic Study of Patients with Bilateral Neck Dissection. Int J Rad Oncol Biol Phys. 2019 Sep 1;105(1):E427. doi: https://doi.org/10.1016/j.ijrobp.2019.06.1543

[18]   Snider JW, Schneider RA, Poelma-Tap D, Stieb S, Murray FR, Placidi L, Albertini F, Lomax A, Bolsi A, Kliebsch U, Malyapa R, Weber DC. Long-Term Outcomes and Prognostic Factors After Pencil-Beam Scanning Proton Radiation Therapy for Spinal Chordomas: A Large, Single-Institution Cohort. Int J Radiat Oncol Biol Phys. 2018 May 1;101(1):226-233. doi: 10.1016/j.ijrobp.2018.01.060. Epub 2018 Feb 2. PMID: 29619966

[19]   Yu NY, Gamez ME, Hartsell WF, Tsai HK, Laramore GE, Larson GL, Simone CB 2nd, Rossi C, Katz SR, Buras MR, Golafshar MA, Vargas CE, Patel SH. A Multi-Institutional Experience of Proton Beam Therapy for Sinonasal Tumors. Adv Radiat Oncol. 2019 Jul 16;4(4):689-698. doi: 10.1016/j.adro.2019.07.008. PMID: 31673662; PMCID: PMC6817523

[20]   Iwata H, Toshito T, Hayashi K, Yamada M, Omachi C, Nakajima K, Hattori Y, Hashimoto S, Kuroda Y, Okumura Y, Mizoe JE, Ogino H, Shibamoto Y. Proton therapy for non-squamous cell carcinoma of the head and neck: planning comparison and toxicity. J Radiat Res. 2019 Oct 23;60(5):612-621. doi: 10.1093/jrr/rrz036. PMID: 31147697; PMCID: PMC6805978

[21]   Chang JY, Zhang X, Knopf A, Li H, Mori S, Dong L, Lu HM, Liu W, Badiyan SN, Both S, Meijers A, Lin L, Flampouri S, Li Z, Umegaki K, Simone CB 2nd, Zhu XR. Consensus Guidelines for Implementing Pencil-Beam Scanning Proton Therapy for Thoracic Malignancies on Behalf of the PTCOG Thoracic and Lymphoma Subcommittee. Int J Radiat Oncol Biol Phys. 2017 Sep 1;99(1):41-50. doi: 10.1016/j.ijrobp.2017.05.014. Epub 2017 May 19. PMID: 28816159

[22]   DeCesaris CM, McCarroll R, Mishra MV, Glass E, Greenwald BD, Carr S, Burrows W, Mehra R, Regine WF, Simone CB 2nd, Choi JI, Molitoris JK. Assessing Outcomes of Patients Treated With Re-Irradiation Utilizing Proton Pencil-Beam Scanning for Primary or Recurrent Malignancies of the Esophagus and Gastroesophageal Junction. J Thorac Oncol. 2020 Jun;15(6):1054-1064. doi: 10.1016/j.jtho.2020.01.024. Epub 2020 Mar 4. PMID: 32145427

[23]   Lin SH, Hobbs BP, Verma V, Tidwell RS, Smith GL, Lei X, Corsini EM, Mok I, Wei X, Yao L, Wang X, Komaki RU, Chang JY, Chun SG, Jeter MD, Swisher SG, Ajani JA, Blum-Murphy M, Vaporciyan AA, Mehran RJ, Koong AC, Gandhi SJ, Hofstetter WL, Hong TS, Delaney TF, Liao Z, Mohan R. Randomized Phase IIB Trial of Proton Beam Therapy Versus Intensity-Modulated Radiation Therapy for Locally Advanced Esophageal Cancer. J Clin Oncol. 2020 May 10;38(14):1569-1579. doi: 10.1200/JCO.19.02503. Epub 2020 Mar 11. PMID: 32160096; PMCID: PMC7213588

[24]   Macdonald OK, Kruse JJ, Miller JM, Garces YI, Brown PD, Miller RC, Foote RL. Proton beam radiotherapy versus three-dimensional conformal stereotactic body radiotherapy in primary peripheral, early-stage non-small-cell lung carcinoma: a comparative dosimetric analysis. Int J Radiat Oncol Biol Phys. 2009 Nov 1;75(3):950-8. doi: 10.1016/j.ijrobp.2009.04.023

[25]   Albertini F, Matter M, Nenoff L, Zhang Y, Lomax A. Online daily adaptive proton therapy. Br J Radiol. 2019 Nov 11:20190594. doi: 10.1259/bjr.20190594

Chapter I: Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing

Abstract

Purpose: Proton therapy is affected by range uncertainty, which is partly caused by an ambiguous conversion from x-ray attenuation to proton stopping power. CT calibration curves, or Hounsfield look-up tables (HLUTs), are institution-specific and may be a source of systematic errors in treatment planning. A range probing method to verify, optimize and validate HLUTs for proton treatment is proposed.

Methods and Materials: An initial HLUT was determined according to the stoichiometric approach. For HLUT validation, three types of animal tissue phantoms were prepared: a pig's head, "thorax" and femur. CT scans of the phantoms were taken and a structure, simulating a water slab, was added on the scan distal to the phantoms to mimic the detector used for integral depth-dose measurements. The CT scans were imported into the TPS to calculate individual pencil beams directed through the phantoms.

The phantoms were positioned at the therapy system isocenter using x-ray imaging. Shoot-through pencil beams were delivered, and depth-dose profiles were measured using a multi-layer ionization chamber. Measured depth-dose curves were compared to the calculated curves and the range error per spot was determined. Based on the water equivalent path length (WEPL) of individual spot, a range error margin was defined. Ratios between measured error and theoretical margin were calculated per spot. The HLUT optimization was performed by identifying systematic shifts of the mean range error per phantom and minimizing the ratios between range errors and uncertainty margins.

Results: After optimization, the ratios of the actual range error and the uncertainty margin over the complete data set did not exceed 0.75 (1.5SD), indicating that the actual errors are covered by the theoretical uncertainty recipe.

Conclusions: The feasibility of using range probing to assess range errors was demonstrated. The theoretical uncertainty margins in the institution-specific setting potentially may be reduced by ~25%.

Introduction

Over the last decades, proton therapy is becoming a more widely available treatment modality. However, range uncertainty is commonly regarded as a significant concern. Major contributors to the range uncertainty are linked to the CT calibration, or conversion from CT number to SPR (direct or via mass density), and handling of lateral and longitudinal heterogeneities by the dose calculation engine of the treatment planning system. By quantifying various contributors, range uncertainty recipes have been proposed and these typically consist of a relative component (relative to the range of the beam) and an absolute component, which is largely influenced by proton beam delivery equipment [1]. The institution-specific range uncertainty margin will depend on the utilized treatment modalities, the performance of the equipment and the choice of dose calculation engine (differentiating between analytical and Monte Carlo engines).

CT numbers are imperfect input for dose calculations due to images being affected by the noise and lack of an unambiguous assignment of

tissue properties based on CT numbers. [2]. As CT calibration is one of the contributors to range uncertainty, typically an effort is invested in this task during the implementation phase of new proton treatment delivery equipment. There are two methodologies frequently used for CT calibration: the tissue-substitute method and the stoichiometric method [3]. The tissue-substitute method relies on establishing a calibration curve based on scans of known density materials, which are meant to substitute specific human-like tissues and correlating those densities to measured CT number. The stoichiometric method also requires scans of tissue-equivalent materials. However, the measured data is used to characterize the CT via three fitting parameters. The obtained parameters are used to pre-calculate expected CT numbers for human-like tissues, considering their elemental composition.

A critical look at the stoichiometric method has recently been taken by Goma et al. [4]. It was pointed out that, in order to accurately perform fitting of the measured data and thereby characterize the CT acquisition properly, the fitting procedure itself must be mathematically constrained. However, the approach to set constraints is not well defined. Furthermore, it was shown that the outcome of the calibration may be affected by the selection of phantom for the initial scan. Goma et al. applied the stoichiometric method for two commercially available CT calibration phantoms of CIRS (Norfolk, VA, USA) and Gammex (Middleton, WI, USA). It was observed that in case of the CIRS phantom, the tissue-substitute method and the stoichiometric method resulted in two different HLUTs, especially in bone-like tissue section. While in case of the Gammex phantom, the tissue-substitute method and stoichiometric method resulted in nearly the same calibration curve, situated between the two calibration curves determined for the CIRS phantom. Due to the absence of a ground truth, it remains unclear, which of the obtained curves would be the most appropriate for actual clinical use.

The purpose of the current study was to propose and apply a methodology that would allow to verify, optimize and validate a HLUT and its performance using range estimations from the treatment planning system (TPS) and range probing (proton radiography) measurements.

Furthermore, the intent is to confirm that the range uncertainty recipe derived from literature is applicable in an institution-specific setting for the use of robust optimization during the treatment planning.

Material and methods

The following methodology was developed and applied to create, verify, optimize and validate a site-specific HLUT prior to implementation in the clinic:

1. Creation of an initial HLUT based on the stoichiometric method and its implementation in the TPS
2. Validation of the HLUT
   a. Acquisition of CT scans of animal tissue samples
   b. Execution of range probing calculations in TPS
   c. Performance of range probing measurements for the tissue samples
3. Optimization of the HLUT
   a. Comparison of range probing measurements with TPS calculations and definition of residual range errors
   b. Minimization of range errors and implementation of the optimized HLUT into the TPS
4. Validation of the optimized HLUT
   a. Re-calculation of range probing data in TPS
   b. Re-evaluation of the residual range errors (including independent sample)
   c. Release of the HLUT for clinical use

Initial HLUT

For proton dose calculation during the treatment planning phase, imaging data (most commonly, patient CT) are generally converted to relative proton stopping power ratio (SPR) maps. However, depending on the treatment planning system and dose calculation engine, the user input for calculation of SPR maps may differ. For example, the conversion of CT number may be defined either directly to proton stopping power or to physical density. RayStation 7 (RaySearch Laboratories, Sweden) requires defined CT numbers versus predetermined physical densities as an input for its CT calibration. Afterwards, during the dose calculation itself, the TPS is using built-in material look-up tables [5] to convert mass density to SPR. According to RayStation Reference manual, built-in material look-up tables have been based on ICRU49, ICRU44 and ICRP23 report.

To create an initial HLUT, the stoichiometric approach was partially used [3]. Since user-provided input only specifies a CT number to mass density curve, but mass density to SPR conversion is performed by TPS, the stoichiometric method was applied only to calculate the CT number to mass density curve. Furthermore, to confirm accurate SPR values calculated by the TPS, the conversion methodology [5] was replicated outside of the TPS and theoretical SPR values were calculated and compared to the values provided by the TPS. No significant variations ($< 0.2\%$) were observed.

For the initial CT number measurements, the tissue substitute phantom by CIRS model 062M was used. The phantom consists of inner and outer cylinders, which can be filled with a set of inserts, made of materials that are substitutes for human-like tissues in terms of density, designed to establish the HLUT.

For every clinically used scan protocol (and reconstruction kernel) the CIRS phantom was scanned in two configurations: (1) only the middle cylinder (representing a small object) and (2) the full phantom, consisting of both cylinders (representing a large object). It was observed that variations between the HU versus mass density curves for large and small objects were most pronounced in calcium rich regions. Overall, the variations due to the choice of reconstruction kernel and size of the scan object were considered acceptable to create an averaged-out curve per scan energy. This also allowed to minimize the number of clinically used curves, simplifying the treatment planning process.

Scans of the tissue phantoms were performed using Somatom Definition AS scanner (Siemens Healthineers, Erlangen, Germany) with following scan settings: 120 kV, reconstruction slice thickness of 2 mm, I40f or I40s reconstruction kernels (according with the applicable clinical scan protocols), with enabled iMAR artefact correction setting. Reconstruction kernels include iterative reconstruction algorithm Safire (used strength: 3) and have intrinsic beam hardening correction for water only.

Although, the stoichiometric method was used to establish the initial HLUT, an additional HLUT (Fig. 2) based on tissue-substitute method was calculated for comparison purpose.

The HLUT was split in three sections: (a) organ-like tissue, (b) fat-like tissue and (c) bone-like tissue. These 3 sections were linearly fitted; i.e.,

Table 1. Summary of utilized CIRS inserts for definition of initial HLUT.

Insert label	Relative electron density	HLUT section	Small configuration, HU	Large configuration, HU
Lung (inhale)	0.19	organ-like	−789	−780
Lung (exhale)	0.489	organ-like	−511	−505
Adipose	0.949	fat-like	−68	−66
Breast (50% gland / 50% adipose)	0.976	fat-like	−32	−31
Muscle	1.043	organ-like	40	38
Liver	1.052	organ-like	52	51
Trabecular bone	1.117	bone-like	230	204
Dense bone	1.456	bone-like	914	845

the HLUT consists of 3 fitted lines and 2 transition areas [6]. A list of used CIRS phantom inserts and their assignment to corresponding section of the HLUT is provided in Table 1.

Validation data set

The measurements were performed using three fresh vacuum-sealed animal tissue phantoms: (a) a pig's head, (b) a "thorax", which consisted of ribs, liver, muscle (with cartilage), fat and (c) a femoral bone with soft tissue. CT scans and measurements were performed over a period of two days. Samples were stored in fridge between the activities.

The used tissue phantoms had approximately following diameters at their thickest slices: head phantom 19 cm, thorax phantom 26 cm, femoral bone 16 cm. These dimensions are comparable with anthropomorphic phantoms and representative for cases, such as, intracranial, head and neck or pediatric indications. Larger dimensions of the phantoms would make range probing (or proton radiography) unfeasible, due to limitations imposed by the maximum available energy of the proton beam in the clinical facility (230 MeV, depth of 32 cm). Additionally, some of the scans for femoral bone were performed by placing the phantom on the solid water plates, in order to introduce additional scattering material in the

field of view. Average water equivalent thickness (WET) values per phantom along the range probe path were as follows: Head phantom 143.2 mm (SD 50.8 mm), Thorax phantom 43.0 mm (SD 20.5 mm), Femoral bone 113.9 mm (SD 40.9 mm).

All phantom samples were scanned with a CT scanner using clinical scanning pre-sets (120 kV) established already during the definition of the initial HLUT. Afterwards scans were imported into the TPS for (a) the placement of the isocenters for the range probing measurements and (b) for the calculation of the individual pencil beams for comparison with the subsequently measured data.

Range probing measurements [7, 8] were performed following the methodology as described by Farace et al. [9]. The approach is based on mapping the samples of interest with individual proton pencil beams of an energy high enough to pass through the sampled area and measuring the exit residual range of individual pencil beams. Measurements were performed with the Giraffe multi-layer ionization chamber (IBA dosimetry, Germany). Due to the size of detector (electrode diameter of 12 cm), a maximum area of 4.5 by 4.5 cm² can be covered in a single measurement frame. Therefore, the tissue samples were covered by multiple measurement frames, where each frame contained of 81 individual pencil beams (spot spacing 0.5 cm). Measurements were performed in a movie measurement mode with a sampling time of 10 ms. The delivery system (Proteus Plus, IBA, Belgium) was intentionally slowed down to ensure enough delay between the two consecutive spots within the same frame (i.e., field). All spots were delivered with 210 MeV energy.

As stated by Farace et al. [9], nominal range accuracy for this measurement technique is ± 0.5 mm. Due to the detector size and selection of the frame size (4.5 × 4.5 cm²), there is no need for high positioning accuracy of the MLIC perpendicular to the beam axis. As verified experimentally, displacements in the order of 5 mm, will not have relevant impact on the shape of the measured integral depth dose curve. Furthermore, positioning accuracy of the MLIC along the beam axis also won't have significant impact on the measurement accuracy, as the distance from the isocenter mainly affect the amount of air between the detector and the tissue sample. Appropriate positioning accuracy of the MLIC under such circumstances can be achieved by employing on-board laser positioning

systems, which typically are an integral part of proton treatment rooms. Accuracy of the experiment will depend on the alignment of the phantom to the isocenter of the proton treatment room. In our case, according to the commissioning date for rigid anthropomorphic phantoms positioning accuracy of less than 0.5 mm can be achieved. Tissue phantoms, if handled properly and over a short time frame, can be considered nearly rigid.

The tissue material was positioned at the isocenter by using the on-board x-ray imaging system: either CBCT or kV-kV imaging. The samples were repositioned between the frames by applying predefined offsets by the robotic patient positioning system. The patient position system has a positioning accuracy < 0.5 mm.

To simulate the range probing measurements in the TPS in each animal tissue CT scan, a slab of homogeneous water-equivalent material ($40 \times 40 \times 50$ cm^3) was added representing the multi-layer ionization chamber (MLIC) measurement device, which is calibrated to output measurements in water-equivalent depth. The used MLIC had an electrode diameter of 12 cm. Every single pencil beam (FWHM 8.2 mm) was calculated individually using a Monte Carlo dose calculation algorithm (version 4.1) with an uncertainty of 0.5% and on a 1 mm isotropic dose grid. Afterwards, dose distributions were integrated on the plane along the beam axis to create integral depth dose curves to compare with the measured integral depth dose distributions. Experimental setup is shown in Figure 1.

Raw measurement data and exported data from TPS were processed and analyzed by a dedicated in-house Matlab tool [10, 11]. The measurements and TPS calculations were transformed into proton radiograms and reconstructed radiograms respectively. Per pencil beam, a shift (*shift*) along the beam axis to reach the best alignment between measured ($MLIC_i$) and calculated (TPS_i) depth dose curve was calculated by solving least square cost function (1). This shift was defined as a residual range error.

$$S = \sum_{i=1}^{n} \big(MLIC_i(depth_i) - TPS_i(depth_i + shift) \big)^2 \tag{1}$$

Based on literature [1] in case of Monte Carlo calculations, the range uncertainty is estimated as 2.4% of beam range + 1.0 mm. For every measured pencil beam, a range error margin was calculated as 2.4%
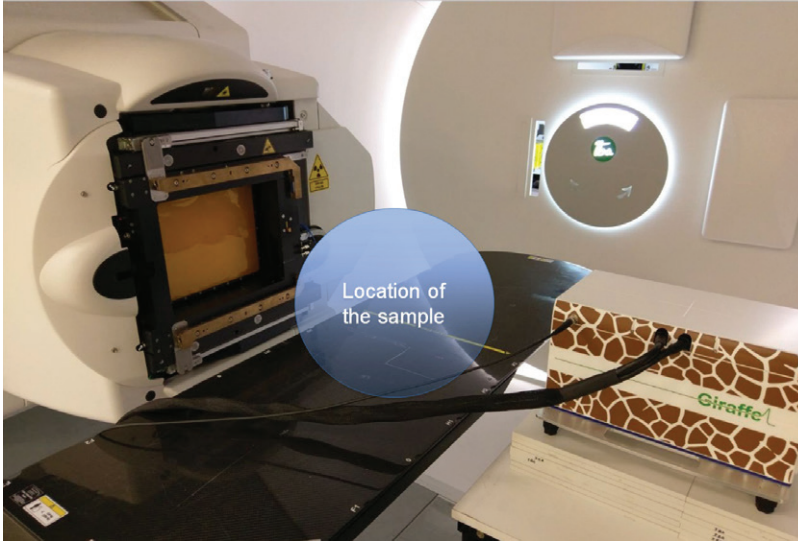
Figure 1. Experimental setup.

of water-equivalent path length (WEPL) + 1.0 mm. The WEPL of the individual pencil beam was used instead of the beam range because the definition of range in water is provided as an input to the TPS for beam modelling. Therefore, it is more appropriate to exclude the contribution of range error in water (or residual range itself as measured by the MLIC) from the uncertainty margin calculation. It was considered that the energy reproducibility of the proton delivery system for the measurements is within the absolute (+ 1.0 mm) component of the range uncertainty margin.

Eventually the range errors, defined as the discrepancy between measurement and calculation, were compared to corresponding range error margins, and for each pencil beam the ratio between error and margin was calculated. A schematic representation of the approach is shown in Figure 2.

If the ratio was larger than 1, the range error exceeds the uncertainty margin. If it is lower, the range error is within the uncertainty margin. In total, over all three samples, approximately 1600 individual pencil beams were measured and evaluated.
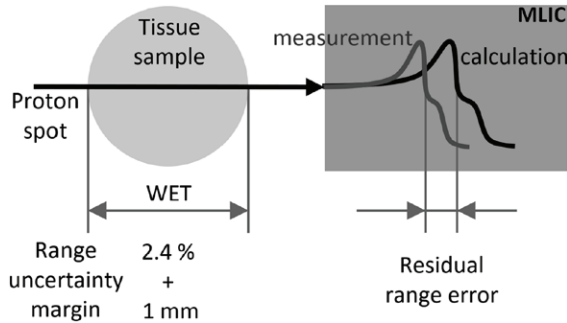
Figure 2. Schematic representation of the performed measurement and analysis. Range uncertainty margin is calculated per every pencil beam as 2.4% of water-equivalent thickness (WET) of the tissue in the beam path plus 1 mm. Multi-layer ionization chamber (MLIC) was used to acquire integral depth dose distally from the tissue sample. MLIC was simulated as a water slab for the purpose of calculations in the treatment planning system.

Optimization of the HLUT

By reviewing range error and uncertainty margin ratio histograms per tissue sample it was possible to identify inaccurate CT curve sections. Since the HLUT consisted of three fitted line segments (organ-like, fat-like and bone-like tissues), each specific segment of the HLUT could be corrected by adjusting the slope and intercept to achieve an optimal agreement between measurements and calculations. Iterations were performed by determining approximate overshoots or undershoots for a specific HLUT region based on the measurement set and afterwards adjusting the corresponding line segment to compensate previously identified overshoots or undershoots.

Afterwards, the optimized HLUT was introduced in the TPS and the entire data set was recalculated using the new HLUT. The analysis as described previously was repeated to validate the modifications. Figure 3 shows histograms of ratios between range error and uncertainty margin per phantom type.

It can be observed that the distribution is skewed for the thorax phantom case, which is likely linked to the composition of tissues (some more predominant than other) in the phantom. Furthermore, an independent measurement set, specifically focusing on the tissue-type (bone-like tissues in our case) in the adjusted area of the HLUT, was included in the analysis.
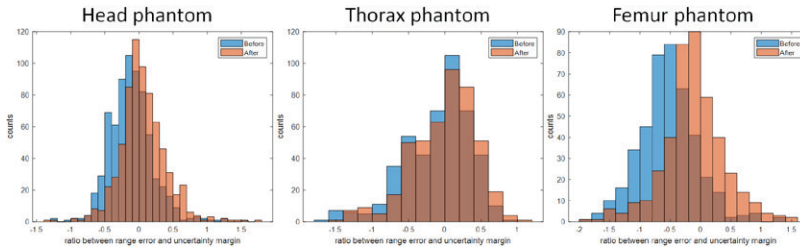
Figure 3. Histograms of ratios between range error and uncertainty margin before and after optimization of HLUT per phantom type.

Results

Using the initial HLUT, created with the stoichiometric method, residual range error maps and histograms were determined for all 3 tissue samples separately.

In case of (a) the pig's head, a mean range error of −0.54 mm with a standard deviation (SD) of 1.5 mm was observed; for (b) the "thorax" phantom, the mean range error was −0.17 mm (SD of 1.0 mm) and for (c) the femoral bone, the mean range error was −2.37 mm (SD 2.0 mm). Based on these observations, it was concluded that bone-like tissues in TPS are seen denser than they are. Therefore, the slope of bone-like tissue section of the HLUT was adjusted to compensate for this effect. The optimized HLUT is shown in Figure 4. The intercept and slope of the segment representing bone-like tissues was adjusted from 1.009 and 0.0007 to 1.029 and 0.0006 respectively.

The optimized HLUT lies in between the initial stoichiometric HLUT and HLUT as calculated by the tissue substitute method. Range error maps were recalculated using the optimized HLUT. After recalculation in case of (a) the pig's head, the mean range error changed to 0.33 mm (SD 1.4 mm); for (b) the thorax phantom, the mean range error dropped to 0.03 mm (SD 0.9 mm) and for (c) the femoral bone, the average range error reduced to −0.43 mm (SD 1.9 mm). Mean range errors as observed prior to and after HLUT optimization are provided in Table 2. The results presented in this section and Table 2 are based on the initial set of tissue phantoms.

The map of the ratios between range errors and uncertainty margin (defined as 2.4% of WEPL + 1.0 mm), as calculated using the optimized
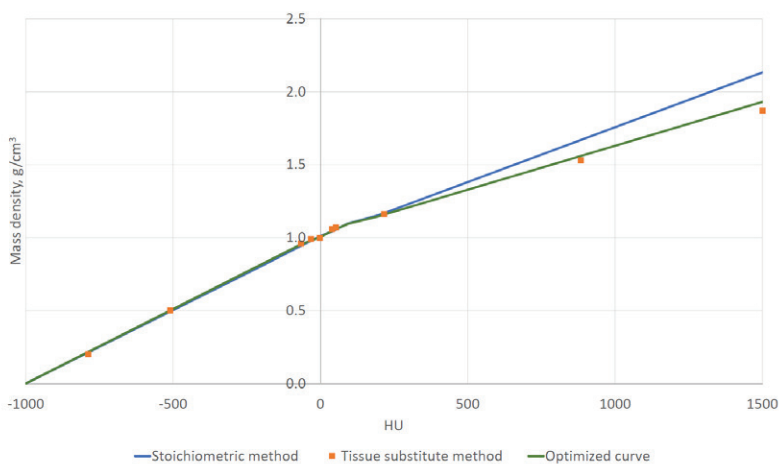
Figure 4. Comparison of obtained HLUTs.

Table 2. Summary of range errors before and after adjustment of HLUT.

	Before HLUT adjustment	After HLUT adjustment
Head phantom	−0.54 (SD 1.5) mm	0.33 (SD 1.4) mm
Thorax phantom	−0.17 (SD 1.0) mm	0.03 (SD 0.9) mm
Femoral bone	−2.37 (SD 2.0) mm	−0.43 (SD 1.9) mm

HLUT for head sample, is shown in Figure 5. Two areas highlighted in Figure 5 (Area A and B) have not been covered in the analysis.

In case of Area A, few of the spots acquired in this area could not have been clearly separated timewise while post-processing the measurements. Therefore, to avoid ambiguous sampling, all spots in Area A were excluded from the analysis. Area B was not included in the analysis and measurements in this area were not performed, because most of the spots in this area would have travelled only through the air. Range error and uncertainty margin ratios for a combined data set (all tissue phantoms, including additional femoral-bone phantom, and all measurement points) are shown in a histogram in Figure 6.
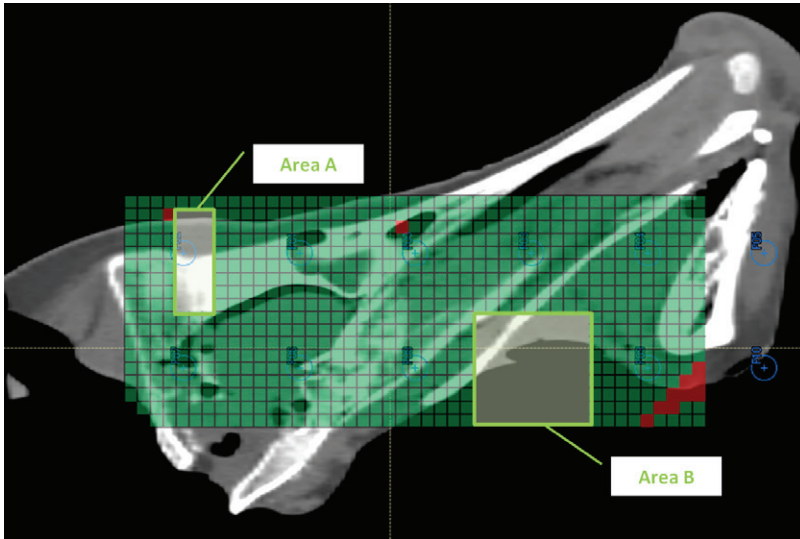
Figure 5. Range error and uncertainty margin ratio map. Red squares indicate spots for which range errors exceeded the uncertainty margin.
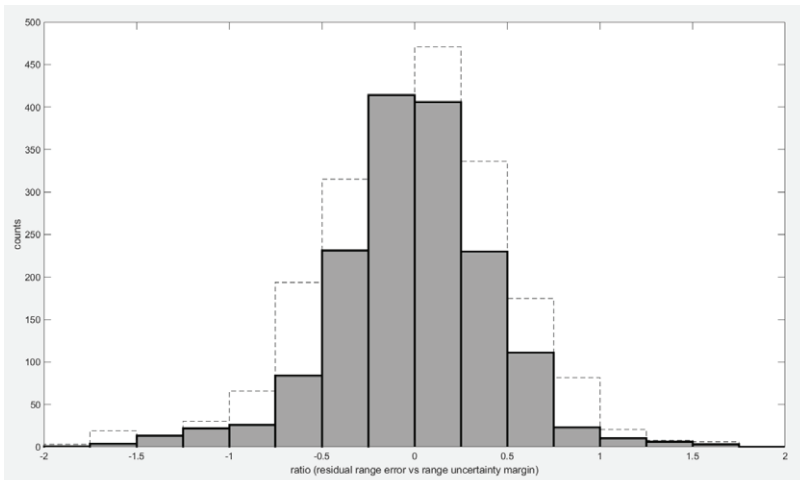
Figure 6. Histogram of ratios between range error and uncertainty margin for a combined data set after adaptation of the HLUT. For 1.5σ of the cases the ratio between the range error and the uncertainty margin is less than 0.75, when assuming an uncertainty margin of 2.4% + 1 mm. The dashed line indicates the upper border of the confidence interval accounting for the uncertainty of the measurement and evaluation.

Discussion

Geometrical localization of range errors

Although the range error distribution seems to be normally distributed, if considering isolated areas of the sample, the magnitude of range errors between these areas varies. For example, increased range errors are well correlated with intersections between materials of clearly different density, such as, high- and low-density bone intersections (see Figure 7).

This observation is also consistent with the literature [9, 12]. As pointed out by Farace et al., increased range errors based on range probing measurements in a head phantom were observed along the skull contour.

This indicates that systematic density scaling, which is broadly used in robust treatment plan optimization, is not the ideal approach to address the range uncertainty problem. Preferably, range uncertainty in the planning process should be applied considering knowledge about the materials in the beam path. Such information as mass density and mass density variation laterally to the beam path should be considered to more realistically account for range uncertainty in the robust optimization process.

In the absence of such algorithms in the TPS, an indication specific range uncertainty recipe may be implemented. For example, it may not be necessary to apply the same magnitude of CT number scaling in robust optimization of an intracranial tumour, when the beam path is intersecting the skull perpendicularly and does not intersect cavities, as compared to a base of skull case, where beam may intersect ear canal or to some extent travel parallel to the brain — bone intersection. Therefore, also the beam angle selection may be used to minimize the range uncertainty, which needs to be accounted for by robust optimization.

Figure 8 shows the range error and uncertainty margin ratio map, for a case where the uncertainty margin of 2.4% + 1.0 mm has been reduced by half.

Spots that are intersecting such areas as brain, soft tissue in the snout or are perpendicular to flat and large bony areas would still have range errors within the uncertainty margin. However, beams, which are traveling close to bone — soft tissue intersections or through cavities, have range errors outside the reduced uncertainty margin.
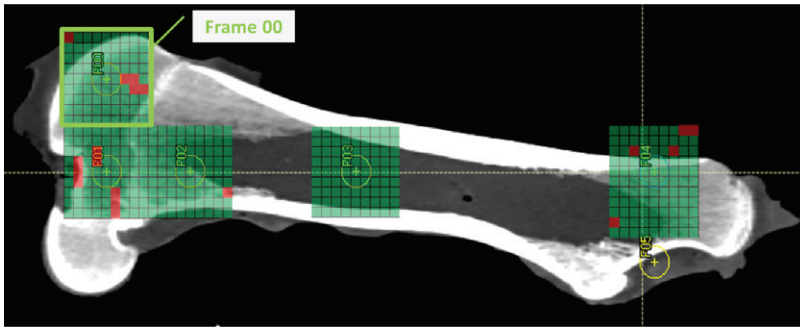
Figure 7. Range error and uncertainty margin ratio map for femoral bone. Red squares indicate spots with range error above uncertainty margin. Four failing spots on the Frame 00 are intersecting a titanium screw, which was attached to the bone.
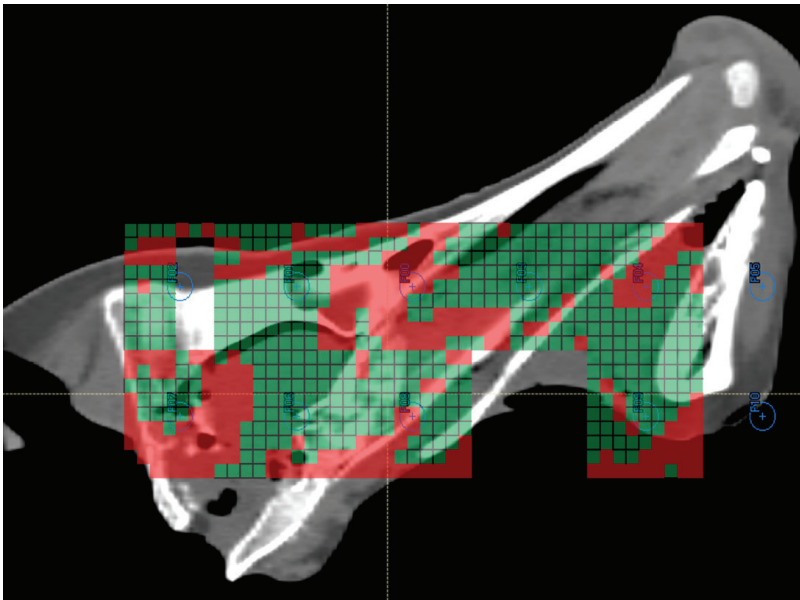
Figure 8. Range error and uncertainty margin ratio map for a case, where uncertainty margin has been reduced by half (1.2% + 0.5 mm).

General

Based on the range error and uncertainty margin ratio histogram, it can be observed that 1.5 SD of the spots has a ratio less than 0.75, which

indicates that the range uncertainty margin, set as 2.4% + 1.0 mm in our institution-specific setting, is an overestimation of the actual range errors encountered for the studied tissue phantom set, if the optimized HLUT is used. In case of applying initial, non-optimized HLUT, 1.5 SD of the spots has a range error to uncertainty margin ratio of less than 0.9. The obtain results are representative for the sample dimensions as presented above. Beam hardening effect may introduce additional uncertainty for patients of significantly larger dimensions. Therefore, ultimately range probing checks should be performed for actual patients.

Following this study based on animal tissue, it is worthwhile to consider the application of the methodology using patient-specific data. This would bring more insight in SPR values for human tissue and assess further gains in potentially employing site-specific or patient-specific HLUTs [13, 14, 15, 16]. Using the range probing in patients would require relatively low doses. Without further adjustments of the technique, the dose at the plateau would be less than 1 cGy per frame. Additionally, range probing could be integrated in the workflow to provide information on anatomical variations and assist in the decision-making process for triggering plan adaptation.

Eventually, range probing could enrich CBCT data and improve the quality and reliability of virtual CTs, created based on the CBCT anatomy. Currently, one of the challenges in creating virtual CTs based on CBCT data sets is the ability to validate the accuracy of CT number retrieval, which commonly is done based on image deformation fields [17, 18] or in combination with other methodologies, such as machine learning [19]. Similarly, to the use of range probing in the scope of this work for validation of range calculation accuracy based on CT data set, it could be used for the validation of virtual CT data sets.

Currently, there is a substantial interest in the field for the application of dual energy CT imaging for proton treatment planning, to provide more accurate SPR data [20, 21, 22]. It is estimated that the use of dual energy CT would allow to reduce range uncertainties to about 2% [23]. By using the proposed range probing methodology for optimization and validation of the HLUTs, it was demonstrated that range uncertainty could be reduced to almost 2% as well. This reduction may be further enhanced when site specific or patient specific HLUTs are used. However, combining the range probing approach with the use of dual energy CT imaging could potentially

allow to reduce the uncertainty even further. Nevertheless, in order to support comparison between range uncertainty estimation for usage of dual energy CT and results of this study, it would be necessary to also perform the range probing-based evaluation for dual energy CT images.

A limitation of the HLUT optimization is, that within the optimization process contribution of absolute component in range uncertainty, which is affecting the experimental data set, is also "minimized". However, in practice adjustment of HLUT cannot reduce the contributions to absolute component of the range uncertainty. By overfitting HLUT to obtain perfect agreement between measurement and simulation, one might incorrectly introduce a compensation for range errors caused by contributors to the absolute component in the HLUT. The inclusion of independent samples and data sets in the evaluation to some extent provides a possibility to assess the impact of the above-mentioned issue.

The methodology is not intended, nor suitable for identifying tissue-specific ground-truth stopping power ratios. It is rather an end-to-end verification, which looks at tissue compositions in an integral manner. Therefore, the methodology should not be used for extensive optimizations of the HLUT, if the problematic area is not obvious (such as in the current case bone-like tissues). By overfitting the HLUT, solutions might be found that give excellent agreement between measurements and simulations, however, due to integral characteristic of the range probe, still incorrect stopping powers might be assigned to individual tissues. To overcome this limitation, more projections could be acquired, which resample an approach towards proton CT.

Possible deformations of the tissue samples that may happen between CT simulation and treatment delivery can impact the comparison between measurements and simulations. This is a limitation of the proposed methodology. It is difficult to numerically asses the possible impact of phantom deformations, as it can vary significantly depending on type of deformation, extent and localization. For our experiments, possible deformations were investigated by extensively reviewing CBCT image overlaid with the CT image and no significant difference were identified.

Currently one of the drawbacks of the experimental measurement technique is the lack of integration between measurement device and beam delivery system. More reliable sampling during measurement, for

example using a trigger mode, would be desirable. This would allow to avoid or limit artefacts in the measurement set and a need to exclude data, as shown in Figure 5 area A.

While the range probing based method for assessment of range accuracy in the treatment delivery process was demonstrated in a single energy CT based setting, in principle the technique can also be applied to perform range accuracy evaluations in departments, which use dual energy CT for patient simulation [24].

In conclusion, it has been demonstrated that range probing is an efficient method for institution-specific validation and optimization of HLUTs prior to their use in the clinic, opening possibilities for reducing literature-based range uncertainty margins. Further range probing studies should evaluate the potential range uncertainty reduction for site- or patient-specific HLUT alone and/or in conjunction with DECT.

References

[1]   Paganetti H., Range uncertainties in proton therapy and the role of Monte Carlo simulations., Phys Med Biol. 2012 Jun 7;57(11):R99-117. doi: 10.1088/0031-9155/57/11/R99. Epub 2012 May 9.

[2]   Unkelbach J, Paganetti H., Robust Proton Treatment Planning: Physical and Biological Optimization., Semin Radiat Oncol. 2018 Apr;28(2):88-96. doi: 10.1016/j.semradonc.2017.11.005.

