# Evaluation of Calculation-Based Patient Specific QA for Online Adaptive Radiotherapy

Patrik Sibolt[i], Jeremy Booth[ii], Trent Aland[iii], Ghirmay Kidane[iv], Richard Popple[v], and Bin Cai[vi]

i. *Dept of Oncology, Copenhagen University Hospital – Herlev and Gentofte, Copenhagen, Denmark*
ii. *Northern Sydney Cancer Centre, Royal North Shore Hospital, Sydney, Australia*
iii. *Icon Group, South Brisbane, Australia*
iv. *Barking Havering and Redbridge University Hospitals NHS Trust, Romford, UK*
v. *University of Alabama, Birmingham, USA*
vi. *Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO, USA*

## I. Abstract

With recent developments in online adaptive radiotherapy (oART), software- and calculation-based quality assurance (QA) solutions are needed in order to fully exploit the potential of this treatment alternative. While measurement- versus calculation-based pre-treatment patient-specific QA has been debated for a long time, the community suffers from a chronic lack of quantitative data and risk analysis on using these approaches with modern RT delivery systems. The aim of the work carried out by a global panel of early adopters of CBCT-based oART was to suggest a model for safe implementation of calculation-based QA in an oART setting, based on failure modes and effects analysis (FMEA) and evaluations of the accuracy and sensitivity of the calculation-based QA incorporated with the specific oART solution. Based on the FMEA analysis, risks associate with the Mobius3D QA process specific for the oART workflow with Ethos were identified. Combining the FMEA analysis with the early experiences from comparison between calculation- and measurement-based QA, safe implementation of calculation-based QA for oART in the specific oART setting using Mobius3D was supported. The panel in this paper has suggested a hybrid model for safe implementation, utilizing routine patient- and machine-specific calculation-based plan QA in general, with supplementary regular measurement-based QA.

## II. Problem Statement

Since the introduction of IMRT there has been consensus around the recommendations of applying a robust pre-treatment patient-specific QA (PSQA) procedure [1–3]. However, there has been much debate, and variability in clinical practices, regarding measurement-based versus calculation-based patient specific quality assurance (QA) for intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) plans [4]. The proponents of measurement-based QA believe that this QA approach is critical in ensuring safe radiotherapy delivery, given the multitude of potential failure modes and risks associated with radiotherapy delivery systems. The adopters of calculation-based processes believe that there is enough redundancy and lack of inherent risks in modern delivery systems where safe delivery can be ensured with this QA approach. While this debate has been ongoing for over 10 years, there is still not a clear and universally accepted standing and direction on this issue. With conventional IMRT and VMAT delivery, both approaches are clinically practical, and users generally chose to use measurement and/or calculation-based QA based on their local regulations, billing requirements, and/or clinical experience and beliefs. The most common approach in modern PSQA of IMRT and VMAT might even already be a hybrid approach; combining independent dose calculation with measurements

[5]. This ability to utilize a QA approach of choice has minimized the importance of the issue of selecting either a measurement- or calculation-based QA approach.

However, with online adaptive radiotherapy (oART) that choice is removed as measurement-based, especially phantom-based pre-treatment QA is not clinically practical while maintaining the proposed benefits of this treatment technique. Therefore, there is a need to quantitatively assess and define risks with individual oART QA approaches in much greater detail than typically performed and accepted in the field for conventional IMRT and VMAT delivery. Recent developments in oART emphasizes the importance of this discussion [6], and highlights the need for additional attention in relation to CBCT-based oART with full re-optimization to the anatomy-of-the-day. On many levels, a key reason that the debate between measurement- and calculation-based QA is still ongoing is the chronic lack of quantitative data and risk analysis on using these approaches with modern RT delivery systems.

As representatives of an Adaptive Intelligence Consortium (AIC), a global panel of six early adopter institutions were defined to discuss these issues (see Appendix A for an introduction to institutes *i*-*vi* and their experiences with oART and calculation-based QA). The aim of the work carried out by this panel was to suggest a model for safe implementation of calculation-based QA in an oART setting, based on failure mode and effect (FMEA) analysis and evaluations of the accuracy and sensitivity of the calculation-based QA incorporated with the specific oART solution.

## III. Risk Control Measures for integrated calculation-based QA of oART

To ensure the safe implementation of calculation-based QA in an oART workflow, a FMEA was conducted on the solution specifically implemented for the Ethos system using Mobius3D (M3D) version 3.0 and later, including both pre-treatment and post-delivery verification. Mobius3D consists of several different modules. Modules applicable to Ethos are: MobiusCalc for DICOM based 3D Plan QA, MobiusFX for delivery logfile-based pre-treatment and post-delivery QA, and MobiusAdapt for online DICOM based 3D Plan QA and adaptive treatment summary. For simplicity, if not specified explicitly, the suite name Mobius3D/M3D will be used throughout the scope of this paper independently of the involved module. One underlying assumption was that a rigid institutional-specific implementation of an offline patient-specific (measurement- or calculation-based) QA process had already been established. Thereby, the focus of this project was to establish new foundations for a robust calculation-based QA process for the oART PSQA process; including the on-couch pre-treatment plan-specific QA and a subsequent post-treatment QA process.

A subset of the global panel was formed by three physicists (institutes *i*, *ii*, and *vi*) and one process specialist which represent the experience from multiple clinics in three continents as well as the developer group from the manufacturer. The oART patient-specific, calculation-based QA process using the M3D suite was divided into five major steps: (A) Data transfer, (B) Mobius3D calculation, (C) Analysis and display, (D) Data storage and record, and (E) Delivery QA; within which critical steps and their related potential failure modes, causes and effects were identified (Figure 1).

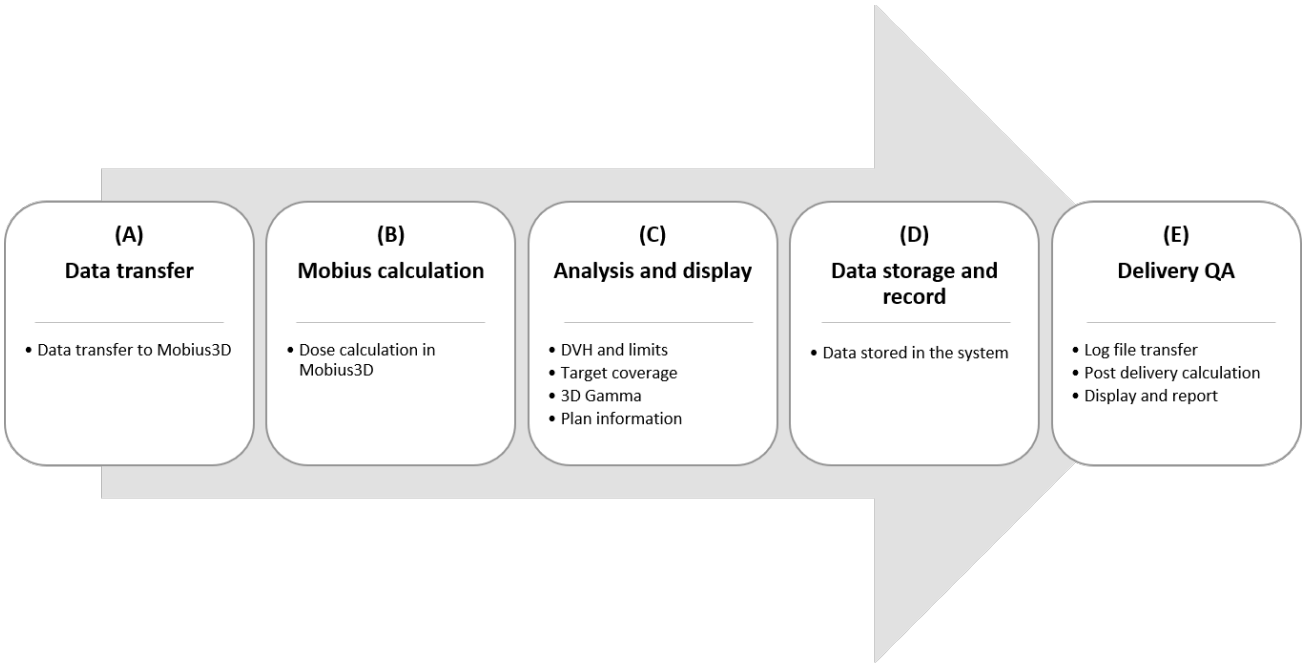


*Figure 1. Schematic illustration of the online adaptive radiotherapy patient-specific, calculation-based quality assurance process divided into five major steps: (A) Data transfer, (B) Mobius3D calculation, (C) Analysis and display, (D) Data storage and record, and (E) Delivery QA; within which critical steps and their related potential failure modes, causes and effects were identified.*

Occurrence (*O*), severity (*S*) and detectability (*D*) was scored based on TG100 recommendations [7], individual clinic's experience and consensus from the panel. The Risk Priority Numbers were then calculated for each failure mode using the following equation:

$$RPN = O \times S \times D$$

with *O*, *S* and *D* ranging from 1 to 10. The full list of failure modes and risk analysis scoring is presented in Appendix B (Table B1 – B5).

