University of Groningen

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

Meijers, Arturs; Marmitt, Gabriel Guterres; Ng Wei Siang, Kelvin; van der Schaaf, Arjen; Knopf, Antje C; Langendijk, Johannes A; Both, Stefan

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Original Article

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

Arturs Meijers [a,*], Gabriel Guterres Marmitt [a], Kelvin Ng Wei Siang [a], Arjen van der Schaaf [a], Antje C. Knopf [a,b], Johannes A. Langendijk [a], Stefan Both [a]

[a] University of Groningen, University Medical Centre Groningen, Department of Radiation Oncology, Groningen, The Netherlands; [b] Division for Medical Radiation Physics, Carl von Ossietzky Universität Oldenburg, Germany

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$_{IDC}$ workflows. 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 ΔTCP/NTCP values between the nominal and the log-plan dose were assessed.
Results: Results of PSQA$_M$ and PSQA$_{IDC}$ from both implemented workflows 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.
© 2020 The Author(s). Published by Elsevier B.V. Radiotherapy and Oncology 150 (2020) 136–141 This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

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.

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 indepen-

dent 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 min vs 30–35 min). 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 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.

Error scenarios 1–6 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

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

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 correlates 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 Fig. 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 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 Fig. 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.

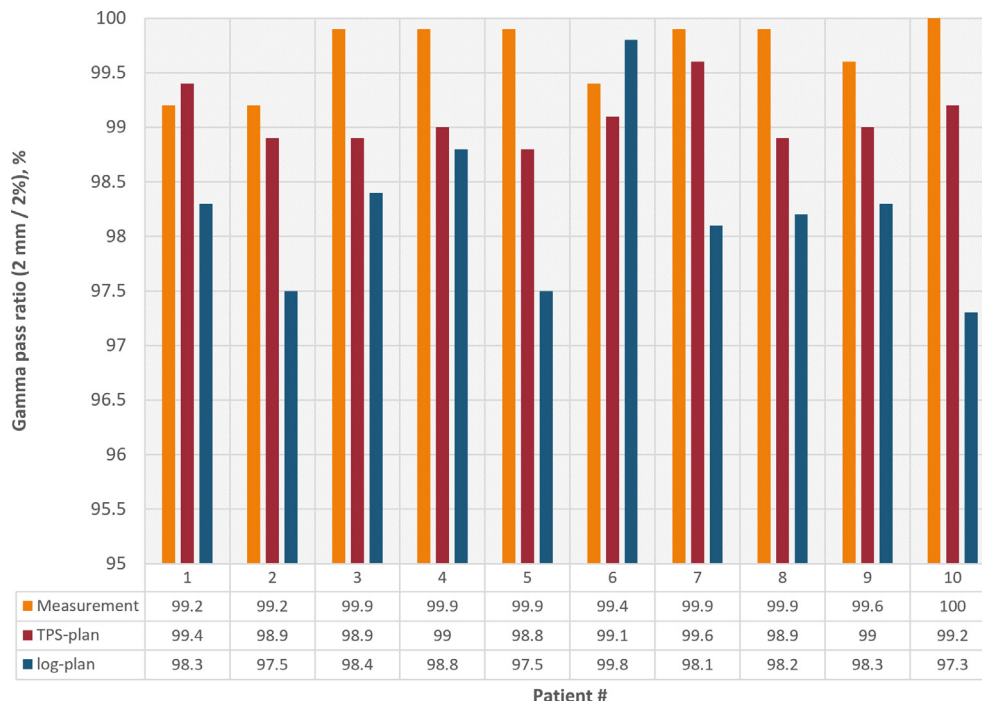| Patient #	1	2	3	4	5	6	7	8	9	10
Measurement	99.2	99.2	99.9	99.9	99.9	99.4	99.9	99.9	99.6	100
TPS-plan	99.4	98.9	98.9	99	98.8	99.1	99.6	98.9	99	99.2
log-plan	98.3	97.5	98.4	98.8	97.5	99.8	98.1	98.2	98.3	97.3

Fig. 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.