[3]   Schneider U, Pedroni E, Lomax A., The calibration of CT Hounsfield units for radiotherapy treatment planning., Phys Med Biol. 1996 Jan;41(1):111-24.

[4]   Gomà C, Almeida IP, Verhaegen F., Revisiting the single-energy CT calibration for proton therapy treatment planning: a critical look at the stoichiometric method., Phys Med Biol. 2018 Nov 26;63(23):235011. doi: 10.1088/1361-6560/aaede5. doi: 10.1120/jacmp.v15i3.4721.

[5]   RayStation 7B Reference manual. RaySearch Labaratories.

[6]   Ainsley CG, Yeager CM., Practical considerations in the calibration of CT scanners for proton therapy., J Appl Clin Med Phys. 2014 May 8;15(3):4721.

[7]   Schneider U, Pedroni E., Proton radiography as a tool for quality control in proton therapy., Med Phys. 1995 Apr;22(4):353-63.

[8]   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 Aug 21;55(16):4771-82. doi: 10.1088/0031-9155/55/16/010. Epub 2010 Aug 3.

[9]   Farace P, Righetto R, Meijers A., Pencil beam proton radiography using a multilayer ionization chamber., Phys Med Biol. 2016 Jun 7;61(11):4078-87. doi: 10.1088/0031-9155/61/11/4078. Epub 2016 May 10.

[10]   Farace P, Righetto R, Deffet S, Meijers A, Vander Stappen F., 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 Dec;43(12):6405.

[11]   Deffet S, Macq B, Righetto R, Vander Stappen F, Farace P., Registration of pencil beam proton radiography data with X-ray CT., Med Phys. 2017 Oct;44(10):5393-5401. doi: 10.1002/mp.12497. Epub 2017 Aug 31.

[12]   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 Aug 7;53(15):4137-51. doi: 10.1088/0031-9155/53/15/009. Epub 2008 Jul 17.

[13]   Schneider U, Pemler P, Besserer J, Pedroni E, Lomax A, Kaser-Hotz B., Patient specific optimization of the relation between CT-hounsfield units and proton stopping power with proton radiography., Med Phys. 2005 Jan;32(1):195-9.

[14]   Doolan PJ, Testa M, Sharp G, Bentefour EH, Royle G, Lu HM., Patient-specific stopping power calibration for proton therapy planning based on single-detector proton radiography., Phys Med Biol. 2015 Mar 7;60(5):1901-17. doi: 10.1088/0031-9155/60/5/1901. Epub 2015 Feb 10.

[15]   Collins-Fekete CA, Brousmiche S, Hansen DC, Beaulieu L, Seco J., Pre-treatment patient-specific stopping power by combining list-mode proton radiography and x-ray CT., Phys Med Biol. 2017 Aug 3;62(17):6836-6852. doi: 10.1088/1361-6560/aa7c42.

[16]   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 Mar 12;64(6):065008. doi: 10.1088/1361-6560/ab03db.

[17]   Landry G, Nijhuis R, Dedes G, Handrack J, Thieke C, Janssens G, Orban de Xivry J, Reiner M, Kamp F, Wilkens JJ, Paganelli C, Riboldi M, Baroni G, Ganswindt U, Belka C, Parodi K., Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation., Med Phys. 2015 Mar;42(3):1354-66. doi: 10.1118/1.4908223.

[18]   Veiga C, Janssens G, Teng CL, Baudier T, Hotoiu L, McClelland JR, Royle G, Lin L, Yin L, Metz J, Solberg TD, Tochner Z, Simone CB, McDonough J, Teo BK., First Clinical Investigation of Cone Beam Computed Tomography and Deformable Registration for Adaptive Proton Therapy for Lung Cancer., Int J Radiat Oncol Biol Phys. 2016 May 1;95(1):549-59. doi: 10.1016/j.ijrobp.2016.01.055. Epub 2016 Feb 4.

[19]   Tappeiner E, Pröll S, Hönig M, Raudaschl PF, Zaffino P, Spadea MF, Sharp GC, Schubert R, Fritscher K., Multi-organ segmentation of the head and

neck area: an efficient hierarchical neural networks approach., Int J Comput Assist Radiol Surg. 2019 May;14(5):745-754. doi: 10.1007/s11548-019-01922-4. Epub 2019 Mar 7.

[20] Hudobivnik N, Schwarz F, Johnson T, Agolli L, Dedes G, Tessonnier T, Verhaegen F, Thieke C, Belka C, Sommer WH, Parodi K, Landry G., Comparison of proton therapy treatment planning for head tumors with a pencil beam algorithm on dual and single energy CT images., Med Phys. 2016 Jan;43(1):495. doi: 10.1118/1.4939106.

[21] Wohlfahrt P, Möhler C, Stützer K, Greilich S, Richter C., Dual-energy CT based proton range prediction in head and pelvic tumor patients., Radiother Oncol. 2017 Dec;125(3):526-533. doi: 10.1016/j.radonc.2017.09.042. Epub 2017 Oct 16.

[22] Wohlfahrt P, Troost EGC, Hofmann C, Richter C, Jakobi A., Clinical Feasibility of Single-Source Dual-spiral 4D Dual-Energy CT for Proton Treatment Planning Within the Thoracic Region., Int J Radiat Oncol Biol Phys. 2018 Nov 15;102(4):830-840. doi: 10.1016/j.ijrobp.2018.06.044. Epub 2018 Jul 10.

[23] Li B, Lee HC, Duan X, Shen C, Zhou L, Jia X, Yang M., Comprehensive analysis of proton range uncertainties related to stopping-power-ratio estimation using dual-energy CT imaging., Phys Med Biol. 2017 Aug 9;62(17):7056-7074. doi: 10.1088/1361-6560/aa7dc9.

[24] Wohlfahrt P, Möhler C, Hietschold V, Menkel S, Greilich S, Krause M, Baumann M, Enghardt W, Richter C., Clinical Implementation of Dual-energy CT for Proton Treatment Planning on Pseudo-monoenergetic CT scans., Int J Radiat Oncol Biol Phys. 2017 Feb 1; 97(2):427-434. doi: 10.1016/j.ijrobp.2016.10.022. Epub 2016 Oct 21.

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

Abstract

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 isocentre 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 a 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.

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], [4], [5], [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 (i) machine related, such as,

reproducibility and stability of the equipment, (ii) physics related, such as, transformation of CT numbers to mass density to SPR, (iii) patient related, such as, anatomical and physiological variations, and (iv) biology related, linked to the end-of-range effect and relative biological effectiveness (RBE) uncertainty [5]. Nonetheless, only (i) and (ii) 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 are 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 in range accuracy predictions in a near-clinical (for example, commissioning phase) or clinical setting. Techniques as proton radiography [12],[13],[14],[15],[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.

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 isocentre. 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 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.

RP in treatment planning
A dedicated treatment field with a gantry angle of 90 degrees is incorporated in the clinical TPS treatment plan. Currently the choice of gantry angle is limited to the lateral orientation (90 or 270 degrees) due to constrains linked to MLIC Giraffe positioning. Positioning the gantry at 90 degrees allows easier access to the patient with the measurement
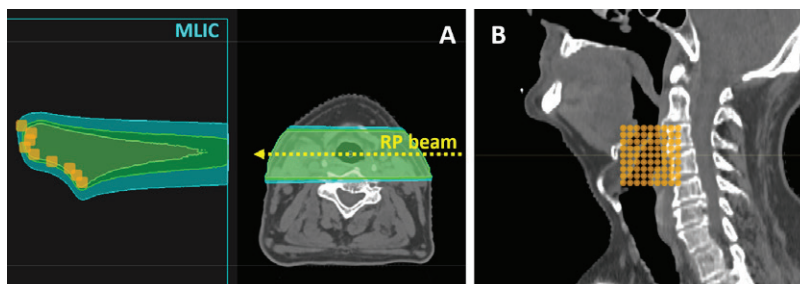
Figure 1. Visualization of a range probing (RP) field for an example patient geometry (Patient 2). (A) The dose distribution of the RP field is shown from a transversal view of the patient, where the RP field is directed from the patient's left to the right (as from a gantry angle of 90 degrees). The 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.

equipment in our site-specific conditions. The field consists of 81 spots, covering a $4 \times 4\,cm^2$ area around the treatment isocentre. 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. A RP field (81 spots) delivers approximately $1\,cGy_{RBE}$ 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 isocentre of 8.2 mm at our facility. In our clinical practice, the treatment isocentre for HNC patients is in the proximity of C3 or C4 vertebrae. Since RP spots are centred around the isocentre, this allows to intersect a broad mixture of tissues (bones, various muscles, fat tissue, nodes and, in some cases, tumour) during the QC procedure. As an example, Figure 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.

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 minutes before or after the treatment fraction, with the patient immobilized as in the treatment position.

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 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 minutes. After the calibration is performed, the patient may enter the TR. The patient is positioned at the TR isocentre 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 isocentre along the beam axis using in-room lasers. The distance between isocentre 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 non-standard modifications to the beam delivery system.
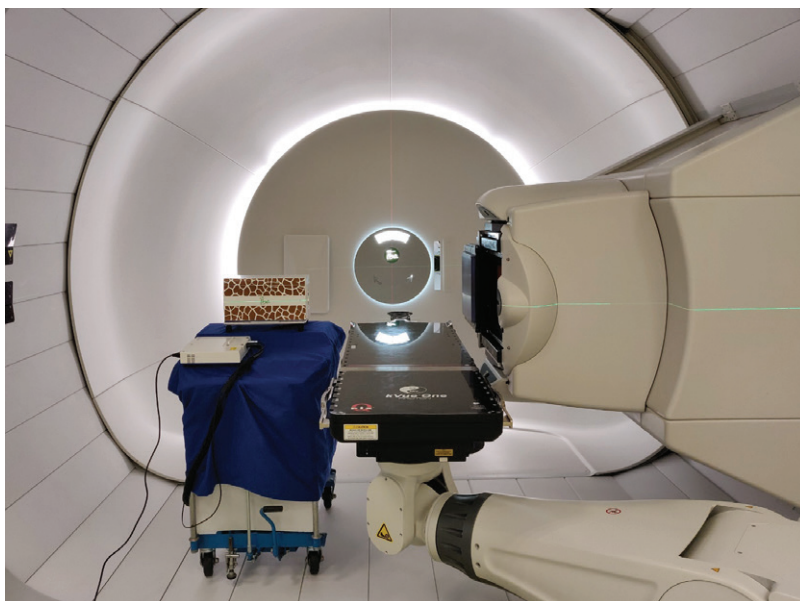
Figure 2. Equipment setup in the treatment room for proton range probing acquisition. The gantry angle is set to 90 degrees and the MLIC is positioned on the trolley on the right side of the patient.

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 Figure 2.

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 minutes, of which approximately 15 seconds are the time of RP delivery.

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.

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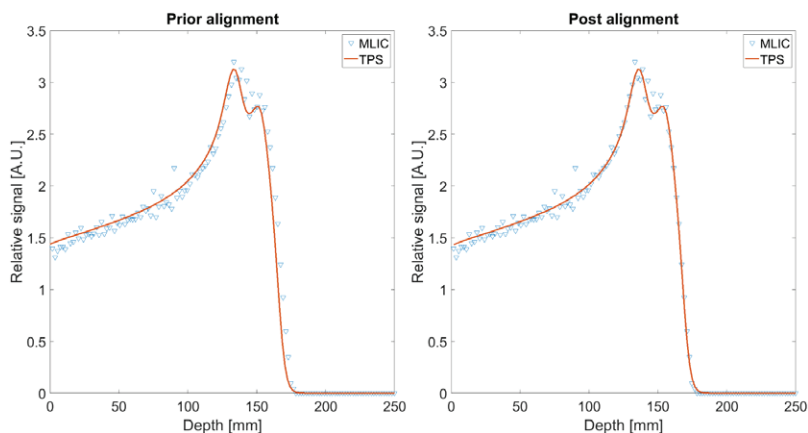
Figure 3. Exemplary IDD curves as measured by MLIC and calculated by TPS for one of the pencil beams. (Left) plot shows raw MLIC and TPS data sets prior the calculation of an optimal offset, while (right) plot shows both data sets aligned (in this case 2.3 mm offset was calculated using the least squares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

Anatomical inconsistency checks

In addition to reporting results for complete measurement frames (including all 81 spots per measurement, resulting in 1134 analysed 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 (1) proximity of trapezius muscle and shoulders, (2) swallowing muscles, (3) 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.

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 Figure 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 isocentre. 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 1 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 1 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 1.

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 1.
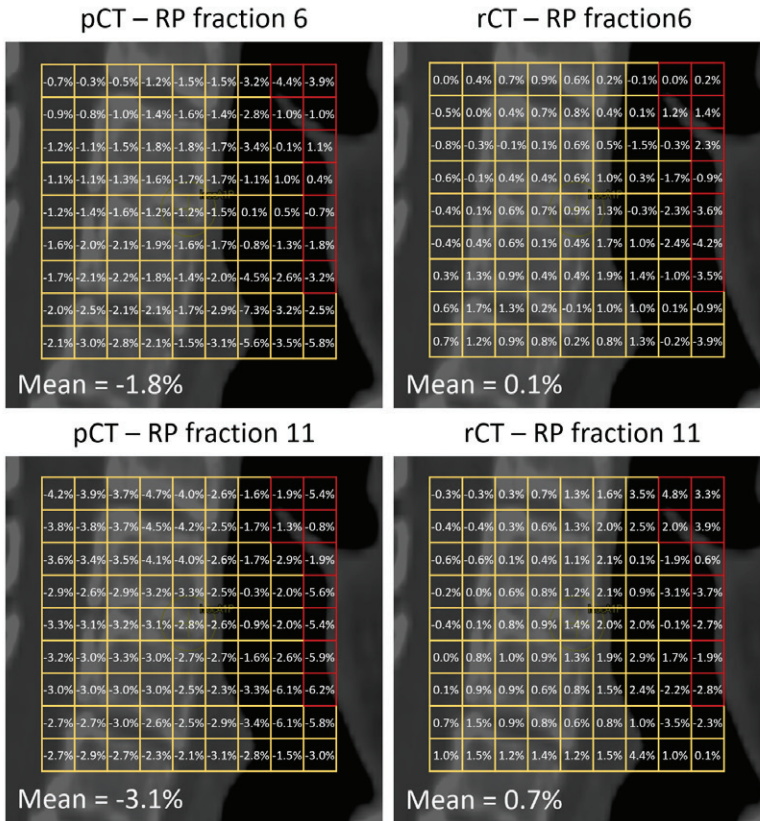
Figure 4. Grids of relative range errors overlaid with the sagittal planes for an exemplary patient (Patient 3). Relative range errors are shown for a comparison between measured IDDs and individually simulated IDDs on the planning CT (left) and the two repeat CTs (right). For the example patient, proton 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.

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.

Table 1. 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.").

		Mean relative range error (1.5 SD) [%]								
		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)

Discussion

A RP QC procedure has been applied in vivo for the first time and results for the first 7 HNC patients have been presented.

It can be noticed based on Table 1, 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 1 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 ($\sim 1\%$) 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 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 isocentre. 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 H_i 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 Figure 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: (1) overall accuracy of range error determination according to the RP QC method (~ 0.5 mm) [19], [20], [21], [25], which includes spot position accuracy, (2) energy fluctuations of the treatment delivery machine (~ 0.1 mm short-term to ~ 0.5 mm long-term), (3) residual setup errors

and intrafraction motion, including, for instance, muscle relaxation, in the treatment room (~ 1 mm), (4) anatomical inconsistencies of the patient between the treatment room and CT imaging room, (5) rigid registration of the planning and repeat CTs. Sources (4) and (5) 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 Figure 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 Figure 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.

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.

Acknowledgements

References

[1]   Wilson RR. Radiological use of fast protons. Radiology. 1946; 47:487–91.

[2]   Kjellberg RN, Koehler AM, Preston WM, Sweet WH. Stereotaxic instrument for use with the Bragg peak of a proton beam. Confin Neurol. 1962;22:183–9.

[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, Yamamori T, Suzuki M, Nagane M, et al. Evaluation of the relative biological effectiveness of spot-scanning proton irradiation in vitro. J Radiat Res. 2016;57:307–11

[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 Nov 21;56(22):7257-71.

[7] Bar E, Lalonde A, Zhang R, Jee KW, Yang K, Sharp G, 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., Wesp, P., Johnson, R.P., Pankuch, M., Bashkirov, V., Rit, S., Volz, L., Schulte, R.W. and Landry, G., 2019. Experimental comparison of proton CT and dual energy x-ray CT for relative stopping power estimation in proton therapy. Physics in Medicine & Biology, 64(16), p.165002.

[9] Meyer, S., Kamp, F., Tessonnier, T., Mairani, A., Belka, C., Carlson, D.J., Gianoli, C. and Parodi, K., 2019. Dosimetric accuracy and radiobiological implications of ion computed tomography for proton therapy treatment planning. Physics in Medicine & Biology, 64(12), p.125008.

[10] Li B, Lee HC, Duan X, Shen C, Zhou L, Jia X, Yang M., Comprehensive analysis of proton range uncertainties related to stopping-power-ratio estimation using dual-energy CT imaging., Phys Med Biol. 2017 Aug 9;62(17):7056-7074.

[11] Yang M, Zhu XR, Park PC, Titt U, Mohan R, Virshup G, Clayton JE, Dong L. Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration. Phys Med Biol. 2012 Jul 7;57(13):4095-115. doi: 10.1088/0031-9155/57/13/4095. Epub 2012 Jun 7. PMID: 22678123; PMCID: PMC3396587.

[12] Johnson RP. Review of medical radiography and tomography with proton beams. Rep Prog Phys. 2018 Jan;81(1):016701

[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 Aug 21;55(16):4771-82. doi: 10.1088/0031-9155/55/16/010. Epub 2010 Aug 3. PMID: 20679697.

[14] Hammi A, Koenig S, Weber DC, Poppe B, Lomax AJ. Patient positioning verification for proton therapy using proton radiography. Phys Med Biol. 2018 Dec 10;63(24):245009. doi: 10.1088/1361-6560/aadf79. PMID: 30188866.

[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 Oct 17;63(20):205016. doi: 10.1088/1361-6560/aae043. PMID: 30203783.

[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 Sep 21;59(18):5501-15. doi: 10.1088/0031-9155/59/18/5501. Epub 2014 Aug 29. PMID: 25170567.

[17] Richter C, Pausch G, Barczyk S, Priegnitz M, Keitz I, Thiele J, Smeets J, Stappen FV, Bombelli L, Fiorini C, Hotoiu L, Perali I, Prieels D, Enghardt

W, Baumann M. First clinical application of a prompt gamma based in vivo proton range verification system. Radiother Oncol. 2016 Feb;118(2):232-7.

[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 Aug 7;53(15):4137-51.

[19]   Meijers A, Free J, Wagenaar D, Deffet S, Knopf AC, Langendijk JA, Both S. Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing. Phys Med Biol. 2020 Feb 4;65(3):03NT02.

[20]   Meijers A, Seller Oria C, Free J, Bondesson D, Rabe M, Parodi K, Landry G, Langendijk JA, Both S, Kurz C, Knopf AC. Assessment of range uncertainty in lung-like tissue using a porcine lung phantom and proton radiography. Phys Med Biol. 2020 May 11.

[21]   Farace P, Righetto R, Meijers A., Pencil beam proton radiography using a multilayer ionization chamber., Phys Med Biol. 2016 Jun 7;61(11):4078-87.

[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 Mar 12;64(6):065008.

[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 Dec 29;63(1):015025. doi: 10.1088/1361-6560/aa9cff. PMID: 29176041.

[25]   Farace P, Righetto R, Deffet S, Meijers A, Vander Stappen F. 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 Dec;43(12):6405. doi: 10.1118/1.4966703. PMID: 27908151.

[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 Nov 1;105(3):504-513. doi: 10.1016/j.ijrobp.2019.06.2529. Epub 2019 Jul 2. PMID: 31271828.

[27]   Wang P, Yin L, Zhang Y, Kirk M, Song G, Ahn PH, Lin A, Gee J, Dolney D, Solberg TD, Maughan R, McDonough J, Teo BK. Quantitative assessment of anatomical change using a virtual proton depth radiograph for adaptive head and neck proton therapy. J Appl Clin Med Phys. 2016 Mar 8;17(2):427-440.

[28]   Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol. 2020 Mar 6.

Supplementary material 1

Table 1. 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.").

		Mean absolute relative range error (1.5SD) [%]								
		QC session 1					QC session 2			
		pCT		rCT		Fx #	pCT		rCT	
Pat #	Fx #	compl.	red.	compl.	red.		compl.	red.	compl.	red.
1	3	2.6	2.0	1.6	1.0	8	3.2	2.8	1.4	0.8
2	6	2.9	2.1	2.2	0.8	11	2.2	2.0	1.9	1.1
3	6	1.9	1.9	0.9	0.7	11	3.1	3.0	1.4	1.2
4	2	4.9	3.5	2.6	0.4	9	4.7	3.1	2.8	1.1
5	2	1.5	1.0	2.3	1.3	7	1.8	1.3	1.1	0.8
6	3	3.9	4.1	1.2	0.9	8	4.3	3.4	3.5	0.6
7	9	2.4	1.0	4.7	1.1	14	9.1	4.3	3.6	1.3

Chapter III: Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: Implementation and proof-of-concept

Abstract

Background and Purpose: Motion induced uncertainties hamper the clinical implementation of pencil beam scanning proton therapy (PBS-PT). Prospective pre-treatment evaluations only provide multi-scenario predictions without giving a clear conclusion for the actual treatment. Therefore, in this proof-of-concept study we present a methodology for a fraction-wise retrospective 4D dose reconstruction and accumulation aiming at the evaluation of treatment quality during and after treatment. ***Material and Methods:*** We implemented an easy-to-use, script-based 4D dose assessment of PBS-PT for patients with moving tumours in a

commercially available treatment planning system. This 4D dose accumulation uses treatment delivery log files and breathing pattern records of each fraction as well as weekly repeated 4D-CT scans acquired during the treatment course. The approach was validated experimentally and was executed for an exemplary data set of a lung cancer patient.

Results: The script-based 4D dose reconstruction and accumulation was implemented successfully, requiring minimal user input and a reasonable processing time (around 10 minutes for a fraction dose assessment). An experimental validation using a dynamic CIRS thorax phantom confirmed the precision of the 4D dose reconstruction methodology. In a proof-of-concept study, the accumulation of 33 reconstructed fraction doses showed a linear increase of D98 values. Projected treatment course D98 values revealed a CTV under dosage after fraction 25. This loss of target coverage was confirmed in a DVH comparison of the nominal, the projected (after 16 fractions) and the accumulated (after 33 fractions) dose distribution.

Conclusions: The presented method allows for the assessment of the conformity between planned and delivered dose as the treatment course progresses. The implemented approach considers the influence of changing patient anatomy and variations in the breathing pattern. This facilitates treatment quality evaluation and supports decisions regarding plan adaptation. In a next step, this approach will be applied to a larger patient cohort to investigate its capability as 4D quality control and decision support tool for treatment adaptation.

Introduction

Treatment of moving tumours in the lung with pencil beam scanning (PBS) proton therapy has been identified as being challenging early on [1]. This has mainly two causes: First, large density changes in the treated area. A tumour movement out of the planned treatment position largely impacts the range of the proton beam. Second, motion of the tumour takes place at the same time scale as the motion of the pencil beam during treatment, causing the so-called interplay effect by beam and motion interference. This effect can lead to highly inhomogeneous

dose distributions and was, up to now, substantially explored in phantom measurements and treatment planning studies [2–6]. These studies identified a multitude of influencing factors, e.g. patient-specific parameters like motion amplitude and tumour position at treatment start as well as machine-specific parameters like pencil beam spot size and the time required for spot repositioning. Most evaluations agree that the impact of the interplay effect is highly individual for a specific set of patient characteristics and machine parameters as well as their specific combinations per treatment fraction. Thus, it is hard to predict the dosimetric consequences of the tumour motion in advance in a pre-treatment evaluation. Such prospective pre-treatment evaluations consist of multiple scenarios, based on nominal patient and treatment machine characteristics. These analyses are clinically relevant as they aim on assessing before treatment whether the worst possible dose distributions would still be acceptable for treatment. Nevertheless, by taking such an approach, the treatment plan may be optimized to be robust against scenarios that might never occur during the actual treatment. Therefore, the actual dose degradation per fraction, which remains unknown in the multi-scenario approach, is also of great importance in the clinical assessment of the treatment quality and has been identified by the community as an "essential [need] for a comprehensive and safe clinical implementation of scanned particle treatment for moving targets" [7]. Subsequently accumulated dose distributions during the course of treatment can be used to support decisions on the treatment adequacy and the need of treatment adaptation. Furthermore, a realistic accumulated dose distribution for the full treatment course is a better basis for correlations with side effects and recurrences. The actual fraction dose distribution of PBS treatments can be evaluated retrospectively by using the required information from the treatment machine, i.e. at which time point precisely which spot was applied at which geometric position, and from the patient, i.e. in which position the patient anatomy was at the time of a particular spot delivery. Richter et al. [8] presented such an approach for liver treatment with carbon ions using an in-house developed software.

An experimental validation of an in-house deforming grid 4D dose calculation implementation for PBS proton therapy was recently presented by Krieger et al. [9]. A single-field plan was delivered to a homogeneous

PMMA phantom and measured by a high-resolution scintillating-CCD system. Various motion scenarios were simulated using a 4D Quasar phantom and logged by an optical tracking system in real-time. It was shown that the deforming grid 4D dose calculation was able to predict the complex patterns of 4D dose distributions with high dosimetric and geometric accuracy. In a paper by Klimpki et al. [10] the utility of this in-house deforming grid 4D dose calculation for the evaluation of different pencil beam scanning techniques in terms of effectiveness and efficiency of rescanning moving targets was demonstrated. A 4D dose calculation routine for PBS proton therapy using machine log files and motion pattern was also recently implemented in RayStation by Pfeiler et al. [11]. In a simple phantom setup it was shown that this implementation could be used to evaluate interplay effects. Our work demonstrates on a complex clinical case not only the possibility to employ current 4D dose calculation methodologies in clinical practice but also adds a novel subsequent dose accumulation, to facilitate decision making in adaptive radiotherapy.

In this paper, we describe a proof-of-concept study investigating the feasibility of fraction-wise retrospective dose reconstruction and subsequent dose accumulation in a commercial treatment planning system using its scripting functionality. For most realistic dose assessment, input data comprised the treatment delivery machine log files of 33 fractions and the patient's breathing patterns, which were acquired during treatment, as well as weekly acquired 4D-CT datasets taken throughout the treatment course. The precision of the 4D dose reconstruction methodology is experimentally validated using a dynamic CIRS thorax phantom.

Materials and Methods

Patient data

The study was performed with imaging and respiratory motion data from a patient with a non-small cell lung cancer stadium IIIB (T1aN3M0) treated within an on-going prospective clinical trial (PRONTOX, ClinicalTrials.gov Identifier NCT02731001 [12]) at OncoRay (Dresden, Germany), who had given informed consent to the scientific use of his data. Clinical target

volume (CTV) was defined with an 8 mm margin based on the internal gross tumour volume (iGTV) delineated on time-resolved computed tomography (4D-CT) images and the lymphnode GTV. Contoured organs at risk (OAR) comprised lungs, spinal cord, heart, brachial plexuses and oesophagus.

Imaging data

4D-CT imaging was performed with an in-room Siemens Somatom Definition AS (Siemens Healthineers, Germany) coupled to a pressure belt system (ANZAI, Japan) for motion surrogate retrieval. One pre-treatment 4D-CT and five weekly control 4D-CT datasets taken in treatment position at fractions 7, 13, 19, 25, and 31 were acquired, each comprising eight breathing phases and an average CT. The initial tumor motion (3D vector of the centroid position of the tumor) in the planning 4D-CT was of 0.7mm, varying between 0.5mm, 0.8mm, 0.3mm, 0.4mm and 1mm in the repeated 4D-CTs of week 1, 2, 3, 4 and 5.

Motion output files were acquired during 22 out of 33 fractions delivered in free breathing, using the same ANZAI belt as during 4D-CT acquisition (Figure 1). For the remaining 11 out of 33 fractions motion monitoring was not possible due to practical reasons in the clinical workflow (e.g. delay in treatment schedule resulting in the decision to skip optional respiratory signal measurement).

Image registration

Six degrees-of-freedom rigid registrations were performed between the planning 4D-CT and the sequential control 4D-CTs. The registrations focused on bony structures, especially the vertebral bodies close to the target, similar to the actual treatment setup. For voxel mapping, deformation vector fields were generated between the planning and control 4D-CTs using the built-in hybrid deformable image registration algorithm ANACONDA of the treatment planning system RayStation 6 (RaySearch, Sweden). ANACONDA combines image information, such as intensities, with anatomical information provided by contoured image sets and its performance was recently assessed in comparison to various other deformable image registration methods [13]. In our data set the lungs were used as controlling region of interest. The quality of

the deformation vector fields was assessed by mapping and manually reviewing the structures from the reference image to the other CT image datasets as well as checking the vector fields visually for consistency and plausibility.

Treatment planning

PBS proton treatment planning was done with RayStation 6 using the CTV as target and the IGTV density assigned to muscle tissue. The PBS plan was designed with two beams (0°, 145°) and optimised aiming for robustness against 5 mm setup and 3.0% range uncertainty. Optimisation focused on adequate target coverage: V_{95}>98% of prescribed dose of 66 Gy(RBE) and minimisation of OAR dose. The Monte Carlo dose calculation engine was used to calculate the final dose. Dose evaluations were based on perturbed dose and were acceptable with nominal plan values for CTV V_{95}=97.4%, spinal cord D_1=29.5 Gy(RBE), heart D_{mean}=2.7 Gy(RBE), lungs D_{mean}=8.2 Gy(RBE) and oesophagus D_{mean}=17.6 Gy(RBE).

Simulation of PBS treatment – 4D calculation

Thirty-three machine log files of the PBS treatment were obtained from dry run deliveries at the clinical proton treatment facility at UMCG (Groningen, Netherlands) equipped with the IBA Proteus®PLUS system (IBA, Belgium).

Post-processing of the output files of the ANZAI motion monitoring system and the IBA machine delivery log files were performed with in-house scripts (Python), which can be executed directly in RayStation via the scripting module. Analysed log file data included spot energy, spot position on IEC x and y axis at iso-centre plane, delivered monitor units per spot and absolute time point of spot delivery.

For each fraction, the dose of each delivered spot was mapped to a 4D-CT phase by combining the log file and a selected part of an ANZAI output file. Figure 1 shows which part of the ANZAI output file was used for specific fraction. For this proof-of-concept, this selection was arbitrary just assuring that the selected part of an ANZAI output file covered the beam delivery duration. Spot delivery times were compared to the ANZAI output file, determining the breathing phase of the patient at spot delivery and assigning the spots to the corresponding 4D-CT phases. For each
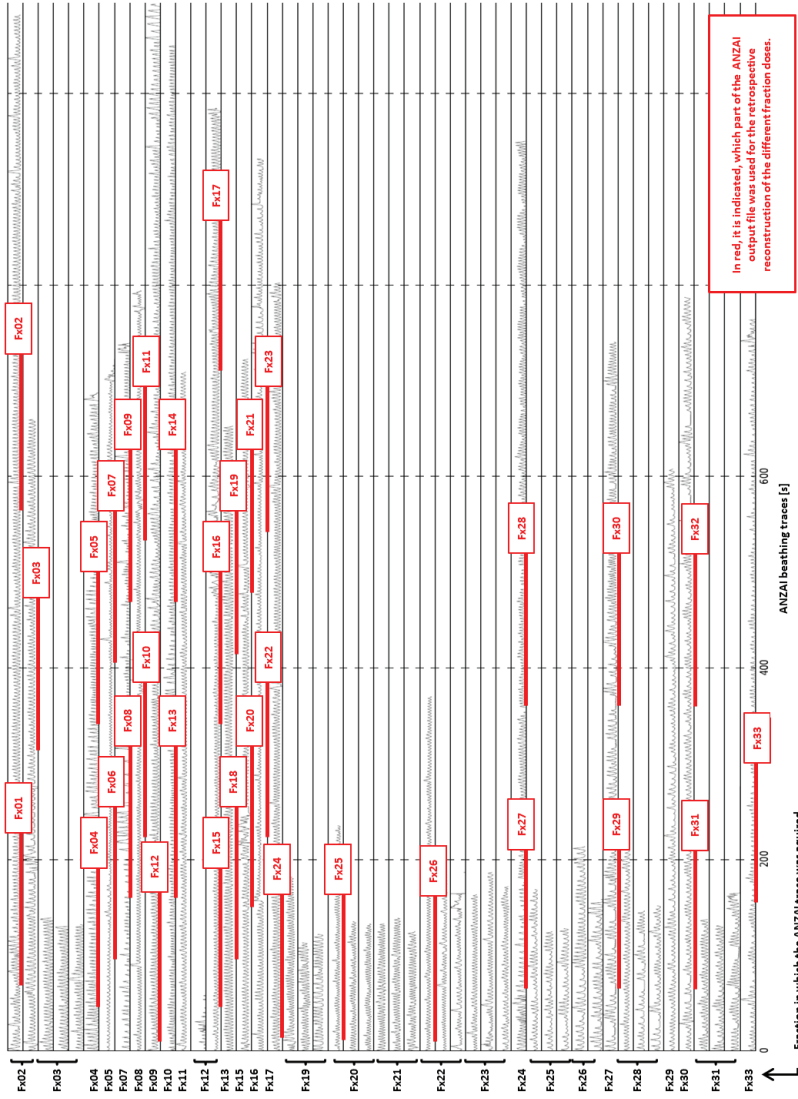
Figure 1. Overview of used motion monitoring data.

In red, it is indicated, which part of the ANZAI output file was used for the retrospective reconstruction of the different fraction doses.

Figure 2. Overview of the data set used to demonstrate the workflow for fraction-wise retrospective 4D dose reconstruction for moving targets.

fraction, the most recent 4D-CT image dataset was used, as illustrated in Figure 2. With the available data, a set of eight sub-plans in DICOM format corresponding to the eight 4D-CT phases was created for each fraction. Each sub-plan contained only the spots that corresponded to the associated phase of the breathing cycle.

Those sub-plans were re-imported in RayStation and recalculated on the corresponding 4D-CTs. The dose distributions of the different respiration phases were warped using deformable image registration, and summed on the reference CT. This procedure resulted in one dose distribution per fraction taking into account weekly anatomical changes, daily varying organ motion, setup changes and the interplay effect. The accumulated treatment course dose was calculated at the end of the treatment by summing all fraction doses. Furthermore, projected full treatment course doses D_n after x of n fractions following formula [1] were calculated after each treatment fraction by summing the available fractions and scaling them to match the total course dose.

$$D_n = \frac{n}{x} \cdot \sum_{i=1}^{x} D_i \tag{1}$$

The entire proposed clinical workflow is sketched in Figure 3.

Experimental validation

To experimentally validate the 4D dose reconstruction approach described in the previous section, a dynamic CIRS thorax phantom (CIRS, Virginia, USA) mimicking an average human thorax in shape, proportion and composition was utilized. A rod, made out of lung equivalent material, contained a spherical high-density target structure of about 9 mm in diameter and was inserted into the lung equivalent lobe of the phantom. The rod was connected to a motion actuator and a cyclic motion trajectory was predefined. Over one simulated breathing cycle, the motion range of the spherical high-density structure was 2 cm in sup-inf, 1 cm in ant-post directions and no motion laterally. An ANZAI belt was connected to the phantom surrogate motion platform to obtain ANZAI motion monitoring output data. A 4D CT of the phantom was acquired and reconstructed into 10 phases. During the 4D CT acquisition breathing cycle of 4 sec was used.
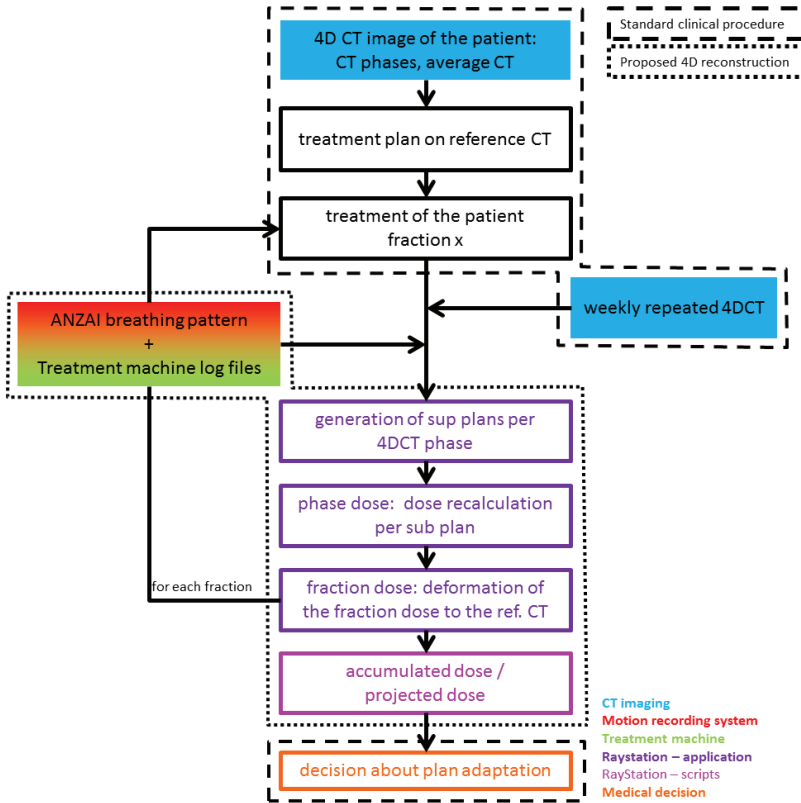
Figure 3. Implemented workflow for fraction-wise retrospective 4D dose reconstruction for moving targets supporting the medical decision in respect to treatment adaptations.