*Assumptions*

One crucial underlying assumption in the evaluation of the gathered failure modes was that a correct commissioning of the TPS as well as the M3D system had been performed. Situations where the TPS and M3D model are both inaccurate but agree are not considered in this white paper. Such limitations, potentially leading to unacceptable dose errors progressing through undetected, needs to be addressed during commissioning. Previous studies on M3D have demonstrated the feasibility of achieving good agreement with both measurements in phantom geometries as well as comparable results to TPS dose calculations [8–10]. For many linac systems, as in the case for Ethos, the TPS as well as the M3D models are fixed, only giving the option for the user to adjust the correlation in between them and with measurements by optimization of a dynamic leaf gap (DLG) offset in M3D. This study assumed a selection of an optimal DLG offset for the intended clinical use of Ethos treatment plans. This can e.g. be achieved by altering the DLG offset in M3D within a range and with a step size enabling the user to distinguish the value rendering in the best agreement between the M3D calculated and corresponding measured dose distributions. As a reference, the DLG offset, optimized for IMRT and/or VMAT plans to be delivered to

targets in the pelvic region, ranged from 0% to 0.7%, with a most common value of 0%, for the participants in this global panel of users.

Mobius3D was assumed to be used as intended by trained and experienced staff. Therefore, e.g. the failure mode related to a wrong or incorrect plan being sent to M3D, occurring due to failure in Ethos sending the wrong plan, would need to be detected by an experienced user who is familiar with the patient plan and is able to detect deviations after a visual inspection (if not failing gamma with M3D). Furthermore, the assumption was also, in case of the wrong plan being sent to M3D, that the adaptive (or scheduled) plan can be delivered without performing an adaptive check in M3D, and that M3D or measurement-based QA would be carried out after the current fraction or before the next fraction. In terms of treatment cancelation, only normo-fractionated non-SBRT treatments were considered.

'Wrong or insufficient dose calculation' pertains to dose calculation in Mobius3D, or more likely, in Ethos. The field design in Ethos was considered unlikely to error, while technical structures could be sent with error and would be challenging to detect in Mobius3D as there are no available tools and would require visual inspection. Similarly, wrong image could be attached in error, which is also challenging to detect with Mobius3D as it is not designed for that purpose.

In general, all evaluations were based on the assumption of knowledgeable staff being present at adaptive treatments, that will visualize and understand all clinical decisions such as contouring, DVH limits per site, etc.

*Outcomes*

The FMEA of oART M3D QA resulted in 32 failure modes with an RPN average of 48.8. Three of the failure modes have the highest risks (RPN=120): "Wrong DVH calculations or display", "Wrong coverage calculations or display" and "Wrong dose volume metrics calculations or display" in the "Analysis and Display" step which could be due to software malfunction. However, in our collective experience, the occurrences of those failure modes are thought to be very low (O=1- 2). The RPN of majority failure modes (78%) are below 60. The failure modes with RPN greater or equal to 60 are listed in Table 1. The distribution of the highest RPN within each failure mode is plotted in Figure 2.

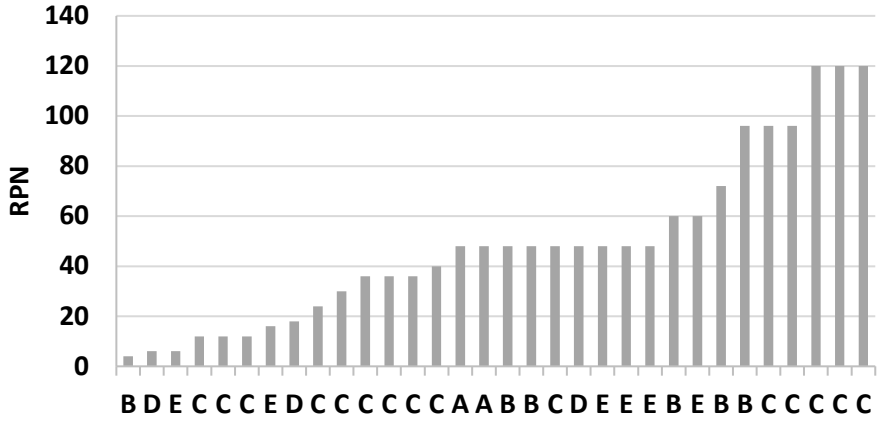


*Figure 2. Distribution of the highest RPN within each oART QA failure mode versus their associated steps (A) Data transfer, (B) Mobius3D calculation, (C) Analysis and display, (D) Data storage and record, and (E) Delivery QA.*

*Table 1. Top nine RPN scores and associated failure modes.*

| Step | Sub-process | Failure mode | $RPN_{max}$ |
|---|---|---|---|
| C | DVH limits calculation and display | Wrong DVH limits calculations or display | 120 |
| C | Target coverage analysis and display | Wrong coverage calculations or display | 120 |
| C | DVH analysis and display | Wrong DVH calculations or display | 120 |
| C | 3D gamma calculation and display | Wrong 3D gamma calculations or display | 96 |
| C | DVH limits calculation and display | DVH limits not reviewed or reviewed incorrectly | 96 |
| B | Dose calculation | Wrong or insufficient dose calculation due to wrong or inappropriate physical property assignment | 96 |
| B | Dose calculation | Wrong or insufficient dose calculation due to wrong or inappropriate imaging used | 72 |
| E | DVH, Gamma, RMS calculations and display | Wrong DVH, Gamma, RMS, D&V limits calculations or display | 60 |
| B | Dose calculation | Wrong or insufficient dose calculation due to wrong or inappropriate delineation | 60 |

(A) Data transfer

The data transfer step involves transfer of data from Ethos to Mobius3D. The two failure modes identified were 'Plan failed to transfer to M3D (from Ethos)' or 'Wrong/incomplete plan was transferred'. In our experience, occasionally there are transfer errors where the plan failed to transfer; this is easily identified, and the plan can be resent. On very rare (non-zero) occasions, a plan may not be sent, or time pressure means that patients may be treated prior to on-couch PSQA and in this circumstance a measurement may be done after treatment. This decision would be based on experience of users with this system and similar treatment plans. The potentially more sinister fault mode is an incomplete or wrong plan sent for QA. This includes, for example, wrong patient on couch, plan doesn't meet protocol or wrong session used or transfer failure that alters the file transmitted. These cases have not been experienced by the panel and are not unique to Ethos-Mobius3D system, but are possible and should be incorporated in full QA.

(B) Mobius3D calculation

The dose calculation step involves M3D independent dose calculation. Six failure modes were identified in the M3D dose calculation step, three of which were rated in top 10 ranking. These were wrong or insufficient dose calculation due to wrong or inappropriate imaging used, due to wrong or inappropriate physical property assignment or due to wrong or inappropriate delineation. These fault modes scored more highly due to difficulty in detection and potential for severe consequences. They all originate from errors that might be made within Ethos and flow through to M3D where functionality/tools are not present (as M3D is not designed to detect and distinguish the cause of these type of errors). Wrong or inappropriate imaging refers to occurrences such as differences between CBCT and synthetic CT (sCT) used to calculate the dose that may be present if the anatomy of the day varies significantly (large geometric change or large HU change) from at the time of

planning. The dose viewed on Ethos is calculated on sCT and shown on CBCT. M3D provides a view of the daily sCT in the report but many institutions may proceed to deliver treatment without review, or use surrogates such as accuracy of bone/body contour. Similarly, the physical property assignment is made in Ethos with no control in M3D relying on processes upstream or experienced treatment staff. Delineation of targets and OAR's on CBCT is executed in Ethos, driving Ethos dose optimization and dose calculation. M3D will compare independent dose calculation against Ethos but will not report on accuracy of contouring that directly affects dose calculation. M3D provides an independent dose calculation with nominal 5% upper bound, but third-party software, offline analysis or experienced treatment staff may be required to catch upstream fault modes.