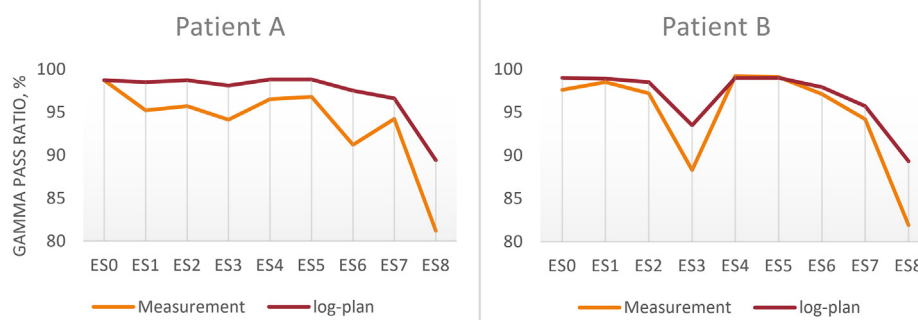| Pat.	TPS-plan dose				log-plan dose			
	ΔNTCP, %			ΔTCP, %	ΔNTCP, %			ΔTCP, %
	Dysph.	Xerost.	Tube feeding		Dysph.	Xerost.	Tube feeding	
1	0.1	0.0	0.0	2.2	−0.1	0.1	0.0	3.1
2	−0.2	0.0	−0.3	1.1	−0.3	0.0	−0.3	2.1
3	0.2	−0,4	0,0	2.0	0,2	−0.4	0,0	2.1
4	−0.1	0.1	−0.1	1.6	−0.1	0.0	−0.1	1.9
5	0.5	−0.3	−0.2	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

Fig. 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.

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, %			ΔTCP, %	ΔNTCP, %			ΔTCP, %
	Dysph.	Xerost.	Tube feeding		Dysph.	Xerost.	Tube feeding	
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

Consistency can be observed between gamma pass ratios (2 mm/2%) for PSQA$_M$ and PSQA$_{IDC}$ as shown by trends in Figs. 1 and 2. Consistent decisions regarding plan quality would be made according to either PSQAM or PSQAIDC (Fig. 1) and lower gamma pass ratios would be observed with either method in case of delivery errors (Fig. 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. Further-

more, 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 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 Fig. 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 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

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

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.

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.

Funding

Disclosures

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

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

Conflict of interest statements

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

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2020.06.027.

References

[1] Sharpe MB. IAEA Technical Reports Series No. 430: Commissioning and quality assurance of computerized planning systems for radiation treatment of cancer. Med Phys 2006;33:561.

[2] Agazaryan N, Solberg TD, DeMarco JJ. Patient specific quality assurance for the delivery of intensity modulated radiotherapy. Med Phys 2003;4:40–50.

[3] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys 2018;45:e53–83.

[4] Kry SF, Molineu A, Kerns JR, Faught AM, Huang JY, Pulliam KB, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. Int J Radiat Oncol Biol Phys 2014;90:1195–201.

[5] Belosi MF, van der Meer R, Garcia de Acilu Laa P, Laa P, Bolsi A, Weber DC, et al. Treatment log files as a tool to identify treatment plan sensitivity to inaccuracies in scanned proton beam delivery. Radiother Oncol 2017;125:514–9.

[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;63:205014.

[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, et al. Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy. Phys Med 2020;70:49–57.

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

[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;16:381–8.

[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;107:267–73.

[12] Widder J, van der Schaaf A, Lambin P, Marijnen CAM, Pignol JP, Rasch CR, et al. The quest for evidence for proton therapy: model-based approach and precision medicine. Int J Radiat Oncol Biol Phys, 2016;95:30–36.

[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;43:1700.

[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;45:846–62.

[15] Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, et al. 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;105:86–93.

[16] Beetz I, Schilstra C, van der Schaaf A, van den Heuvel ER, Doornaert P, van Luijk P, et al. 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;105:101–6.

[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;75:385–92.

[18] Eisbruch A, Kim HM, Feng FY, Lyden TH, Haxer MJ, Feng M, et al. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. Int J Radiat Oncol Biol Phys 2011;81:e93–9.

[19] Feng FY, Kim HM, Lyden TH, Haxer MJ, Feng M, Worden FP, et al. 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;68:1289–98.

[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;126:56–67.

[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;27:285–99.

[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 American Society for Radiation Oncology 61st Annual Meeting. September 1, 2019 Vol. 105, Supplement, 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, et al. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol 2020.