A monoenergetic squared anterior — posterior directed PBS proton field of 115 MeV and 5 cm × 5 cm size was delivered. The energy of the beam was chosen in order to ensure that the protons would pass through the phantom when not passing through the high density target structure and vice versa. The size and position of the field was chosen to cover the entire area of the projected movement of the spherical target structure. The phantom was aligned to the treatment isocentre by using on-board x-ray imaging system. A dosimetric measurement was performed by placing a Gafchromic EBT3 film (Ashland Inc., Covington, USA) posteriorly below the phantom. Two measurements were performed: First, the target was
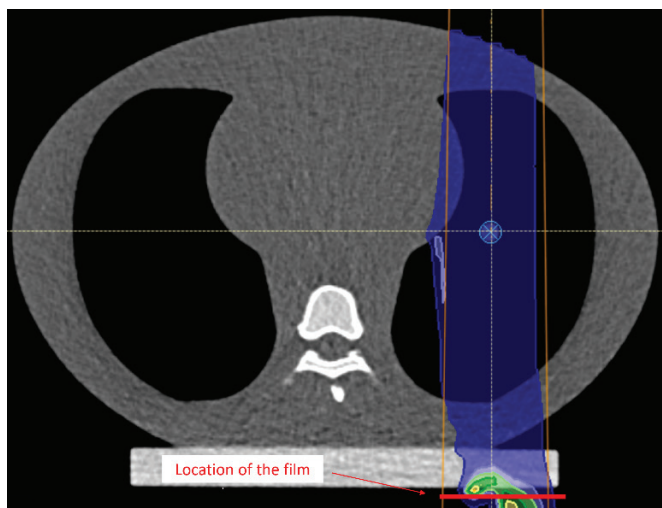
Figure 4. Transversal slice of the dynamic CIRS thorax phantom. An anterior — posterior monoenergetic beam is used to cover the area enclosing the motion range of the spherical target structure. The film is placed posteriorly below the phantom, perpendicular to the beam axis.

moved with a simulated breathing cycle of 4 s. Secondly, a breathing cycle of 8 s was applied. During the beam delivery the "breathing motion" of the phantom was registered with the ANZAI system. The beam-on time was 2.8 s, therefore only a few pencil beams traversed the spherical high-density structure, resulting in a low dose shadow on the Gafchromic EBT3 film. The position of this shadow in respect to the field edges depends on the timing of the beam delivery (scanning) in respect to the timing of the motion of the spherical high-density structure, in other words, on the interplay. The experimental setup is shown in Figure 4.

To validate the proposed 4D dose reconstruction approach, the dose profiles of the measured films were compared to the 4D reconstructed dose, using the 4D CT, the machine log file and the ANZAI output signal as described in the previous section.

Results

The timeline of the PBS delivery was correlated with a different part of the motion monitoring data for each fraction (Figure 1). The correlation for one exemplary fraction is shown in Figure 5. Thirty-three fraction doses were recalculated using different parts of the ANZAI output files, the six different 4D-CT datasets and machine log file of 33 dry run irradiations. The dose distributions of the log-based sub-plans and the resulting fraction dose is shown for an exemplary fraction in Figure 6.

Fraction doses were subsequently accumulated. Figure 7 shows the linear increase of D98 values over the entire treatment course. Following formula [1] projected full treatment course doses were calculated after the delivery of each fraction. After an initial projected increase of D98 values, a drop after fraction 6 was recognized (Figure 7). At fraction 25 the projected D98 values drop below the D98 tolerance level, indicating an under-dosage of the CTV.

A comparison of the nominal dose distribution, the projected full treatment course dose after 16 fractions and the 33 accumulated reconstructed fraction doses reveals a CTV under dosages of the projected and accumulated dose distribution compared to the nominal treatment dose (Figure 8). Almost no changes in spinal cord dose, mean heart and lung doses were observed.
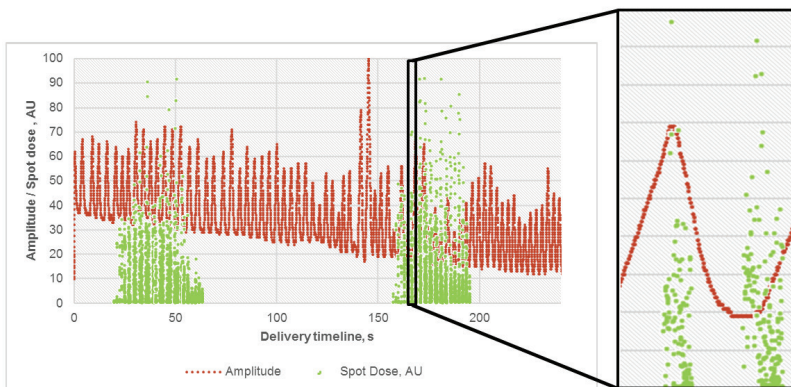
Figure 5. PBS plan delivery timeline based on delivery machine log file (green dots) overlaid on patient's breathing pattern (red dots). On the right, a zoom in can be seen.
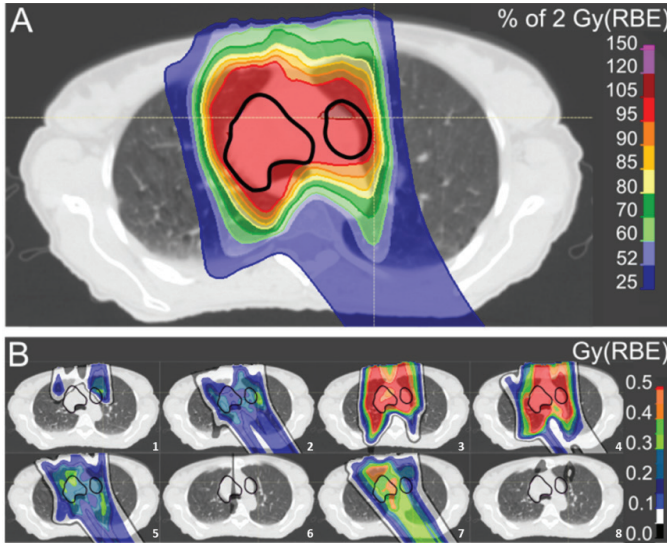
Figure 6. 4D dose of one fraction (A) calculated with the presented method based on sub-plan doses for eight 4D-CT phases (B). CTV is shown in black.
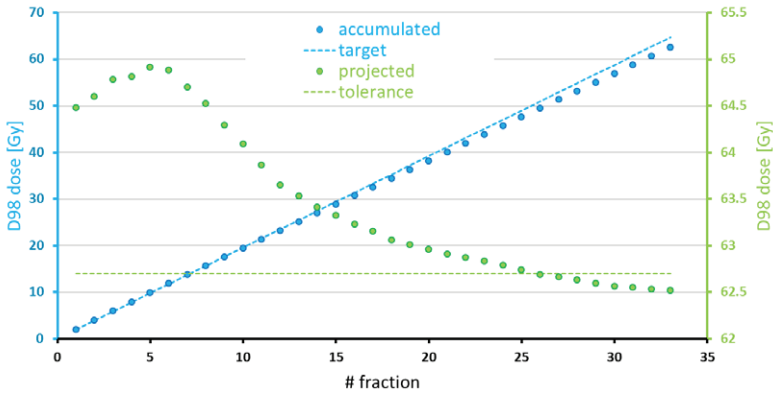
Figure 7. Accumulated (blue) and projected (green) D98 dose after each fraction.

Required time for processing of the machine log file and the ANZAI output file was 30 seconds for one fraction. We estimate that these calculations require between 20 and 120 seconds for any arbitrary patient, depending on the number of treatment fields and spots. Sub-plan import
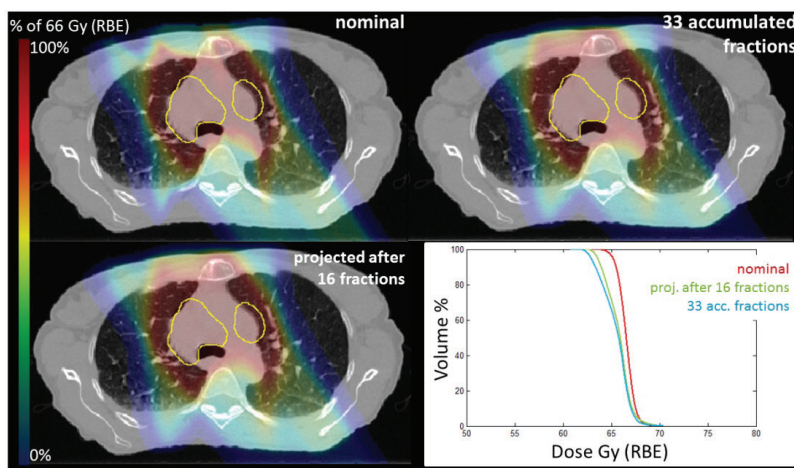
Figure 8. Comparison of nominal dose, the projected dose after delivering 16 fractions and the accumulated dose at the end of treatment course. The CTV is displayed in yellow. On the lower right, DVH plots for the nominal (red), the projected (green) and the accumulated (blue) dose distributions for the CTV can be seen.
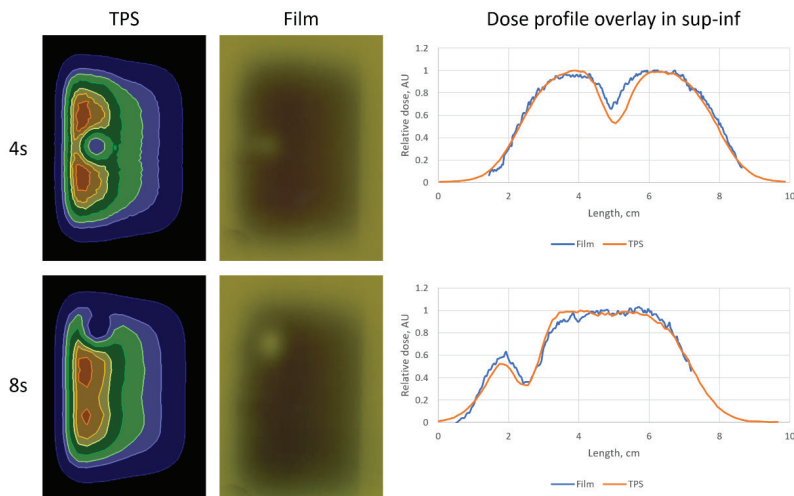
Figure 9. Comparison between dose planes as experimentally acquired on Gafchromic film and as retrieved from 4D reconstructed dose distribution by using treatment delivery log files and motion information monitored with an ANZAI system. Profile analysis shows agreement between position of the target structure as according to the measurement and as according to the 4D dose reconstruction within 1 mm.

and dose accumulation for one fraction takes around 4–8 minutes, depending on the number of sub-plans (or in other words the number of 4D-CT phases). In addition, a manual data preparation is required, namely placing the current input files (log files, ANZAI output) in a dedicated folder and specifying relevant computation settings in a setup file. Calculation of projected course dose or total course dose based on fraction doses need additionally less than five minutes for manual input in RayStation.

Experimental validation

Figure 9 shows a comparison of the Gafcromic EBT3 film measurement and the 4D reconstructed dose. The position of the low dose shadow in respect to the field edges agrees with a millimetric accuracy, demonstrating the accuracy of the proposed 4D dose reconstruction methodology using machine log files and ANZAI output data.

Discussion

In this proof-of-concept study, we showed the feasibility of fraction-wise retrospective assessment of the actually delivered dose distribution within a commercial treatment planning system. Our approach incorporates treatment machine log files and recorded breathing patterns for each fraction as well as multiple 4D-CT datasets acquired throughout the treatment course. The 4D dose reconstruction per fraction is performed completely in RayStation 6 via its scripting module. With such an easy-to-use method, the accumulated delivered dose at any time point during the treatment course or a projected full treatment course dose can be quickly assessed by summing up the doses to the current treatment fraction and for the projection, scaling it with the number of total fractions. The accumulated delivered dose or the projected treatment course dose is relevant for monitoring the quality of the treatment delivery while progressing through the treatment and can support decisions regarding the necessity of treatment plan adaptation. The feature of considering actual instead of nominal quantities for tumour motion, beam delivery and the subsequent consideration of the actual interplay cannot be overrated, even if still

estimations apply. These are essential contributions to the high-quality implementation of PBS proton therapy of moving targets.

Experimental validation

The precision of the 4D dose reconstruction methodology was experimentally validated using a dynamic CIRS thorax phantom. Two shoot-through irradiations of the phantom, operating in a slow and in a fast mode, resulting in different interplay between the motion and the delivery timeline, were performed. In order to create a sharp low dose shadow, the delivery time was minimized. Given the high sensitivity of the interplay effect on the motion and delivery timeline, the millimetric accuracy of the measured and reconstructed low dose shadow in both experiments validates the 4D dose reconstruction methodology. However, more experiments should be performed using realistic delivery times and iso-centric dose measurements to confirm the accuracy for clinical settings. Achieving this in a anthropomorphic 4D phantom remains however challenging.

4D dose reconstruction quality

The frequency of the input 4D-CT acquisition has an impact on the dose reconstruction quality. More scans which are acquired as close as possible to the actual treatment will increase the quality of the reconstructed dose. This frequency as well as its off-line nature is a limitation in the precision of dose reconstructions. The main assumption is that the breathing motion pattern and anatomical situation of the patient is the same during image acquisition and treatment, which is not always valid. Using weekly 4D-CT datasets, as in our proof-of-concept study, is a reasonable compromise between the clinically manageable frequency of 4D-CT acquisitions, the additional imaging dose to the patient and the highest dose reconstruction quality. As the differences in motion pattern and anatomical changes for our exemplary patient case were rather small between the consecutive 4D-CTs, the chosen frequency was sufficient. For other patients another frequency might be appropriate which could be assessed by comparing several consecutive 4D-CT datasets or 4D-CBCTs. The implementation of the script-based dose reconstruction is, however, independent of the number of images, and can be used "as is" including as many images as available. It is also possible to include synthetic daily CT data which could

be created based on daily 4D cone-beam CT or magnet-resonance (MR) information [14–16]. These developments aim on the almost or indeed online assessment of motion and anatomical changes. Although this approach would reduce the uncertainty in potentially outdated anatomy and motion data, there is still a lot of translational work to be done like implementing reliable dose calculations for PT based on cone-beam CT and MR, before this could be actually used in a clinical setting.

Additional uncertainty arises from the precision of the applied deformable image registration in RayStation. The vector fields were assessed qualitatively with tools in RayStation and rated as sufficiently precise for the given purpose. It is highly important to evaluate the vector field quality for each deformable image registration to assess the uncertainties introduced in the dose calculation as highlighted in Ribeiro et al. [13]. Quantitative evaluation, however, is largely time consuming and complicated (e.g. based on manually choosing landmarks in both registered datasets) [17] and currently not featured in RayStation.

Current technical issues

A practical limitation in the presented retrospective 4D dose reconstruction is the limited capability of RayStation to work with a large number of dose distributions. This might be avoided, if only fraction doses are used inside RayStation without storing the sub-plan doses, decreasing the number of handled dose cubes by a factor of 9. An adaption by RaySearch to enable the handling of larger amounts of data is desirable.

A difficult step in the general implementation of the retrospective dose assessment is the synchronisation between the time stamps of the treatment machine log files and the breathing curve output of the ANZAI system. Richter et al. described the remaining synchronisation accuracy in their system to be about 250 ms, despite their efforts [8]. This remaining uncertainty had a large influence on the estimated dose distributions, with up to 25% variation in parameters like V_{95} and V_{107} of the target. This underlines the importance of a precise synchronisation process. They concluded that the synchronisation accuracy needs to be better than 25 ms to achieve acceptable 4D dose reconstruction. However, this problem largely depends on the individual software setup and IT infrastructure and needs to be addressed individually in each implementation.

In our approach, we used treatment machine log files from an IBA proton machine and the breathing pattern output from the ANZAI system. However, in principle, the scripts can be adapted to the output of any respiratory motion management system and any log file output. Furthermore, different CT systems use different reconstruction algorithms (in general, phase-based versus amplitude-based reconstruction). The breathing pattern analysis and 4D-CT assignment would need to be adapted to another input system if required. Although, the scripts were implemented for RayStation 6, the adaptation of these scripts to fit the new requirements of RayStation 7 or later should be feasible without major changes.

In summary, the retrospective 4D dose reconstruction and accumulation, now implemented in Groningen and Dresden, allows for an approximation of the actually delivered dose and is an extremely powerful tool in supporting the quality assessment of PBS proton therapy for moving tumours. With a time duration of 10 minutes per fraction assessment, the approach is clinically feasible, contrasting a time consuming prospective 4D evaluations considering a comprehensive parameter space of different organ motions, various interplay scenarios, different anatomical variations and multiple setup scenarios.

In future, the retrospective 4D dose reconstruction and accumulation is foreseen to be used for assurance of the treatment delivery quality and to support decisions on plan adaptation. To enable this, (site–specific) treatment quality indicators and action levels must be defined to trigger plan adaptation if required. One possible adaptation strategy could be the introduction of an additional motion mitigation technique (for example, breath-hold) at a given time point of fractioned treatment, for example when an increasing target motion amplitude would compromise the treatment quality. Alternatively, one could re-optimize the treatment plan using the 4D reconstructed dose as background dose and eventually use the existing control CTs for robust optimization (anatomical robustness). Any adapted treatment course should then be further monitored, assuring that further deviations shown in the subsequently reconstructed and accumulated 4D treatment are within acceptable margins.

The described proof-of-concept mainly focuses on 4D treatments, however the methodology can be generalized and applied to indications

across the body. To reconstruct and accumulate dose for treatment sites not affected or minimally affected by the organ motion, the amount of required input data would simply be smaller and the number of calculations would be decreased.

Conclusion

The development of a 4D-dose-accumulation treatment-assessment tool, now ready for clinical application, allows for assessing the quality of the delivered dose throughout the treatment course, taking appropriate actions, e.g. plan adaptations, in case of significant deviations. We implemented and experimentally validated this tool in a commercially available treatment planning system using its scripting functionality for an easy-to-use procedure feasible in a clinical setting. This is an essential step towards safe clinical implementation of PBS proton treatments for moving targets. In the next step, the general clinical relevance of the approach will be evaluated on a broader patient population and compared to prospective dose evaluations.

Acknowledgments:

References

[1] Phillips, M. H., Pedroni, E., Blattmann, H., *et al.* Effects of respiratory motion on dose uniformity with a charged particle scanning method. *Phys. Med. Biol.* 1992; **37:** 223–233.

[2] Bert, C., Grözinger, S. O. & Rietzel, E. Quantification of interplay effects of scanned particle beams and moving targets. *Phys Med Biol* 2008; 53: 2253–2265.

[3]   Kraus, K. M., Heath, E. & Oelfke, U. Dosimetric consequences of tumour motion due to respiration for a scanned proton beam. *Phys. Med. Biol.* 2011; 56: 6563–81.

[4]   Grassberger, C., Dowdell, S., Lomax, A., *et al.* Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2013; 86: 380–386.

[5]   Jakobi, A., Perrin, R., Knopf, A., Richter, C., Feasibility of proton pencil beam scanning treatment of free-breathing lung cancer patients. *Acta Oncologica* 2017; 57: 203-210.

[6]   Knopf, A., Hong, T.S., 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(22):7257-71.

[7]   Knopf, A.-C., Stützer, K., Richter, C., *et al.* Required transition from research to clinical application: Report on the 4D treatment planning workshops 2014 and 2015. *Phys. Medica* 2016; 32: 874–882.

[8]   Richter, D., Saito, N., Chaudhri, N., *et al.* Four-dimensional patient dose reconstruction for scanned ion beam therapy of moving liver tumors. *Int. J. Radiat. Oncol. Biol. Phys.* 2014; 89: 175–181.

[9]   Krieger, M., Klimpki G., Fattori G., Hrbacek J., Oxley D., Safai S., Weber D. C., Lomax A. J. Experimental validation of a deforming grid 4D dose calculation for PBS proton therapy. *Phys. Med. Biol.* 2018; 63: 055005 (11pp).

[10]   Klimpki G., Zhang Y., Fattori G., Psoroulas S., Weber D.C., Lomax A. and Meer D. The impact of pencil beam scanning techniques on the effectiveness and efficiency of rescanning moving targets. *Phys. Med. Biol.* 2018; 63: 145006 (13pp).

[11]   Pfeiler T., Bäumer C., Engwall E., Geismar D. Spaan B., Timmermann B. Experimental validation of a 4D dose calculation routine for pencil beam scanning proton therapy. *Z Med Phys* 2018; 28: 121–133.

[12]   Zschaeck, S., Simon, M., Löck, S., *et al.* PRONTOX — proton therapy to reduce acute normal tissue toxicity in locally advanced non-small-cell lung carcinomas (NSCLC): study protocol for a randomised controlled trial. *Trials* 2016; 17: 543.

[13]   Ribeiro, C.O., Knopf, A., Langendijk, J.A., Weber, D.C., Lomax, A.J., Zhang, Y.. Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours. *Radiother Oncol.* 2018; S0167-8140(18); 30133-6.

[14]   Kurz, C., Nijhuis, R., Reiner, M., *et al.* Feasibility of automated proton therapy plan adaptation for head and neck tumors using cone beam CT images. *Radiat. Oncol.* 2016; 11: 64.

[15]   Veiga, C., Janssens, G., Teng, C.-L., *et al.* First Clinical Investigation of Cone Beam Computed Tomography and Deformable Registration for Adaptive

Proton Therapy for Lung Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2016; 95: 549–559.

[16]   Zhang, Y., Knopf, A., Tanner, C., Boye, D. & Lomax, A. J. Deformable motion reconstruction for scanned proton beam therapy using on-line x-ray imaging. *Phys. Med. Biol.* 2013; 58: 8621–8645.

[17]   Stützer, K., Haase, R., Lohaus, F., Barczyk, S., Exner, F., Löck, S., Rühaak, J., Lassen-Schmidt, B., Corr, D., Richter, C. Evaluation of a deformable registration algorithm for subsequent lung computed tomography imaging during radiochemotherapy. *Medical Physics* 2016; 43(9); 5028-39.

Chapter IV: Assessment of range uncertainty in lung-like tissue using a porcine lung phantom and proton radiography

Abstract

Purpose: Thoracic tumours are increasingly considered indications for pencil beam scanned proton therapy (PBS-PT) treatments. Conservative robustness settings have been suggested due to potential range straggling effects caused by the lung micro-structure. Using proton radiography (PR) and a 4D porcine lung phantom, we experimentally assess range errors to be considered in robust treatment planning for thoracic indications.
Methods and Materials: A human-chest-size 4D phantom hosting inflatable porcine lungs and a corresponding 4DCT were used. Five PR frames were planned to intersect the phantom at various positions. Integral depth-dose curves (IDDs) per proton spot were measured using a multilayer ionisation chamber (MLIC). Each PR frame consisted of 81 spots with an assigned energy of 210 MeV (FWHM 8.2 mm). Each frame was delivered 5 times while simultaneously acquiring the breathing signal of

the 4D phantom, using an ANZAI load cell. The synchronised ANZAI and delivery log file information was used to retrospectively sort spots to their corresponding breathing phase. Based on this information, IDDs were simulated by the treatment planning system (TPS) Monte Carlo dose engine on a dose grid of 1 mm. In addition to the time-resolved TPS calculations on the 4DCT phases, IDD were calculated on the average CT. Measured IDDs were compared with simulated ones, calculating range error for each individual spot.

Results: In total 2025 proton spots were individually measured and analysed. Range error of a specific spot is reported relative to its water equivalent path length (WEPL). The mean relative range error was 1.2% (1.5 SD 2.3%) for the comparison with the time-resolved TPS calculations and 1.0% (1.5 SD 2.2%) when comparing to TPS calculations on the average CT.

Conclusions: Determined mean relative range errors justify the use of 3% range uncertainty for robust treatment planning in clinical setting for thoracic indications.

Introduction

Proton therapy is increasingly employed for the treatment of thoracic indications, such as non-small cell lung cancer (NSCLC) (Macdonald *et al* 2009). In thoracic patients, proton therapy may reduce radiation-induced complications, such as radiation pneumonitis, which is linked to the lung dose (Appelt *et al* 2014, Palma *et al* 2013), dysphagia, which correlates with dose to the oesophagus (Wijsman *et al* 2015, Zhu *et al* 2010, Belderbos *et al* 2005, Gomez *et al* 2012, El Naqa *et al* 2006) or a reduced life expectancy (2-year mortality), which is correlated with the dose to the heart (Wang *et al* 2017).

The adoption of proton therapy for the treatment of thoracic tumours, especially when applied with Pencil Beam Scanning (PBS), has been hampered by numerous sources of uncertainty, which can potentially have severe dose distribution degrading effects. Sources of uncertainty are linked to organ motion, interplay effects (Bert and Rietzel 2007, Paganetti 2012) and range uncertainty, among others. Regarding the range

IV

uncertainty, literature suggests that uncertainty as high as 5% might be applicable for lung tissue.

In addition, it has been speculated that due to the micro-structure of the lung, which cannot be properly detected with computed tomography (CT) scans, proton beams may be affected by an increased range straggling effect, which would enlarge range uncertainty further (Uries *et al* 1986, Titt *et al* 2015, España and Paganetti 2011, Baumann *et al* 2019, Perles *et al* 2011, Sell *et al* 2012).

Most of the recent work is based on Monte Carlo simulations, as experimental range measurements in lung-like tissue are not straightforward to perform. Practically, it is rather challenging to maintain lung tissue air-ventilated in *ex vivo* conditions. Therefore, when assessing range accuracy in lung-like tissue or performing CT calibration for proton dose calculations, simplified tissue substitutes are often used. These are materials similar to lung tissue in terms of physical density, however, they do not resemble the elemental composition or structure of actual lung tissue. As a result, lung-like tissue is associated with increased range uncertainty (Paganetti 2012). Nonetheless, artiChest (PROdesign, Heiligkreuzsteinach, Germany) (Biederer and Heller 2003, Etzold 2020) is a phantom, which allows to ventilate porcine lungs within a plastic shell while mimicking breathing motion.

Furthermore, proton radiography (PR) is a proton imaging technique introduced and investigated by numerous groups (Parodi 2019). It relies on detecting high energy protons passing through a sample-of-interest, positioned in the beam path. The detected signal can give insight into the water equivalent thickness (WET) of the materials in the beam path. This can be useful, for instance, for patient positioning or to provide an input for assessment of range uncertainty in the proton treatment planning process.

In our clinic we have adopted a PR technique (also referred to as range probing) as described by Farace et al. (Farace *et al* 2016). It relies on measuring individual integral depth-dose curves (IDDs) per shoot-through proton pencil beams distally from the sample-of-interest, using a multi-layer ionisation chamber (MLIC). The technique has been demonstrated to be useful for validation and optimisation of CT calibration curves by Meijers et al. (Meijers *et al* 2020). In this previous study range errors were assessed in bone- and soft tissue-like phantoms. In particular, one of the

phantoms was designed as a "thorax" phantom, consisting of ribs, liver, fat and muscle attached to a low-density styrofoam block, which allowed to assess range uncertainties in tissue types found in thorax, except lung tissue itself. In the previous study 1.5 SD of range errors was found to be within theoretical range uncertainty margin of 2.4% + 1 mm.

The objective of this study was to experimentally assess range errors in an air ventilated porcine lung using PR and, in combination with the previous study, which investigated bone- and soft tissue-like phantoms (Meijers *et al* 2020), to justify the choice of range uncertainty for robust thorax treatment planning to be used in a clinical setting. In this context, range errors were determined as discrepancy between treatment planning system (TPS) predicted ranges versus measured ranges through the porcine lung tissue phantom.

Material and methods

Preparation

A 4D phantom artiCHEST, which consists of two water-filled plastic shells in the shape and dimensions of a human thorax and which can host *ex vivo* porcine lungs, was used (see Figure 1). *Ex vivo* porcine lungs, including parts of the trachea, were inspected for their intactness and were prepared for the experiment within 24 h after extraction. The 4D phantom was configured to mimic a 4 s long breathing cycle. This was accomplished by a 2-way air pumping system. A first pump ensures vacuum in the chest cavity while the bronchia is open to atmospheric pressure, letting the porcine lungs inflate. A second pump drives an air-operated diaphragm, which can produce cyclical motion, following a user-defined breathing curve. A phase-binned 4DCT scan (Somatom Definition AS, Siemens Healthineers, Erlangen, Germany) was performed and reconstructed into scans of 10 breathing phases and an average scan, all imported into TPS (RayStation 8B, RaySearch Laboratories, Stockholm, Sweden). The breathing signal of the 4D phantom was collected by using an ANZAI load cell (Anzai, Tokyo, Japan). The clinical CT calibration curve in the TPS was used to correlate mass densities to the CT numbers of the 4DCT scan. The CT calibration curve was defined following the stoichiometric
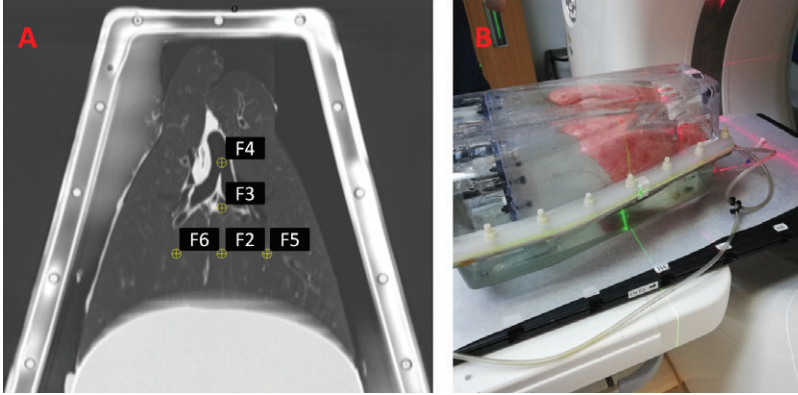
Figure 1. (A) Coronal view of the 4D phantom (average CT) showing the centres of the five proton radiography frames. (B) Setup of the 4D phantom at the CT scanner.

method (Schneider *et al* 1996) with minor deviations and validation as described by Meijers et al. (Meijers *et al* 2020). Proton stopping power ratios (SPR) are calculated by the TPS based on the provided mass density to CT number curve and internal material look-up table.

Five PR fields (i.e., frames) in anterior-posterior (AP) direction were planned to intersect the 4D phantom at various positions. The centres of the frames (F2–F6), as well as the setup of the 4D phantom are shown in Figure 1. Areas, which were covered by PR frames, were subject to a motion of up to 9 mm. The highest extent of the motion was observed in frame 2.

The AP beam direction selection was driven by two aspects: (1) currently the acquisition is technically feasible only along cardinal axis and (2) the AP direction allowed to design some of the frames (F5 and F6) in a way that only lung tissues are intersected, without soft tissue being in the beam path, this way allowing to focus on range errors in lung tissues.

Each PR field consisted of 81 spots separated by 5 mm and covering an area of $4 \times 4 \, cm^2$. All spots were assigned an energy of 210 MeV and 1 MU/spot. On the used system full width at half maximum (FWHM) in air at the isocentre for spots with 210 MeV energy is 8.2 mm, according to the spot size measurements performed during the commissioning phase.

Focusing on porcine tissue sample anatomy, proton spots in frames 5 and 6 were intersecting lung tissue only, while frames 2, 3 and 4 were
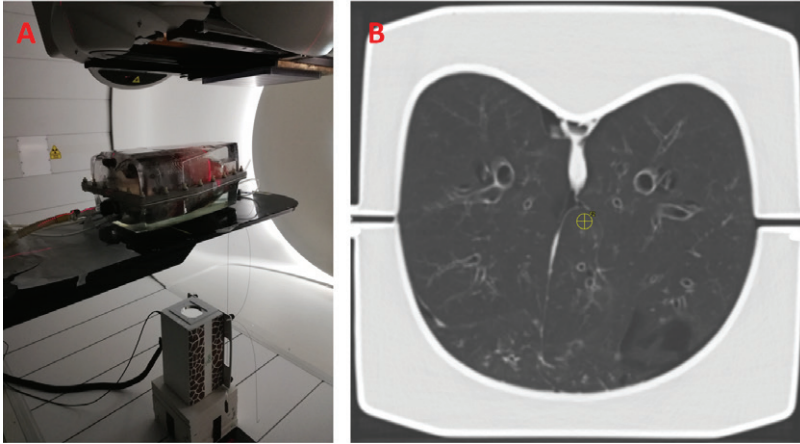
Figure 2. (A) Setup of the 4D phantom and measurement equipment in the treatment room. (B) Transversal view of the 4D phantom.

IV

located in mediastinum area, therefore proton spots were intersecting soft tissue, vessels and tracheas in addition to lung tissue.

Measurements

In the proton treatment room (Proteus Plus, IBA, Louvain-la-Neuve, Belgium) the 4D phantom was positioned on the treatment table supported by 6D robotic positioning system and aligned to the treatment room isocentre by using the on-board cone-beam computed tomography (CBCT) system. Positioning accuracy of the robotic arm is < 0.5 mm, as demonstrated by recurrent quality assurance (QA) procedures.

The MLIC (Giraffe, IBA Dosimetry, Schwarzenbruck, Germany) was positioned distally from the 4D phantom below the treatment table 63 cm from the isocenter on a slab of solid water, as shown in Figure 2. The MLIC was aligned to the isocentre along the beam axis.

Each of the five PR frames was delivered 5 times in order to ensure that various phases of the breathing cycle were sampled by each proton spot. Delivery time of a single frame is approximately 12.5 seconds. The phantom was repositioned between the frames by applying an offset with the robotic arm. The breathing signal of the 4D phantom was acquired using ANZAI load cell during the acquisition of the PR frames. Cell was

connected via an adaptor to the hose, which was operating the diaphragm. Before the experiment the ANZAI system and proton therapy system (PTS) were synchronised to the same time server to ensure consistent absolute time reference for both systems. After delivery of the PR frames, treatment delivery log files of the PTS were collected.

Since measurements were performed for fields in AP direction, proton beams were passing through the treatment table (Qfix Standard insert, Qfix, PA, USA). During the clinical commissioning of the patient positioning devices the WET of the table insert was determined to be 5 mm. This value was taken into account during data analysis, by shifting measured IDDs accordingly.

During measurements, the proton spots were passing through various materials in the 4D phantom in the following order (see Figure 2. B): plastic shell, water, plastic shell, porcine lung, plastic shell, water and plastic shell. The plastic shell of the 4D phantom is made of thermoplastic copolyester (physical density according to the material data sheet is 1.27 g/cm³). The WET of the shell was experimentally verified and agreed to the theoretical calculation within 0.5 mm, which is the uncertainty of the Giraffe-based WET measurements (Farace *et al* 2016). The data set for the experimental verification consisted of PR acquisition for frame 2 for an empty phantom (without porcine lung).

The inner surface of the plastic shell (the one in contact with lung tissue) was covered with a thin layer of ultrasound gel, to reduce friction and avoid tearing of the lung tissue. In terms of CT numbers, the gel corresponded to water-like material. No additional corrective actions were taken.

Analysis

PR frames were acquired during simulated breathing of the 4D phantom. *A priori,* it was not known which spots will be delivered during what specific breathing phase. Therefore, treatment delivery log files (referred to as logs) and ANZAI breathing patterns, acquired during beam delivery, were used to sort out spots to their corresponding breathing phase. For this purpose, the method as described by Meijers et al. and in-house built scripts were used (Meijers *et al* 2019). In addition to other data, logs contain information regarding position, monitor units (MU), energy and delivery timeline for every delivered spot. From the ANZAI breathing
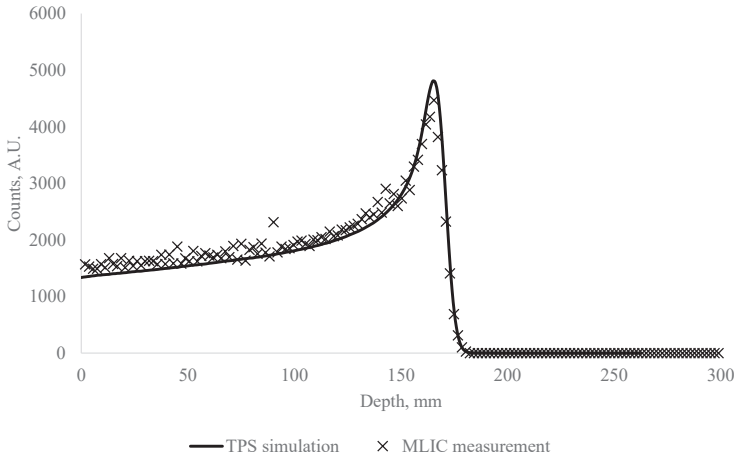
Figure 3. Example IDDs as calculated and measured through lung tissue for one of the example spots. IDDs are shown in the initial position prior to shift optimization. For the shown spot 0.7 mm shift (range error) was calculated using the least square method.

patterns, it is possible to determine in what phase the 4D phantom was at a specific moment in time. By combining these two datasets, a set of 10 DICOM sub-plans was created, where each sub-plan contained only the spots corresponding to a specific breathing phase. These sub-plans per measured frame (5 frames × 5 repetitions per frame) were imported in the TPS and each individual spot was calculated on the corresponding phase of the 4DCT image set. The end point of a spot delivery was considered as a time stamp for this spot. Assignment to the phase was done based on this time stamp. Approximate delivery time of a spot was 5 ms. Depending on the timing, it is possible that some of the spots were partially delivered in one phase and ended in the next one. However, this effect was considered negligible and no additional corrective actions were taken. In order to simulate the MLIC, a water slab was added distally to the 4D phantom. The TPS Monte Carlo dose calculation engine of RayStation with an accuracy of 0.5% on a dose grid of 1 mm was used to calculate TPS-predicted IDDs. For the purpose of dose calculations, the plastic shell of the phantom was contoured and overridden with a density of 1.27 g/cm³. TPS-calculated dose distributions were integrated along the beam axis to create an IDD per spot.

A set of Matlab scripts based on openREGGUI open source code (Deffet *et al* 2017, openreggui.org 2020) was used to compare measured IDDs with the simulated ones and calculate the range error for every individual spot. The range error is obtained by the least squares method and corresponds to the optimal shift between two IDDs. Figure 3 shows measured (MLIC) and calculated (TPS) IDDs for one of the example spots passing through lung tissue (frame 5).

In addition to TPS calculations on the corresponding phase of the 4DCT, calculations were also performed on the average CT, since it is common practice to perform clinical treatment planning on average CTs.

Results

In total 2025 proton spots were individually measured and analysed. The mean water equivalent path length (WEPL) through the 4D phantom was 143.7 mm (1 SD 32.5 mm), as determined based on MLIC measurements. All 10 phases of the 4DCT were well represented in the measurement data set. On average 203 spots (1 SD 18 spots) were associated to an individual phase. The lowest number of spots (171) was associated with 70% phase and the highest number of spots (240) was associated with 0% phase.

Although range uncertainty recipes, as suggested in literature (Paganetti 2012), typically consist of relative and absolute components, in practice most of the commercial TPSs allow to specify only the relative component for the purposes of robust optimisation or robustness evaluation. Therefore, further on the range error of each specific spot is evaluated relative to the WEPL of this spot through the 4D phantom, which is determined based on MLIC measurement.

Figure 4 shows a histogram of relative range errors for all 2025 measured spots. For this data set the TPS dose calculation per spot was performed on the corresponding phase (0 — 90%) of the 4DCT. The mean relative range error is 1.2% (1.5 SD 2.3%).

Figure 5 shows the histogram of relative range errors for the complete measurement data set, however, in this case, the TPS dose calculation for all spots was performed on the average CT. The mean relative range error here is 1.0% (1.5 SD 2.2%).