(C) Analysis and display

The analysis and display step involves the display of M3D calculated dose and analysis against the Ethos calculation, within the MobiusAdapt module of M3D. The top five failure modes for online PSQA were all from this analysis and display step. These modes scored highly due to high score in detectability and potential for severe consequences. The failure modes identified include wrong DVH calculations or display due to software malfunction, wrong target coverage calculations due to software malfunction, wrong DVH limits displayed due to software malfunction, DVH limits not reviewed/reviewed incorrectly due to human error, and wrong 3D gamma configuration (wrong tolerances/criteria) due to human error. The risk of software malfunction may be low but was scored highly due to the extreme difficulty to detect such an error (with M3D and based on our assumptions) and potential for an underlying dose error to reach the patient. Software malfunction and human error are not unique to this Ethos-Mobius3D system. The urgency of oART, even with the automation and purposeful lowering of cognitive load in the software design, can elevate the chance of human error. With an appropriately commissioned M3D and a well-trained adaptive team, this FMEA can provide guidance on a sufficient comprehensive PSQA for oART beyond the tools provided in M3D alone.

(D) Data storage and record

The data storage step involves the transfer of M3D data to a storage location for subsequent recall. There were no failure modes in this step rated inside the top 10. The highest rated failure mode was data corrupted or lost due to human error.

(E) Delivery QA

This step includes a post-treatment independent calculation of the delivered dose based on log files within the MobiusFX (MFX) module of M3D; providing DVH, 3D Gamma, dose volume metrics, and MLC positioning errors. The highest rating failure mode was incorrect display of DVH, gamma, RMS due to software malfunction. This rated highly due to challenges to detect incorrect display, the severity was rated lower in this offline setting compared to similar failure modes discussed above for the online setting.

Based on FMEA results, QA and QC measures can be proposed to reduce risks for high RPN steps. Recommendations are discussed in the following section (section V)

## IV. Measurement-based compared with calculation-based QA

### a. Early data and analysis on measurement-based compared with calculation-based QA

In order to assess the reliability of the M3D based independent dose calculation solution for oART, institute *i* conducted comparisons between their different measurement-based pre-treatment QA systems. These early experiences, during the world's first commissioning of the Ethos system, included a set of forty-eight automatically generated treatment plans (nine- and twelve-field IMRT, two- and three-arc VMAT) for bladder and rectum patients, in both cases including simultaneous treatment of pelvic lymph nodes, created using a pre-clinical release of the Ethos system for pre-treatment plan generation [11]. Plans were exported to the M3D system (v. 2.2) for independent dose calculation and measured using the Delta4+ phantom (D4) (ScandiDos AB, Uppsala, Sweden) as well as with portal dosimetry (Halcyon v2.0). Two- and three-dimensional gamma evaluations were conducted according to institutional standard protocols for the three investigated QA systems. In all cases a 10% dose threshold was applied.

The M3D calculations demonstrated average gamma passing rates above 97% (3%/2mm, global gamma) for all cases, above current clinical pass criteria of 95% and in general comparing well with corresponding results for D4 measurements (3%/2mm, local gamma, with 95% and 90% tolerance for VMAT and IMRT, respectively) (Table 2). Similarly, all plans were accepted based on portal dosimetry, with an average passing rate of 100% (3%/2mm, local gamma). Overall, the results demonstrate high deliverability of the automatically generated Ethos plans for oART independently of plan complexity (for the range evaluated in this set of plans), as demonstrated by evaluation of the gamma passing rates as a function of MU/Gy (Figure 3).

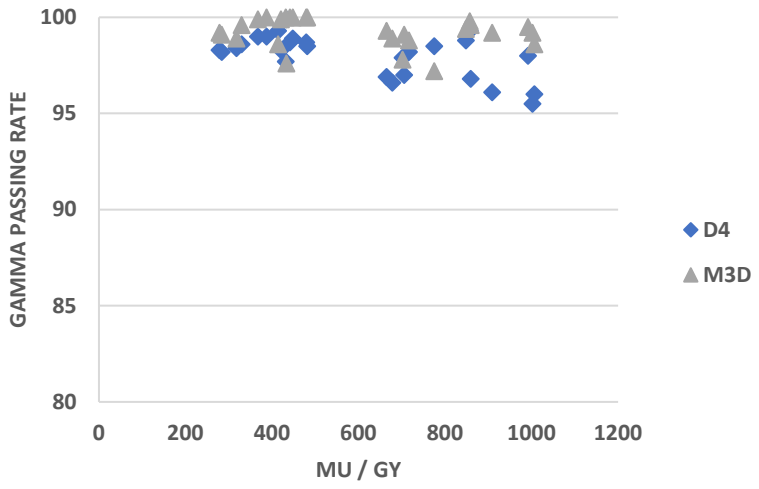


*Figure 3. Gamma passing rate as a function of MU/Gy (as a measure for plan complexity) for measurements-based (3%/2mm local gamma, Delta4, D4, in blue diamonds) versus calculation-based (3%/2mm global gamma, Mobius3D, M3D, in grey triangles) QA of the forty-eight first plans evaluated during the world's first clinical implementation of the Ethos therapy treatment suite at institute i.*

*Table 2. Summary of the average gamma passing rates and standard deviations achieved using Delta4 (D4), portal dosimetry (PD) and Mobius3D (M3D) for patient-specific QA of for forty-eight treatment plans generated in ETHOS for six bladder and six rectum cancer patients at institute i.*

| | Bladder | | | Rectum | | |
|---|---|---|---|---|---|---|
| | *D4* | *PD* | *M3D* | *D4* | *PD* | *M3D* |
| 9-field IMRT | *98.3 ± 0.9* | *99.9 ± 0.0* | *99.5 ± 0.3* | *96.9 ± 0.7* | *100.0 ± 0.0* | *98.8 ± 0.6* |
| 12-field IMRT | *97.3 ± 1.2* | *99.9 ± 0.1* | *98.6 ± 1.0* | *97.4 ± 1.0* | *100.0 ± 0.0* | *98.5 ± 0.8* |
| 2-arc VMAT | *98.7 ± 0.4* | *99.7 ± 0.1* | *98.7 ± 0.9* | *98.5 ± 0.5* | *99.6 ± 0.2* | *100.0 ± 0.0* |
| 3-arc VMAT | *98.9 ± 0.6* | *99.9 ± 0.1* | *97.8 ± 1.4* | *98.7 ± 0.3* | *99.7 ± 0.2* | *100.0 ± 0.0* |
| ALL | *98.3 ± 1.0* | *99.9 ± 0.1* | *98.6 ± 1.1* | *97.9 ± 1.0* | *99.8 ± 0.2* | *99.3 ± 0.8* |

Similarly to the pre-clinical QA assessment, the first forty-five patient-specific clinical adapted treatment plans, re-optimized on the anatomy of the day for the four first patients ever treated with oART using the Ethos platform (all nine- or twelve-field IMRT bladder treatments), were assessed during the online adaptive sessions using M3D and compared to D4-based measurements conducted post-treatment (Figure 4). These results indicate a correlation between gamma passing rates and plan complexity for the measurement-based QA. The MU/Gy was applied as a surrogate for plan complexity, while corresponding gamma evaluation data is also displayed as a function of the total PTV volume (cc) in order to demonstrate the impact of the target volume on the number of monitor units as well as, especially, the D4 gamma passing rates.

Similar correlation between gamma passing rates and plan complexity, as observed for D4, was also indicated by the 2%/2mm based analysis of the M3D calculation-based QA; with the specific plan failing the 90% passing rate criteria of the measurement-based solution also failing the 95% criteria for the calculation-based approach. Corresponding correlation with plan complexity is more difficult to distinguish for a M3D 3%/2mm global gamma criteria, even if that same plan was demonstrating lower gamma passing rate than the rest also for 3%/2mm. The remaining plans all passed the respective gamma passing rate criteria.