IV
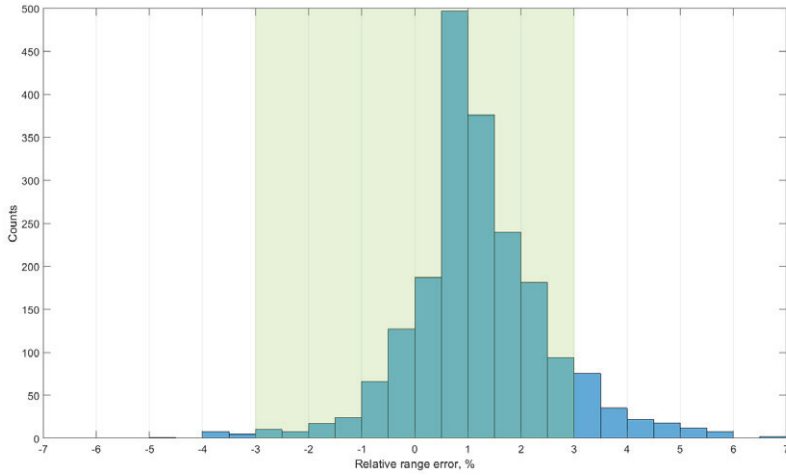
Figure 4. Histogram of range errors relative to the WEPL of the corresponding spot. Dose calculations in the TPS per spot are performed on the corresponding 4DCT phase. Positive shift indicates that measured range (MLIC) exceeds calculated range (TPS). The green area indicates ± 3% range uncertainty.
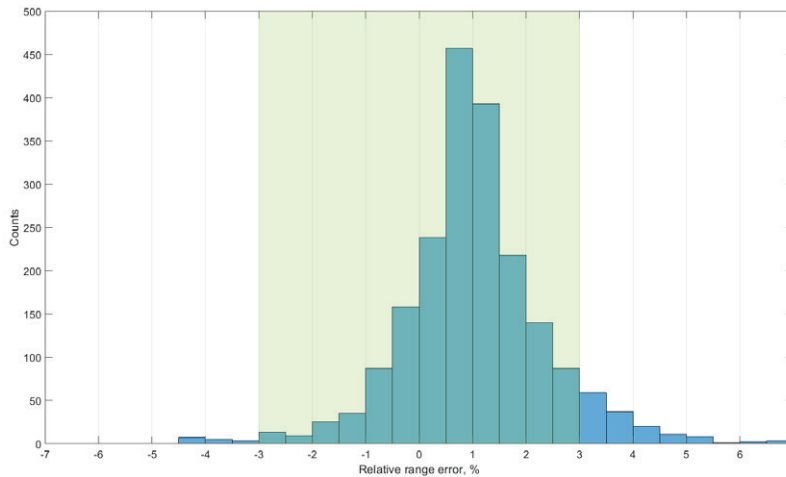
Figure 5. Histogram of range errors relative to the WEPL of the corresponding spot. Dose calculations in the TPS per spot are performed on the average 4DCT. Positive shift indicates that measured range (MLIC) exceeds calculated range (TPS). The green area indicates ± 3% range uncertainty.
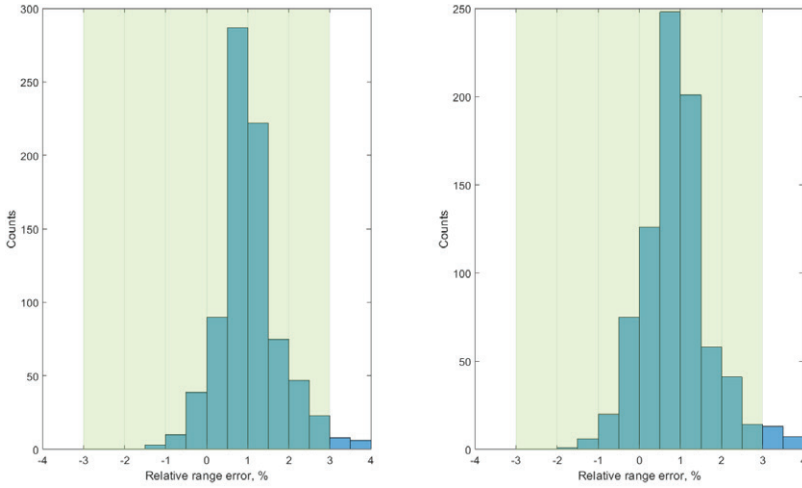
Figure 6. Histograms of range errors relative to the WEPL of the corresponding spot for frames 5 and 6 only. (Left) shows the comparison to TPS calculated doses on corresponding 4DCT phases, while (right) shows comparison to calculations performed on the average CT. The green area indicates ± 3% range uncertainty.

Since frames 5 and 6 were intersecting lung tissue only (see Figure 1), relative range error histograms for these frames are separately shown in Figure 6.

Mean relative range error for comparison to calculations on 4DCT phases is 1.0% (1.5 SD 1.1%), while for comparison to calculations performed on average CT it is 0.8% (1.5 SD 1.2%).

Discussion

Comparison of the TPS-predicted and experimentally measured ranges for shoot-through proton spots directed through a porcine lung in the 4D phantom showed good agreement, especially for spots travelling through lung tissue only (frames 5 and 6).

The introduction of soft / lung tissue intersections in the beam path creates high WET gradients that are in addition moving with the breathing cycles. Therefore, larger relative range errors are observed in frames located in mediastinum region (especially frames 3 and 4). The sensitivity
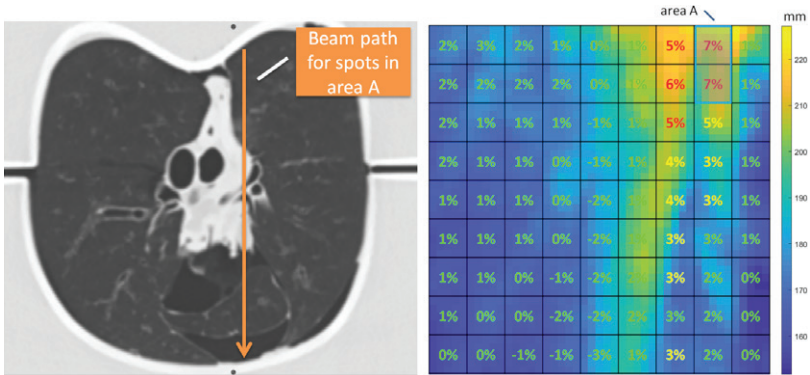
Figure 7. (Left) A transversal view of the phantom in the area of frame 3. (Right) Overlay of the WET map of the phantom along the proton path (coronal view PR) and relative range errors, as measured for proton spots included in the frame 3. Every square represents a proton spot. For instance, the WET gradient over the area A is 30.5 mm.

IV

of the range accuracy towards high WET gradients in presence of setup uncertainty is an observation that has been reported in literature already several times (Farace *et al* 2016, Meijers *et al* 2020, Knopf *et al* 2008). As an example, Figure 7 shows the transversal view of the phantom in the area of frame 3. One can observe rapid changes of tissue material in the lateral direction along the beam path of a spot (indicated by an orange line) that results in high WET gradients. Such WET gradients are correlating with increased range errors (area A in Figure 7). Increase in range errors may be caused by local setup errors, resulting from positioning misalignments and breathing motion. This type of errors has random component, therefore, in case of fractionated treatments, effects on the edges may smear out, however further investigations should confirm this.

For technical and practical reasons, the 4D phantom was continuously kept under motion (simulating breathing) throughout the experiment: from the assembly of the phantom, during CT scan, till the end of measurement session, which is a time period of about 4 to 5 hours. During this time spatial configuration of the porcine tissue at some areas locally may have changed, resulting in an increased setup error of some tissue versus the proton beam. In addition, it is not possible neither to guarantee, nor to verify that spatial configuration of tissue is exactly reproducible from one breathing cycle to another.
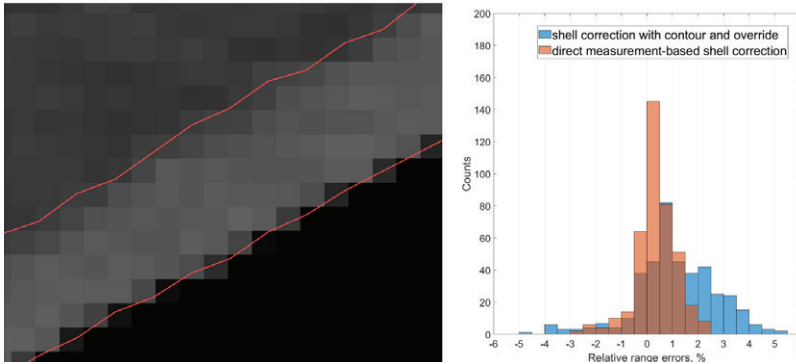
Figure 8. (Left) Magnification of the phantom in a transversal view and shell contour (red), for which density override was assigned. DICOM image has pixel spacing of 1.07 mm. (Right) Histograms of relative range errors for frame 2 measurements, comparing two methods for shell density correction.

It can be observed that the evaluation performed on the average CT scan results in slightly smaller mean range error and SD, compared to the evaluation performed on the phases of 4DCT scan. Although the difference is relatively small, one possible explanation for this observation might be the blurrier edges of individual structures in the average CT image. This might make the comparison less sensitive to small local setup inconsistencies, which may be caused by spatial configuration variations as discussed in the previous paragraph. However, further investigations are necessary to confirm this.

The current study is primarily focussing on range error assessment. However, the data set potentially could be used for future investigations, in order to evaluate effect of range mixing in more details. Potentially the goodness of fit may be used as a measure for such assessments.

A small, but systematic shift of relative range errors (about 1%) was observed, meaning that the TPS systematically overestimates the density of some materials in the beam path. This observation is potentially caused by an ambiguity in the contour definition of the shell of the phantom. The density of the thermoplastic shell was underestimated on the CT images (1.12 instead of 1.27 g/cm^3), therefore contouring and override of the shell was necessary. However, the shell has a curved shape, it does not have an uniform thickness and the exact edges of the surfaces on the CT images are ambiguous as illustrated in Figure 8. For this reason, the definition

of the shell is prone to uncertainty. As mentioned before, in order to obtain WET information of the shell, the PR of frame 2 was acquired for an empty phantom (without porcine lung). This measurement was used to directly correct for the shell density discrepancy of the PR of frame 2 acquired for the assembled phantom (with porcine lung), instead of using override structure. For this direct correction, the mean relative range error for frame 2 (5 repeated measurements) shifts towards zero (mean relative range error 0.3% instead of 1.1%), as demonstrated in Figure 8. This suggests that ambiguity in the override approach to at least some extent may have contributed to the observed systematic shift.

The PR method used in this study is not suitable to establish ground truth proton stopping power ratios (SPR) per tissue type. Individual proton spots are intersecting various mixtures of materials and tissues in lateral and longitudinal directions. Therefore, every acquired IDD is intrinsically integral by nature. In order to extract SPR information per tissue / material type, acquisitions of multiple projections would be necessary (Meyer *et al* 2017, Krah *et al* 2015), which would be a step towards proton CT (Dedes *et al* 2018).

The main purpose of the proposed method of investigation is to assess the range calculation accuracy of the TPS in a near-clinical setting using an end-to-end approach. In case of major observed deviations between TPS-calculated and measured ranges, further investigations to identify root cause would be necessary.

From the perspective of assessing individual tissue types corresponding to individual segments of CT calibration curve, it is advantageous that PR acquisitions have been performed separately for lung-like tissue and soft / bone-like tissue (Meijers *et al* 2020). However, from the point of view of end-to-end testing, the absence of bone-like tissue in the current experimental setup may be considered as a limitation. The ability to perform PR measurements *in vivo* would provide additional confirmation for the made observations.

In conclusion, the range accuracy assessment in an *ex vivo* lung tissue using 4D phantom showed good agreement between calculated ranges in the TPS and the proton radiography measurements. Mean relative range error for a complete data set was 1.2% (1.5 SD 2.3%), when TPS dose calculations were performed on the corresponding phases of the 4DCT,

and 1.0% (1.5 SD 2.2%), when TPS dose calculations were performed on the average 4DCT. Reported results are for the evaluation, where override for the shell has been used. This investigation, in combination with a previously conducted study (Meijers *et al* 2020) investigating bone- and soft tissue-like phantoms, allowed to confirm the use of 3% range uncertainty (1.5 SD interval) for robust 4D CT and Monte Carlo based treatment planning in clinical setting for thoracic indications as adequate.

Acknowledgement

References

Appelt A L, Vogelius I R, Farr K P, Khalil A A and Bentzen S M 2014 Towards individualized dose constraints: Adjusting the QUANTEC radiation pneumonitis model for clinical risk factors *Acta Oncol. (Madr).* **53** 605–12

Baumann K S, Flatten V, Weber U, Lautenschläger S, Eberle F, Zink K and Engenhart-Cabillic R 2019 Effects of the Bragg peak degradation due to lung tissue in proton therapy of lung cancer patients *Radiat. Oncol.* **14** 1–15

Belderbos J, Heemsbergen W, Hoogeman M, Pengel K, Rossi M and Lebesque J 2005 Acute esophageal toxicity in non-small cell lung cancer patients after high dose conformal radiotherapy *Radiother. Oncol.* **75** 157–64

Bert C and Rietzel E 2007 4D treatment planning for scanned ion beams *Radiat. Oncol.* **2** 1–10

Biederer J and Heller M 2003 Artificial thorax for MR imaging studies in porcine heart-lung preparations *Radiology* **226** 250–5

IV

Dedes G, Johnson R P, Pankuch M, Detrich N, Pols W M A, Rit S, Schulte R W, Parodi K and Landry G 2018 Experimental fluence-modulated proton computed tomography by pencil beam scanning *Med. Phys.* **45** 3287–96

Deffet S, Macq B, Righetto R, Vander Stappen F and Farace P 2017 Registration of pencil beam proton radiography data with X-ray CT *Med. Phys.* **44** 5393–401

España S and Paganetti H 2011 Uncertainties in planned dose due to the limited voxel size of the planning CT when treating lung tumors with proton therapy *Phys. Med. Biol.* **56** 3843–56

Etzold, M., 2020. Home. [online] Artichest.com. Available at: <http://www.artichest.com/> [Accessed 15 April 2020].

Farace P, Righetto R and Meijers A 2016 Pencil beam proton radiography using a multilayer ionization chamber *Phys. Med. Biol.* **61** 4078–87

Gomez D R, Tucker S L, Martel M K, Mohan R, Balter P A, Guerra J L L, Liu H, Komaki R, Cox J D and Liao Z 2012 Predictors of high-grade esophagitis after definitive three-dimensional conformal therapy, intensity-modulated radiation therapy, or proton beam therapy for non-small cell lung cancer *Int. J. Radiat. Oncol. Biol. Phys.* **84** 1010–6 Online: http://dx.doi.org/10.1016/j.ijrobp.2012.01.071

Knopf A, Parodi K, Paganetti H, Cascio E, Bonab A and Bortfeld T 2008 Quantitative assessment of the physical potential of proton beam range verification with PET/CT *Phys. Med. Biol.* **53** 4137–51

Krah N, Testa M, Brons S, Jäkel O, Parodi K, Voss B and Rinaldi I 2015 An advanced image processing method to improve the spatial resolution of ion radiographies *Phys. Med. Biol.* **60** 8525–47

Macdonald O K, Kruse J J, Miller J M, Garces Y I, Brown P D, Miller R C and Foote R L 2009 Proton Beam Radiotherapy Versus Three-Dimensional Conformal Stereotactic Body Radiotherapy in Primary Peripheral, Early-Stage Non-Small-Cell Lung Carcinoma: A Comparative Dosimetric Analysis *Int. J. Radiat. Oncol. Biol. Phys.* **75** 950–8

Meijers A, Free J, Wagenaar D, Deffet S, Knopf A C, Langendijk J A and Both S 2020 Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing *Phys. Med. Biol.* **65**

Meijers A, Jakobi A, Stützer K, Guterres Marmitt G, Both S, Langendijk J A, Richter C and Knopf A 2019 Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: Implementation and proof-of-concept *Med. Phys.* **46** 1140–9

Meyer S, Gianoli C, Magallanes L, Kopp B, Tessonnier T, Landry G, Dedes G, Voss B and Parodi K 2017 Comparative Monte Carlo study on the performance of integration- and list-mode detector configurations for carbon ion computed tomography *Phys. Med. Biol.* **62** 1096–112

El Naqa I, Bradley J, Blanco A I, Lindsay P E, Vicic M, Hope A and Deasy J O 2006 Multivariable modeling of radiotherapy outcomes, including dose-volume and clinical factors *Int. J. Radiat. Oncol. Biol. Phys.* **64** 1275–86

Openreggui.org. 2020. [online] Available at: <https://openreggui.org/> [Accessed 15 April 2020].

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

Palma D A, Senan S, Tsujino K, Barriger R B, Rengan R, Moreno M, Bradley J D, Kim T H, Ramella S, Marks L B, De Petris L, Stitt L and Rodrigues G 2013 Predicting radiation pneumonitis after chemoradiation therapy for lung cancer: An international individual patient data meta-analysis *Int. J. Radiat. Oncol. Biol. Phys.* **85** 444–50 Online: http://dx.doi.org/10.1016/j.ijrobp.2012.04.043

Parodi K 2019 Latest developments in in- vivo imaging for proton therapy *Br. J. Radiol.*

Perles, L.A., Mirkovic, D., Sawakuchi, G.O. and Titt, U., 2011. Monte Carlo investigation of rebinning material density distributions of lung parenchyma phantoms in proton therapy. Nuclear technology, 175(1), pp.22-26.

Schneider U, Pedroni E and Lomax A 1996 The calibration of CT Hounsfield units for radiotherapy treatment planning *Phys. Med. Biol.*

Sell, M., Titt, U., Perles, L., Mirkovic, D., Mohan, R., Bangert, M. and Oelfke, U., 2012. WE-E-BRB-02: Evaluation of Analytical Proton Dose Predictions with a Lung–Like Plastic Phantom. Medical physics, 39(6Part27), pp.3956-3956.

Titt U, Sell M, Unkelbach J, Bangert M, Mirkovic D, Oelfke U and Mohan R 2015 Degradation of proton depth dose distributions attributable to microstructures in lung-equivalent material *Med. Phys.* **42**

Uries M, Goiteins M, Holleyo W R and Cheno G T Y 1986 Degradation of the Bragg peak due to inhomogeneities *Phys. Med. Biol.*

Wang K, Eblan M J, Deal A M, Lipner M, Zagar T M, Wang Y, Mavroidis P, Lee C B, Jensen B C, Rosenman J G, Socinski M A, Stinchcombe T E and Marks L B 2017 Cardiac toxicity after radiotherapy for stage III non-small-cell lung cancer: Pooled analysis of dose-escalation trials delivering 70 to 90 Gy *J. Clin. Oncol.* **35** 1387–94

Wijsman R, Dankers F, Troost E G C, Hoffmann A L, Van Der Heijden E H F M, De Geus-Oei L F and Bussink J 2015 Multivariable normal-tissue complication modeling of acute esophageal toxicity in advanced stage non-small cell lung cancer patients treated with intensity-modulated (chemo-)radiotherapy *Radiother. Oncol.* **117** 49–54 Online: http://dx.doi.org/10.1016/j.radonc.2015.08.010

Zhu J, Zhang Z C, Li B S, Liu M, Yin Y, Yu J M, Luo L M, Shu H Z and Crevoisier R De 2010 Analysis of acute radiation-induced esophagitis in non-small-cell lung cancer patients using the Lyman NTCP model *Radiother. Oncol.* **97** 449–54 Online: http://dx.doi.org/10.1016/j.radonc.2010.09.025

Chapter V: Evaluation of interplay and organ motion effect by means of 4D dose reconstruction and accumulation

Abstract

Purpose: Pencil beam scanned proton therapy (PBS-PT) treatment quality might be compromised by interplay and motion effects. Via fraction-wise reconstruction of 4D dose distributions and dose accumulation, we assess the clinical relevance of motion related target dose degradation in thoracic cancer patients.

Methods and Materials: For the ten thoracic patients (Hodgkin lymphoma and non-small cell lung cancer) treated at our proton therapy facility, daily breathing pattern records, treatment delivery log-files and weekly repeated 4DCTs were collected. Patients exhibited point-max target motion of up to 20 mm. They received robustly optimized treatment plans, delivered with five-times rescanning in fractionated regimen. Treatment delivery records were used to reconstruct 4D dose distributions and the accumulated treatment course dose per patient. Fraction-wise target dose degradations

were analyzed and the accumulated treatment course dose, representing an estimation of the delivered dose, was compared with the prescribed dose. *Results:* No clinically relevant loss of target dose homogeneity was found in the fraction-wise reconstructed 4D dose distributions. Overall, in 97% of all reconstructed fraction doses, D98 remained within 5% from the prescription dose. The V95 of accumulated treatment course doses was higher than 99.7% for all ten patients.

Conclusions: 4D dose reconstruction and accumulation enables the clinical estimation of actual exhibited interplay and motion effects. In the patients considered here, the loss of homogeneity caused by interplay and organ motion did not show systematic pattern and smeared out throughout the course of fractionated PBS-PT treatment. Dose degradation due to anatomical changes showed to be more severe and triggered treatment adaptations for five patients.

V

Introduction

Pencil beam scanned proton therapy (PBS PT) treatments of intrathoracic targets are associated with significant uncertainty. Treatment quality may be compromised by setup errors, range uncertainties, respiratory motion baseline shifts, anatomical changes, delivery inaccuracies and motion of various sources (e.g. respiration, cardiac motion, swallowing, etc.). More specifically, the relative motion between a thoracic target volume and the scanning proton beam can cause deviation of the delivered dose from the planned dose and is referred to as the interplay effect. The interplay effect has been of concern since the clinical introduction of PBS PT [1] and has hampered its wide range clinical deployment for the treatment of thoracic indications [2].

To generate PBS PT plans that are robust against possible uncertainties, robust optimization techniques have been introduced and are increasingly used in clinical routine, especially when treating moving targets. Outcomes of the robust optimization must be evaluated to check that robustness objectives are met. Robustness evaluation is commonly performed through simulations of multiple error scenarios [3]. The outcome of robustness evaluation depends on the sampling of uncertainties. The

more realistic and comprehensive a robustness evaluation is, the more time consuming and computationally expensive it gets, eventually making it unfeasible for deployment in clinical routine on regular basis.

Especially the assessment of the clinical impact of the interplay effect is difficult due to the large parameter space affecting it and has been subject of many in-silico simulation studies [4], [5], [6], [7], [8]. Three sets of parameters determine the interplay effect: (i) delivery characteristics in the sense of the delivery timeline, the start of delivery with respect to the organ motion (starting phase), the scanning path, the spot size, the number of applied rescans, the type of rescanning (e.g. layered or volumetric rescanning) and the use of, e.g., gating, (ii) plan characteristics such as the field directions, the number of fields and layer spacing and (iii) patient characteristics such as the patient and fraction specific motion pattern in terms of amplitude, frequency and variability and the target volume size and location. Due to this large amount of variables, the impact of the interplay effect for a specific patient so far has only been assessed in a probabilistic manner by simulating several possible delivery scenarios with varying input parameters (e.g. the starting phase, the number of rescans, etc.). In this way, possible deviations of the delivered dose from the planned dose can be estimated. However, it remains unknown what is the probability of one scenario over another and there is lack of insight on what dose actually is delivered to the patient. Consequently, when correlating outputs to the planned treatment dose, this correlation is impeded by the unknown deviation of the actual delivered dose from the planned dose.

Accounting for dose delivery uncertainties by robust optimization, considering many of possible treatment scenarios, comes at the cost of integral dose and dose to healthy tissues. However, for every individual patient only one scenario occurs. By creating overly robust plans, the normal tissue may be unnecessarily overdosed to compensate for situations that may never happen in practice.

Contrasting the concept of "overly conservative" robust optimization is the concept of adaptive treatment delivery [9]. Here, much more conformal treatment plans can be delivered on the basis that deviations from the nominal scenario are accounted for by a plan adaptation. Adaptation can either be triggered based on continuous accumulation and evaluation of the delivered dose distribution or executed on a daily basis or even

real-time in the future. However, that requires high degree and reliability of automation. A triggered adaptation relies on an assessment of the actual delivered fraction dose, which in general is challenging to obtain, especially for thoracic indications due to motion affecting the treatment delivery.

We have developed a methodology aimed at gaining more insight on fraction-wise delivered dose distributions. This methodology is based on retrospective reconstruction of four-dimensional (4D) dose distributions and accumulation providing means to continuously assess treatment course quality [10]. 4D dose reconstruction for carbon ion has previously been reported on by Richter et al. [11]. In the current study commercially available solutions have been used to implement the dose reconstruction workflow, different indications and much longer fractionation schemes with multiple repeat CTs have been investigated. Furthermore, investigated treatment plans have been prepared using robust optimization and evaluation planning techniques.

Here we present initial results of the application of 4D dose reconstruction and accumulation (4DREAL) of 10 consecutive patients with thoracic indications (Hodgkin lymphoma (post-chemotherapy) and non-small cell lung cancer (NSCLC)) treated with PBS PT at our facility. The focus of this study is to primarily evaluate fraction-wise and consecutive accumulated target volume doses (high dose area). The objective is to assess the impact of the interplay and organ motion on the target dose homogeneity and to investigate the consequences of fraction-wise loss of homogeneity on the total accumulated course dose.

Material and methods

In our facility the treatment of targets, which are affected by respiratory motion, is performed following a procedure based on four principles: (1) motion assessment, (2) robust planning, (3) robustness evaluation and (4) retrospective 4D dose reconstruction.

(1) Motion assessment
Planning 4DCT (phase-based reconstruction) is used to assess the magnitude of the motion. Phase-based reconstructions are more suitable for

dose calculations due to equidistant spacing in time, however, may suffer from anatomy-induced artifacts, which could be less pronounced in amplitude-based reconstructions [12]. End-of-inhale and -exhale phases are defined by a medical doctor (MD) in consultation with a medical physicist (MP). Afterwards, the MD defines the target volumes on the selected phases, which are later used for the definition of the CTV and ITV as per ICRU62. ITV was obtained as a union of the CTVs of the end-of-inhale and exhale phases. Deformable image registration (DIR) is performed between the selected phases using Anatomy Constrained Deformation Algorithm (ANACONDA). Deformation vector fields are evaluated by a MP to quantify the extent of motion within target volume. Motion is assessed in terms of average motion of the target volume and maximum motion of any voxel (point-max). Depending on the extent of the motion the approach to treatment planning and delivery is fine-tuned. For example, decisions are made regarding the field selection and design, enlargement of the spot size [13], exact rescanning strategy [4], [6], etc.

(2) Robust planning

Currently all patients in our clinic receive 3D robust optimized treatments plans. Specifically, for NSCLC patients this decision was based on a preclinical study, in which 3D and 4D robust optimization techniques were compared in terms of achievable plan robustness [14]. This pre-clinical study was conducted utilizing 4D robustness evaluation method (4DREM) introduced by Ribeiro et al. [3]. 3D robust optimization is performed on a single image set (average CT of the planning 4DCT) and accounts for setup and range uncertainty. Optimization for lung cancer patients is performed assuming 6 mm setup uncertainty, while optimization for Hodgkin lymphoma is performed assuming 5 mm setup uncertainty. This is due to immobilization differences and setup reproducibility. Hodgkin lymphoma patients are typically immobilized with a 5-point thermoplastic mask as opposed to NSCLC patients, who are immobilized on a wing board. Estimated range uncertainty for all above mentioned indications is 3%, based on the experimental evaluation as shown by Meijers et al. [15], where the evaluation of range accuracy has been performed for average CT-based calculations, as well as for phase-based calculations. For our patient cohort, all plans incorporated

5-times rescanning (in-layer scaled rescanning). For all lymphoma patients spot size was intentionally enlarged by retracting the range shifter, while for NSCLC patients no intentional spot size enlargement was done. The decision regarding the enlargement of spot size for NSCLC was also based on the pre-clinical study mentioned above, which showed that robust plans can be achieved without enlargement of the spot size, however, during the plan optimization ITV, defined on the average CT, for NSCLC cases was overridden with a muscle tissue density. All lymphoma patients were treated with anterior, anterior-oblique beam arrangement (minimum of 2 fields) and all NSCLC cases were planned with 3-field arrangement. For NSCLC patients anterior, lateral and / or posterior beam directions were used depending on the exact location of the target volume.

Proton spot size in our facility is 3 to 6.5 mm (sigma) as a function of proton energy, which varies from 230 to 70 MeV respectively. Spot spacing as a function of the spot full width at the half maximum (FWHM) in water with a ratio of 0.8 to 1 was used during plan optimization. Energy layers were spaced as a function of peak width with a ratio of 0.8 to 1. Peak width in our facility is 8.7 to 1.7 mm as a function of energy (230 to 70 MeV respectively). Layer switching time is approximately 0.7 to 0.9 s depending on the energy step.

V

(3) Robustness evaluation

Robustness evaluation was performed to assess the outcome of robust optimization. Robustness evaluation was performed in 3D, simulating a set of scenarios with pre-defined setup and range errors. Twenty-eight scenarios are calculated, simulating range errors of ± 3% in combination with setup errors of ± 6 or ± 5 mm, depending on the indication as mentioned above. Robustness of the plan was assessed on the basis of voxel-wise minimum and voxel-wise maximum dose distributions. In our facility plans are considered robust if V95 of the target volume on voxel-wise minimum distribution exceeds 98% [16].

(4) Retrospective 4D dose reconstruction

4D dose reconstruction is performed following the method described by Meijers et al. [10]. This method makes use of treatment delivery log files,

patient's breathing signals and most recent available 4DCT information throughout the treatment course.

Fraction-wise breathing patterns of each patient are acquired using the Anzai belt system (Anzai Medical, Tokyo, Japan). After the delivery of each fraction, treatment delivery log files are collected. Among other data, log files contain information for every delivered spot regarding its position, dose in terms of monitor units (MU), energy and timing. Based on the timing information of the breathing signal, delivered spots, as retrieved from the log files, are sorted into corresponding breathing phases. Afterwards, the spots are written into a set of DICOM sub-plans, where each sub-plan contains only the spots associated with a specific breathing phase. These sub-plans are imported into the treatment planning system (TPS), and each sub-plan is calculated on the corresponding phase of the 4DCT. In this step, the most recent available 4DCT is used. In our current clinical practice, patients that are subject to respiratory motion receive weekly repeated 4DCTs. An exception are Hodgkin lymphoma patients, for whom 4DCT in the last week of the treatment may be skipped, if no observations with clinical consequences were made based on daily CBCTs and previous 4DCTs.

Deformation vector fields between a fraction-wise reference phase and all other phases of the 4DCT data set are defined. In the majority of cases, the reference phase is the end-of-exhale phase. Deformation vector fields are used to warp dose contributions of all sub-plans per fraction to the fraction-wise reference phase, where they are summed. Afterwards, deformation vector fields between a course-wise reference phase and fraction-wise reference phase are defined and the fraction-wise summed dose is warped to the course-wise reference phase. On the course-wise reference phase, individual fraction doses are accumulated. The same course-wise reference phase (and target volume) is used throughout the treatment course (also in case plan adaptations have been made). Course-wise reference phase, for example, could be the end-of-exhale phase of the planning CT. Schematically the workflow is shown in Figure 1.

Reconstructed fraction doses and accumulated course doses are calculated for the 10 consecutive patients affected by motion and treated at our proton facility. Table 1 summarizes some of the planning and target characteristics of these cases.
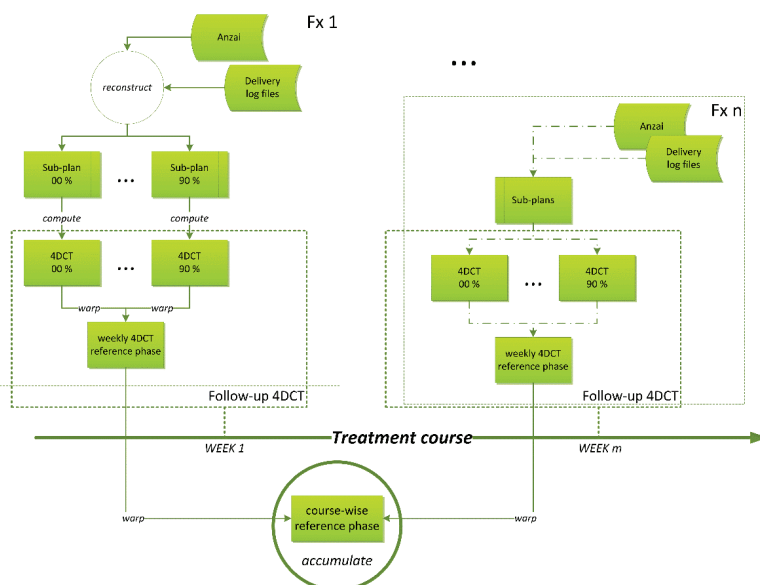
Figure 1. Schematic representation of the 4D dose reconstruction and accumulation workflow for n fractions.

Table 1. Characteristics of the treatment course preparation specifics for the 10 patients. The point-max motion corresponds to the maximum motion observed for any voxel in the target volume based on the planning 4DCT.

Pat. #	Indication	Prescription	ITV volume, cm³	Mean motion, mm	Point-max motion, mm	ITV V95$_{vox\text{-}min}$, %
01	Lymphoma	15×2.0 Gy$_{RBE}$	166	0.9	< 5	98.13
02	NSCLC	25×2.4 Gy$_{RBE}$	44	2.2	< 6	99.42
03	NSCLC	25×2.4 Gy$_{RBE}$	243	0.9	< 6	99.77
04	NSCLC	25×2.4 Gy$_{RBE}$	131	0.7	< 5	99.82
05	NSCLC	25×2.4 Gy$_{RBE}$	357	0.7	< 7	99.06
06	NSCLC	25×2.4 Gy$_{RBE}$	217	1.1	< 5	99.74
07	NSCLC	25×2.4 Gy$_{RBE}$	298	1.2	< 9	98.41
08	NSCLC	25×2.4 Gy$_{RBE}$	202	1.4	< 8	98.41
09	NSCLC	25×2.4 Gy$_{RBE}$	336	0.6	< 6	99.49
10	Lymphoma	15×2.0 Gy$_{RBE}$	339	2.1	< 20	99.67

ITV V95$_{vox\text{-}min}$ is V95 based on the voxel-wise minimum dose distribution, derived in the process of robustness evaluation. This parameter is a measure of the robustness for the initial treatment plan.

Results

Due to the difference in indications, various fractionation schemes and patient-related events, the available data per patient and some treatment characteristics varied among patients. The number of acquired repeat CTs, the number of treatment plan adaptations, the number of fractions, for which breathing signal was acquired and the smallest and the largest observed point-max motions between the end-of-inhale and -exhale phases and mean motions as observed on repeat CTs are listed in Table 2.

For some of the fractions, the acquisition of the breathing signal was skipped either due to logistical issues or patient-related issues. For patient 2 one of the fractions was delivered on a linac due to pending evaluation of the repeat CT, where large anatomical variations were observed. For patient 5 the last two fractions were delivered on a linac in a satellite site due to hospitalization of the patient unrelated to the radiotherapy treatment itself.

All plan adaptations were necessary due to anatomical changes. Changes in the volume of postoperative fluid caused adaptations for patients 2 and 3. For patient 6 adaptation was triggered by a tumor shrinkage, which

Table 2. Available data and treatment characteristics for the 10 considered patients. Mean and point-max motions are shown as the range between the smallest and the largest motion values, as observed on any one specific repeat CT.

Pat. #	Repeat CTs	Plan adaptations	Breathing signals, [available / total]	Mean motion (min-max), mm	Point-max motion (min-max), mm
01	2	0	15 / 15	(1.1–1.2)	< (6–6)
02	5	2	24 / 25	(1.8–2.2)	< (7–9)
03	5	1	23 / 25	(1.0–2.1)	< (7–10)
04	5	0	22 / 25	(0.8–1.0)	< (5–7)
05	5	0	22 / 25	(0.7–1.1)	< (7–8)
06	5	1	25 / 25	(1.8–2.1)	< (7–11)
07	5	1	25 / 25	(0.5–0.9)	< (7–12)
08	5	0	25 / 25	(0.9–1.5)	< (7–8)
09	5	1	25 / 25	(0.7–1.2)	< (5–7)
10	3	0	15 / 15	(1.1–1.5)	< (11–14)

Table 3. CTV dose statistics based on reconstructed fraction doses and accumulated course dose. Mean D98 and D2 doses and V95 are mean over all the fractions. In the brackets the lowest and the highest values, as observed for any one of the fractions, are provided. A single course-wise CTV, as defined on the reference image set, has been used for the DVH analysis per patient.

Pat. #	Prescription, Gy_{RBE}	Mean D98 (min-max), Gy_{RBE}	Mean D2 (min-max), Gy_{RBE}	Mean V95 (min-max), %	Acc. D98, Gy_{RBE}	Acc. D2, Gy_{RBE}	Acc. V95, %
01	30	29.7 (29.5−29.8)	31.1 (30.8−31.5)	100 (100−100)	29.9	30.9	100.0
02	60	59.1 (58.3−59.9)	62.2 (61.4−63.3)	100 (100−100)	60.0	61.4	100.0
03	60	57.9 (56.4−58.8)	61.9 (61.4−62.8)	99.1 (96.5−100)	58.4	61.1	99.8
04	60	59.5 (58.9−59.8)	61.6 (61.2−61.9)	100 (100−100)	59.8	61.4	100.0
05	60	59.4 (59.0−59.8)	62.3 (61.6−63.4)	100 (99.9−100)	59.7	62.0	100.0
06	60	58.9 (57.3−59.9)	62.1 (61.3−62.6)	99.9 (99.1−100)	59.8	61.4	100.0
07	60	58.8 (56.4−59.8)	62.3 (61.2−63.1)	99.4 (97.5−100)	59.5	61.9	99.7
08	60	59.2 (58.4−59.7)	61.9 (61.5−62.4)	100 (100−100)	59.7	61.4	100.0
09	60	58.9 (58.0−59.6)	62.0 (61.5−62.6)	99.9 (99.6−100)	59.4	61.6	100.0
10	30	29.6 (29.4−29.7)	30.6 (30.4−31.2)	100 (100−100)	29.7	30.4	100.0

V

resulted in an unacceptable dose to OARs. Weight gain required a plan adaption for patient 7. While disappearance of pleural effusion caused plan adaptation for patient 9.

For all follow up 4DCTs the motion evaluation was performed according to the previously described methodology. For example, for patient 6, due to tumor shrinkage, during the course maximum amplitude of the motion increased significantly, reaching 11 mm as opposed to 5 mm observed in the initial planning CT.

Table 3 shows the summary of observations made on the basis of fraction-wise and accumulated treatment course dose distributions. Mean D98 and D2 (with SD) for the CTV are listed per patient for all
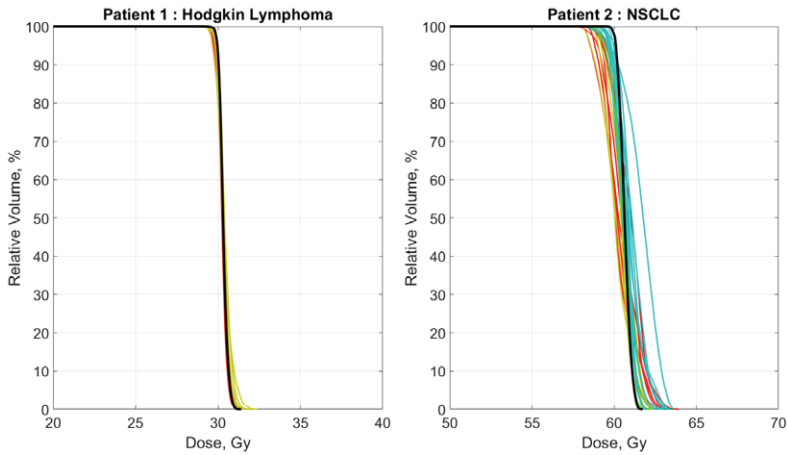
Figure 2. DVHs of CTV for reconstructed fraction-wise 4D dose distributions and accumulated course dose. DVHs are corresponding to cases 1 and 2. Fraction doses are shown in color, while accumulated course dose is shown in black. Fraction doses that have been calculated on the same 4DCT also share the same color. Assigned colors are red, yellow, green, light blue and blue, corresponding from an earlier 4DCT to a more recent in this order.

reconstructed fractions of the treatment course. In addition, D98, D2 and V95 of the accumulated treatment course dose distribution are shown.

As an example, Figure 2 shows all reconstructed 4D fractions and accumulated course dose for patients 1 and 2. Additionally supplementary material 1 shows all reconstructed and accumulated doses for all patients.

Out of total 221 reconstructed fractions combined for all 10 patients presented in this study dose to the target volume (D98) remained within 5% from the prescription dose, with only 6 fractions being an exception. In no case, accumulated treatment course dose distributions showed major variations from the prescription dose.

Organ at risk (OAR) doses are summarized in supplementary material 2 for illustrative purposes. However, due to the limitations of accumulated doses, as discussed further, and the set scope of the work, OAR doses will not be discussed in detail.

Discussion

It can be observed that interplay effects and organ motion introduce loss of dose homogeneity in the target volume on fraction basis. Furthermore, this scales with the degree of target motion. One may notice that the loss of homogeneity for Hodgkin lymphoma is smaller than the loss of homogeneity for NSCLC. For the included NSCLC patients, whole target volumes in the lung were mobile, while for lymphoma patients large parts of the treatment volume are relatively immobile, as parts are located cranially with respect to the lung. Therefore, the organ motion on average is affecting this area to a lesser extent.

In the current data set it was not observed that loss of homogeneity induced by motion effects follow a systematic pattern. Systematic patterns generally were caused by anatomical variations, such as, changes in postoperative fluids, patient's weight changes or tumor shrinkage. In all cases, although loss of homogeneity was present on fractional basis, homogeneity was recovered when performing dose accumulation. Local hot and cold spots did not occur in the same location in different fractions.

Although fractionation likely smears out interplay and organ motion effects over the course of radiotherapy, these effects should be considered differently when moving towards hypo-fractionation. In such case, loss of homogeneity very likely may have clinical implications. Nevertheless, it should be noted that for the initial 10 patient data set the observed average target volume motion is relatively low. That is due to the introduced guideline to initially limit the point-max motion below 10 mm during the patient selection (although this was not strictly followed anymore for the patient 10). This illustrates how the 4DREAL methodology can be applied in clinical practice to gradually expand patient inclusion criteria, while ensuring close daily monitoring of the treatment course.

Limitations
We would like to point out several important assumptions that are made, when performing 4D dose reconstruction and accumulation as described in this study.

(1) Planning and repeat 4DCTs are assumed to be good representations of patient's anatomy. By reviewing daily CBCTs and comparing them to

repeat CTs, one can judge whether the repeat CT is representing the daily anatomy, however this is a subjective evaluation. In case of non-minor inconsistencies, it is difficult to estimate the actual impact of these observations on the calculated dose distributions. In the future, this limitation might be overcome by introducing post-processed synthetic 4DCTs based on daily CBCTs suitable for proton dose calculations. There are multiple examples for developments towards introduction of 4DCBCTs [17] and synthetic CTs [18], which do have improved CT number accuracy that might make synthetic 4DCTs suitable for proton dose calculation. The initial investigations on the use of artificial intelligence for reduction of motion-induced artifacts in the 4DCT also show promising results [19]. This may further improve 4DCT quality.

(2) It is assumed that 4DCT, which captures the average patient motion derived from multiple breathing cycles, is representative of the patient's 4D anatomy. Patients exhibiting irregular breathing may be identified calculating the ratio of extreme inhalation amplitude and regular tidal inhalation amplitude. [20] However, breathing cycles are not constant over time. Therefore, potentially a better accuracy of dose reconstruction could be achieved by introducing, so called, "5DCTs". 5DCT can be obtained by combining 4DCTs of variable motion characteristics, each representing an individual breathing cycle. Developments [21] are ongoing, which aim at modeling variations of subsequent breathing cycles. These models can be used to animate 4DCT and generate 5DCTs, incorporating breathing cycle variability. However, validation and, therefore, quantification of accuracy, remains a major challenge for these approaches and eventually these images would provide only an estimation about the motion. Consequently, the added value of 5DCT in dose reconstruction and the impact of its uncertainty remains a topic for further investigations.

(3) Our dose reconstruction method heavily depends on dose warping. The accuracy and physical meaning of the warped doses is a topic for further investigations. For example, the effect of voxel volume deformation on the meaning of dose must be further clarified. Also, the radiobiological effect on addition of fractionated and varying doses per voxel should be further investigated. For variations of up to $\pm 10\%$ in high dose area the additional radiobiological effect has been estimated to be minimal [22].

However, variations per voxel in low dose areas might be much larger and their radiobiological consequences for dose addition is a subject for further clarifications. Due to this, within the scope of current evaluation, we did not investigate low dose areas or dose to OARs, but primarily focused on high dose areas and uniformity of the dose within target volume. To some extent geometric accuracy of the dose warping in the phantom study was investigated during the development of the methodology [10], by being able to reconstruct with a sub-millimeter accuracy the shadow caused by a moving ball bearing in the beam path. However, it has been shown that the accuracy of dose warping reduces when the magnitude of the deformation increases [23]. Also, further investigation on dosimetric consequences caused by the use of different DIR algorithm are required [24].

V

Currently the smoothness of deformation vector fields is assessed by visual inspection of the deformation grids and accuracy is assessed by manual review of the mapped contours. In case of major anatomical changes, which cannot be attributed to the deformations, deformable image registration would fail. For none of the cases presented here this was the case. However, such scenario is highly probable in clinical routine, in which case corrective actions would be necessary. Otherwise, the meaning of warped doses would become even more questionable. To some extent such situations might be corrected by manual adjustment of control contours and use of them as control ROIs during DIR.

(4) Currently our implementation of 4D dose reconstruction method does not allow to account for residual patient setup errors or residual beam delivery discrepancies. However, we use 4D dose reconstruction primarily as a tool for gaining insights into interplay and organ motion effects.

In addition to 4D dose reconstruction we perform robustness evaluation on repeat CTs to assess robustness against residual setup errors and range uncertainties. Combined effects of interplay effect, organ motion, setup errors, range uncertainty and changing anatomy indeed might cause some additional dose perturbations. However, since we do not observe systematic patterns linked to interplay from fraction to fraction, it is likely that these additional perturbations would not have severe effects on treatment course dose due to fractionation.

Due to these assumptions and in an absence of further evaluations, we do not recommend considering reconstructed and accumulated dose distributions as "clinical" doses. Therefore, at this stage we would not use accumulated dose distributions, for example, as a background (in other words, "already delivered") dose to be used in plan optimization in case of plan adaptations. However, it would be an attractive use case for the accumulated doses if some of the limitations would be addressed or proven not relevant. This way accumulated hot or cold spots or unintended dose to OARs could be directly mitigated during the plan adaptation process.

The proposed 4D dose calculation workflow can also be employed prospectively by using simulated breathing patterns and log files from the dry-runs. In such way for cases, when motion amplitude exceeds predefined acceptable levels, a set of simulated fractions can be generated to assess if fraction-wise hot / cold spots have systematic behavior. If this is not the case, accumulated DVH would converge towards steeper curve.

Furthermore, a promising future application of accumulated dose distributions could be its correlation with treatment outcomes. This could clarify the clinical relevance of fraction dose variations and could help to reduce uncertainties in the dose parameters enclosed in TCP and NTCP models.

Eventually, the use of daily treatment related information (delivery log files, imaging data, breathing signals, etc.,) could be automatically retrieved and processed in a dose accumulation workflow. By introducing warning and action levels for accumulated doses or even using accumulated dose to track TCP and / or NTCP values, it would be possible to implement a whole new layer of quality control longitudinally throughout the treatment course. By using adaptive loops in this process, it would be possible to ensure and gain more confidence that initial clinical goals are met at the end of the treatment course.

In conclusion, the developed methodology for fraction-wise 4D dose reconstruction was applied to 10 consecutive thoracic patients, subject to respiratory motion. Contrary to findings in prospective simulation studies, we did not observe any clinically relevant loss of target dose homogeneity due to interplay and motion effects. Fraction-wise loss of target dose homogeneity due to interplay and organ motion showed no systematic pattern and smeared out with fractionation. Dose degradation

caused by anatomical changes showed to be more severe and caused treatment adaptations in five out of ten patients. Although, warped dose distributions should be interpreted with caution, this study provides more realistic incremental insight into the effect of breathing related organ motion on PBS-PT dose delivery.

References

[1]   Bert C, Rietzel E. 4D treatment planning for scanned ion beams. Radiat Oncol. 2007 Jul 3;2:24.

[2]   Macdonald OK, Kruse JJ, Miller JM, Garces YI, Brown PD, Miller RC, Foote RL. Proton beam radiotherapy versus three-dimensional conformal stereotactic body radiotherapy in primary peripheral, early-stage non-small-cell lung carcinoma: a comparative dosimetric analysis. Int J Radiat Oncol Biol Phys. 2009 Nov 1;75(3):950-8. doi: 10.1016/j.ijrobp.2009.04.023.

[3]   Ribeiro CO, Meijers A, Korevaar EW, Muijs CT, Both S, Langendijk JA, Knopf A. Comprehensive 4D robustness evaluation for pencil beam scanned proton plans. Radiother Oncol. 2019 Jul;136:185-189. doi: 10.1016/j.radonc.2019.03.037. Epub 2019 Apr 20.

[4]   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 Nov 21;56(22):7257-71. doi: 10.1088/0031-9155/56/22/016. Epub 2011 Oct 28.

[5]   Schätti A, Zakova M, Meer D, Lomax AJ. Experimental verification of motion mitigation of discrete proton spot scanning by re-scanning. Phys Med Biol. 2013 Dec 7;58(23):8555-72.

[6]   Bernatowicz K, Lomax AJ, Knopf A. Comparative study of layered and volumetric rescanning for different scanning speeds of proton beam in liver patients. Phys Med Biol. 2013 Nov 21;58(22):7905-20. doi: 10.1088/0031-9155/58/22/7905. Epub 2013 Oct 29.

[7]   Grassberger C, Dowdell S, Sharp G, Paganetti H. Motion mitigation for lung cancer patients treated with active scanning proton therapy. Med Phys. 2015 May;42(5):2462-9. doi: 10.1118/1.4916662.

[8]   Engwall E, Glimelius L, Hynning E. Effectiveness of different rescanning techniques for scanned proton radiotherapy in lung cancer patients. Phys Med Biol. 2018 May 2;63(9):095006. doi: 10.1088/1361-6560/aabb7b.

[9]   Albertini F, Matter M, Nenoff L, Zhang Y, Lomax A. Online daily adaptive proton therapy. Br J Radiol. 2019 Nov 11:20190594. doi: 10.1259/bjr.20190594

V

[10]   Meijers A, Jakobi A, Stützer K, Guterres Marmitt G, Both S, Langendijk JA, Richter C, Knopf A. Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: Implementation and proof-of-concept. Med Phys. 2019 Mar;46(3):1140-1149. doi: 10.1002/mp.13371. Epub 2019 Feb 4.

[11]   Richter D, Saito N, Chaudhri N, Härtig M, Ellerbrock M, Jäkel O, Combs SE, Habermehl D, Herfarth K, Durante M, Bert C. Four-dimensional patient dose reconstruction for scanned ion beam therapy of moving liver tumors. Int J Radiat Oncol Biol Phys. 2014 May 1;89(1):175-81. doi: 10.1016/j.ijrobp.2014.01.043. PMID: 24725700.

[12]   Wink N, Panknin C, Solberg TD. Phase versus amplitude sorting of 4D-CT data. J Appl Clin Med Phys. 2006;7(1):77-85. doi:10.1120/jacmp.v7i1.2198

[13]   Grassberger C, Dowdell S, Lomax A, Sharp G, Shackleford J, Choi N, Willers H, Paganetti H. Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. Int J Radiat Oncol Biol Phys. 2013 Jun 1;86(2):380-6. doi: 10.1016/j.ijrobp.2013.01.024.

[14]   Ribeiro CO, Visser S, Meijers A, Korevaar EW, Wijsman R, Muijs C, Both S, Langendijk JA, Knopf AC. Towards the Clinical Implementation of Pencil Beam Scanned Proton Therapy (PBS-PT) for Thoracic Indications: A Comprehensive 4D Robustness Evaluation Based on Patient and Machine Specific Information. VOLUME 105, ISSUE 1, SUPPLEMENT, E754-E755, SEPTEMBER 01, 2019. https://doi.org/10.1016/j.ijrobp.2019.06.779

[15]   Meijers A, Seller Oria C, Free J, Bondesson D, Rabe M, Parodi K, Landry G, Langendijk JA, Both S, Kurz C, Knopf AC. Assessment of Range Uncertainty in Lung-Like Tissue Using a Porcine Lung Phantom and Proton Radiography. Phys Med Biol. 2020 May. doi: 10.1088/1361-6560/ab91db

[16]   Korevaar EW, Habraken SJM, Scandurra D, Kierkels RGJ, Unipan M, Eenink MGC, Steenbakkers RJHM, Peeters SG, Zindler JD, Hoogeman M, Langendijk JA. Practical robustness evaluation in radiotherapy — A photon and proton-proof alternative to PTV-based plan evaluation. Radiother Oncol. 2019 Dec; 141:267-274. doi: 10.1016/j.radonc.2019.08.005.

[17]   Niepel K, Kamp F, Kurz C, Hansen D, Rit S, Neppl S, Hofmaier J, Bondesson D, Thieke C, Dinkel J, Belka C, Parodi K, Landry G. Feasibility of 4DCBCT-based proton dose calculation: An ex vivo porcine lung phantom study. Z Med Phys. 2019 Aug;29(3):249-261. doi: 10.1016/j.zemedi.2018.10.005.

[18]   Botas P, Kim J, Winey B, Paganetti H. Online adaption approaches for intensity modulated proton therapy for head and neck patients based on cone beam CTs and Monte Carlo simulations. Phys Med Biol. 2018 Dec 19;64(1):015004. doi: 10.1088/1361-6560/aaf30b.

[19]   Werner R, Sentker T, Madesta F, Gauer T, Hofmann C. Intelligent 4D CT sequence scanning (i4DCT): Concept and performance evaluation. Med Phys. 2019 Aug;46(8):3462-3474.

[20]   Zeng C, Plastaras JP, Tochner ZA, White BM, Hill-Kayser CE, Hahn SM, Both S. Proton pencil beam scanning for mediastinal lymphoma: the impact of interplay between target motion and beam scanning. Phys Med Biol. 2015 Apr 7;60(7):3013-29. doi: 10.1088/0031-9155/60/7/3013. Epub 2015 Mar 19. PMID: 25789418.

[21]   Zhang Y, Knopf A, Tanner C, Boye D, Lomax AJ. Deformable motion reconstruction for scanned proton beam therapy using on-line x-ray imaging. Phys Med Biol. 2013 Dec 21;58(24):8621-45. doi: 10.1088/0031-9155/58/24/8621.

[22]   Bortfeld T, Paganetti H. The biologic relevance of daily dose variations in adaptive treatment planning, Int J Radiat Oncol Biol Phys. 2006 Jul 1;65(3):899-906.

[23]   Spautz S, Peters N, Meijers A, Jakobi A, Knopf AC, Troost EGC, Richter C, Stützer K. Accuracy and robustness of 4D logfile-based dose reconstruction. Conference proceedings 4D workshop 2019, Krakow.

[24]   Ribeiro CO, Knopf A, Langendijk JA, Weber DC, Lomax AJ, Zhang Y. Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours., Radiother Oncol. 2018 Jul;128(1):174-181. doi: 10.1016/j.radonc.2018.03.001. Epub 2018 Mar 20.

V

Supplementary material 1

Patient 1 : Hodgkin Lymphoma

Patient 2 : NSCLC

VI

Patient 3 : NSCLC

Patient 4 : NSCLC

Patient 5 : NSCLC

Patient 6 : NSCLC

Patient 7 : NSCLC

Patient 8 : NSCLC

VI

Patient 9 : NSCLC

Patient 10 : Hodgkin Lymphoma
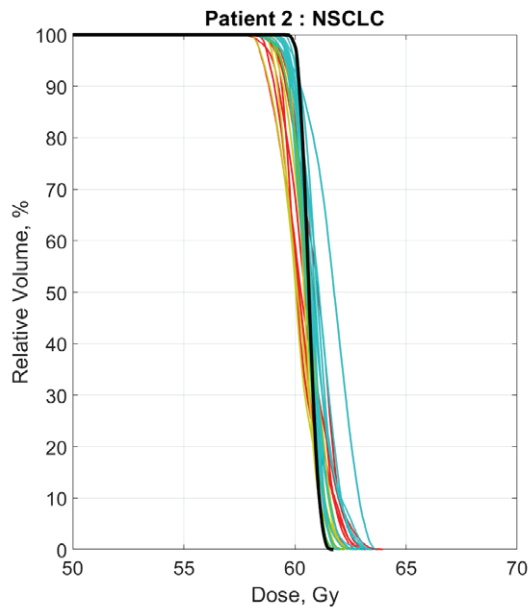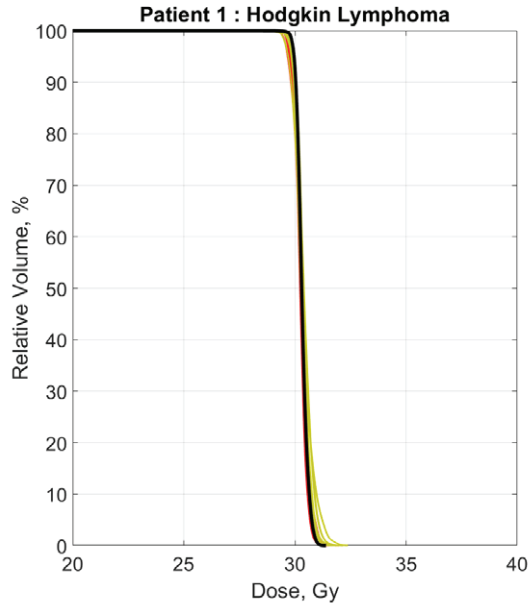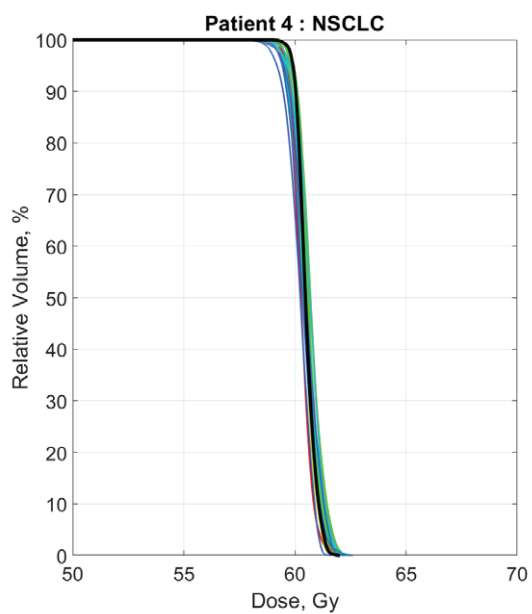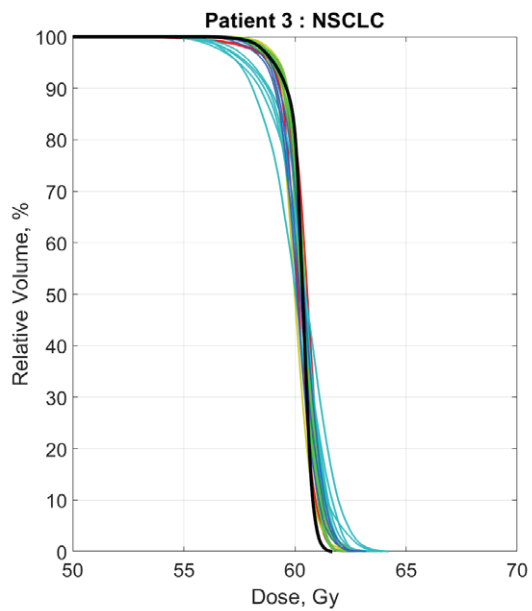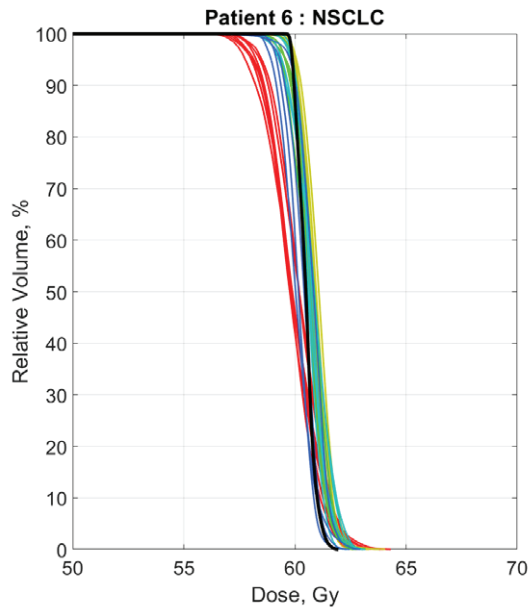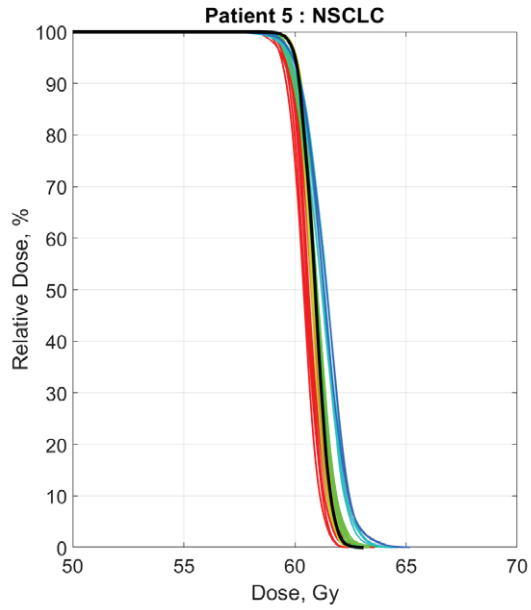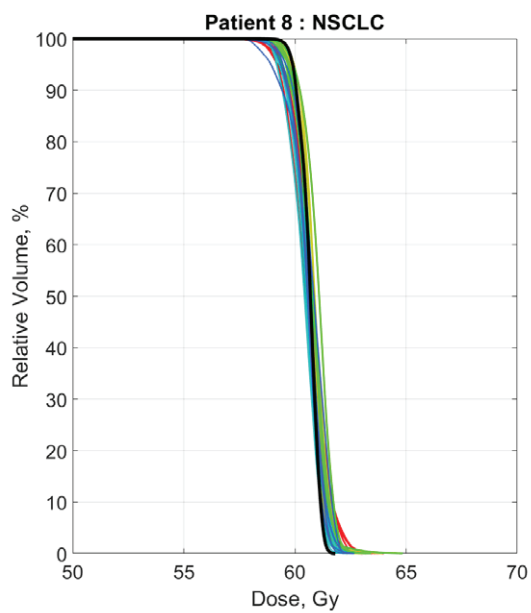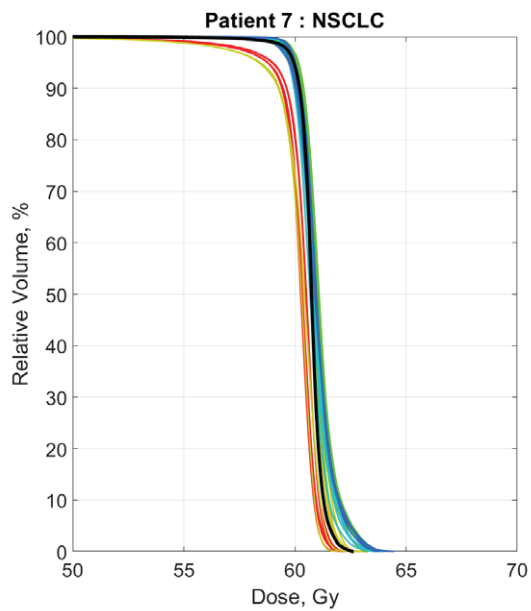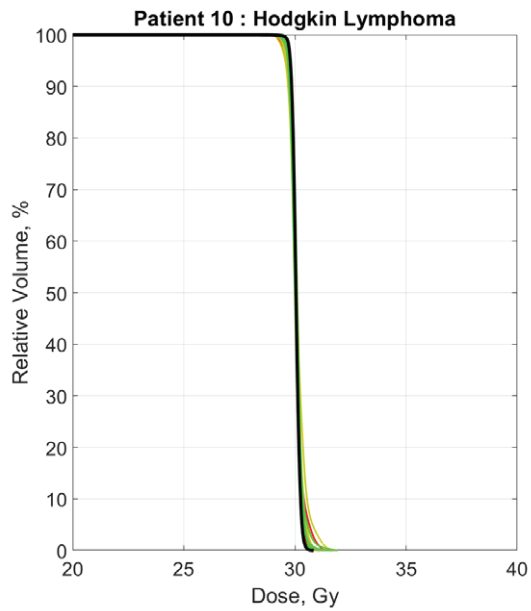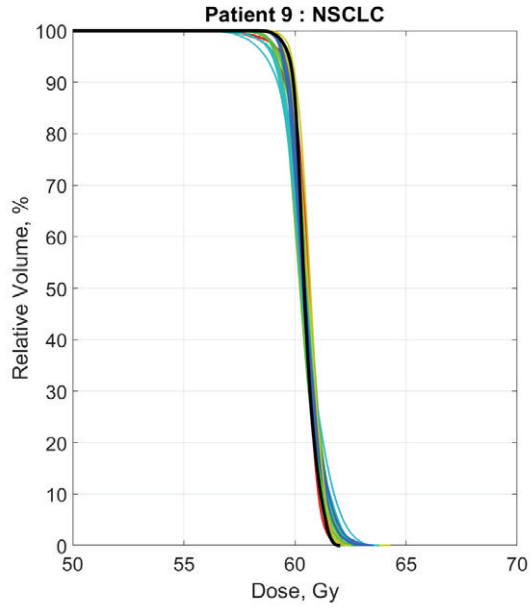
Supplementary material 2

Supplementary material 2 summarizes organ at risk (OAR) doses for the 10 patients. The doses are provided for illustrative purposes.

Table 1. Overview of mean heart doses (MHD) for the 10 patients. Nominal MHD is provided for the initial plan (plan adaptations are not reflected).

Pat. #	Indication	Prescription	Fraction-wise mean MHD (min-max), GyRBE	Accumulated MHD, GyRBE	Nominal MHD, GyRBE
01	Lymphoma	15 × 2.0 GyRBE	2.3 (2.1-2.6)	2.3	2.5
02	NSCLC	25 × 2.4 GyRBE	3.0 (2.6-3.3)	3.0	3.1
03	NSCLC	25 × 2.4 GyRBE	11.6 (9.4-12.9)	11.6	10.4
04	NSCLC	25 × 2.4 GyRBE	6.7 (6.5-7.1)	6.7	6.8
05	NSCLC	25 × 2.4 GyRBE	7.6 (7.2-8.5)	7.6	7.6
06	NSCLC	25 × 2.4 GyRBE	2.0 (1.5-3.3)	2.0	1.4
07	NSCLC	25 × 2.4 GyRBE	1.4 (0.8-2.0)	1.4	0.9
08	NSCLC	25 × 2.4 GyRBE	0.4 (0.3-0.6)	0.4	0.4
09	NSCLC	25 × 2.4 GyRBE	5.1 (4.4-5.9)	5.1	4.4
10	Lymphoma	15 × 2.0 GyRBE	3.6 (3.4-4.0)	3.6	4.0

VI

Table 2. Overview of mean lung doses (MLD) for the 10 patients. Nominal MLD is provided for the initial plan (plan adaptations are not reflected).

Pat. #	Indication	Prescription	Fraction-wise mean MLD (min-max), GyRBE	Accumulated MLD, GyRBE	Nominal MLD, GyRBE
01	Lymphoma	15 × 2.0 GyRBE	4.0 (4.0-4.1)	4.0	4.1
02	NSCLC	25 × 2.4 GyRBE	6.4 (6.0-6.9)	6.4	6.2
03	NSCLC	25 × 2.4 GyRBE	9.6 (8.9-9.9)	9.6	9.5
04	NSCLC	25 × 2.4 GyRBE	10.6 (10.4-10.8)	10.6	10.8
05	NSCLC	25 × 2.4 GyRBE	12.8 (12.3-13.9)	12.8	12.3
06	NSCLC	25 × 2.4 GyRBE	9.8 (8.9-11.7)	9.8	10.8
07	NSCLC	25 × 2.4 GyRBE	12.0 (10.7-12.9)	12.0	11.3
08	NSCLC	25 × 2.4 GyRBE	9.9 (9.5-10.6)	9.9	9.9
09	NSCLC	25 × 2.4 GyRBE	10.2 (9.9-10.5)	10.2	9.9
10	Lymphoma	15 × 2.0 GyRBE	7.1 (6.8-7.2)	7.1	6.8

Table 3. Overview of lung V5Gy$_{RBE}$ (V5) for the 10 patients. Nominal V5 is provided for the initial plan (plan adaptations are not reflected).

Pat. #	Indication	Prescription	Fraction-wise mean V5 (min-max), %	Accumulated V5, %	Nominal V5, %
01	Lymphoma	15 × 2.0 GyRBE	20.8 (20.5-21.6)	20.9	20.9
02	NSCLC	25 × 2.4 GyRBE	22.2 (21.4-23.1)	22.7	21.7
03	NSCLC	25 × 2.4 GyRBE	33.1 (30.7-34.3)	33.2	31.6
04	NSCLC	25 × 2.4 GyRBE	43.3 (42.4-44.0)	43.3	42.5
05	NSCLC	25 × 2.4 GyRBE	33.8 (31.9-37.3)	33.8	32.0
06	NSCLC	25 × 2.4 GyRBE	33.0 (30.2-38.9)	34.6	35.7
07	NSCLC	25 × 2.4 GyRBE	28.5 (26.4-30.0)	28.7	28.6
08	NSCLC	25 × 2.4 GyRBE	24.4 (23.4-26.0)	24.7	24.5
09	NSCLC	25 × 2.4 GyRBE	22.3 (21.8-22.9)	22.4	21.5
10	Lymphoma	15 × 2.0 GyRBE	37.1 (35.6-38.6)	37.1	35.5

Table 4. Overview of mean esophageal doses (MED) for the 10 patients. Nominal MED is provided for the initial plan (plan adaptations are not reflected).

Pat. #	Indication	Prescription	Fraction-wise mean MED (min-max), GyRBE	Accumulated MED, GyRBE	Nominal MED, GyRBE
01	Lymphoma	15 × 2.0 GyRBE	13.9 (13.6-14.5)	13.9	14.3
02	NSCLC	25 × 2.4 GyRBE	15.0 (13.6-17.4)	15.0	16.8
03	NSCLC	25 × 2.4 GyRBE	40.9 (38.3-42.1)	40.9	38.6
04	NSCLC	25 × 2.4 GyRBE	36.9 (36.6-37.4)	36.9	36.6
05	NSCLC	25 × 2.4 GyRBE	15.0 (14.0-16.2)	15.0	14.5
06	NSCLC	25 × 2.4 GyRBE	9.6 (9.1-10.6)	9.6	8.1
07	NSCLC	25 × 2.4 GyRBE	23.8 (18.2-26.7)	23.8	22.1
08	NSCLC	25 × 2.4 GyRBE	0.3 (0.2-0.5)	0.3	0.3
09	NSCLC	25 × 2.4 GyRBE	19.2 (16.0-21.0)	19.2	16.4
10	Lymphoma	15 × 2.0 GyRBE	15.4 (15.3-15.5)	15.4	15.2

Chapter VI: Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy

Abstract

For radiation therapy, it is crucial to ensure that the delivered dose matches the planned dose. Errors in the dose calculations done in the treatment planning system (TPS), treatment delivery errors, other software bugs or data corruption during transfer might lead to significant differences between predicted and delivered doses. As such, patient specific quality assurance (QA) of dose distributions, through experimental validation of individual fields, is necessary. These measurement based approaches, however, are performed with 2D detectors, with limited resolution and in a water phantom. Moreover, they are work intensive and often impose a bottleneck to treatment efficiency. In this work, we investigated the potential to replace measurement-based approach with a simulation-based patient specific QA using a Monte Carlo (MC) code as independent dose calculation engine in combination with treatment log files. Our developed QA

platform is composed of a web interface, servers and computation scripts, and is capable to autonomously launch simulations, identify and report dosimetric inconsistencies. To validate the beam model of independent MC engine, in-water simulations of mono-energetic layers and 30 SOBP-type dose distributions were performed. Average Gamma passing ratio 99 ± 0.5 % for criteria 2%/2 mm was observed. To demonstrate feasibility of the proposed approach, 10 clinical cases such as head and neck, intracranial indications and craniospinal axis, were retrospectively evaluated via the QA platform. The results obtained via QA platform were compared to QA results obtained by measurement-based approach. This comparison demonstrated consistency between the methods, while the proposed approach significantly reduced in-room time required for QA procedures.

Introduction

The number of proton therapy centers is further growing, permitting the delivery of highly conformal dose distributions through the use of pencil beam scanning (PBS) [1, 2]. For a PBS treatment plan, the weight of several thousand of pencil beams is iteratively optimized to achieve a conformal high dose region while sparing organs at risk [3, 4]. Multi-field optimization resulting in intensity modulated proton therapy (IMPT) plans is seen as state-of-the-art. The achievement of homogeneous target dose distribution with minimum and optimally balanced normal tissue doses for IMPT plans generally leads to highly complex in-homogeneous, per-field target dose distributions [4]. Sub-optimal treatment plans were shown to help account for the uncertainties during these optimizations [5].

Treatment planning systems (TPS) that calculate such plans are complex software systems [6], which makes comprehensive testing, commissioning and quality assurance inevitable. In addition to the optimized fluence map, the delivery of a PBS treatment plan requires at least two more transformations. In the first step, it needs to be converted into machine readable files and in a second step, these files have to be correctly interpreted and delivered by the treatment machine. Both of these transformations are potential sources of errors which may be difficult to detect, especially given the complexity of the treatment plans.

VI

As such, patient specific quality assurance (PSQA) of absolute dose distributions, through experimental validation of individual fields, is currently necessary and commonly done. Multiple experimental approaches for patient specific QA have been reported [7, 8, 9, 10]. Excepting few 3D measurement approaches [11], these measurement are performed with 2D detectors, with limited resolution and in a solid water phantom. Moreover, they are work intensive and often impose a bottleneck for the throughput of a treatment room, or limit the ability to adapt a treatment plan in a timely manner.

In order to decrease the PSQA measurement beam-time, work has been done for the use of delivery system control files (hereafter referred to as treatment log files) instead [12, 13, 14]. Meier et al. [15] have shown the use of log files for independent dose calculation systems, with the intention of detecting problems or differences in TPS dose computations. In order to achieve a greater independence, the Monte Carlo dose engine used for QA dose calculation should be based on independent algorithms with completely separated code bases.

After each delivery, files containing details of the machine parameters are generated by the Proton Therapy System (PTS). Treatment log files can either be obtained prior to the start of a treatment course by performing a dry-run irradiation or will be generated inherently during each delivery of a fraction. In order to apply this method successfully, log files must contain information on the delivered spot position, dose and energy recorded. This information may then be used to create a plan and reconstruct the dose that was actually delivered, which then could be compared to the prescribed planning dose.

The aim of this work is to describe the implementation of a platform for the execution of PSQA workflows, and present an extended validation of its many components. Such a platform should require minimal human intervention, relying on automated simulations when data is available. Also, it should be flexible enough in order to fit future applications, such as adaptive planning and 4D dose accumulation. Initially, two workflows were designed and implemented: TPS-plan-based QA, which uses an independent Monte Carlo engine for checking of the TPS dose calculation; and Log-based QA, which reconstructs the dose based on the machine logs. Dose calculation by QA platform is performed on patient's geometry using the planning CT.

Materials and methods

In order to integrate incoming data processing, computation, visualization and report, a software platform following a server-client architecture was developed. The main building blocks of the application are part of the OpenPATh initiative [16] created to support open-source software applications for research in proton-therapy. Open-source enables researchers to reuse and build upon existent code to avoid rewriting from scratch. In addition, a multi-party contribution to the development and usage of said applications improves the robustness and the trustworthiness of research software for proton-therapy [17]. The open-source modules used in this research include OpenREGGUI, Orthanc, MCsquare, and CAPTAIN and are presented in the following subsections.

1. OpenREGGUI

OpenREGGUI [18] is an image processing software featuring various registration methods, filtering methods, segmentation tools and other radiotherapy dedicated functions such as dose volume histogram computation and others. It is a powerful application interface that helps clinicians to monitor patient information, and to compare planned treatments with actual measurements when running clinical studies in research projects.

The use of OpenREGGUI requires MATLAB [19]. It offers a graphical interface to visualize DICOM images and to operate many image processing functions. It allows defining complex workflows that can be triggered directly from the MATLAB command line as well. The toolkit also provides many desired functionalities: a) a wrapper function that formats DICOM files to the input files required by MCsquare; b) Interpolation of dose maps to guarantee matching grid sizes between TPS and MCsquare dose simulations; c) Evaluation of clinical goals after dose computation (DVH computation). In this work, such workflows were used as data processing libraries.

2. Orthanc

The project known as Orthanc [20] was used as a standalone DICOM server. What makes Orthanc a compelling choice is the fact that it provides a comprehensive Application Programming Interface (API), making

VI

it possible to access it from any computer language. Orthanc receives a request for data download each time a task requires it, and uploads the resulting data when such process finishes.

3. MCsquare

Monte Carlo dose recalculations were performed using MCsquare [21]. It is an open-source fast multipurpose Monte Carlo algorithm, optimized for exploiting massively parallel central processing unit (CPU) architectures. Simulations were performed with 12 calculation threads, in an Intel Xeon server with 48 processing units. The 64 GB of RAM available are shared when multiple simulations are launched simultaneously, each allocating approximately 10 GB.