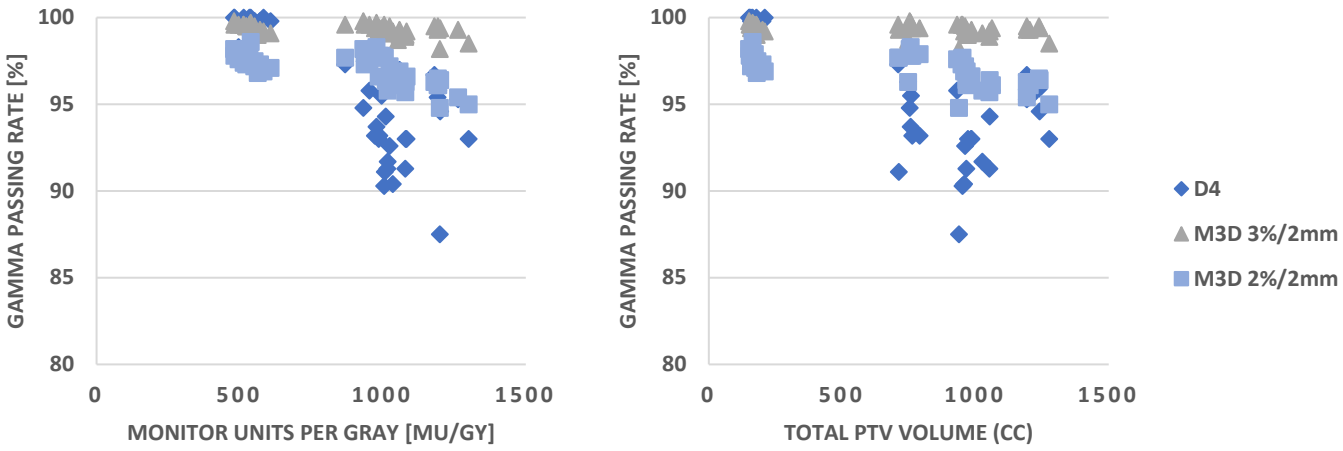


*Figure 4. Gamma passing rates as a function of MU/Gy (left) and total planning target volume (right) for measurements-based (3%/2mm local gamma, Delta4, D4, in blue diamonds) versus calculation-based (3%/2mm and 2%/2mm global gamma, Mobius3D, M3D, in grey triangles and light blue squares, respectively) QA of the forty-five first clinical adapted bladder plans evaluated after the world's first clinical implementation of ETHOS at institute i.*

When considering dose accuracies by evaluating calculations to measurements, possible interplay of dose-to-medium and dose-to-water should be considered, particularly in cases like for M3D where the model is tweaked to match doses measured with chambers and/or diodes/arrays calibrated against chambers. However, this is a non-Ethos related issue not discussed further in the white paper and users are encouraged to acknowledge the issue and follow internal guidelines.

Differences in sensitivity observed here (Figure 3 and Figure 4) are mainly related to the use of local or global gamma evaluation, and if the evaluation is performed on 2D or 3D dose distributions. These results support the need to optimize gamma criteria and tolerance levels depending, not only on the type of plans to be evaluated, but also on the applied evaluation method. In this case, results are indicating the need for tighter tolerances using M3D for independent dose calculation compared to D4 for phantom-based measurements.

*Table 3. Patient-specific QA measures from measurement- (ArcCHECK, AC) and calculation-based (Mobius3D, M3D) QA of thirty-three treatments plans generated in Ethos for a variety of disease sites at clinic iii. Corresponding plan complexity measure (MU/Gy) as well as field technique information is provided together with the QA results, which includes gamma passing rates and target coverage parameters.*

| Anatomical site | Technique | Fields | MU/Gy | AC 3%/2mm | M3D 3%/2mm | M3D 2%/2mm | Mean PTV (%) | 90% PTV coverage (%) |
|---|---|---|---|---|---|---|---|---|
| Bladder | IMRT | 7 | 538.3 | 99.4 | 99.7 | 98.5 | -0.1 | -1.4 |
| Sacrum | IMRT | 12 | 700.8 | 99.7 | 98.9 | 96.3 | 0.1 | -0.6 |
| Lung | IMRT | 12 | 371.8 | 98.3 | 99.5 | 98.9 | -1.8 | -4.6 |
| Rectum | IMRT | 12 | 1001 | 95.5 | 99.6 | 97.8 | -0.2 | -1 |
| Oesophagus | IMRT | 9 | 695.0 | 98.8 | 98.5 | 95.9 | -0.5 | -3.4 |
| Sternum | IMRT | 7 | 435.5 | 99.8 | 99.3 | 97.5 | 0.3 | -1.3 |
| Spine | IMRT | 7 | 460.3 | 99.5 | 98.3 | 94.9 | 0.2 | -0.1 |
| Spine | IMRT | 9 | 509.5 | 98.7 | 99.8 | 99.2 | -0.4 | -1.2 |
| Pelvis | IMRT | 9 | 727.2 | 94.1 | 99.5 | 97.5 | -0.5 | -1 |
| Brain Whole | IMRT | 9 | 401.0 | 99.9 | 98.7 | 95.8 | 0.7 | 0 |
| Abdomen | IMRT | 7 | 379.0 | 99.6 | 99.7 | 99.1 | -0.3 | -0.7 |
| Larynx | VMAT | 2 | 235.5 | 100 | 97.8 | 96.3 | 1 | -3.1 |
| Brain | IMRT | 9 | 407.1 | 99.8 | 99.7 | 99.3 | 0.1 | -0.7 |
| Kidney | IMRT | 7 | 429.75 | 100 | 99.9 | 99.4 | 0.6 | 0.2 |
| ProsB_N | IMRT | 12 | 984.0 | 99.4 | 99.4 | 97.0 | -0.3 | -0.3 |
| Anus | IMRT | 9 | 1088.3 | 98.7 | 98.6 | 95.5 | 0.2 | -0.7 |
| ProsB_N | IMRT | 12 | 1033.0 | 100 | 99.6 | 98.3 | 0.1 | 0.3 |
| Pancreas | IMRT | 9 | 616.1 | 97.7 | 99.9 | 99.0 | 0.5 | 0 |
| Prostate_N | VMAT | 3 | 263.4 | 100 | 96.8 | 88.8 | 2.2 | 1.9 |
| Brain | IMRT | 9 | 643.3 | 99.5 | 98.5 | 90.0 | 0.4 | -0.1 |
| Pelvis | IMRT | 7 | 893.9 | 99.6 | 99.4 | 96.4 | 0.7 | 0 |
| Spine | IMRT | 9 | 649.2 | 100 | 99.2 | 97.0 | 1.5 | 1.2 |
| Prostate_N | IMRT | 9 | 836.5 | 100 | 99.7 | 98.4 | 0.6 | 0.2 |
| Prostate_N | IMRT | 9 | 790.8 | 99.4 | 99.5 | 97.5 | 1.3 | 0.3 |
| Brain | IMRT | 9 | 508.2 | 98.8 | 99.0 | 86.8 | 1 | 0 |
| Brain | IMRT | 9 | 454.8 | 98.3 | 99.9 | 99.6 | 0.5 | -0.4 |
| ProstateBed | VMAT | 2 | 286 | 100 | 98.2 | 91.1 | 2.2 | 1.9 |
| Pelvis | IMRT | 12 | 966.1 | 98.4 | 99.7 | 98.3 | 0.6 | -0.2 |
| ProstateSV | IMRT | 12 | 665.7 | 99.4 | 99.9 | 99.6 | 0.4 | 0.3 |
| Sacrum | IMRT | 7 | 583.5 | 95.6 | 99.6 | 98.3 | 0.4 | 0.2 |
| T Spine | IMRT | 7 | 428 | 99.6 | 99.1 | 96.4 | 0.4 | 0 |
| Pelvis | IMRT | 12 | 800.25 | 99.5 | 99.4 | 97.9 | -0.5 | -1.5 |
| Scapula | VMAT | 2 | 235 | 100 | 99.3 | 93.5 | 1.5 | 0.7 |

In contrary to the indicated correlations observed for the failing plan detected by both M3D and D4, two plans just barely passing the 95% M3D tolerance level resulted in passing rates well above the tolerance level for the D4. This is an indication of the need for further

sensitivity and specificity evaluation using introduced intentional errors. It is furthermore an indication of the limitations of only basing PSQA on gamma evaluation. In general, gamma statistics should be checked in a structure-by-structure basis and three-dimensional (or two-dimensional) gamma passing rates should only be used as an indicator of potential problems and not as the single indicator of plan quality.

Use of target coverage or DVH limits, as available in M3D, will assist in further evaluation of the potential clinical impact of any detected deviations. Evaluation of thirty-three offline generated reference plans (clinic iii) for a variety of disease sites indicated that such values can assist in confirming an observed failing gamma evaluation, adding information on the potential clinical impact, but also serve as an independent barrier for detecting clinically suboptimal plans, potentially not detected by neither measurement- nor calculation-based gamma evaluation (Table 3).

For these data, comparing measurements using the ArcCHECK (AC) (Sun Nuclear corporation, FL), the correlation with plan complexity was not as strong as for the above discussed D4 results for the bladder treatments. However, further gathering of similar data from a range of QA equipment, to benchmark QA results for plans generated on the Ethos platform, could serve as a powerful tool for commissioning and clinical routine use of M3D as the primary PSQA method for oART.