MCsquare was configured to run all simulations with a MC statistics of 1×10^8 particles, which is equivalent to a standard deviation between 1 and 2% calculated inside the 50% higher dose region for all clinical plans tested. Its inputs are the DICOM files for the plan and the CT coming from the TPS. The method described by Schneider et al. [22] was used to convert from HU to human tissues, which includes elemental composition, weights and densities. The elastic and inelastic nuclear interactions are sampled from cross sections in the ICRU 63 report. In order to be able to compare it with the TPS dose, the dose-to-medium exported MCsquare is later converted into dose-to-water by applying the appropriate Stopping Power Ratio to each voxel in the dose map [23].

MCsquare uses different algorithms, code base and physical tables from RayStation MC. However, they were shown to have similar accuracy in simulation and experimental validations [24].

4. CAPTAIN

The CAPTAIN project [18] was built with a series of industry standard web technologies and is a Free and Open Source Software (FOSS) released under the Apache 2 license. In short, CAPTAIN is an automated workflow manager. Its main feature is the autonomous launching of computation workflows without human intervention. CAPTAIN is developed based on Node.js [25] technology.

Fig. 1 shows a scheme of the platform, which can be separated into three layers:
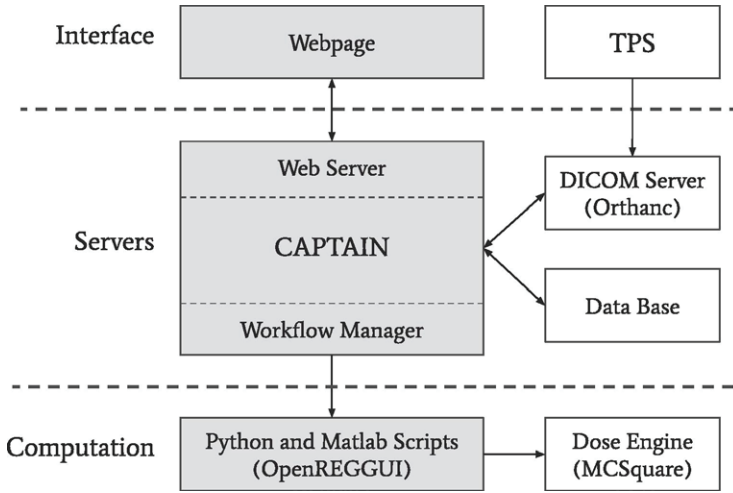
Figure 1. Scheme of the platform architecture, its core units (in grey) and external components (in white).

1. The user interface composed of a web-site with access to workflow configuration and results.
2. The servers, composed of three isolated processes: CAPTAIN main server including the Workflow Manager, a dedicated DICOM server and a database server.
3. The computation layer, where a series of OpenREGGUI and Python scripts are called in order to complete the assigned tasks. These scripts are also responsible for launching the Monte Carlo simulations.

These layers are further discussed on the subsections below.

4.1. Interface

The user interface with CAPTAIN server is done through its website. It was written in Typescript and HTML5. It follows a centralized model where the client task is only interfacing with the servers.

The home page has an index of all patients, and from each patient entry one may configure workflow setups and preview results. Each workflow has its dedicated configuration page, where the default parameter values may be changed and/or new data may be uploaded. During a treatment,

VI

as new data become available, different configurations might be set for the workflows.

The result page of a workflow displays a list of all calculation tasks, their execution status, and links leading to reports. The report page has detailed information on a workflow run, such as configuration parameters and result values of the calculations. DICOM objects may also be downloaded from the report page, both from configuration and results fields.

4.2 Servers

The CAPTAIN main server is responsible for the data exchange occurring between clients, data storage servers and the computation scripts, introduced below. It was generated with the Angular Full-Stack Generator [26], is managed by Gulp.js [27] and written in javascript.

The server offers four Application Programming Interfaces (APIs) to connect to it and act on the database (DB). These four APIs are:

- User API: entry point from the user interface to configure the user access to the system.
- Dicom-server API: exchange of data between the CAPTAIN server and the DICOM server Orthanc.
- patient API: entry point from the user interface that is used to access patient data, to configure the workflows parameters.
- patientResults API: manages the access from the user-interface to the workflow results.

Apart from the dedicated DICOM server, described at Sub-Section 2, a second database server is responsible for holding state information of the platform. MongoDB [28] is a FOSS cross-platform document-oriented database that runs on NoSQL and uses JSON-like documents. This DB contains three collections: users, patients and results. These collections hold documents with information on user authorization and permissions, patient meta-data, and workflow result values or DICOM meta-data, respectively.

It is imperative that both data servers are kept synchronized to each other. To that end, a system of triggers and parser were implemented. Upon the arrival of new data at the DICOM Server, a signal is sent to CAPTAIN server requesting parsing of the received patient data. Meta-data is

extracted from the DICOM objects, and then saved as a Patient document in the DB. These documents are lightweight structure data, holding only descriptions and links pointing to relevant instances inside the DICOM Server.

4.3 Computation

In this work, workflow denotes a complete computation chain, starting at a set of initial parameters and finalizing at a set of result values. The computation between these two states is divided in small chunks, here called tasks, which perform more specific calculations. Therefore, a workflow is the recipe of what tasks to execute, in which order and with what parameters. It is defined by the following objects:

- Check function, which verifies if all necessary parameters are available, and queries the DB for previously calculated results.
- List of Tasks to be performed, each completing a specific computation step on the available data.
- Task input recipe, encoding which parameters to use in each of the corresponding tasks.
- Task output recipe, encoding which results to save from each of the corresponding tasks.

VI

A single task may be used by multiple workflows, which lower complexity and increases consistency between different workflows. Examples of implemented tasks include: the conversion of treatment log files into a RT plan, independent dose recalculation and evaluation of Gamma analysis between two dose maps. Each task is composed of the following steps:

- Preparation, which creates temporary folders for computation, queries the DICOM server for data and the DB for other input parameters.
- Launching, responsible for the start of the computational step, which may include multiple script launches and/or external calls.
- Exporting, which saves the output data to the DB and/or DICOM objects to the DICOM server.

In order to function autonomously, the Workflow Manager receives a notification when new patient data is received by the DICOM Sever. The Manager then triggers the start of all available workflows, starting with the Check for available data. Many of the required input parameter are

automatically set, such as the plan, structure set and dose received from TPS. Others may be adjusted, such as gamma analysis distance and dose tolerances.

5. Workflows for QA

For an envisioned automatic plan QA, two workflows were created: (A) a TPS-plan-based patient QA and (B) a Log-based patient QA workflow. A scheme of these workflows is showed in Fig. 2.

In (A) TPS-based plan QA, a secondary Monte Carlo dose calculation is executed with the same plan input as the TPS but with a completely independent implementation (using a different programming language, different algorithms, different physical models, and with the code being
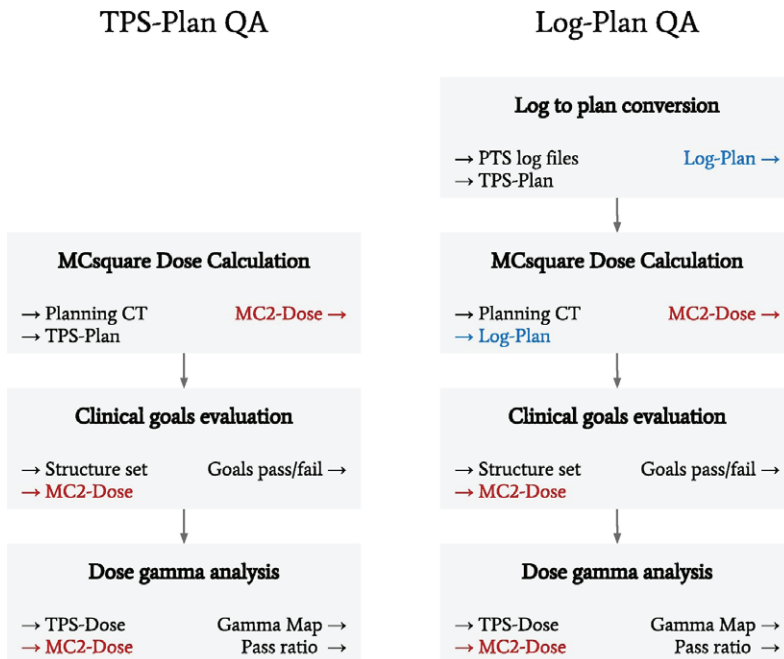
Figure 2. Schemes of the TPS- and Log-based Plan QA workflows. Each task has its input parameters showed on the left, and output on the right. The outputs of one task may be linked to the input of another inside a workflow chain, such as the MCsquare recomputed dose map (MC2-Dose, in red) and the log reconstructed plan (Log-Plan, in blue).

written by a different developer). Hence, this allows for a redundant check of the treatment planning dose calculation.

In (B) Log-based plan QA, a secondary Monte Carlo dose calculation is made, taking the field information from a different source than the TPS. A new plan is created from treatment log files, which become available after each beam delivery. This can be used as a consistency check between the expected TPS calculated dose and the PTS effective delivery dose.

In clinical routine, these workflows are foreseen to be executed consecutively. Planning is performed with RayStation's Monte Carlo algorithm, which generates the original (TPS) dose map. After plan approval in TPS, data is exported to a dedicated DICOM server that triggers the QA workflow (A). An independent dose recalculation automatically takes place and Gamma analysis results — comparing original and recalculated dose maps — become available at the website.

After review, a dry-run is performed by delivering the plan in air in order to generate treatment log files. The upload of the logs triggers the Log-based plan QA workflow (B). A Log-based plan is reconstructed from the retrieved treatment log files, and results become automatically available for review.

VI

5.1 Log to plan conversion

The structure and content of treatment log files are vendor specific or even equipment model specific, however generally treatment log files will contain chronological list of events, which were registered by therapy control system during the delivery of specific treatment prescription, and a list of readouts that were acquired by sensors, which are integral part of the delivery control system. Sensor readouts may contain such information as potentiometer positions, hall probe readouts, set points on power supplies, charge on strips or wires of ionization chambers, etc. Although the readouts from the sensors may not directly be meaningful in any way to what type of prescription has been delivered to the patient, this information may be used to recontact what prescriptions have been delivered into clinically meaningful way. After all, therapy control system as an input uses clinical prescription (treatment plans), to define state of various components of the system to achieve delivery of a specific prescription. By reversing this process clinical prescription itself can be

reconstructed based on the state of the machine, which is indicated by the output of various built-in sensors. Currently, the availability of such log files are subjected to specific agreements with the manufacturer of the delivery system.

Log file interpretation script retrieves from the treatment log files a set of spots that has been delivered during the specific session and assigns their position in X and Y directions at the isocenter plane by using readouts from the strips of the in-nozzle ionization chamber, energy based on the position of degrader wheel and MU based on readout from the integral pads of the in-nozzle ionization chambers. Additionally, a set of corrections, such as, virtual source-axis-distance (VSAD) and IC-to-isocentre distance correction, temperature and pressure correction, etc, needs to be applied.

After the content of the delivered prescription is reconstructed, it is written into a DICOM ion plan object and log-based plan is created. Compatibility to DICOM standard significantly eases usability of the data for other purposes; not only being limited to the use for independent dose re-calculation within the scope of patient QA platform, but also enables use cases like import and re-calculation in TPS, etc.

The quality of the reconstructed plan is dependent on the accuracy of log files recorded parameters, therefore constant validation of its performance is required. The procedure adopted for log files validation in this work is discussed in Sub-Section 6.1.

5.2. Structures overrides

Clinical plans often have overridden structures created in the TPS. Their main uses are: (a) cover patient support devices, (b) create regions with uniform water phantoms and (c) define boundaries for dose calculations within external contour. In order to perform accurate calculations, these exceptions are handled by python scripting. It automatically identifies override tags in the structure set DICOM header, such as the presence of a Material ID in the structure description or an Interpreted Type of 'External'. For (a) and (b) the density of the linked material is converted to Hounsfield Units (HU) by consulting the CT calibration curve, then overwriting the structure space inside the CT with its interpolated value. For (c), the volume outside the external structure is overwritten with air.

Therefore, a new CT image with all overrides applied to it is created and then used as input for the MC simulations.

5.3. Clinical goals evaluation

In clinical practice treatment plans are commonly assessed on basis of clinical goals. Clinical goals are defined as a set of dosimetric criteria that should be met to achieve intended clinical outcome while risks of developing complications are maintained reasonably low. Clinical goals usually are either defined by user in TPS and checked automatically or are manually checked by the user during the plan review based on dose-volume histogram (DVH).

In case of RayStation (RaySearch, Sweden) TPS, clinical goal templates can be defined in TPS and in our clinical practice they are commonly used during plan review process. A dedicated script was developed that can be executed in TPS on patient specific basis to extract a list of defined patient-specific clinical goals. Afterwards the list is written in a specific format to a JSON file and may be imported into the patient QA platform via dedicated interface. By using the imported file, a set of clinical goals is populated in the QA platform.

This makes it possible to assess independently re-calculated dose distribution against clinical goals in a similar way as it is done during plan review in TPS. One of the main reasons to use clinical goals in combination with more commonly employed Gamma analysis for assessment of QA dose distribution is because often it is not exactly straight forward to interpret Gamma analysis results in a clinically meaningful way. In other words, failing pixels or voxels in Gamma analysis cannot be easily interpreted in the sense of their clinical relevance. In some cases, localization of failing Gamma analysis points may be highly important clinically. The expectation is that evaluation of the QA dose distribution against clinical goals will help to identify these situations, even if Gamma passing ratio, especially globally, would not seem alarming.

5.4. Gamma analysis

In this study, dose distributions are compared using a 3D version of the global Gamma analysis [29, 30]. The calculations are performed by an external Python script, which is based on the npgamma library [31].

VI

Both reference (TPS) and evaluated (MCsquare) dose maps are Monte Carlo simulations, which may skew passing rate evaluations due to the statistical uncertainties [32]. For this reason, MC statistical error is kept less than 2% in all calculations presented here.

6. Validation and testing

Due to complexity of the platform an extensive testing plan was put in place to ensure expected functionality. Testing and validation effort may be mainly split in four sections:

6.1. Log recording of plan delivery

Treatment log files from QA tests were collected at the IBA proton therapy system in the Groningen Proton Therapy Center and then used to reconstruct the delivered plans, here called log-based plans. These are plans from standard daily Morning QA. The plan is composed of 1580 spots, from which the recorded spot positions and MU were used to retrospectively evaluate agreement between log content (output) and delivery prescription (input). Following the deliveries through half a year period provided information on the accuracy and consistency of the delivery system. Split between the two delivery rooms, 60 log files sets were analyzed in this manner.

   Additionally, independent external measurements from standard monthly QA procedure with a Lynx detector were used to assess the errors during delivery. The plans consisted of 5 spots per field, delivered with energies between 70 and 225 MeV and measured in gantry angles between 50 and 315°. Log files from these deliveries were collected and analyzed. The relative position of the spots in reference to a central spot was then calculated for measurement and log files recordings.

6.2. Low level component testing

As introduced earlier, workflows make use of multiple lower level computational modules in order to generate results. Performance of such modules was tested by performing calculations under controlled conditions, where data input is well defined and expected output is known or can be easily predicted. Testing of this type was applied to such modules, and the procedure varied per component:

- 3D Gamma analysis module was tested by introducing errors of known magnitude in the synthetic dose object. For the testing purpose geometrical and dose errors were introduced.
- The log to plan converter was tested in two steps: (a) by reconstructing spot energy, position and MU and comparing it directly to the planned ones and (b) by importing log-based DICOM plans back into the TPS, using TPS dose engine to recalculate the dose and comparing it to the original planned dose.
- REGGUI's implementation of clinical goals evaluation was validated against the planning TPS evaluation. A set of 5 plans were calculated with both methods, and the clinical goals evaluations were grouped into 4 categories: target volume coverage in CTV; Mean dose in ROIs; maximum and minimum dose in ROIs above 100 cGy cutoff; maximum and minimum dose in ROIs below 100 cGy cutoff.

6.3. Beam model validation

Accuracy of the independent Monte Carlo engine and quality of the beam model are crucial to the proposed QA workflow; therefore, particular attention was paid to validation of this component. Validation of the beam model was performed in several tests, where the complexity of the testing method gradually increases. Initially in-water calculations were performed for several mono-energetic layers. The energy of these layers was varied between 70 and 225 MeV. The main objective of these calculations was to determine range calculation accuracy in-water. Further a set of 30 SOBP-type fields, which was a sub-set of validation data that was earlier used for the purpose of TPS commissioning, was calculated in water. Range of SOBP fields was varied between $32\,g/cm^2$ and $4.1\,g/cm^2$ and modulation — between 2 and 4 cm. SOBP calculations performed by independent MC engine were compared to calculations performed by clinically used TPS, which has been already commissioned, by using 3D Gamma analysis with criteria of 2%/2 mm. The Gamma analysis was performed in absolute dose procedure which is sensitive to dose ratio discrepancies, and calculated range discrepancies were analyzed. Eventually, an experiment using animal tissues was set up to evaluate accuracy of MC calculations taking into account lateral

VI

and longitudinal heterogeneities. A vacuumed pig's head, positioned on top of solid water slabs, was scanned on a CT. Multiple treatment fields (SOBP and mono-energy layers) in anterior-posterior direction were prepared and a dose plane at the depth of 3.1 mm (WET) in solid water below the animal tissues was selected for comparison with the measurements. In the proton treatment room head was aligned using CBCT and measurements at the selected depth were performed with an ionization chamber array MatriXX PT (IBA dosimetry, Germany). Proton beam measurements were performed on a pigs head phantom using one SOBP and three mono-energetic beams with energies between 175 and 225 MeV, and an array of ionization chambers was used for measuring 2D dose distributions at different depths of 3 and 7 mm. Calculations were performed by TPS and MCsquare dose engines with MC statistic tuned for 0.5% uncertainty, adjusting the number of particles per plan accordingly. Measured dose planes were compared to calculations by using 2D Gamma analysis with the criteria of 3%/3 mm.

6.4. Functional workflow testing

To test the functionality of the QA platform in clinical setting 10 patient cases were evaluated by proposed QA method retrospectively. These clinical cases included such indications as head and neck, intracranial indications and cranio-spinal axis. Via functional testing full data flow was considered: beginning with data transfer from TPS to QA platform and ending with creation of the reports. As part of this testing phase timing of the workflows was performed.

The gamma analysis comparison for TPS- and Log-QA workflows are performed with an acceptance criteria of higher than 95% Passing Ratio (2 mm/2%) calculated in the 3D volume of the dose map. The criteria values were tuned for increasing the error sensitivity of the workflows. For validation purposes, its results are then compared to the standard measurement-based QA — which is currently performed with an acceptance criteria of higher than 95% Passing Ratio (3 mm/3%) in 2D dose maps measured and calculated at 3 different depths per field.

Results

The results are laid out in increasing complexity. We start from the validation of the log file recording consistency, followed by the validation of the MCsquare beam model and the application of our proposed PSQA workflows to a set of clinical plans. Then, some of the clinical plans are used for validation of the log to plan reconstruction algorithm and the OpenREGGUI implementation for clinical goals evaluation.

1. Log file consistency

Log-files consistency was validated, as per Sub-Section 2.6.1. When comparing the reconstructed spot position and MU to 60 morning QA plans prescriptions over half a year period, the observed average spot position error and standard deviation in X was -0.0339 ± 0.380 mm, and in Y was 0.268 ± 0.470 mm. The maximum accumulated MU error over one entire delivery was 1.75 MU, which corresponds to 0.4% of the prescribed dose (418 MU). A detailed list of the error analysis performed is presented in Table 1.

When analyzing the log reconstructed spot position to measurements from monthly QA plans, relative position errors were calculated. Considering deliveries in Room 1, the average and standard deviation in X was 0.0373 ± 0.221 mm and in Y was 0.0940 ± 0.217 mm; with a maximum position error of 0.514 mm. In Room 2 the results were similar,

VI

Table 1. Comparison between log files recordings and plan prescriptions over 60 deliveries split between two delivery rooms. Statistics for average and standard deviation of position and dose errors are given together with the maximum (max.) observed value in the data set.

Error description	Room 1 (30 deliveries)	Room 2 (30 deliveries)
Spot position shift in x (mm)	-0.0891 ± 0.476 max. 1.15	0.0213 ± 0.250 max. 1.15
Spot position shift in y (mm)	-0.0961 ± 0.347 max. 1.01	-0.439 ± 0.568 max. 1.12
Accumulated MU error (MU)	0.772 ± 0.316 max. 1.38	1.16 ± 0.259 max. 1.75

the average and standard deviation in X was 0.0727 ± 0.155 mm and in Y was 0.0494 ± 0.235 mm; with a maximum position error of 0.386 mm.

2. Beam model validation

As discussed in Sub-Section 2.6.3, the beam model was validated via MC-based calculations of treatment fields performed with MCsquare. The comparison of 30 SOBPs containing ranges between 4.1 and 32 g/cm^2 and modulation between 20 and 40 mm showed a 99% ± 0.5% Gamma passing ratio. For the full energy spectrum from 70 to 225 MeV range discrepancy in water was <1 mm.

The dose calculation accuracy of the MC dose engine was also evaluated using heterogeneous real animal tissues [33]. Comparisons between 2D dose distributions from measurement and simulations with Gamma criterion of 3%/3 mm are provided in Table 2. Gamma pass ratios are approximately 95% or greater for all cases. Deviations are found at high density gradient regions (soft tissue/bones and air/tissue interfaces) and high dose gradients regions, which could be explained by the different material tables used to convert the CT image to chemical compositions in the two engines. MCsquare simulations performed with greater MC statistics (1×109 particles) showed no measurable improvement, indicating lower statistical noise compared to other sources of uncertainties.

Table 2. Accuracy comparison for TPS and MCsquare dose calculations in heterogeneous animal tissue showing Gamma percentage pass ratios for different beam energies and depths. TPS dose maps and MCsquare dose maps are evaluated with 2D Gamma analysis (3%/3 mm).

Beam energy, solid-water depth	TPS dose (γ pass ratio %)	MCsquare dose (γ pass ratio %)
SOBP	98.2	94.9
225 MeV, 3 mm spacing	98.3	98.7
225 MeV, 7 mm spacing	97.9	98.9
200 MeV, 3 mm spacing	99.0	96.9
200 MeV, 7 mm spacing	98.0	94.6
175 MeV, 3 mm spacing	98.5	98.9
175 MeV, 7 mm spacing	99.2	98.3

3. Workflow testing

The feasibility of the TPS-based and Log-based QA workflows was tested against standard measurement-based QA, as referred in SubSection 2.6.4. A comparison between measurement-based, TPS-based and Log-based QA for 10 clinical cases, including craniospinal axis, intracranial and head and neck cases, is summarized in Table 3. An example of calculated dose distributions for a breast cancer case is shown in Fig. 3.

Table 3. Gamma pass ratios for measurement-based plan QA, and the proposed TPS- and Log-based plan QA.

Patient	Measurement based (γ pass ratio %)	TPS-plan recalculation (γ pass ratio %)	Log-plan recalculation (γ pass ratio %)
1	100	97.43	97.71
2	99.81	98.75	96.94
3	100	98.60	91.56
4	100	98.43	98.59
5	98.31	95.37	93.30
6	98.55	96.05	95.54
7	99.44	99.05	97.98
8	99.12	99.14	98.99
9	99.52	99.30	99.14
10	99.31	96.02	94.97
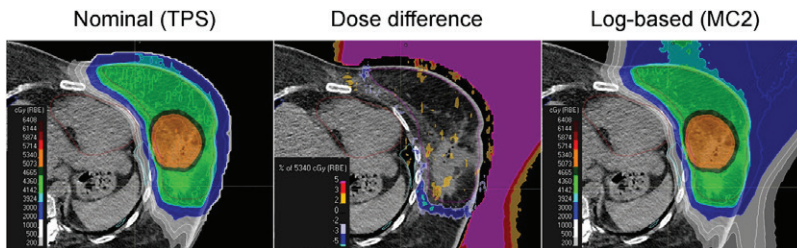
Figure 3. Dose maps for the TPS calculated plan (left), the log-based recalculation (right) and the dose difference (center). For visualization purpose, recalculated dose maps have been imported in TPS. The dose maps are compared with Gamma analysis (2 mm, 2%).
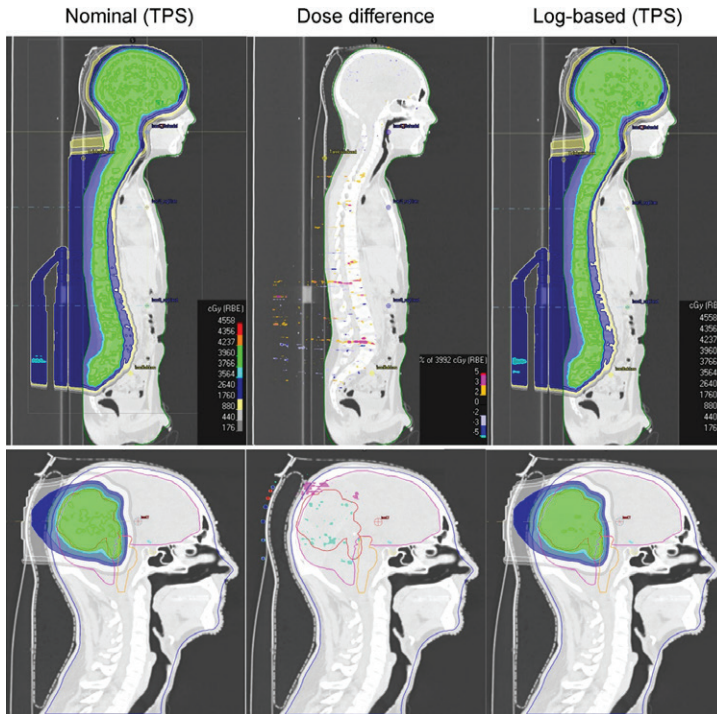
Figure 4. TPS calculated dose maps: planning dose (left), log-based recalculation (right) and the dose difference (center). Isodose curves showed in color. The dose maps are compared with Gamma analysis (2 mm, 2%).

Independent MC calculations for these cases require 15–20 min calculation time per treatment plan. Patient specific QA according to the proposed TPS-based and Log-based methodology requires about 10 min of in-treatment-room time per patient for log file acquisition. That compares to 40 min in-treatment-room time per patient for measurement-based QA.

4. Component testing with clinical plans

Some of the aforementioned clinical plans were used for the validation of CAPTAIN's low-level components, as described in Sub-Section 2.6.2. Based on calculated DVH, no statistical meaningful differences between the two plans are found — example showcased in Fig. 4.

Discussion

The overall trend observed in the workflow validation is that TPS-based plan QA showed lower Gamma pass ratios than measurement-based QA. This is expected, since the former compares the whole 3D dose distribution volume with CT compositions and the later relies on 2D Gamma analysis at different depths of solid-water phantoms. In turn, the Log-based plan QA shows a Gamma analysis pass ratio marginally lower than the aforementioned cases. Since both TPS- and Log-based plan QAs use the same dose computation procedure, the difference can be attributed to discrepancies in delivered spot position or dose, which were recorded on the treatment log files and used during plan reconstruction.

The automated PSQA is managed and executed from inside the CAPTAIN server, a multipurpose and flexible platform. In order to provide specific functionalities, it integrates with other open source projects, such as OpenREGGUI and MCsquare. The use of these modules was crucial to speed up development and guarantee performance, however it also demands comprehensive testing of its components. First, MCsquare beam model profile was validated by comparing simulations for a wide array of SOBPs in solid water phantoms, and in inhomogeneous animal tissue phantoms. Over a wide range of energies and modulations, good agreement was found between MCsquare and TPS dose distributions. Secondly, OpenREGGUI evaluation of clinical goals was validated against the TPS evaluation method for a set of clinical plans. The results indicate that in ROIs subjected to doses higher than 100 cGy, the average difference between the two implementations was lower than 1%. Noticeably, the discrepancies are larger for small doses and small volumes; since different algorithms for DVH computation are used in TPS and OpenREGGUI, processes such as interpolation and voxelization play a greater role in these cases. The evaluation of clinical goals is dependent on the approach taken for calculation of DVHs. As there are different approaches possible, it may introduce bias in clinical goal evaluation. Consistency of the two methods is necessary for a complete re-evaluation of the reconstructed plan from treatment log files, where TPS evaluation of clinical goals are directly compared to OpenREGGUI's.

VI

PSQA procedures based on independent dose calculations, such as ours, have already been introduced for PBS ion therapy in some facilities [12, 34, 35]. The main advantages are the reduced time required for the measurements and the high resolution 3D dose distributions provided in the patient geometry, which improves the verification procedure of both the planned and delivered dose. Recently, at PSI [15], a toolkit for independent dose calculations was developed, which allows for dose reconstructions at several points in the treatment workflow. Still, the implementation of such workflows in clinical environments is restricted to few examples. Our implementation of Log-based PSQA brings a reactive and automated platform for ease of use in clinical workflows. Since all presented here is open source, an interested reader should also be able to implement similar automated worflows in other clinics.

Due to its architecture, the CAPTAIN platform is capable to be extended for other purposes that would also benefit from its modularity. Workflows are simple computation recipes, and existing solutions are easily refitted into workflows' tasks. Based on fraction wise patient information this platform could accomplish automated daily dose reconstruction and accumulation. With an automated comparison of the accumulated dose against the expected dose, a request for adaptation could be automatically triggered in case of deviations. A concept for fraction-wise retrospective 4D dose reconstruction and accumulation was recently published [36]. Within a corresponding automated workflow in our introduced platform, supplied treatment log files, motion records and acquired repeated 4D CTs/CBCTs would trigger a dose reevaluation and would trigger an adaptation if significant treatment quality deviations are observed, such as dose discrepancy due to anatomical changes. We have plans to extend the platform to include such functionality in the future. In this case, the time required for the plan dry-run delivery and log files collection may become a hindrance for a fast online treatment tracking. Since adaptive workflow are generally complex and labor intensive, automatizing work-flows as much as possible as proposed here will be essential for a clinical implementation of adaptive proton therapy.

The benefit from this kind of analysis mainly depends on the accuracy of the log-file values, therefore a preliminary assessment of the uncertainty in the recorded parameters of the scanning pencil beams is recommended.

For example, Li et al. [37] have compared the planned and recorded values with dedicated measurements, to ensure that the monitor and the recording system work properly, and that the log-files are accurate enough to be used for evaluating the uncertainties in the delivered dose, caused by variations in the beam characteristics. Henceforth, similar routine machine QAs and further validation of CAPTAIN's components will be necessary for the clinical realization of the workflows here proposed.

Conclusions

A new PSQA workflow was developed using an automated web platform. Low level components were validated, such as the log to plan converter and clinical goals evaluation. MCsquare beam model was validated in solid water and animal tissue phantoms, displaying dose distributions comparable to others simulated by the Monte Carlo algorithm available in the TPS. The proposed patient specific automated QA shows consistency between the measurement- and Log-based QA for a wide range of clinical plans. This supports potential replacement of measurements with MC-based treatment plan QA in future. The use of this platform in clinical routine has the potential of significantly reducing the required in-treatment-room time for PSQA. Furthermore, the implemented platform has the potential to also automatize other clinical procedures as for example fraction wise dose reconstruction and accumulation, which may provide input for decision support regarding plan adaptation.

VI

References

[1]   Mohan R, Grosshans D. Proton therapy — present and future. Adv Drug Deliv Rev 2017;109:26–44. Radiotherapy for cancer: present and future.

[2]   Lomax A. What will the medical physics of proton therapy look like 10 years from now? A personal view. Med Phys 2018;45(11):e984–93.

[3]   Lomax A. Intensity modulation methods for proton radiotherapy. Phys Med Biol 1999;44(1):185.

[4]   Unkelbach J, Paganetti H. Robust proton treatment planning: physical and biological optimization. Seminars in Radiation Oncology 2018;28(2):88–96. Proton Radiation Therapy.

[5]   Knopf AC, Lomax A. In vivoproton range verification: a review. Phys Med Biol 2013;58(15):R131–60.

[6]   Saini J, Traneus E, Maes D, Regmi R, Bowen SR, Bloch C, et al. Advanced proton beam dosimetry part I: review and performance evaluation of dose calculation algorithms. Transl Lung Cancer Res 2018;7(2).

[7]   Arjomandy B, Sahoo N, Ciangaru G, Zhu R, Song X, Gillin M. Verification of patientspecific dose distributions in proton therapy using a commercial two-dimensional ion chamber array. Med Phys 2010;37(11):5831–7.

[8]   Lin L, Kang M, Solberg TD, Mertens T, Baumer C, Ainsley CG, et al. Use of a novel two-dimensional ionization chamber array for pencil beam scanning proton therapy beam quality assurance. J Appl Clin Med Phys 2015;16(3):270–6.

[9]   Lomax AJ, Böhringer T, Bolsi A, Coray D, Emert F, Goitein G, et al. Treatment planning and verification of proton therapy using spot scanning: initial experiences. Med Phys 2004;31(11):3150–7.

[10]   Trnková P, Bolsi A, Albertini F, Weber DC, Lomax AJ. Factors influencing the performance of patient specific quality assurance for pencil beam scanning IMPT fields. Med Phys 2016;43(11):5998–6008.

[11]   Henkner K, Winter M, Echner G, Ackermann B, Brons S, Horn J, et al. A motorized solid-state phantom for patient-specific dose verification in ion beam radiotherapy. Phys Med Biol 2015;60(18):7151.

[12]   Zhu X, Li Y, Mackin D, Li H, Poenisch F, Lee A, et al. Towards effective and efficient patient-specific quality assurance for spot scanning proton therapy. Cancers 2015;06(7):631–47.

[13]   Winterhalter C, Fura E, Tian Y, Aitkenhead A, Bolsi A, Dieterle M, et al. Validating a Monte Carlo approach to absolute dose quality assurance for proton pencil beam scanning. Phys Med Biol 2018;63(17):175001.

[14]   Matter M, Nenoff L, Meier G, Weber DC, Lomax AJ, Albertini F. Alternatives to patient specific verification measurements in proton therapy: a comparative experimental study with intentional errors. Phys Med Biol 2018;63(20):205014.

[15]   Meier G, Besson R, Nanz A, Safai S, Lomax AJ. Independent dose calculations for commissioning, quality assurance and dose reconstruction of PBS proton therapy. Phys Med Biol 2015;60(7):2819–36.

[16]   Colaborators. The OpenPATh initiative; 2016. https://openpath.software/.

[17]   Colaborators. Open-source; 2019.https://opensource.org/strategic.

[18]   openREGGUI consortium. Image processing open-source platform for adaptive proton therapy in cancer treatment; 2016.https://openreggui.org/.

[19]   MATLAB. The MathWorks Inc, Natick, MA, USA; 1994–2019.https://mathworks.com/.

[20]   Jodogne S. Orthanc v1.4.2; 2012–2018.https://www.orthanc-server.com.

[21]   Souris K, Lee JA, Sterpin E. Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core CPU architectures. Med Phys 2016;43(4):1700–12.

[22]   Bazalova M, Beaulieu L, Palefsky S, Verhaegen F. Correction of CT artifacts and its influence on Monte Carlo dose calculations. Med Phys 2007;34(61):2119–32.

[23]   Paganetti H. Dose to water versus dose to medium in proton beam therapy. Phys Med Biol 2009;54(14):4399–421.

[24]   Sorriaux J, Testa M, Paganetti H, de Xivry JO, Lee JA, Traneus E, et al. Experimental assessment of proton dose calculation accuracy in inhomogeneous media. Phys Med 2017;38:10–5.

[25]   Hammond S, Cantrill B. Node.js v10.11.0; 2009–2018.https://nodejs.org.

[26]   contributors G. Angular Full-Stack generator v5.0.0; 2018.https://github.com/angular-fullstack/generator-angular-fullstack.

[27]   Mao J, Schmitt M, Stryjewski T, Holt CC, Lubelski W. gulp.js v4.0.0; 2013–2018. https://github.com/gulpjs/gulp.

[28]   Ittycheria D, Merriman D, Horowitz E. MongoDB v4.0.2; 2009–2018.https://www.mongodb.com.

[29]   Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys 1998;25(5):656–61.

[30]   Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. Med Phys 2003;30(9):2455–64.

[31]   Biggs S, contributors. npgamma; 2015–2018.https://pypi.org/project/npgamma.

[32]   Graves YJ, Jia X, Jiang SB. Effect of statistical fluctuation in Monte Carlo based photon beam dose calculation on gamma index evaluation. Phys Med Biol 2013;58(6):1839–53.

[33]   Siang KNW. Validating Monte Carlo Calculations of Clinical Proton Beams in Animal Tissue Phantoms. Rijks Universitiet Groningen; 2019.

[34]   Mackin D, Li Y, Taylor MB, Kerr M, Holmes C, Sahoo N, et al. Improving spotscanning proton therapy patient specific quality assurance with HPlusQA, a secondcheck dose calculation engine. Med Phys 2013;40(12):121708.

[35]   Molinelli S, Mairani A, Mirandola A, Freixas GV, Tessonnier T, Giordanengo S, et al. Dosimetric accuracy assessment of a treatment plan verification system for scanned proton beam radiotherapy: one-year experimental results and Monte Carlo analysis of the involved uncertainties. Phys Med Biol 2013;58(11):3837–47.

[36]   Meijers A, Jakobi A, Stützer K, Guterres Marmitt G, Both S, Langendijk JA, et al. Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: implementation and proof-of-concept. Med Phys 2019.

[37]   Li H, Sahoo N, Poenisch F, Suzuki K, Li Y, Li X, et al. Use of treatment log files in spot scanning proton therapy as part of patient-specific quality assurance. Med Phys 2013;40(2):021703.

VI

Chapter VII: Feasibility of patient specific quality assurance for proton therapy based on independent dose calculation and predicted outcomes

Abstract

Purpose: Patient specific quality assurance (PSQA) is required to verify the treatment delivery and the dose calculation by the treatment planning system (TPS). The objective of this work is to demonstrate the feasibility to substitute resource consuming measurement based PSQA ($PSQA_M$) by independent dose recalculations ($PSQA_{IDC}$), and that $PSQA_{IDC}$ results may be interpreted in a clinically relevant manner using normal tissue complication probability (NTCP) and tumor control probability (TCP) models.

Methods and Materials: A platform for the automatic execution of the two following $PSQA_{IDC}$ workflows was implemented: (i) using the TPS generated plan and (ii) using treatment delivery log files (log-plan). 30 head and neck cancer (HNC) patients were retrospectively investigated.

$PSQA_M$ results were compared with those from the two $PSQA_{IDC}$ work-flows. TCP / NTCP variations between $PSQA_{IDC}$ and the initial TPS dose distributions were investigated. Additionally, for two example patients that showed low passing $PSQA_M$ results, eight error scenarios were simulated and verified via measurements and log-plan based calculations. For all error scenarios ∆TCP / NTCP values between the nominal and the log-plan dose were assessed.

Results: Results of $PSQA_M$ and $PSQA_{IDC}$ from both implemented work-flows agree within 2.7% in terms of gamma pass ratios. The verification of simulated error scenarios shows comparable trends between $PSQA_M$ and $PSQA_{IDC}$. Based on the 30 investigated HNC patients, $PSQA_{IDC}$ observed dose deviations translate into a minor variation in NTCP values. As expected, TCP is critically related to observed dose deviations.