**b.** Evaluation of calculation vs. measurement-based QA in situations with introduced known intentional errors

In order to investigate the ability of the M3D solution to detect deviating plan deliveries, four different known intentional minor errors to the beam configuration (Table 4) were introduced to two different clinical treatment plans generated in Ethos at clinic *i*; one dual half-arc VMAT plan generated for a palliative rib treatment and one nine field IMRT plan generated for a curative bladder cancer treatment. Together with the unaltered plan (None) and a plan altered on all parameters (All), a total of twelve anonymized plans were imported in to M3D as well as delivered to the D4 phantom using the linear accelerator in file mode, for subsequent comparison to the dose distributions of the unaltered plans.

*Table 4. Description of known international minor errors introduced to two different clinical treatments plans (one dual half-arc VMAT palliative plan and one nine field IMRT curative bladder plan) generated in Ethos at clinic i.*

| Error | Modifier | Value | Unit | Description |
|---|---|---|---|---|
| **Gant** | Gantry angle | 1 | degree | Gantry angle was changed with 1 degree for all fields |
| **Coll** | Collimator angle | 1 | degree | Collimator angle was changed with 1 degree for all fields |
| **MLC** | MLC1 A Leaf bank | 1 | millimeter | All MLCs on all banks were shifted 1 millimeter for all fields |
| | MLC1 B Leaf bank | 1 | millimeter | |
| | MLC2 A Leaf bank | 1 | millimeter | |
| | MLC2 B Leaf bank | 1 | millimeter | |
| **MU** | MU scaling | 2 | percent | The MU per control point was increased 2% for all fields |

*Table 5. Resulting QA measures for Mobius3D and Delta4 based QA of two treatment plans generated at clinic i, one dual half-arc VMAT palliative plan (Ribs left) and one nine field IMRT curative plan (Bladder) with introduced known intentional minor errors.*

| | | Mobius3D | | | | | Delta4 |
|---|---|---|---|---|---|---|---|
| | Error | *3% / 3mm* | *3% / 2mm* | *2% / 2mm* | *PTV $D_{mean}$* | *PTV $D_{95\%}$* | *3% / 2 mm* |
| Ribs Left | None | *99.7* | *99.2* | *98.1* | *1.1* | *-1.3* | *97.9* |
| | Coll | *99.7* | *99.2* | *98.1* | *1.1* | *-1.4* | *97.0* |
| | Gant | *99.6* | *99.0* | *97.6* | *1.3* | *-1.2* | *100* |
| | MLC | *99.6* | *97.8* | *94.9* | *0.9* | *-2.2* | *100* |
| | MU | *97.9* | *97.3* | *96.5* | *3.0* | *0.5* | *96.6* |
| | All | *97.7* | *96.4* | *93.8* | *3.0* | *-0.3* | *96.8* |
| Bladder | None | *100* | *99.3* | *97.6* | *0.1* | *-1.5* | *99.8* |
| | Coll | *100* | *99.3* | *97.5* | *0.1* | *-1.5* | *99.8* |
| | Gant | *98.7* | *93.2* | *88.2* | *0.0* | *-1.6* | *99.8* |
| | MLC | *99.9* | *98.3* | *95.3* | *0.0* | *-1.8* | *99.2* |
| | MU | *99.1* | *98.6* | *96.8* | *2.0* | *0.3* | *97.0* |
| | All | *96.7* | *89.4* | *83.7* | *1.9* | *0.1* | *96.8* |

This initial investigation indicated that both measurement- and calculation-based QA are insensitive to the introduction of the minor changes to beam configurations applied here (Table 5). While measurement-based QA in this case was observed to only detect changes to the MU per control point, there are indications that the calculation-based QA, depending on the selected gamma criteria, might detect other errors to a greater degree. While, when using a 3%/3mm gamma criteria and a 95% passing rate tolerance, calculation-based QA is as insensitive as the current clinical approach for measurement-based QA, there are in certain cases indications that a selection of a stricter criteria should be considered. These results thereby emphasize the importance of optimizing these criteria for the specific clinical situation as a part of the commissioning and clinical implementation process.

The additional information provided with the M3D solution, giving direct feedback on e.g. the PTV mean dose and D95%, helps in the following evaluation of the clinical impact of the detected deviations. Similar information about the dosimetric impact for OAR are furthermore provided in the M3D solution. In that sense the calculation-based QA here might serve as a more powerful tool than certain measurement-based solutions. However, these initial results demonstrated a need for further investigation of the sensitivity and specificity for additional treatment scenarios.

Additional larger intentional known errors were introduced to the beam configuration of a separate clinical treatment plan generated in Ethos at institute *iv* (Table 6). Such, in some cases severe, errors are unlikely to occur without the internal control system of the linear accelerator detecting them and preventing improper patient treatment. Plans with these introduced errors were nonetheless evaluated in M3D, using the same set of gamma criteria. Results indicate a correlation between reduced gamma passing rates with increasing severity of the introduced errors (Table 6). These results indicate that calculation-based QA could serve as a valuable tool for PSQA. However, further evaluation of the correlation of these results with measurements are needed.

*Table 6. Introduced known intentional larger errors together with corresponding resulting QA measures for Mobius3D based QA of one treatment plan generated at institute iv.*

| ***Error description*** | | | | ***Mobius3D gamma passing rate*** | | |
|---|---|---|---|---|---|---|
| *Error* | *Modifier* | *Value* | *Unit* | *3% / 3mm* | *3% / 2mm* | *2% / 2mm* |
| **None** | None | None | None | 100 | 99.7 | 98.7 |
| **Coll1** | Collimator angle | 1 | Degree | 100 | 99.7 | 98.5 |
| **Coll2** | Collimator angle | 2 | Degree | 100 | 99.2 | 97.7 |
| **Coll3** | Collimator angle | 3 | Degree | 99.9 | 98.2 | 95.8 |
| **Coll4** | Collimator angle | 4 | Degree | 99.5 | 96.4 | 93.3 |
| **Coll5** | Collimator angle | 5 | Degree | 98.7 | 94.1 | 90.2 |
| **MLC1** | MLC positions | 5 | millimeter | 99.2 | 96.9 | 93.3 |
| **MLC2** | MLC positions | 10 | millimeter | 98.3 | 91.9 | 84.3 |
| **MLC3** | MLC positions | 20 | millimeter | 89.4 | 74.4 | 61.3 |
| **MLC4** | MLC positions | 30 | millimeter | 72.3 | 57.5 | 43.9 |
| **MLC5** | MLC positions | 40 | millimeter | 58.1 | 44.2 | 32.9 |
| **MLC6** | MLC positions | 50 | millimeter | 46.5 | 34.3 | 26.3 |
| **MLC7** | MLC leaf stuck | 1 | leaf | 99.2 | 96.9 | 93.3 |
| **MLC8** | MLC slow | 1 | leaf | 99.8 | 99.4 | 98.1 |
| **MU1** | MU scaling | 5 | Percent | 100 | 99.8 | 98.9 |
| **MU2** | MU scaling | 10 | Percent | 100 | 99.8 | 99 |
| **MU3** | MU scaling | 15 | Percent | 99.6 | 99.4 | 97.6 |
| **MU4** | MU scaling | 20 | Percent | 98.5 | 98.1 | 96.6 |
| **MU5** | MU scaling | 25 | Percent | 97.8 | 97.4 | 95.2 |
| **MU6** | MU scaling | -5 | Percent | 97.4 | 96.9 | 93.9 |
| **MU7** | MU scaling | -10 | Percent | 99.9 | 97.9 | 96.6 |
| **MU8** | MU scaling | -15 | Percent | 99.6 | 97.9 | 92.6 |
| **MU9** | MU scaling | -20 | Percent | 98.7 | 95.5 | 85.8 |
| **MU10** | MU scaling | -25 | Percent | 96.5 | 91.5 | 79.3 |

## V. Safe implementation of Mobius3D calculation-based QA

The M3D software has been adopted by many users for regular pre-treatment IMRT or VMAT PSQA. The workflow and fundamental calculation algorithm is relatively mature. In the scenario of oART, it utilizes a similar workflow, same calculation algorithm and similar display interface. Therefore, it is not a completely new approach or utilization for pre-treatment PSQA of oART.

The commissioning process of M3D is still critical. The system shall be appropriately configured and thoroughly validated on an institution-specific level before release to clinical use. The commissioning approaches and experiences on M3D for regular non-adaptive systems should be easily transferred for oART system. The standard beam data and preconfigured beam model utilized by Ethos platform makes the commissioning process relatively straightforward and allows for system benchmarking. Routine QA of M3D is also recommended to test the functionality, connectivity, and consistency of the system.