Conclusions: We demonstrated a feasibility to substitute $PSQA_M$ with $PSQA_{IDC}$. In addition, we showed that $PSQA_{IDC}$ results can be interpreted in clinically more relevant manner, for instance using TCP / NTCP.

Introduction

The preparation of radiotherapy treatments and their delivery is affected by several sources of uncertainty. Furthermore, radiotherapy treatments require the acquisition, exchange, storage and processing of large amount of digitized data, which can become corrupted. To ensure that treatments are delivered within clinically acceptable tolerances, patient specific quality assurance (PSQA) has always been an essential component of the treatment delivery process.

VII

Historically, first for 2D, and later 3D conformal radiotherapy, PSQA was based on independent dose recalculation and in-vivo dose output measurements. Corresponding recommendations were for example given in IAEA TRS430 [1], which provided guidelines for the implementation of quality assurance (QA) programs in radiotherapy departments. Within the scope of this study we are focusing on PSQA aspects, such as, monitor unit (MU), in a broader sense, dose calculation and delivery check, data transfer and integrity check, but omit such topics as planning process and plan check.

However, with the introduction of intensity modulated radiotherapy (IMRT) and later volumetric modulated arc therapy (VMAT), independent MU recalculations, often performed manually, became non-feasible due to the complexity of the calculations. Therefore, upon adoption of IMRT in the clinic, dose calculations mostly were done by treatment planning systems (TPS). Furthermore, beam modulation required the transfer of large amount of data to the delivery equipment, which demands complex and precise functional performance. In order to gain confidence and to verify the performance of new and relatively non-transparent automated treatment delivery modalities such as IMRT, in-beam measurement-based PSQA procedures became an integral part of QA programs in radiotherapy departments [2], replacing independent MU recalculation and in-vivo dosimetry. Since then $PSQA_M$ procedures have evolved and been addressed by various task groups, for example, AAPM Task Group No. 218. [3]

Since the introduction of particle therapy in clinical practice, PSQA has been mainly based on an approach requiring in-beam measurements ($PSQA_M$). In-beam measurements were a necessity for passively scattered or uniformly scanned proton treatment fields in order to perform field calibration on a routine basis, as TPS was usually providing only relative dose. However, in the recent years with a wide-spread adoption of pencil beam scanning, the usefulness and value of continuous $PSQA_M$ procedures have been questioned [4].

Focusing on particle therapy, numerous groups have proposed, to investigate and implement PSQA procedures that are based on independent dose recalculation ($PSQA_{IDC}$), additionally proposing a use of treatment delivery log files and/or use treatment machine steering files [5], [6] in this process. This topic is of particular interest for particle therapy centers because of the high cost of treatment beam time, in which case maximizing clinical throughput allows treatments to be more accessible to the public. In addition, these novel methods facilitate the deployment of daily adaptive proton therapy (PT).

At our institution, we co-developed and implemented an open source workflow automation platform CAPTAIN [7], on basis of which we deployed a $PSQA_{IDC}$ procedure that relies on independent Monte Carlo (MC) calculations [8] and enables input of treatment delivery log files.

Within the current PSQA$_{IDC}$ process, the evaluation of independently recalculated dose distributions is performed using 3D gamma analysis [9], [10] and the assessment of clinical goals, which are defined and calculated based on dose volume histograms (DVHs).

The currently deployed PSQA$_{IDC}$ workflow consists of two stages: (i) an independent dose recalculation based on the treatment plan as received from the TPS (TPS-plan) and (ii) an independent dose recalculation based on the treatment plan as reconstructed from treatment delivery log files (log-plan), which are obtained from the proton delivery system (PTS) after a dry-run. Although dry-run requires some beam time, in our practice so far time required is significantly lower than for a complete PSQA$_M$ procedure (5-7 minutes vs 30-35 minutes). The calculations are performed in the patient geometry. The independence in the PSQA$_{IDC}$ approach is achieved through an entirely independent implementation of secondary dose calculation engine from the primary TPS dose calculation engine. In addition, TPS and IDC uses different material lookup tables for determining elemental composition related to CT numbers.

In the Netherlands, in accordance with a national consensus, for most indications patient selection for PT is made following a model-based approach [11], [12]. The underlying principle of the model-based approach is to select a treatment (protons or photons) on patient-specific basis that would allow to minimize risk of therapy induced complications. This is done by calculating normal tissue complication probability (NTCP) according to approved models for photon and proton treatment plans with identical target coverage and determining the difference in NTCP (ΔNTCP) between these two plans. If ΔNTCP is above a certain nationally agreed threshold, the patient is referred for PT. In the framework of a Model Based Clinic (MBC), a secondary application of PSQA$_{IDC}$ could be an additional confirmation of the decision-making process underlying patient selection, where NTCP values may be recalculated based on QA dose distributions.

The purpose of this study is to further explore PSQA procedures based on automation and independent dose recalculation (PSQA$_{IDC}$) within the unique environment of the MBC. Specifically, we investigate feasibility to link PSQA$_{IDC}$ with clinically relevant measures adopted in the MBC, while also providing means to enclose model-based patient selection

VII

process within the overall PSQA procedure. In addition, the sensitivity of various indicators towards delivery errors is evaluated.

Methods and Materials

A group of 30 consecutive head and neck cancer (HNC) patients was retrospectively evaluated in this study. For these patients NTCP values were calculated based on the dose distributions as calculated in the TPS (RayStation 8B, RaySearch, Sweden) by its clinical dose calculation algorithm (Monte Carlo v.4.4). In addition, both dose distributions (TPS-plan and log-plan) calculated by an independent MC dose calculation engine (MCsquare) were used to recalculate NTCP values. MCsquare is an open-source Monte Carlo proton dose calculation engine [13], [14], which utilizes multi-threaded processing to ensure fast calculation times. Furthermore, $PSQA_M$ results were retrieved and compared to $PSQA_{IDC}$ results in terms of gamma pass ratios. The $PSQA_M$ procedure for the presented cases has been performed at 3 measurement depth (1 cm and two additional in high dose region varying per field). The presented gamma pass ratio per patient was calculated as a ratio between the number of all passing measurement points versus the total number of measurement points (all fields, all depths combined).

Additionally, two patients with relatively low gamma pass ratios as shown in the currently employed nominal $PSQA_{IDC}$ workflow were selected. To establish a consistency baseline for log file-based calculations, treatment delivery log files for 5 clinical fractions were collected and QA doses were calculated using the log-plan based workflow. Afterwards, for these two patients, multiple error scenarios (ES) of the nominal plan were created. A python script to alter spot positions and MU in DICOM ion plans was created. It was used to introduce offsets to the prescribed spot positions and MU for the selected treatment plans. To introduce errors for each spot, offsets were randomly sampled from normal distributions. Maximum allowed offsets (2 sigma) were predefined per ES and are listed in Table 1. In this context, the absolute error is a fixed offset applied to the whole layer and the relative error is an offset applied to an individual spot.

Table 1. Summary of maximum introduced errors per spot (2 sigma) per error scenario.

	Absolute position error, mm	Relative position error, mm	MU error, %
ES1	0.5	0.5	0
ES2	1.0	1.0	0
ES3	1.0	2.0	0
ES4	0	0	1
ES5	0	0	2.5
ES6	1	2	2.5
ES7	2	2	3
ES8	2	4	5

VII

Error scenarios 1 to 6 are designed such that introduced offsets are within tolerances set in the treatment control system, which monitors the proton beam delivery online, therefore, such offsets could in principle appear also in the delivery log files. In contrast, scenarios 7 and 8 are rather theoretical. If such offsets would occur during beam delivery, the delivery would be interrupted by the treatment control system.

For the selected two additional cases (error scenario cases) the nominal plans and all error scenario plans were delivered by the PTS, while performing PSQA$_M$ procedure with a 2D ionization chamber array MatriXX PT (IBA Dosimetry, Schwarzenbruck, Germany). The array was positioned at 1 cm depth, in order to capture all energy layers within the field. Furthermore, a measurement at only one depth per field for the error scenarios was done to limit beam time usage. Each treatment plan consisted of 4 treatment fields. Measured dose distributions were analyzed using global 2D gamma analysis with 2 mm / 2% criteria and a cutoff value of 10%.

Furthermore, log files were collected for these deliveries. Using the deployed PSQA$_{IDC}$ workflow, independent MC dose calculations were performed using the log files from the nominal plan and the error scenarios. Based on these nominal and error scenario doses, the following quality control parameters were calculated: gamma pass ratios (criteria 2 mm / 2%) and the variations in TCP and NTCP values.

NTCP values were calculated for grade 2 xerostomia [15], [16] and dysphagia [17], [18], [19] and for grade 3 tube feeding dependence [20]. In addition to the risk factors, the probability of xerostomia in the used model is correlated with the mean dose to the contralateral parotid gland. The probability of dysphagia is correlated with mean dose to the oral cavity and to the superior pharyngeal constrictor muscle (PCM), while the probability of tube feeding dependence is correlated with the mean dose to the superior PCM, inferior PCM, contralateral parotid gland and cricopharyngeal muscle.

TCP values were calculated based on the model proposed by Lühr et al. [21]. Model parameters (tumor control dose D_{50} and slope γ_{50}) were not calibrated to reflect tumor control probability in our clinical practice. Values for these parameters were chosen identical to estimations made by Lühr et al. In the proposed model TCP corelates with the DVH of the primary gross tumor volume (GTV), primary clinical tumor volume (CTV) and elective CTV. TCP values were calculated purely for illustrative purposes.

Results

The results for the measurement based and the two independent dose recalculations based PSQA procedures for the first ten HN patients are shown in Figure 1. The results include 2D gamma pass ratios (2 mm / 2%) for $PSQA_M$ and 3D gamma pass ratios (2 mm / 2%) for independent dose recalculation based on the TPS-plan and the log-plan. Most of the plans consisted of 4 treatment fields, with 2 exceptions (pat. 1 and 2), where treatment plans had 5 fields.

Table 2 summarizes the results for variations in NTCP and TCP as calculated based on initial TPS dose distributions compared to recalculated dose distributions based on either the TPS-plan or the log-plan. Appendix I summarizes the ΔNTCP data for all 30 patients.

Overall, for the entire 30 patients cohort, average ΔNTCP of 0.2% (SD 0.2%) was observed for dysphagia, when comparing nominal dose distribution to TPS-plan based QA dose distribution, and 0.1% (SD 0.2%), when comparing to log-plan based QA dose distribution. Average ΔNTCP
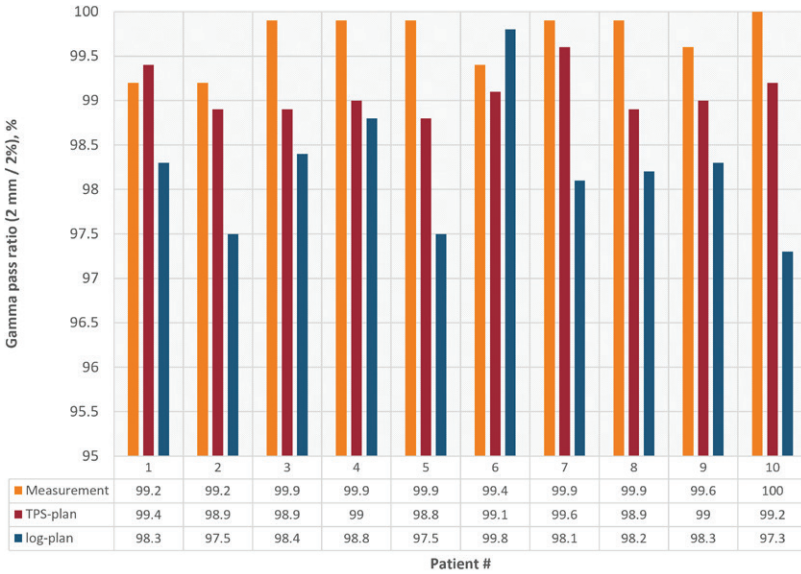
Figure 1. Overview of gamma pass ratios (2 mm / 2%) for 10 head and neck proton therapy patients. Results are shown for the $PSQA_M$ and $PSQA_{IDC}$ procedures.

Table 2. Overview of NTCP and TCP variations between nominal and QA dose distributions for 10 head and neck patients. Variations in NTCP are shown for grade 2 dysphagia and xerostomia and grade 3 tube feeding dependence at 6 months post radiotherapy.

VII

	TPS-plan dose				log-plan dose			
	ΔNTCP, %				ΔNTCP, %			
Pat.	Dysph.	Xerost.	Tube feeding	ΔTCP, %	Dysph.	Xerost.	Tube feeding	ΔTCP, %
1	0.1	0.0	0.0	2.2	−0.1	0.1	0.0	3.1
2	−0.2	0.0	−0.3	1.1	−0.3	0.0	−0.3	2.1
3	0.2	−0,4	0,0	2.0	0,2	−0,4	0,0	2.1
4	−0.1	0.1	−0.1	1.6	−0.1	0.0	−0.1	1.9
5	0.5	−0.3	−0.2	1.7	0.0	−0.5	−0.4	3.0
6	0.1	−0.3	−0.2	1.4	−0.2	−0.2	−0.3	2.9
7	0.2	−0.4	0.0	1.5	0.1	−0.1	0.0	2.6
8	0.1	−0.5	−0.1	2.5	0.2	−0.3	−0.1	3.3
9	0.2	−0.1	0.0	2.1	0.1	−0.1	0.0	2.3
10	0.1	−0.3	−0.2	2.1	−0.3	−0.5	−0.4	3.5

Figure 2. Trends of gamma pass ratios for 2 error scenario cases, for whom a set of 8 error scenarios was generated and evaluated according to $PSQA_M$ and $PSQA_{IDC}$ procedures. ES0 corresponds to the nominal plan, where no offsets to the prescribed spot positions or MU have been introduced.

of −0.1% (SD 0.3%) was observed for xerostomia, when evaluating TPS-plan QA dose distribution, and −0.1% (SD 0.3%), in case of log-plan QA dose distribution. While for tube feeding dependence average ΔNTCP of 0.0% (SD 0.1%) was observed for evaluation of TPS-plan QA dose distribution and −0.1% (SD 0.2%) for log-plan QA dose distribution.

The consistency check for the log file-based calculations, as performed using log files from 5 clinical fractions for the 2 error scenario cases, showed SD of 0.1% for gamma pass ratios. In addition, results from the error scenarios test are shown in Figure 2.

Results include 2D gamma pass ratios for the measurements performed at 1 cm depth with MatriXX ionization chamber array and 3D gamma pass ratios for dose recalculated based on treatment delivery log files as collected from deliveries of treatment plans with introduced offsets to spot positions and prescribed MU.

In Table 3 the effect of introduced errors is reflected in the changes of NTCP and TCP values for the error scenario cases. The shown difference in TCP / NTCP is determined by comparing TCP / NTCP values as calculated for nominal dose distributions and TCP / NTCP values as calculated for dose distributions, which were obtained by recalculating log-file based treatment delivery plans.

Table 3. Overview of TCP and NTCP variations between nominal and QA dose distributions for error scenarios, which were generated for treatment plans of the two HNC patients (error scenario cases). Scenario ES0 represents $PSQA_{IDC}$ of the unaltered plan.

| | Patient A | | | | Patient B | | | |
| | ΔNTCP, % | | | | ΔNTCP, % | | | |
	Dysph.	Xerost.	Tube feeding	ΔTCP, %	Dysph.	Xerost.	Tube feeding	ΔTCP, %
ES0	0.1	−0.5	−0.2	2.5	0.2	0.3	0	2.6
ES1	0.3	−0.1	−0.1	2.6	0.2	0.3	0	1.4
ES2	−0.2	−1	−0.4	1.7	0	−0.4	−0.1	1.8
ES3	−0.1	−0.3	−0.3	2.3	−1	0	−0.1	−0.3
ES4	0.1	−0.4	−0.2	2.2	0.3	0.4	0	0.6
ES5	0.2	−0.4	−0.2	2.3	0.2	0.2	0	1.4
ES6	0.4	−0.2	−0.1	2.1	0.3	0.6	0	0.8
ES7	0.4	0.1	−0.1	3.7	0	0	0	1.4
ES8	−0.2	−1	−0.5	4.0	−0.9	0.4	−0.1	2.5

Discussion

VII

Consistency can be observed between gamma pass ratios (2 mm / 2%) for $PSQA_M$ and $PSQA_{IDC}$ as shown by trends in Figure 1 and 2. Consistent decisions regarding plan quality would be made according to either PSQAM or PSQAIDC (Figure 1) and lower gamma pass ratios would be observed with either method in case of delivery errors (Figure 2). In most cases, gamma analysis performed for measurements done at 3 depths per field scores higher gamma pass ratios than for independent dose recalculation based PSQA approach. It is not unexpected that gamma pass ratios between $PSQA_M$ and $PSQA_{IDC}$ will not match, because the two PSQA methods have some major fundamental differences, such as the testing medium. $PSQA_M$ is based on water-like medium, while $PSQA_{IDC}$ is based on patient geometry depicted in the planning CT. Furthermore, in $PSQA_M$ steep gradient regions, especially in longitudinal direction, such as distal dose falloff, are often avoided. Gradient regions would usually score lower gamma pass ratios, if included. Furthermore, the number of evaluation points in case of $PSQA_{IDC}$ is much larger, as in case of $PSQA_M$, only a limited number of dose planes in sampled.

When comparing the two recalculation-based PSQA results, the log-plan typically scores slightly lower gamma pass ratios than the TPS-plan. This can be explained by the fact that the log-based plan also includes delivery discrepancies in spot position and delivered MU per spot compared to the nominal plan (TPS-plan). In this sense Patient 6 is an outlier. There were no unusual specifics noticed in the plan design. This behavior might be explained by statistical noise in the MC calculations in combination with already relatively high gamma pass ratios for this case.

For all 30 clinical cases plans scored high gamma pass ratios according to applied PSQA procedures. Therefore, also no major variations were observed in the NTCP values as calculated for the three selected complication models. Average observed ΔNTCP values were close to zero (0.2% for dysphagia) and the standard deviation remained small (0.3% for xerostomia). By reviewing ΔNTCP values in the process of PSQA supplementary to the gamma analysis, one can make better judgement on the clinical relevance of the observed variations.

Furthermore, by investigating ΔNTCP values between nominal and QA dose distributions, it can be ensured that patient selection for photon or proton therapy, in the context of MBC, is covered within the PSQA program and decision making is reliable and consistent. In fact, observed maximum variations for discussed 30 clinical cases do not exceed the uncertainty of the NTCP value itself [15], [16] and they are small compared to the clinical decision thresholds (currently set in the Netherlands at 10% for Common Terminology Criteria for Adverse Events (CTCAE) grade 2 and 5% for grade 3 toxicities). Therefore, it may be considered that decisions made regarding patient selection have been robust against the sources of errors covered by QA process itself.

It should be noted that the used NTCP models are limited to specific complications and do not cover all possible radiation induced complications. Furthermore, NTCP models vary greatly in terms of their quality, availability of validation and may need a population-specific calibration. In the absence of comprehensive selection of NTCP models, clinical goals based on DVH statistics might be employed. For instance, by monitoring mean dose to such structures as the oral cavity, PCMs, cricopharyngeal muscle and parotids, one might identify cases when out-of-tolerance deviations occur. For Patient B ES3 (dysphagia ΔNTCP 1%) mean dose

Table 4. Dose statistics for selected organs at risk of exemplary case Patient B. Mean doses are shown for dose distribution as calculated by TPS, dose distribution as reconstructed based on delivery log files of the nominal plan (ES0), and log files of the error scenario 3 (ES3).

	TPS dose, Gy_{RBE}	ES0 log-dose, Gy_{RBE}	ES3 log-dose, Gy_{RBE}
PCM superior	33.5	33.3	34.8
Oral cavity	16.4	16.2	17.4
PCM inferior	40.5	40.2	39.8
Cricopharyngeal m.	16.0	16.5	17.0
Contralateral parotid	20.0	19.7	20.0

increase of 1.3 Gy_{RBE} to PCM superior and 1.0 Gy_{RBE} to oral cavity was observed. As an example, dose statistics for this case are shown in Table 4.

The evaluation of eight error scenarios for two exemplary patient cases revealed consistency of trends for gamma pass ratios between $PSQA_M$ and $PSQA_{IDC}$ procedures. For instance, ES3 for Patient B shows drop in gamma pass ratio for both QA methods as can be seen in Figure 2. Some of the error scenarios (such as, ES2, ES3, ES8) resulted in larger deviations of NTCP values between nominal and QA dose distributions, reaching as much as 1% variations. An example of inconsistency between gamma pass ratio and clinical implications can be observed in xerostomia ΔNTCP values for Patient A. By comparing ES2 and ES8 metrics, one can observe that gamma pass ratios for these scenarios are 98.7% and 89.4% respectively ($PSQA_{IDC}$ method), however both scenarios result in the same 1% increase in probability of xerostomia. These discrepancies may originate from different sources. First, dose deviations with different signs may cancel out in an organ at risk with no relevant change in the mean OAR dose and the NTCP as a result. Otherwise, dose deviations may be spatially located outside of organs at risk as recognized by the used NTCP models. This may be a sign that these dose deviations are not relevant, or, that the NTCP models are incomplete. Therefore, the use of comprehensive NTCP profiles that include multiple toxicities and multiple organs at risk will be paramount for the clinical interpretation of the QA results. Due to a recent worldwide increase in data registration

VII

programs and implementation of MBCs it is expected that more and better models for such profiles will emerge in coming years. In our institution we are working on a comprehensive profile for HNC patients that includes 22 toxicities at several time points and describes dose-effect relationships in 14 distinct organs at risk (preliminary results presented by van den Bosch et al. [22]). Furthermore, as models become more individualized, the dose-effect relationships may become steeper, allowing increasingly critical evaluation of dose deviations.

It can be observed that gamma pass ratios in case of $PSQA_M$ are slightly higher than $PSQA_{IDC}$ for the shown 10 clinical cases, while the opposite behavior can be noticed for error scenario analysis. This is linked to the fact that measurements were performed at three depths for the 10 clinical cases, while for error scenario analysis only one proximal depth of 1 cm was chosen to capture all layers and be more sensitive to the introduced errors, resulting in lower gamma pass ratios. Although evaluations at 1 cm depth might be associated with increased dose calculation uncertainties due to the dose calculation engine, these effects are more pronounced for analytical engines. Based on the commissioning process (average gamma pass ratio 99.6% (SD 0.8%)), the 1 cm depth has been used as a standard depth of measurement in our clinic for shallow depth region. Overall a good agreement between TPS dose and measurements has been observed. To provide a baseline value, for clinical plans (based on 30 patient cohort) the mean gamma pass ratio of measurements at 1 cm depth is 99.7% (SD 0.6%).

The used TCP model highly correlates with the DVH of the GTV. In our case, the independent dose calculation engine systematically overestimates dose to the target volume by about 1% compared to the clinical TPS dose calculation engine. Therefore, about 2% TCP increase for QA doses can be systematically observed (see Table 2). Furthermore, as mentioned earlier, model parameters were not calibrated to represent our clinical experience. Nonetheless, increase in TCP may indicate formation of hot areas (see Table 2, Pat. A, ES8) and decrease would indicate formation of cold areas (see Table 2, Pat. B, ES3). In absence of calibrated and reliable TCP models, one might introduce clinical goals derived from the DVHs, similarly as was suggested for coping with the lack of NTCP models. For instance, CTV D2 for Patient A ES8 increased by 2.1 Gy_{RBE}, while CTV D98 for Pat. B ES3 decreased by 1.9 Gy_{RBE}.

There is a major role for $PSQA_M$ procedures during the launch of a new facility or introduction of a treatment modality or new indication. However, in long term such procedures cost enormous amount of beam time, while bringing rather limited added value. Transition towards adaptive radiotherapy, where adaptations are performed over increasingly shorter time frames, will make $PSQA_M$ procedures obsolete. If the primary objectives of PSQA are to (i) verify TPS calculation accuracy (avoiding software bugs in specific conditions), (ii) verify accuracy of treatment delivery equipment and (iii) confirm integrity of data during their transfer process, it might be possible to perform these PSQA tasks with a process that does not rely on in-beam measurements. For instance, TPS calculations can be verified by independent dose recalculation, accuracy of the treatment delivery equipment should be checked during thorough machine QA procedures, while data transfer integrity from TPS to PTS and consistency with the prescription can be checked prospectively by performing analysis of the machine steering files, while retrospectively the check of treatment delivery log files can be done. By allowing PTS to translate the plan into machine steering files as a part of PSQA also partially would allow to check plan deliverability, since in practice it may occur that PTS is unable to translate a plan into machine steering files. However, situations, when plan is not deliverable due to technical failures of the hardware, would not be detected. Eventually, interpreting QA results in a clinically meaningful manner will facilitate decision making regarding the quality of the treatment course.

VII

With an availability to retrieve and process daily delivery related information, such as treatment delivery log files, daily imaging data [23], etc., in an automated way and being able to link the outcome of the analysis to clinically meaningful parameters, such as clinical goals, TCP and NTCP, as one of the possible future directions for PSQA might be a process that would allow to continuously monitor treatment course and rise warnings, when deviations from physician's intent occur.

In conclusion, we demonstrated the feasibility to implement a $PSQA_{IDC}$ procedure that allows to check TPS calculation accuracy, deliverability and consistency with the prescription, while providing means to interpret PSQA results in a more clinically relevant manner by means of TCP / NTCP. As a secondary outcome, MBC may benefit from the proposed approach, which may be used for QA of the patient selection process.

Acknowledgements

References

[1]   Sharpe MB. IAEA Technical Reports Series No. 430: Commissioning and Quality Assurance of Computerized Planning Systems For Radiation Treatment Of Cancer. Medical Physics, 2006; 33(2):561.

[2]   Agazaryan N, Solberg TD, DeMarco JJ. Patient specific quality assurance for the delivery of intensity modulated radiotherapy. Medical Physics, 2003; 4(1):40-50.

[3]   Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, Li H, Wijesooriya K, Shi J, Xia P, Papanikolaou N, Low DA. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys. 2018 Apr;45(4):e53-e83.

[4]   Kry SF, Molineu A, Kerns JR, Faught AM, Huang JY, Pulliam KB, Tonigan J, Alvarez P, Stingo F, Followill DS. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. Int J Radiat Oncol Biol Phys. 2014 Dec 1;90(5):1195-201.

[5]   Belosi MF, van der Meer R, Garcia de Acilu Laa P, Bolsi A, Weber DC, Lomax AJ. Treatment log files as a tool to identify treatment plan sensitivity to inaccuracies in scanned proton beam delivery. Radiother Oncol. 2017 Dec;125(3):514-519.

[6]   Matter M, Nenoff L, Meier G, Weber DC, Lomax AJ, Albertini F. Alternatives to patient specific verification measurements in proton therapy: a comparative experimental study with intentional errors. Phys Med Biol. 2018 Oct 17;63(20):205014.

[7]   https://openpath.software/software_package/captain.html *Last visit: 30.03.2020.*

[8]   Guterres Marmitt G, Pin A, Ng Wei Siang K, Janssens G, Souris K, Cohilis M, Langendijk JA, Both S, Knopf A, Meijers A. Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy. Phys Med. 2020 Feb;70:49-57.

[9]   Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. Med Phys. 2003 Sep;30(9):2455-64.

[10]   Chang C, Poole KL, Teran AV, Luckman S, Mah D. Three-dimensional gamma criterion for patient-specific quality assurance of spot scanning proton beams. J Appl Clin Med Phys. 2015 Sep 8;16(5):381–388.

[11]   Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol. 2013 Jun;107(3):267-73.

[12]   Widder J, van der Schaaf A, Lambin P3 Marijnen CAM, Pignol JP, Rasch CR, Slotman BJ, Verheij M, Langendijk JA. The Quest for Evidence for Proton Therapy: Model-Based Approach and Precision Medicine. Int J Radiat Oncol Biol Phys. 2016 May 1;95(1):30-36.

[13]   Souris K, Lee JA, Sterpin E. Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core CPU architectures. Med Phys. 2016 Apr;43(4):1700.

[14]   Barragán Montero AM, Souris K, Sanchez-Parcerisa D, Sterpin E, Lee JA. Performance of a hybrid Monte Carlo-Pencil Beam dose algorithm for proton therapy inverse planning. Med Phys. 2018 Feb;45(2):846-862.

[15]   Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, Chouvalova O, Leemans CR, de Bock GH, Christianen ME, van der Laan BF, Vissink A, Steenbakkers RJ, Langendijk JA. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. Radiother Oncol. 2012 Oct;105(1):86-93.

[16]   Beetz I, Schilstra C, van der Schaaf A, van den Heuvel ER, Doornaert P, van Luijk P, Vissink A, van der Laan BF, Leemans CR, Bijl HP, Christianen ME, Steenbakkers RJ, Langendijk JA. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. Radiother Oncol. 2012 Oct;105(1):101-6.

[17]   Dirix P, Abbeel S, Vanstraelen B, Hermans R, Nuyts S. Dysphagia after chemoradiotherapy for head-and-neck squamous cell carcinoma: dose-effect relationships for the swallowing structures. Int J Radiat Oncol Biol Phys. 2009 Oct 1;75(2):385-92.

[18]   Eisbruch A, Kim HM, Feng FY, Lyden TH, Haxer MJ, Feng M, Worden FP, Bradford CR, Prince ME, Moyer JS, Wolf GT, Chepeha DB, Ten Haken RK. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. Int J Radiat Oncol Biol Phys. 2011 Nov 1;81(3):e93-9.

[19]   Feng FY, Kim HM, Lyden TH, Haxer MJ, Feng M, Worden FP, Chepeha DB, Eisbruch Intensity-modulated radiotherapy of head and neck cancer aiming to reduce dysphagia: early dose-effect relationships for the swallowing structures. Int J Radiat Oncol Biol Phys. 2007 Aug 1;68(5):1289-98.

[20]   Wopken K, Bijl HP, Langendijk JA. Prognostic factors for tube feeding dependence after curative (chemo-) radiation in head and neck cancer: A systematic review of literature. Radiother Oncol. 2018 Jan;126(1):56-67.

VII

[21] Lühr A, Löck S, Jakobi A, et al. Modeling tumor control probability for spatially inhomogeneous risk of failure based on clinical outcome data. Z. Med. Phys. 2017;27:285–299.

[22] Van den Bosch L, van der Schaaf A, van der Laan HP, Steenbakkers R, Both S, Shuit E, Wijers O, Hoebers F, Langendijk JA. Comprehensive NTCP-Profiling to Predict Radiation-Induced Side Effects in Head and Neck Cancer Patients Treated with Definitive Radiotherapy. Proceedings of the Amercian Society for Radiation Oncology 61st Annual Meeting. September 1, 2019 Volume 105, Issue 1, Supplement, Pages E384–E385. doi: https://doi.org/10.1016/j.ijrobp.2019.06.1648.

[23] Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol. 2020 Mar 6.

Appendix I

NTCP variations between nominal and QA dose distributions for 30 consecutive head and neck cancer patients.

Pat.	TPS-plan dose			log-plan dose		
	ΔNTCP, %			ΔNTCP, %		
	Dysph.	Xerost.	Tube feeding	Dysph.	Xerost.	Tube feeding
1	0.1	0.0	0.0	−0.1	0.1	0.0
2	−0.2	0.0	−0.3	−0.3	0.0	−0.3
3	0.2	−0.4	0.0	0.2	−0.4	0.0
4	−0.1	0.1	−0.1	−0.1	0.0	−0.1
5	0.5	−0.3	−0.2	0.0	−0.5	−0.4
6	0.1	−0.3	−0.2	−0.2	−0.2	−0.3
7	0.2	−0.4	0.0	0.1	−0.1	0.0
8	0.1	−0.5	−0.1	0.2	−0.3	−0.1
9	0.2	−0.1	0.0	0.1	−0.1	0.0
10	0.1	−0.3	−0.2	−0.3	−0.5	−0.4
11	0.0	0.6	0.0	0.0	0.4	−0.1
12	0.4	0.0	0.1	0.3	0.1	0.1
13	0.2	0.2	−0.1	0.0	0.4	−0.1
14	0.4	0.2	0.0	0.3	0.1	0.0
15	0.1	0.3	−0.1	−0.3	0.7	0.0
16	0.4	0.0	0.1	0.4	0.0	0.1
17	0.0	−0.3	0.0	0.0	−0.1	0.0
18	0.0	−0.1	0.0	0.0	0.1	0.0
19	0.3	−0.2	0.2	0.4	0.0	0.3
20	0.3	−0.2	0.1	0.3	−0.1	0.0
21	0.1	−0.2	0.1	0.1	−0.1	0.1
22	0.2	0.0	0.1	0.2	0.2	0.1
23	0.2	0.0	0.0	0.2	0.2	0.0
24	0.7	−0.1	0.1	0.5	−0.2	0.0
25	0.3	−0.1	0.1	0.2	0.1	0.1
26	0.2	−0.1	0.0	−0.3	−0.1	−0.2
27	0.1	−0.5	−0.2	−0.1	−0.4	−0.2
28	0.4	−0.5	−0.2	0.1	−0.6	−0.5
29	0.0	−0.4	0.0	−0.3	−0.6	−0.4
30	0.4	−0.3	−0.2	0.3	−0.3	−0.2

VII

Discussion

Proton range uncertainty is broadly considered to be a major limiting factor for utilizing proton dosimetric advantages to their full potential. In Chapters I "Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing" and IV "Assessment of range uncertainty in lung-like tissue using a porcine lung phantom and proton radiography", it is demonstrated that actual range errors can be kept within a 3% uncertainty margin, when using the Monte Carlo dose calculation engine for treatment planning. This can be achieved even for CT simulation workflows that are based on single energy CT protocols as shown by range accuracy validation procedures, which utilize range probing measurements for animal tissue samples. Range accuracy is validated for tissue types found across the body, including lung tissue, organ- and fat-like tissue, as well as bone tissue. Currently, significant effort is being dedicated to a further reduction in range uncertainty in the treatment planning process by introducing dual energy CT protocols for patient simulation. It is expected that DECT protocols might reduce range uncertainty margins to a level of around 2% [1]. Validation procedures using range probing as proposed in this thesis could further be applied to validate range accuracy for DECT-based patient simulation protocols in the future.

Furthermore, the 3% range accuracy level was also confirmed *in-vivo* as shown in Chapter II "First report on an in vivo proton radiography quality control procedure for scanned proton beam therapy in head and neck cancer patients". The range probing-based quality control procedure was applied in clinical practice to monitor range accuracy in head and neck cancer patients on fraction-specific basis, which was done twice per

individual treatment course. However, in studies of patients, in contrast to any phantom or animal tissue studies, it can be observed that the main contribution to range errors is not linked to the physical aspects of range uncertainty, but to the patients themselves. 3% range accuracy level can be observed in y relatively stable parts of the body. However, significantly higher range inaccuracies (up to 10% or more) can be caused due to anatomical changes of the patient compared to the patient's model, which was acquired for the treatment planning purposes. These anatomical changes can be grouped into two categories: (1) variations in shape caused by different anatomy itself, such as weight gain or weight loss, tumor shrinkage, swelling, etc., and (2) variations in position caused by different position of the anatomical structures in relation to other structures, such as different positions of the tongue, swallowing muscles, neck muscles, etc. In practice, repositioning of the patient to compensate for variations in position is not necessarily feasible, as conscious control over the base of tongue or swallowing muscles with regard to their position is fairly limited. These observations are once again a strong advocate for the introduction of adaptive therapy, where streamlined offline adaptive therapy would allow for better handling of relatively slower anatomical variations, which have an observable trend over time. Online adaptive therapy would allow faster pace anatomical variations to be observed on the day of treatment itself. It is worth mentioning that it is certainly beneficial to further reduce uncertainties in the patient's model by addressing physics-related sources of uncertainty, such as working towards adoption of DECT protocols in the clinical workflows. However, a highly certain patient's model in the physical sense does not necessarily mean a highly accurate patient's model overall. As stated earlier and reported in Chapter II, range errors caused by anatomical variations can reach or exceed the 10%-level, which greatly surpasses 3% (SECT) or potential 2% (DECT) range uncertainty linked to the physical sources of the error. In addition to investigating range uncertainty margins, a setup error recipe for head and neck cancer patients was investigated [2]. Furthermore, during the timeframe of the current work, comprehensive 4D robustness evaluations for lung and esophageal indications were performed [3].

A major improvement in the quality of treatment could be expected if it were possible to react to the observed anatomical inconsistencies through

adaptive proton therapy. With this approach, frequent updates or renewals of the patient's model suitable for treatment planning purposes would be necessary. Recent advancements in the field of Artificial Intelligence show the potential of using neural networks to create synthetic CTs (also investigated during the timeframe of the current work [4, 5]), suitable for proton dose calculation, based on the daily imaging data (such as, CBCTs). However, processing of the images by neural network algorithms is sort of a "black box" problem for an end user. Therefore, when considering such tools for clinical use, comprehensive quality control procedures should be applied. In the future, the proposed range probing-based quality control method could also be used to validate such synthetic CTs for their use in the clinical adaptive therapy workflow.

A different online range verification method based on prompt gamma imaging has been proposed, investigated, and reported on in literature [6]. Although prompt gamma imaging can be performed for therapeutic proton pencil beams (unlike range probing, no transmission beams are required), the accuracy of prompt gamma measurement is highly correlated with the dose delivered by pencil beams of interest. Furthermore, uses of prompt gamma cameras require high positioning accuracy, which translates into a significant engineering effort required in order to make cameras accessible for broader use in clinics.

When considering adaptive therapy, the opportunity to reconstruct and accumulate the previously delivered treatment doses becomes desirable. Furthermore, knowledge of the doses delivered so far can even guide the decisions about the need of plan adaptations throughout the treatment course. From such a perspective, dose reconstruction becomes a great quality control tool, allowing monitoring progress of the treatment course and reacting, when necessary. Having reliable knowledge of the overall accumulated treatment dose provides the opportunity to link this dose to the outcomes of the therapy. In this way, the accuracy of predictive outcome models (TCP / NTCP) may potentially be improved in the near future. Dose reconstruction and accumulation is challenging due to multiple reasons; these include: (1) the lack of accurate daily anatomical information of the patient in the treatment room and (2) radiobiological uncertainties in dose addition from multiple fractions. Dose reconstruction for IMPT therapy indications, affected by breathing motion,

is particularly complicated due to the interplay effect between the time structure of the pencil beam scanning and organ motion. In Chapter III "First report on an in vivo proton radiography quality control procedure for scanned proton beam therapy in head and neck cancer patients" a methodology to perform 4D dose reconstructions using treatment delivery log files and fraction-specific breathing patterns was proposed. The validation of the method using 4D phantoms revealed that sub-millimetric spatial reconstruction accuracy can be achieved. The methodology was further applied in clinics by performing fraction-specific 4D dose reconstructions for patients undergoing proton therapy treatments for indications in thoracic region. Outcomes were reported in Chapter V. Performed reconstructions made it possible to observe that organ motion does cause loss of dose homogeneity in the target volume on a fractional basis. However, local cold and hot spots do not have systematic behavior and adequate overall target dose homogeneity can be maintained. This observation confirmed proton beam rescanning, spot size enlargement and robust optimization as effective means to mitigate organ motion. The dose reconstruction method made it possible to gradually expand motion limits, which were used as a guideline for determining whether proton therapy can be administered safely. By performing close monitoring of the treatment course progress through fraction-wise dose reconstructions, it is possible to evaluate the quality of treatment course and determine if corrective actions for the treatment are necessary. The study also reported that half of the patients required at least one re-planning during the delivery of the treatment course. However, all adaptations were required due to anatomical changes of the patients. This once again highlights the importance of adaptive proton therapy and underlines the fact that anatomical variations are some of the greatest sources of uncertainty (also investigated during the timeframe of the current work [7, 8]), in this case, having a greater degrading effect on the dose distributions than the ones caused by organ motion.

A limitation for the study reported in Chapter V "Evaluation of interplay and organ motion effect by means of 4D dose reconstruction and accumulation" was a lack of information about each daily patient's anatomy, as reconstructions were performed on weekly 4DCTs. As a future improvement, dose reconstruction methodology could be applied

on synthetic CTs, as created on the basis of daily CBCTs and validated by use of the proposed range probing quality control procedure. This would increase further reliability and accuracy of daily reconstructed dose distributions and facilitate the introduction of dose reconstruction into patient treatment workflow as one of the most important and powerful quality control cornerstones. Dose reconstruction and accumulation would guide the decision-making process for plan adaptations and facilitate reporting of the outcomes.

The importance of adaptive proton therapy has been demonstrated multiple times. However, major technological hurdles remain, especially for the introduction of online adaptive therapy. Some of these hurdles are linked to the execution of radiotherapy workflow; for instance, the processing of daily patient imaging data by automated, reliable segmentation of the structures; automated treatment planning; sufficiently fast plan optimization and dose calculation. Other hurdles are linked to data handling and software engineering itself; for example, the way that imaging data are stored and transferred or shared between various involved subsystems. Further challenges are linked to the execution of currently applied Patient Specific Quality Assurance (PSQA) procedures, which are not necessarily compatible with online adaptive environment. In Chapter VI "Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy" an alternative approach for performing PSQA has been discussed. The approach is based on the use of treatment delivery log files and independent dose recalculation. The introduction of a PSQA workflow significantly reduced QA time in clinical practice, which is especially linked to the in-room time required previously to perform in-beam QA measurements. In contrast to the previously required 30 to 45 minutes per treatment plan, a QA procedure relying on the use of treatment delivery log files takes not more than 5 to 10 minutes per treatment plan. Such a reduction in time has a significant impact on the throughput of the facility and reduces the QA bottleneck, which sometimes delays the start of the treatment on an adapted treatment plan due to offline adaptive workflows. Although the currently implemented workflow requires a dry-run of the treatment plan to acquire treatment delivery log files, in future this might be overcome by additionally incorporating interfaces to generate machine steering

files and performing analysis of those prior to the delivery of the actual treatment fraction. In this way, the check on data transfer integrity between involved subsystems (treatment planning system, oncology information system and proton therapy system) could be conducted without a dry run. With the addition of such a highly comprehensive PSQA procedure, which would be entirely compatible with any online adaptive workflow that could be deployed, the removal of the patient from the treatment room in order to conduct any PSQA activities would not be required.

Eventually, PSQA procedure could become integrated with dose reconstruction processes, as discussed in Chapters III and V, providing the basis for a longitudinal quality assurance process of the treatment course on fraction-wise basis. By evaluating the treatment progress on the basis of accumulated doses as proposed in Chapter VII "Feasibility of patient specific quality assurance for proton therapy based on independent dose calculation and predicted outcomes" and interpreting results in terms of predicted outcomes, the quality of the treatment could be continuously monitored and required corrective actions taken as necessary.

In conclusion, the introduction of more adaptive treatment procedures and the availability of online adaptive workflows in proton therapy might be the next major advancement for the proton therapy field. Implementation of adaptive workflows imposes technological and logistic challenges. The performed studies support the need for development and introduction of adaptive therapy workflows and propose solutions to several of the challenges, primarily focusing on the quality control aspects.

References

[1]   Li B, Lee HC, Duan X, Shen C, Zhou L, Jia X, Yang M., Comprehensive analysis of proton range uncertainties related to stopping-power-ratio estimation using dual-energy CT imaging., Phys Med Biol. 2017 Aug 9;62(17):7056-7074. doi: 10.1088/1361-6560/aa7dc9

[2]   Wagenaar D, Kierkels RGJ, van der Schaaf A, Meijers A, Scandurra D, Sijtsema NM, Korevaar EW, Steenbakkers RJHM, Knopf AC, Langendijk JA, Both S. Head and Neck IMPT probabilistic dose accumulation: feasibility of a 2 mm setup uncertainty setting. Radiother Oncol. 2020

Sep 5:S0167-8140(20)30770-2. doi: 10.1016/j.radonc.2020.09.001. Epub ahead of print. PMID: 32898561

[3]   Ribeiro CO, Meijers A, Korevaar EW, Muijs CT, Both S, Langendijk JA, Knopf A. Comprehensive 4D robustness evaluation for pencil beam scanned proton plans. Radiother Oncol. 2019 Jul;136:185-189. doi: 10.1016/j.radonc.2019.03.037. Epub 2019 Apr 20. PMID: 31015123

[4]   Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol. 2020 Apr 28;65(9):095002. doi: 10.1088/1361-6560/ab7d54. PMID: 32143207

[5]   Thummerer A, de Jong BA, Zaffino P, Meijers A, Marmitt GG, Seco J, Steen-bakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of the suitability of CBCT- and MR-based synthetic CTs for daily adaptive proton therapy in head and neck patients. Phys Med Biol. 2020 Aug 24. doi: 10.1088/1361-6560/abb1d6. Epub ahead of print. PMID: 33179874

[6]   Richter C, Pausch G, Barczyk S, Priegnitz M, Keitz I, Thiele J, Smeets J, Stappen FV, Bombelli L, Fiorini C, Hotoiu L, Perali I, Prieels D, Enghardt W, Baumann M. First clinical application of a prompt gamma based in vivo proton range verification system. Radiother Oncol. 2016 Feb;118(2):232-7. doi: 10.1016/j.radonc.2016.01.004. Epub 2016 Jan 13. PMID: 26774764.

[7]   den Otter LA, Kaza E, Kierkels RGJ, Meijers A, Ubbels FJF, Leach MO, Collins DJ, Langendijk JA, Knopf AC. Reproducibility of the lung anatomy under active breathing coordinator control: Dosimetric consequences for scanned proton treatments. Med Phys. 2018 Dec;45(12):5525-5534. doi: 10.1002/mp.13195. Epub 2018 Oct 19. PMID: 30229930; PMCID: PMC6334635

[8]   den Otter LA, Anakotta RM, Weessies M, Roos CTG, Sijtsema NM, Muijs CT, Dieters M, Wijsman R, Troost EGC, Richter C, Meijers A, Langendijk JA, Both S, Knopf AC. Investigation of inter-fraction target motion variations in the context of pencil beam scanned proton therapy in non-small cell lung cancer patients. Med Phys. 2020 Jun 23. doi: 10.1002/mp.14345. Epub ahead of print. PMID: 32573792

Appendices

Summary

The thesis focuses on development and clinical implementation of comprehensive and overlaying quality control process aimed at supporting introduction of high precision adaptive IMPT workflows. The thesis consists of seven chapters, covering topics on quality control for proton range accuracy, reconstruction, and accumulation of delivered dose distributions longitudinally throughout the proton therapy course and independent dose recalculation/predictive outcome-based patient specific quality assurance procedures.

A proton range probing method as a quality control tool for range accuracy validation has been proposed and applied for range accuracy assessments in animal tissue samples covering a broad range of tissue types. A 3% range uncertainty margin has been shown as adequate for single energy CT based workflows, when Monte Carlo dose calculation is employed. The procedure has been incorporated in clinical practice as an in vivo quality control tool. Results for head and neck cancer patients confirmed 3% uncertainty margin, however also revealed that range errors introduced by anatomical inconsistencies may significantly surpass (10% or above) the ones which has physics-related origin. This supports the transition towards adaptive therapy workflows.

A fraction-wise 4D dose reconstruction and accumulation procedure utilizing treatment delivery log files and patient-specific daily breathing patterns has been proposed and implemented in clinical practice. Validation of the procedure in controlled conditions with a 4D phantom revealed ability to spatially reconstruct the dose distributions with submillimeter accuracy. The dose reconstructions for the first 10 patients undergoing treatments for thoracic indications showed that organ motion does cause

loss of target dose homogeneity, however local cold and hot spots do not have systematic behavior, therefore with fractionation dose homogeneity is recovered for the investigated range of motion (point maximum < 20 mm). In addition, it was observed that about half of the patients required replanning at least once throughout the treatment course. However, in all cases replanning was required due to the anatomical changes, once again outlining the need for clinical adoption of adaptive proton therapy.

Eventually, an alternative approach for in-beam measurement-based patient specific quality assurance (PSQA) procedure has been investigated, developed, and introduced in clinical practice. The approach is based on the use of treatment delivery log files and independent dose recalculation. The method is intrinsically compatible with online adaptive therapy workflows. In addition, the introduction of approach allowed to significantly reduce the PSQA bottleneck by shortening in-room times from 30–45 minutes to 5–10 instead. This resulted in increased treatment room throughput and shorted time required to switch the treatment to the new plan in case of replanning.

By incorporating the developed range probing QC procedure as a validation tool for synthetic CTs and utilizing developed dose reconstruction and accumulation workflow, it enables possibility to establish a comprehensive longitudinal patient specific quality control process to monitor the treatment delivery in an environment of adaptive proton therapy. By evaluating the treatment progress based on accumulated doses and interpreting results in terms of predicted outcomes, the quality of the treatment could be continuously assessed and required corrective actions taken, as necessary. In this context, predicted outcomes allow to interpret QC data in more clinically relevant manner.

In conclusion, introduction of more adaptive treatment procedures and availability of online adaptive workflows in proton therapy might be the next major advancement needed to take full advantage of the physical characteristics of the proton beam. Implementation of adaptive workflows imposes technological and logistic challenges. The performed studies support the need for development and introduction of adaptive therapy workflows and propose solutions to several of the challenges, primarily focusing on the quality control aspects.

Samenvatting

Dit proefschrift richt zich op de ontwikkeling en de klinische implementatie van een alomvattend kwaliteitscontroleproces dat bedoeld is om behandelprocedures te ondersteunen, die gericht zijn op adaptieve protonentherapie. Dit proefschrift bestaat uit zeven hoofdstukken waarin verschillende onderwerpen aan de orde komen, zoals de kwaliteitscontrole op onnauwkeurigheden in het bereik van protonen (*range errors*) en de reconstructie en accumulatie van de afgegeven dosisverdeling gedurende de protonenbehandeling. Daarnaast wordt ingegaan op procedures die de kwaliteit van patiëntspecifieke onafhankelijke herberekeningen van de bestralingsdosis borgen.

Een kwaliteitscontrolemethode, waarbij gebruik werd gemaakt van een protonen *range probing*-techniek, is toegepast om het protonenbereik in dierlijke specimen te valideren. Het dierlijk specimen bestond uit een breed scala aan weefseltypen. Een *range error*-marge van 3% bleek adequaat voor behandelprocedures die zijn gebaseerd op een CT-scan, verworven met een enkele scanenergie, waarbij werd uitgegaan van op Monte Carlo gebaseerde dosisberekeningen. Deze procedure werd opgenomen in de kliniek en diende als hulpmiddel voor *in vivo* kwaliteitscontrole. De resultaten van patiënten met hoofd-halskanker bevestigden een adequate toepassing van de 3% onzekerheidsmarge, maar deze lieten tegelijkertijd ook zien dat de *range errors* door anatomische inconsistenties significant groter kunnen zijn (≥ 10%) dan de fouten die een fysische oorzaak hebben. Deze bevinding onderschrijft het belang van en de transitie naar adaptieve behandelprocedures.

Een fractionele 4D dosisreconstructie- en accumulatieprocedure waarbij gebruik werd gemaakt van machine-logbestanden van de behandeling

en patiëntspecifieke ademhalingssignalen is voorgesteld en geïmplementeerd in de kliniek. Validatie van de procedure onder gecontroleerde omstandigheden en met behulp van een 4D-fantoom, toonde aan dat het mogelijk was om de dosisverdeling te reconstrueren met sub-millimeter nauwkeurigheid. Dosisreconstructies van de eerste 10 klinische patiënten die werden behandeld voor een thoracale tumor, lieten zien dat orgaanbeweging een verlies van dosishomogeniteit veroorzaakte. Lokale over- en onderdoseringen zijn daarentegen niet systematisch; vanwege de behandelfractionering zal de dosishomogeniteit herstellen. Daarnaast viel op te merken dat ongeveer de helft van alle patiënten een planaanpassing nodig had gedurende de behandeling. In alle gevallen was deze planaanpassing nodig vanwege anatomische veranderingen. Dit onderschrijft nogmaals de noodzaak voor het introduceren van adaptieve protonenbehandelingen.

Tot slot is een, op *in-beam* metingen gebaseerde, alternatieve aanpak voor patiëntspecifieke kwaliteitscontrole (PSQA) onderzocht, ontwikkeld en geïntroduceerd in de kliniek. Deze benadering is gebaseerd op machine-logbestanden van de bestraling en een onafhankelijke herberekening van de dosis. Deze methode is intrinsiek compatibel met online adaptieve behandelprocedures. Bovendien leidde deze implementatie tot een significante vermindering van het PSQA-knelpunt, waarbij de tijd die benodigd is in de behandelruimte, werd verkort van 30–45 minuten naar 5–10 minuten. Daardoor werd de doorstroming per behandelruimte vergroot en werd de benodigde tijd verkort om over te schakelen naar een nieuw plan in het geval van een planaanpassing.

Door de ontwikkelde *range probing*-kwaliteitscontroleprocedure te gebruiken als validatie-instrument voor synthetische CT's en door gebruik te maken van de ontwikkelde dosisreconstructie- en accumulatieprocedures, is het mogelijk om een uitgebreide, longitudinale patiëntspecifieke kwaliteitscontrole op te zetten. Dit kwaliteitscontroleproces kan de behandeling bewaken binnen een werkwijze die is gericht op adaptieve protonentherapie. Door de voortgang van de behandeling te evalueren op basis van de geaccumuleerde dosisverdeling ten opzichte van de beoogde behandeldoelen, kan de kwaliteit van behandeling worden gemonitord. Op basis van deze continue kwaliteitsbewaking kunnen corrigerende maatregelen worden genomen.

Concluderend kan worden gesteld dat de introductie van adaptieve behandelmethodes en de beschikbaarheid van online adaptieve procedures bij protonentherapie, de volgende grote vooruitgang zou kunnen zijn op het gebied van protonentherapie. De implementatie van adaptieve procedures brengt een aantal technologische en logistieke uitdagingen met zich mee. De uitgevoerde onderzoeken onderschrijven de noodzaak voor de ontwikkeling en de introductie van adaptieve behandelprocedures en stellen oplossingen voor een aantal van de uitdagingen voor, waarbij de nadruk primair ligt op de aspecten van kwaliteitscontrole.

Acknowledgements

I would like to thank my promoters Stefan Both and Johannes A Langendijk.

I would like to acknowledge my colleagues, who provided support during the projects, especially, Antje Knopf and her group — Cassia Oraboni Ribeiro, Carmen Seller Oria, Adrian Thummerer and Sabine Visser — and Jeffrey Free. Their support in executing and analyzing experiments is highly appreciated.

I would like to acknowledge Guillaume Janssen, Arnaud Pin, Kevin Souris and openREGGUI community overall for developing open source tools and automation platform that was instrumental in establishing automated patient specific quality assurance program based on use of independent dose recalculation.

Bibliography

Publications in peer reviewed journals

1. Farace P, Righetto R, **Meijers A**. Pencil beam proton radiography using a multilayer ionization chamber. Phys Med Biol. 2016 Jun 7;61(11):4078-87. doi: 10.1088/0031-9155/61/11/4078. Epub 2016 May 10. PMID: 27164479.

2. Farace P, Righetto R, Deffet S, **Meijers A**, Vander Stappen F. 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 Dec;43(12):6405. doi: 10.1118/1.4966703. PMID: 27908151.

3. Chang JY, Zhang X, Knopf A, Li H, Mori S, Dong L, Lu HM, Liu W, Badiyan SN, Both S, **Meijers A**, Lin L, Flampouri S, Li Z, Umegaki K, Simone CB 2nd, Zhu XR. Consensus Guidelines for Implementing Pencil-Beam Scanning Proton Therapy for Thoracic Malignancies on Behalf of the PTCOG Thoracic and Lymphoma Subcommittee. Int J Radiat Oncol Biol Phys. 2017 Sep 1;99(1):41-50. doi: 10.1016/j.ijrobp.2017.05.014. Epub 2017 May 19. PMID: 28816159.

4. den Otter LA, Kaza E, Kierkels RGJ, **Meijers A**, Ubbels FJF, Leach MO, Collins DJ, Langendijk JA, Knopf AC. Reproducibility of the lung anatomy under active breathing coordinator control: Dosimetric consequences for scanned proton treatments. Med Phys.

2018 Dec;45(12):5525-5534. doi: 10.1002/mp.13195. Epub 2018 Oct 19. PMID: 30229930; PMCID: PMC6334635.

5. **Meijers A**, Jakobi A, Stützer K, Guterres Marmitt G, Both S, Langendijk JA, Richter C, Knopf A. Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: Implementation and proof-of-concept. Med Phys. 2019 Mar;46(3):1140-1149. doi: 10.1002/mp.13371. Epub 2019 Feb 4. PMID: 30609061.

6. Ribeiro CO, **Meijers A**, Korevaar EW, Muijs CT, Both S, Langendijk JA, Knopf A. Comprehensive 4D robustness evaluation for pencil beam scanned proton plans. Radiother Oncol. 2019 Jul;136:185-189. doi: 10.1016/j.radonc.2019.03.037. Epub 2019 Apr 20. PMID: 31015123.

7. Wagenaar D, Tran LT, **Meijers A**, Marmitt GG, Souris K, Bolst D, James B, Biasi G, Povoli M, Kok A, Traneus E, van Goethem MJ, Langendijk JA, Rosenfeld AB, Both S. Validation of linear energy transfer computed in a Monte Carlo dose engine of a commercial treatment planning system. Phys Med Biol. 2020 Jan 17;65(2):025006. doi: 10.1088/1361-6560/ab5e97. PMID: 31801119.

8. **Meijers A**, Free J, Wagenaar D, Deffet S, Knopf AC, Langendijk JA, Both S. Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing. Phys Med Biol. 2020 Feb 4;65(3):03NT02. doi: 10.1088/1361-6560/ab66e1. PMID: 31896099.

9. Guterres Marmitt G, Pin A, Ng Wei Siang K, Janssens G, Souris K, Cohilis M, Langendijk JA, Both S, Knopf A, **Meijers A**. Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy. Phys Med. 2020 Feb;70:49-57. doi: 10.1016/j.ejmp.2019.12.018. Epub 2020 Jan 20. PMID: 31968277.

10. Thummerer A, Zaffino P, **Meijers A**, Marmitt GG, Seco J, Steenbakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol. 2020 Apr 28;65(9):095002. doi: 10.1088/1361-6560/ab7d54. PMID: 32143207.

11. **Meijers A**, Seller Oria C, Free J, Bondesson D, Rabe M, Parodi K, Landry G, Langendijk JA, Both S, Kurz C, Knopf AC. Assessment of

range uncertainty in lung-like tissue using a porcine lung phantom and proton radiography. Phys Med Biol. 2020 Jul 31;65(15):155014. doi: 10.1088/1361-6560/ab91db. PMID: 32392543.

12. Brodbek L, Kretschmer J, Willborn K, **Meijers A**, Both S, Langendijk JA, Knopf AC, Looe HK, Poppe B. Analysis of the applicability of two-dimensional detector arrays in terms of sampling rate and detector size to verify scanned intensity-modulated proton therapy plans. Med Phys. 2020 Jun 23. doi: 10.1002/mp.14346. Epub ahead of print. PMID: 32574383.

13. den Otter LA, Anakotta RM, Weessies M, Roos CTG, Sijtsema NM, Muijs CT, Dieters M, Wijsman R, Troost EGC, Richter C, **Meijers A**, Langendijk JA, Both S, Knopf AC. Investigation of inter-fraction target motion variations in the context of pencil beam scanned proton therapy in non-small cell lung cancer patients. Med Phys. 2020 Jun 23. doi: 10.1002/mp.14345. Epub ahead of print. PMID: 32573792.

14. **Meijers A**, Guterres Marmitt G, Ng Wei Siang K, van der Schaaf A, Knopf AC, Langendijk JA, Both S. Feasibility of patient specific quality assurance for proton therapy based on independent dose calculation and predicted outcomes. Radiother Oncol. 2020 Jun 21;150:136-141. doi: 10.1016/j.radonc.2020.06.027. Epub ahead of print. PMID: 32579999.

15. **Meijers A**, Knopf AC, Crijns APG, Ubbels JF, Niezink AGH, Langendijk JA, Wijsman R, Both S. Evaluation of interplay and organ motion effects by means of 4D dose reconstruction and accumulation. Radiother Oncol. 2020 Aug 5:S0167-8140(20)30714-3. doi: 10.1016/j.radonc.2020.07.055. Epub ahead of print. PMID: 32768509.

16. Zeng J, Badiyan SN, Garces YI, Wong T, Zhang X, Simone CB 2nd, Chang JY, Knopf AC, Mori S, Iwata H, **Meijers A**, Li H, Bues M, Liu W, Schild SE, Rengan R; International Particle Therapy Cooperative Group Thoracic Subcommittee. Consensus Statement on Proton Therapy in Mesothelioma. Pract Radiat Oncol. 2020 May 24:S1879-8500(20)30117-X. doi: 10.1016/j.prro.2020.05.004. Epub ahead of print. PMID: 32461036.

17. Wagenaar D, Kierkels RGJ, van der Schaaf A, **Meijers A**, Scandurra D, Sijtsema NM, Korevaar EW, Steenbakkers RJHM, Knopf AC, Langendijk JA, Both S. Head and Neck IMPT probabilistic dose

accumulation: feasibility of a 2 mm setup uncertainty setting. Radiother Oncol. 2020 Sep 5:S0167-8140(20)30770-2. doi: 10.1016/j.radonc.2020.09.001. Epub ahead of print. PMID: 32898561.

18. den Otter LA, Chen K, Janssens G, **Meijers A**, Both S, Langendijk JA, Rosen LR, Wu HT, Knopf AC. Technical note: 4D cone-beam CT reconstruction from sparse-view CBCT data for daily motion assessment in pencil beam scanned proton therapy (PBS-PT). Med Phys. 2020 Oct 4. doi: 10.1002/mp.14521. Epub ahead of print. PMID: 33011990.

19. Seller Oria C, Marmitt GG, Both S, Langendijk JA, Knopf AC, **Meijers A**. Classification of various sources of error in range assessment using proton radiography and neural networks in head and neck cancer patients. Phys Med Biol. 2020 Oct 13. doi: 10.1088/1361-6560/abc09c. Epub ahead of print. PMID: 33049722.

20. Ribeiro CO, Terpstra J, Janssens G, Langendijk JA, Both S, Muijs CT, Wijsman R, Knopf AC, **Meijers A**. Evaluation of continuous beam rescanning versus pulsed beam in pencil beam scanned proton therapy for lung tumours. Phys Med Biol. 2020 Oct 29. doi: 10.1088/1361-6560/abc5c8. Epub ahead of print. PMID: 33120367.

21. Thummerer A, de Jong BA, Zaffino P, **Meijers A**, Marmitt GG, Seco J, Steenbakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of the suitability of CBCT- and MR-based synthetic CTs for daily adaptive proton therapy in head and neck patients. Phys Med Biol. 2020 Aug 24. doi: 10.1088/1361-6560/abb1d6. Epub ahead of print. PMID: 33179874.

22. Bondesson D, **Meijers A**, Janssens G, Rit S, Rabe M, Kamp F, Niepel K, den Otter LA, Both S, Brousmiche S, Dinkel J, Belka C, Parodi K, Knopf AC, Kurz C, Landry G. Anthropomorphic lung phantom based validation of in-room proton therapy 4D-CBCT image correction for dose calculation. Zeitschrift für Medizinische Physik. 2020. ISSN 0939-3889. https://doi.org/10.1016/j.zemedi.2020.09.004.

23. **Meijers A**, Seller Oria C, Free J, Langendijk JA, Knopf AC, Both S. Technical Note: First report on an in vivo range probing quality control procedure for scanned proton beam therapy in head and neck cancer patients. Med Phys. 2021 Jan 11. doi: 10.1002/mp.14713. Epub ahead of print. PMID: 33428795.

Abstracts in peer reviewed conferences

1. Lorentini S, **Meijers A**, Schwarz M. Dosimetric equivalence of proton therapy gantry rooms featuring pencil beam scanning mode. PTCOG. Poster.

2. **Meijers A**, Richter C, Dessy F, Widder J, Korevaar EW, Jakobi A, Ribeiro C, Langendijk JA, Knopf AC. Thoracic tumor treatment course assessment based on 4D dose accumulation for scanned proton therapy. ESTRO 36 Vienna. Oral communication.

3. Ribeiro C, **Meijers A**, Janssens G, Widder J, Langendijk JA, Korevaar EW, Knopf AC. Comprehensive prospective evaluation tool for treatments of thoracic tumours with scanned protons. ESTRO 36 Vienna. Poster.

4. Korevaar EW, Kierkels RGJ, **Meijers A**, van Dijk LV, Brouwer CL, van 't Veld AA, Langendijk JA, van der Schaaf A, Knopf AC. A single dose distribution per plan to represent possible target underdosage in a multi-scenario robustness evaluation. PTCOG 2017 Chiba & KCC Kanagawa. Poster.

5. den Otter LA, Kierkels RGJ, Kaza E, **Meijers A**, Leach MO, Collins DJ, Langendijk JA, Knopf AC. Reproducibility of the lung anatomy using Active Breathing Control: Dosimetric consequences for scanned proton treatments. PTCOG 2017 Chiba & KCC Kanagawa. Poster.

6. Dumont D, Ribeiro C, Janssens G, Geets X, Knopf AC, Sterpin E, **Meijers A**. Evaluation of repainting for moving targets treated with continuous or pulsed scanned proton beams. ESTRO 37 Barcelona. Poster.

7. Ribeiro C, **Meijers A**, Korevaar EW, Langendijk JA, Knopf AC. Comprehensive Evaluation Tool for 4D Pencil Beam Scanned Proton Plans Delivery. ASTRO San Antonio, TX. Moderated poster session.

8. Marmitt GG, van Goethem M, **Meijers A**, Branderburg S, Knopf AC. Experimental Validation of magnetic field deflections of proton beams for online MR-guided PT. 5th MR in RT Symposium. Utrecht, Netherlands. Poster.

9. Wagenaar D, Kierkels RGJ, van der Schaaf A, Sijtsema NM, Korevaar EW, Rodrigues Reis M, **Meijers A**, Knopf AC, van den Hoek A, Langendijk JA, Both S. Margin evaluation using dose accumulation

in head and neck intensity modulated proton therapy. ESTRO 38 Milano. Oral communication.

10. **Meijers A**, Free J, Wagenaar D, Deffet S, Knopf AC, Langendijk JA, Both S. A novel range probing-based optimization of CT calibration curve for Proton Therapy. ESTRO 38 Milano. Oral communication.

11. den Otter LA, Chen GK, Janssens G, **Meijers A**, Both S, Langendijk JA, Wu TH, Knopf AC. First steps towards 4DCBCT reconstruction for moving targets at a scanned proton gantry system. ESTRO 38 Milano. Poster.

12. Marmitt GG, Pin A, Langendijk JA, Both S, Knopf AC, **Meijers A**. Automated proton plan QA via independent Monte Carlo simulations. ESTRO 38 Milano. Poster.

13. **Meijers A**, Free J, Wagenaar D, Deffet S, Knopf AC, Langendijk JA, Both S. Proton radiography: optimization and validation of the CT calibration curve for proton therapy. RKF scientific project day. Oral communication.

14. Wagenaar D, Kierkels RGJ, van der Schaaf A, **Meijers A**, Scandurra D, Sijtsema NM, Korevaar EW, Knopf AC, van den Hoek A, Langendijk JA, Both S. Head/neck margin reduction based on biological parameters. RKF scientific project day. Oral communication.

15. Wagenaar D, Kierkels RGJ, van der Schaaf A, **Meijers A**, Scandurra D, Sijtsema NM, Korevaar EW, Knopf AC, van den Hoek A, Langendijk JA, Both S. Validation of a Monte Carlo LET calculation in a commercial TPS. NVKF congress. Oral communication.

16. den Otter LA, Kurz C, Stanislawski M, Bondesson D, Landry G, Janssens G, **Meijers A**, Both S, Langendijk JA, Knopf AC. Optimization of 4DCBCT image reconstruction at proton gantry systems. PTCOG 58 Manchester. Poster.

17. Wagenaar D, Kierkels RGJ, van der Schaaf A, **Meijers A**, Scandurra D, Sijtsema NM, Korevaar EW, van den Hoek A, Langendijk JA, Both S. Robustness evaluation using dose accumulation in head and neck IMPT. PTCOG. Poster.

18. Wagenaar D, Tran L, **Meijers A**, Bolst D, James B, Traneus E, van Goethem M, Langendijk JA, Rozenfeld A, Both S. First report on validation of a Monte Carlo LET dose engine in a commercial treatment planning system. PTCOG. Oral communication.

19. Ribeiro C, Dumont D, Terpstra J, Janssens G, Langendijk JA, Both S, Sterpin E, Knopf AC, **Meijers A**. A comparison of various pencil beam scanned proton therapy delivery strategies for the treatment of moving targets. PTCOG 58 Manchester. Moderated poster session.

20. Brodbek L, Kretschmer J, Dulkys A, Le TLT, Espelage T, Guterres Marmitt G, **Meijers A**, Looe HK, Knopf AC, Poppe B. IMPT verification with Gafchromic EBT3 films and the OCTAVIUS detector array 1500XDR. DEGRO. Poster.

21. Guterres Marmitt G, Pin A, Janssens G, Langendijk JA, Both S, Knopf AC, **Meijers A**. A Novel Automated Clinical Proton Therapy Patient Quality Assurance via Monte Carlo Simulations and Treatment Log Files. ICCR. Oral communication.

22. Knopf AC, Muijs CT, Wijsman R, **Meijers A**, Ribeiro C, Visser S, den Otter LA, van der Laan HP, Anakotta M, Roos Q, Korevaar EW, Sijtsema NM, Both S, Langendijk JA. THE REACT-STUDY — a comprehensive data set of Repeated 4D CTs and CBCTs to prepare for pencil beam scanned proton therapy (PBS-PT) in lung and esophagus patients. ASTRO. Moderated poster session.

23. **Meijers A**, Both S, Langendijk JA, Knopf AC. Experimental validation of a log file based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy. ASTRO. Moderated poster session.

24. Ribeiro C, Visser S, **Meijers A**, Korevaar EW, Muijs CT, Wijsman R, Both S, Langendijk JA, Knopf AC. Towards the Clinical Implementation of Pencil Beam Scanned Proton Therapy (PBS-PT)for Thoracic Indications: A Comprehensive 4D Robustness Evaluation based on Patient and Machine Specific Information. ASTRO. Moderated poster session.

25. **Meijers A**, Marmitt G, Pin A, Janssens G, Langendijk JA, Knopf AC, Both S. Implementation of automated patient QA for intensity modulated proton therapy. ASTRO. Moderated poster session.

26. Wagenaar D, Kierkels RGJ, van der Schaaf A, **Meijers A**, Scandurra D, Sijtsema NM, Korevaar EW, van den Hoek A, Knopf AC, Langendijk JA, Both S. Robustness evaluation using dose accumulation in head and neck IMPT. ASTRO. Poster.

27. Brodbek L, Kretschmer J, Dulkys A, Le TLT, Espelage T, Guterres Marmitt G, **Meijers A**, Wilborn K, Looe HK, Langendijk JA, Knopf AC, Poppe B. A Frequency analysis of scanned intensity modulated proton therapy treatment plans. AAPM. Poster.

28. **Meijers A**, Free J, Wagenaar D, Deffet S, Knopf AC, Langendijk JA, Both S. Range uncertainty margin evaluation and reduction through range probing based optimization of CT calibration curve in Proton Pencil Beam Scanning. AAPM. Moderated poster session.

29. Wagenaar D, Linh Tran T, **Meijers A**, Bolst D, James B, Povoli M, Kok A, Traneus E, van Goethem M, Langendijk JA, Rosenfeld A, Both S. First Report on Methodology Development and Validation of LET Computed in a Monte Carlo Dose Engine of a Commercial Treatment Planning System. AAPM. Oral communication.

30. Spautz S, **Meijers A**, Jakobi A, Thiele J, Knopf AC, Troost EGC, Richter C, Stützer K. Protonentherapie von Lungentumoren: Log-file-basierte Dosisrekonstruktion zur klinischen Implementierung des Pencil-Beam-Scanning-Verfahrens. DGMP. Poster.

31. Brodbek L, Kretschmer J, Dulkys A, Le TLT, Espelage T, Marmitt GG, **Meijers A**, Willborn K, Stelljes TS, Looe HK, Langendijk JA, Knopf AC, Poppe B. IMPT verification with Gafchromic EBT3 films and the OCTAVIUS detector array 1500XDR. Strahlentherapie und onkologie. Poster.

32. Brodbek L, Kretschmer J, Dulkys A, Le TLT, Espelage T, Marmitt GG, Stelljes TS, Looe HK, Langendijk JA, **Meijers A**, Knopf AC, Poppe B. A frequency analysis of scanned intensity modulated proton therapy treatment plans. Annual meeting of the German association of medical physicists. Poster.

33. Spautz S, Peters N, **Meijers A**, Jakobi A, Knopf AC, Troost EGC, Richter C, Stützer K. Accuracy and robustness of 4D logfile-based dose reconstruction. 11th 4D workshop. Poster.

34. Bondesson D, Janssens G, Rabe M, Stanislawsky M, Niepel K, **Meijers A**, Both S, Rit S, Dinkel J, Belka C, Kamp F, Parodi K, Landry G, Knopf AC, Kurz C. Validation of 4D accumulated dose at a proton therapy CBCT scanner using MA-ROOSTER and a porcine lung phantom. 11th 4D workshop. Poster.

35. Visser S, Neh H, Ribeiro C, Korevaar EW, **Meijers A**, Poppe B, Both S, Langendijk JA, Muijs CT, Knopf AC. Can we use the same planning strategies for diaphragm motion in IMPT as we use in VMAT? 11th 4D workshop. Poster.

36. Ribeiro C, Visser S, Korevaar EW, Sijtsema NM, Both S, Langendijk JA, Wijsman R, Muijs CT, **Meijers A**, Knopf AC. Patient and machine specific evaluation of 3D versus 4D robust optimization in intensity-modulated proton therapy (IMPT) for thoracic indications. PTCOG 2020. Oral communication.

37. Thummerer A, Zaffino P, **Meijers A**, Marmitt GG, Seco J, Steenbakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of CBCT- and MR-based synthetic CTs for adaptive proton therapy in head and neck cancer patients. PTCOG 2020. Oral communication.

38. Kuijpers M, Batin E, van Faassen-van Loenen T, **Meijers A**, Crijns APG, Langendijk JA. Introduction of surface imaging as part of control CT procedure for breast cancer patients treated with proton therapy. ESTRO 2020. Poster.

39. van Faassen-van Loenen TL, Batin E, Kuijpers M, **Meijers A**, Crijns APG, Langendijk JA. Do two-dimensional kilovoltage images add value in the online position verification of breast cancer patients treated with proton therapy. ESTRO 2020. Poster.

40. **Meijers A**, Seller Oria C, Free J, Bondesson D, Rabe M, Dinkel J, Parodi K, Langendijk JA, Both S, Kurz C, Knopf AC. Range uncertainty assessment in lung-like tissue using a porcine lung and proton radiography. AAPM. Oral communication.

41. Seller Oria C, Guterres Marmitt G, Both S, Langendijk JA, Knopf AC, **Meijers A**. Deconvolution of different range error sources using proton radiography and neural networks. ESTRO 2020. Poster.

42. Brodbek L, Kretschmer J, **Meijers A**, Looe HK, Knopf AC, Poppe B. Analysis on the error detection capability of different detector arrays in intensity modulated proton therapy. DGMP 2020. Poster.

Grants

Curriculum Vitae

Arturs Meijers
arturs.meijers@icloud.com

Education

Sep 2018-	**PhD student in Medical Physics** **University of Groningen,** Netherlands
Sep 2010-Jul 2011	**Professional Master degree in Medical Physics** **Riga Technical University**, Latvia
Sep 2005-Jul 2010	**Bachelor degree in Medical Physics and Engineering qualification** **Riga Technical University,** Latvia

Experience

Apr 2020-	**Senior Product Marketing Manager, Varian Medical Systems,** USA
Jan 2019-Apr 2020	**Proton Physics Team Leader, University Medical Centre Groningen,** Netherlands
Apr 2016-Dec 2018	**Senior Medical Physicist, University Medical Centre Groningen,** Netherlands
Jan 2014-Mar 2016	**Product Development Manager, IBA,** Belgium
Aug 2012-Jan 2014	**Medical Physicist, IBA,** Belgium
Sep 2010-Jul 2012	**External Consultant,** on behalf of **Varian Medical Systems**, USA
Mar 2010-Jul 2012	**Medical Physicist, Riga East University Hospital,** Latvia
Jul 2008-Mar 2010	**Physicist, Riga East University Hospital,** Latvia

Qualifications

2003	International English Language Testing System (IELTS) score: 7.5
2006	A, B category driving license
2009	Radiation safety and quality assurance, Riga Technical University
2010	Eclipse administration and physics, Varian Medical Systems
	Eclipse in clinical practice, Varian Medical Systems
	Eclipse operations, Varian Medical Systems
	OBI physics, Varian Medical Systems
Language skills	Latvian: mother tongue
	English: fluent
	Russian: fluent
Awards	**Award of Riga City Golden Fund**
	For academic achievements
	Membership of Riga Technical University Alumni Golden Fund
	For academic achievements

Scientific work

Nov 2011	Scientific Assistant, Riga Stradins University, Latvia
Oct 2010-Nov 2010	Scientific Assistant, Riga Stradins University, Latvia Research work within National research programme "Research of biomaterials for treatment of cardiovascular diseases"
Jul 2008-Nov 2008	Senior Laboratory Assistant, Riga Technical University, Latvia Research work within project "Markerless tracking of patient's breathing motion for breathing adapted radiotherapy"