The FMEA analysis identified risky steps associated with M3D QA process of oART. Based on these results, control strategies should be developed to mitigate the risks. In general, as presented in Table 7, suitably trained staff members, utilization of checklists, enforcement of secondary checks by another team member and utilization of independent software for secondary checking are good quality control (QC) and QA measures which can be implemented at each sub-step accordingly. A continued monitoring of the workflow and a periodic re-assessment could further help improve the process.

The introduced error tests at least indicated that the sensitivity of M3D system is not worse than the regular measurement-based QA. The clinical meaningful QA criteria and action levels should be established based on national guidelines, like TG218 [12] and/or institutional QA policy, together with institutional based sensitivity testing.

*Table 7. Top nine failure modes and potential QC strategies.*

| *Failure mode* | *Potential QC strategies* |
|---|---|
| Wrong DVH limits calculations or display | Carefully evaluate the results by well-trained ART personnel; Compare the DVH limits with the values reported by ETHOS; Use third party or in-house developed DVH limits calculation software as secondary check; |
| Wrong coverage calculations or display | Carefully evaluate the results by well-trained ART personnel; Compare the coverage with the values reported by ETHOS; Use third party or in-house developed dose calculation software as secondary check; |
| Wrong DVH calculations or display | Carefully evaluate the results by well-trained ART personnel; Compare the DVH plot with the values reported by ETHOS; Use third party or in-house developed DVH calculation software as secondary check; |
| Wrong 3D gamma calculations or display | Carefully evaluate the results by well-trained ART personnel; Use third party or in-house developed gamma calculation software as secondary check; |
| DVH limits not reviewed or reviewed incorrectly | Only have well-trained ART personnel to perform the task; Utilize QA checklist and document the DVH limits review activity; Assign a secondary QA personnel for DVH limits review. |
| Wrong or insufficient dose calculation due to wrong or inappropriate physical property assignment | Only have well-trained ART personnel to perform the task; Utilize QA checklist and document the physical property assignment activity; Assign a QA personnel for physical property assignment check; Use third party or in-house developed software to check physical property assignment. |
| Wrong or insufficient dose calculation due to wrong or inappropriate imaging used | Carefully evaluate the image dataset by well-trained ART personnel before proceeding to delineation and re-plan; Utilize QA checklist and document the image used. |
| Wrong DVH, Gamma, RMS, D&V limits calculations or display | Only have well-trained ART personnel to perform the task; Utilize QA checklist and document the image used; Use third party or in-house developed log file QA software as secondary check; |
| Wrong or insufficient dose calculation due to wrong or inappropriate delineation | Only have well-trained ART personnel to perform the delineation; Carefully evaluate the delineation by well-trained ART QA personnel; Utilize contour checklist and document the check results; Use third party or in-house developed contour check software. |

Based on FMEA analysis together with the comparison between calculation- and measurement-based QA, safe implementation of calculation-based QA for oART in the specific oART setting using M3D as described in this paper is supported. Assuming a rigid validation process of the calculation-based QA together with an extensive machine-specific QA program are in place, the panel in this paper suggests a model for routine patient- and machine-specific plan QA using calculation-based QA in general with regular (monthly or quarterly) measurement-based QA of standard and randomly selected clinical treatment plans (Figure 6).

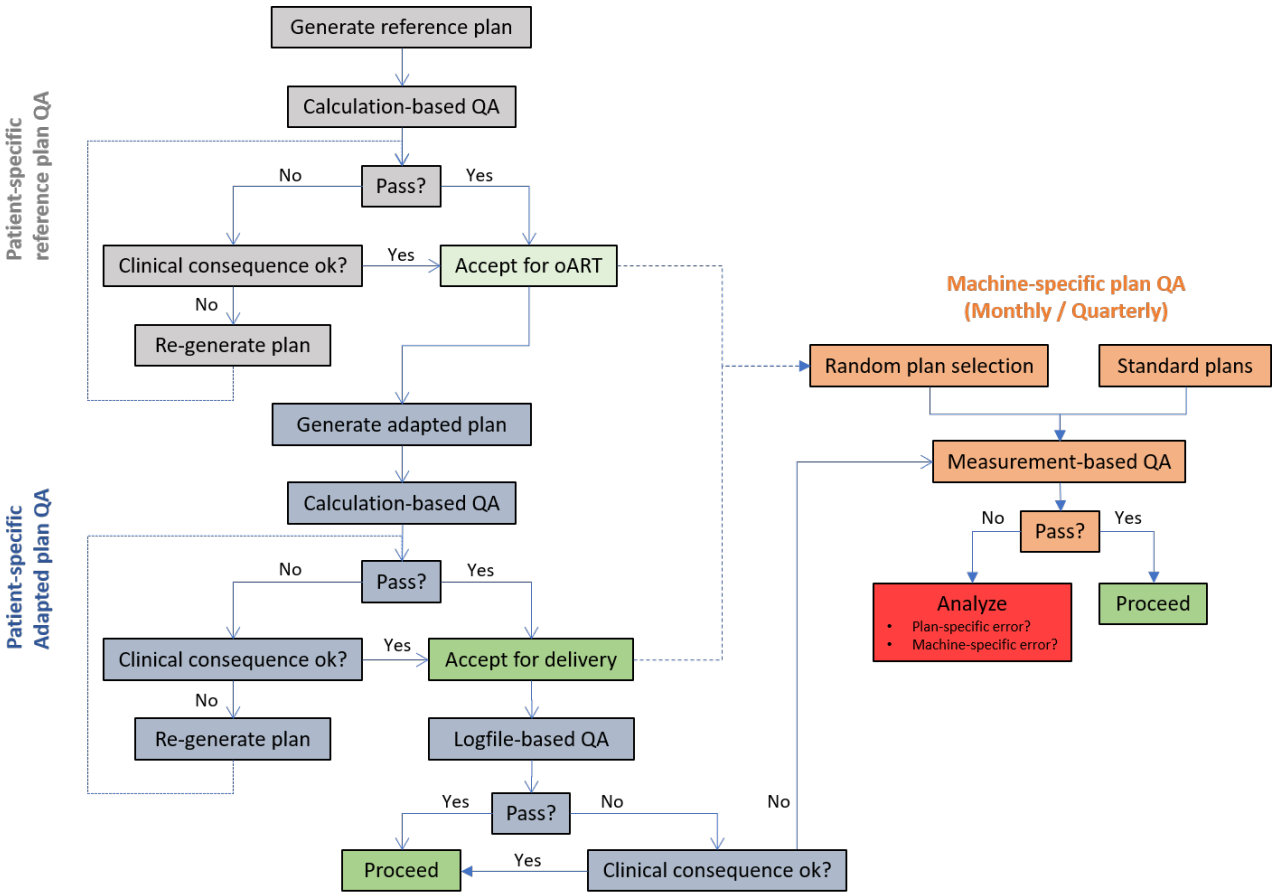


*Figure 5. Proposed model for routine patient- and machine-specific plan QA using calculation-based QA in general with regular (monthly or quarterly) measurement-based QA of standard and randomly selected clinical treatment plans.*

In summary, here are some recommendations for safe implementation of M3D system for oART patient specific pretreatment QA:

1. Identify a commissioning and implementation team for M3D with a key physicist taking the lead role.

2. Following M3D standard commission process and institutional specific commissioning tasks, thoroughly validate the system with various plans ranging from simple field to IMRT/VMAT plans. Evaluate and optimize the selection of the DLG offset in M3D carefully.

3. Based on validation tests as well as institution's QA guideline, determine QA criteria, passing threshold, and action levels.

4. Develop and implement QA/QC measures in oART process to reduce risks of calculation-based pre-treatment QA.

5. Determine routine QA program for Mobius3D and continuously monitor and re-assess the M3D QA process.

6. Establish a training program to have suitably trained staff, with abilities to perform adequate evaluation of displayed M3D results.

## VI. Conclusion

Risks associate with the M3D QA process specific for the oART workflow were identified and based on the results of the FMEA analysis, control strategies should be developed to mitigate the risks. In general, well trained staff member and implementation of a thorough QA program is recommended. Combining the FMEA analysis with the early experiences from comparison between calculation- and measurement-based QA, safe implementation of calculation-based QA for oART in the specific oART setting using M3D as described in this paper is supported. The supporting data indicates the potential added benefit of calculation-based QA in the patient-specific geometry, making evaluation of clinical impact of potential detected deviations accessible. The panel in this paper has suggested a model for routine patient- and machine-specific plan QA using calculation-based QA in general with supplementary regular (monthly or quarterly) measurement-based QA.

## VII. Acknowledgements

As members of the Adaptive Intelligence™ Consortium, the authors receive financial compensation from Varian, a Siemens Healthineers Company, for their original research.

## VIII. References


[1] Ezzell GA, Galvin JM, Low D, Palta JR, Rosen I, Sharpe MB, et al. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee. Med Phys 2003;30:2089–115. https://doi.org/10.1118/1.1591194.

[2] Galvin JM, Ezzell G, Eisbrauch A, Yu C, Butler B, Xiao Y, et al. Implementing IMRT in clinical practice: A joint document of the American Society for Therapeutic Radiology and Oncology and the American Association of Physicists in Medicine. Int J Radiat Oncol Biol Phys 2004;58:1616–34. https://doi.org/10.1016/j.ijrobp.2003.12.008.

[3] Ezzell GA, Burmeister JW, Dogan N, Losasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. Med Phys 2009;36:5359–73. https://doi.org/10.1118/1.3238104.

[4] Siochi RAC, Molineu A, Orton CG. Patient-specific QA for IMRT should be performed using software rather than hardware methods. Med Phys 2013;40. https://doi.org/10.1118/1.4794929.

[5] Visser R, Wauben DJL, De Groot M, Godart J, Langendijk JA, Van'T Veld AA, et al. Efficient and reliable 3D dose quality assurance for IMRT by combining independent dose calculations with measurements. Med Phys 2013;40. https://doi.org/10.1118/1.4774048.

[6] Cai B, Green OL, Kashani R, Rodriguez VL, Mutic S, Yang D. A practical implementation of physics quality assurance for photon adaptive radiotherapy. Z Med Phys 2018;28:211–23. https://doi.org/10.1016/j.zemedi.2018.02.002.

[7] Huq MS, Fraass BA, Dunscombe PB, Gibbons JP, Ibbott GS, Mundt AJ, et al. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management. Med Phys 2016;43:4209–62. https://doi.org/10.1118/1.4947547.

[8] Fontenota JD. Evaluation of a novel secondary check tool for intensity-modulated radiotherapy treatment planning. J Appl Clin Med Phys 2014;15:207–15. https://doi.org/10.1120/jacmp.v15i5.4990.

[9] Nelson CL, Mason BE, Robinson RC, Kisling KD, Kirsner SM. Commissioning results of an automated treatment planning verification system. J Appl Clin Med Phys 2014;15:57–65. https://doi.org/10.1120/jacmp.v15i5.4838.

[10] Jolly D, Dunn L, Kenny J. A clinical database to assess action levels and tolerances for the ongoing use of Mobius3D. J Appl Clin Med Phys 2017;18:59–65. https://doi.org/10.1002/acm2.12009.

[11] Sibolt P, Andersson LM, Calmels L, Sjöström D, Bjelkengren U, Geertsen P, et al. Clinical implementation of artificial intelligence-driven cone-beam computed tomography-guided online adaptive radiotherapy in the pelvic region. Phys Imaging Radiat Oncol 2021;17:1–7. https://doi.org/10.1016/j.phro.2020.12.004.

[12] 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. https://doi.org/10.1002/mp.12810.

**Appendix A. Early Adopter Experiences**

This appendix gives a brief overview of the background of the centers involved in writing this paper and their early experiences and plans for implementation of patient specific QA for online adaptive radiotherapy.

**i.** Dept of Oncology, Copenhagen University Hospital - Herlev and Gentofte, Copenhagen, Denmark

Herlev Hospital implemented the first clinical Ethos system in the fall of 2019 [11] and has since treated >350 patients and >5500 fractions on the platform, out of which >24 patients and >360 fractions with oART. Herlev Hospital has also been an early adopter of M3D for independent dose calculation QA with an installment of the solution in 2013. Prior to the clinical implementation of M3D in 2013, all dynamic treatment plans were initially measured using the D4 system. However, the rapid increase in the use of IMRT and VMAT plans after the introduction of dynamic treatments in 2011 created a need to find a less cumbersome and less time-consuming approach. After a transitional period of using D4, portal dosimetry for electronic portal imaging device-based plan QA and M3D, the latter solution, including logfile-based delivery QA, was fully implemented as a replacement of measurement-based QA for all standard treatment plans. At Herlev Hospital, M3D is not only used for plan delivery evaluation but furthermore used as a method for plan quality assessment with the possibilities to evaluate dose as calculated in the actual patient anatomy, enabling also 3D DVH parameter evaluation. This provides valuable information of the origin of potential dose differences, enabling further investigation into the potential clinical impact of the QC results as compared to most measurement-based solutions.

After the initial period of clinical implementation of oART with Ethos, Herlev Hospital has moved over completely to relying on a M3D-driven workflow for plan-specific QA. This includes the online pre-treatment dose calculation verification as well as the logfile-based MFX calculation of the delivered dose to the patient anatomy, as evaluated for each fraction. Based on the experiences from the pre-clinical study on the plan deliverability and sensitivity of M3D, additional requirements to plan complexity has been added in terms of change in the number of MU from the pre-treatment generated reference plan to the online created adapted plan. In general, the aim is to remain below 1300 MU/Gy for a standard treatment and below a maximum increase of 250 MU/Gy from the reference to the adapted plan. These are local recommendations and dependent on the amount of anatomical changes (enlargements of target structures and proximity to organs at risk).

As for all standard linear accelerators at Herlev Hospital (one Ethos, two Halcyon v.3.0, five TrueBeam) additional machine-specific QA was introduced with the removal of plan-specific measurement-based QA. This includes additional daily, weekly and monthly machine-specific verifications, but most importantly quarterly additional measurement-based plan evaluations of a set of standard plans for machine consistency control as well as verification of a set of newly generated treatment plans. This combination has been implemented at Herlev Hospital since the initial full transition to M3D-driven calculation-based QA and since then proven reliable.

**ii.** Northern Sydney Cancer Centre, Royal North Shore Hospital, Sydney, Australia

Northern Sydney Cancer Centre is a 3 linac all Varian public hospital. Ethos was installed in February 2020 with first adaptive treatments from March of that year. Mobius3D was a new product for this team. Commissioning of the software incorporated guidance from key references such as IAEA TRS430, AAPM MPPG 5a, IAEA TECDOC 1583 and AAPM TG119. Initial commissioning of the Mobius3D software included comparison of measurements (Delta4+ and point dose) and calculation (Eclipse 15606, Ethos v1, Mobius3D v3.1) of simple static fields (e.g. PDD, off-axis, asymmetry, obliquity, SSD variation etc.), couch and accessories, dynamic MLC, inhomogeneity and a range of clinical cases. Our typical case mix for Ethos treatment includes male/female pelvis, head and neck, palliative and GI. The DLG was set at 0.25mm, determined from IMRT and VMAT deliveries using Ethos and Eclipse generated plans for our specific case mix. The Mobius3D gamma criteria of 95% of points with 3%/3mm (dose threshold = 10%) and mean dose difference ±5% was determined for the case mix allocated to Ethos, which doesn't include SRS/SBRT.

The initial Ethos planned (IGRT and Adaptive) plans were quality assured with Mobius3D (calculation and logfile-based delivery), Delta4+ and point dose measurements. To date no plans have failed tolerance criteria. The initial adapted fractions were measured in Delta4+ and point dose, this currently includes >80 fractions, all with acceptable agreement.

Mobius3D contains a visualization of the sCT and our practice is initially to review the sCT prior to accepting an adapted plan. There have been (rare) cases where this synthetic CT is not representative of the CBCT and further investigation is in progress. We are also evaluating the suitability of utilizing surrogates for sCT fidelity (such as bone and body contours) within the target evaluation step. Review of each adaptive technical plan report compares the adapted plan MU to the reference plan, and this is used to understand (relative) plan changes. We are evaluating trends of MU changes with software and measurement based QA results.

**iii.** Icon Group, South Brisbane, Australia

Icon Group currently have two Ethos systems installed and clinical at their Wahroonga facility in Sydney, Australia, and at their Greenslopes facility in Brisbane, Australia. They have treated numerous oART patients to date, with most of these being pelvis based. In addition to this, they have also treated numerous patients using the alternate IGRT mode.

They have extensive experience with M3D and MFX for independent dose calculation as well as delivery QA and are using these products routinely in their clinics. Prior to the implementation of the Mobius3D products, their patent specific QA for modulated treatments consisted of independent phantom based measurements using either ion chambers and/or 2D array devices such as SNC ArcCheck. Whilst M3D is used as a dose calculation check, Varian Portal Dosimetry or MFX is used for delivery QA. Prior to the sole use of Varian Portal Dosimetry or MFX, extensive validation was performed against existing phantom based QA systems.

M3D and MFX are now being used for patient specific QA on both Ethos systems. Additional machine-based QA is also performed on a monthly basis and consists of the delivery and QA of a cohort of standard plans to verify both the machine functionality and delivery as well as the QA system functionality and calculation.

**iv.** Barking Havering and Redbridge University Hospitals NHS Trust, Romford, UK

Queen's Hospital is an all-Varian department with three linear accelerators: one EDGE, one Halcyon v2.0 and one Ethos v1.0, which was upgraded from a Halcyon v2.0 in August 2020. We started using Ethos with oART for prostate cases, with 7-field IMRT. To date, we have treated 13 oART cases on Ethos, including bladder and abdomen.

Prior to the Ethos upgrade, we used Portal Dosimetry, IBA Compass and ArcCHECK for PSQA on all linacs. The PSQA for EDGE treatment plans is calculation-based using the IBA compass system. For the Halcyon all the plans are measured using Portal Dosimetry. Full machine-specific QA including DMLC QA is performed on the machine to monitor consistency. We measure plans-of-the-month by selecting different treatment sites for patients treated on all machines. The centre is also equipped with Sun Nuclear DoseCHECK and PerFRACTION, the latter being commissioned for transit dosimetry for EDGE.

M3D v3.1 was commissioned in conjunction with Ethos as an independent dose calculation QA system. During the clinical implementation of Ethos our QA approach was two-fold:
i) The treatment plan generated by Ethos TPS is exported to M3D for dose distribution evaluation in the patient anatomical dataset. If the plan meets local passing criteria, then a verification plan is created on the ArcCHECK phantom to evaluate the plan deliverability. The trajectory log-files generated by Ethos during measurement-based QA are sent to Mobius3D for the calculation and evaluation of the dose distribution by MFX.
ii) For the on-couch adaptation plan verification process, the structures of the day, the adapted plan and the scheduled plan are sent to Mobius3D. MobiusAdapt automatically performs a plan check using the criteria set for the reference plan. The log files are then received by Mobius3D after treatment is completed, to be evaluated by the MFX

Measurement-based QA has become a cumbersome task with increasing numbers of oART and IGRT patients treated on Ethos. In order to streamline the QA process and eliminate measurement-based QA using ArcCHECK, we have set up a class solution such that pre-treatment QA of the treatment plans is performed using ArcCHECK only for that plans failed to meet local passing tolerances. These tolerances were derived using the Bland-Altman statistical method to evaluate the results obtained from Mobius3D and MobiusFX, using ArcCHECK as our reference PSQA device for VMAT and IMRT.

**v.** University of Alabama, Birmingham, USA

The Department of Radiation Oncology at the University of Alabama at Birmingham (UAB) is a part of the O'Neal Cancer Center at UAB, which is a National Cancer Institute designated Comprehensive Cancer Center. The Department of Radiation Oncology uses Varian treatment planning, delivery, and information systems, having 4 linear accelerators at its central clinic. Installation of Ethos is underway, with pelvis as the initial site for use of oART.

For PSQA of modulated plans, UAB uses ArcCHECK, Delta4, ionization chamber measurements, and Portal Dosimetry. They are transitioning to exclusive use of Portal Dosimetry, with supplementary ionization chamber measurements of randomly selected plans for process control. For Ethos, they will use ArcCheck and ionization chamber measurements.

Mobius3D is used at UAB for pre-treatment independent dose calculation. The Mobius3D system was commissioned and validated by comparing point dose measurements for simple fields and representative modulated plans with M3D calculations. For Ethos, simple fields, the test suite described in the report of Task Group 119, and simulated patient plans will be used for commissioning and validation of M3D. Ionization chamber measurements collected as part of PSQA will be routinely compared with Mobius3D for ongoing process control of both the Ethos treatment planning system and Mobius3D.

**vi.** Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO, USA

The team at Washington University in St. Louis commissioned the first ETHOS system in the United States of America. Since the first treatment in 03/2020, more than 150 cancer patients received radiotherapy treatment with more than 2300 fractions successfully delivered. There are more than 10 patients with 50+ fractions delivered using online adaptive radiotherapy (oART). The majority of the adaptive radiotherapy treatment were stereotactic radiotherapy (SBRT) with hypofractionations. Prior to the implementation of ETHOS system, measurement-based QA is the standard for any IMRT or VMAT plan verification on other linacs. Portal dosimetry is the primary QA approach for regular IMRT or VAMT plan and the film and ionization chamber measurements are often performed for SBRT cases.

The Mobius3D system was new to the WashU team. Three physicists were involved in the initial commissioning and validation process with one of them being assigned as the contact person for any on-going issues and periodic QA. The Mobius3D beam model is built using the pre-configured Ethos beam data. Simple photon calculations for reference conditions, large fields, and small fields were performed. Calculations were performed at a variety of SSDs, depths, and off axis positions and compared against the treatment planning system. A customizable factor, DLG, was determined by calculating 32 IMRT and VMAT plans generated in the Ethos Treatment planning system with several DLG values and analyzing the 3D gamma passing rate. The online adaptive module (MobiusAdapt) was tested by running simulated online adaptive plan on phantom. After the comprehensive test, the system was released for clinical use. Gamma passing rate of 90% with criteria 3%/2mm was used as the threshold based on TG218 recommendations and internal QA guideline.

Currently, for all ETHOS plans, the pretreatment QA are performed with both measurement-based QA using Octavious3D (PTW) and calculation-based QA in Mobius3D. This allows the team to gain more experience with Mobius3D system and to collect more data points for comparison. The online plan specific QA for oART of each fraction is performed only with Mobius3D module.

## Appendix B. Failure modes and effects analysis scoring

*Table B1. Occurrence (O), severity (S) and detectability (D), scored based on TG100 recommendations, and resulting Risk Priority Numbers (RPN) for the critical steps and their related potential failure modes, causes and effects within step (A) Data transfer and storage, of the oART patient-specific, calculation-based QA process using Mobius3D. The RPN were calculated for each failure mode as RPN=O×S×D, with O, S and D ranging from 1 to 10.*

| Steps | Potential Failure Modes | Potential Causes of Failure | Potential Effects of Failure | *O* | *S* | *D* | *RPN* |
|---|---|---|---|---|---|---|---|
| Plan transfer to Mobius3D | Plan failed to transfer | Network issues | Treatment Delay | 2 | 3 | 2 | 12 |
| | | | Treatment cancellation | 2 | 4 | 2 | 16 |
| | | Human configuration errors | Treatment Delay | 1 | 3 | 3 | 9 |
| | | | Treatment cancellation | 1 | 4 | 3 | 12 |
| | | Software malfunction | Treatment delay | 6 | 3 | 2 | 36 |
| | | | Treatment cancellation | 6 | 4 | 2 | 48 |
| | Wrong plan or incomplete plan being sent | Software malfunction | Treatment delay | 1 | 4 | 7 | 28 |
| | | | QA with incomplete data - suboptimal analysis | 1 | 4 | 6 | 24 |
| | | | Toxicity and underdose | 1 | 8 | 6 | 48 |
| | | Network issues | Treatment delay | 1 | 2 | 6 | 12 |
| | | | QA with incomplete data - suboptimal analysis | 1 | 4 | 6 | 24 |
| | | | Toxicity and underdose | 1 | 8 | 6 | 48 |
| | | Data lost or corrupted | Treatment delay | 1 | 2 | 6 | 12 |
| | | | Incomplete data - suboptimal analysis | 1 | 4 | 6 | 24 |
| | | | Toxicity and underdose | 1 | 8 | 6 | 48 |

*Table B2. Occurrence (O), severity (S) and detectability (D), scored based on TG100 recommendations, and resulting Risk Priority Numbers (RPN) for the critical steps and their related potential failure modes, causes and effects within step (B) Mobius3D calculation, of the oART patient-specific, calculation-based QA process using Mobius3D. The RPN were calculated for each failure mode as RPN=O×S×D, with O, S and D ranging from 1 to 10.*

| Step | Potential Failure Modes | Potential Causes of Failure | Potential Effects of Failure | *O* | *S* | *D* | *RPN* |
|---|---|---|---|---|---|---|---|
| Dose calculation | Failed to calculate dose | Software malfunction | Treatment delay | 1 | 3 | 1 | 3 |
| | | | Treatment cancellation | 1 | 4 | 1 | 4 |
| | Wrong or insufficient dose calculation due to wrong or inappropriate imaging used | Human errors | Treatment delay | 3 | 3 | 4 | 36 |
| | | | Treatment cancellation | 3 | 4 | 4 | 48 |
| | | | Toxicity and underdose | 3 | 6 | 4 | 72 |
| | Wrong or insufficient dose calculation due to wrong or inappropriate delineation | Human errors | Treatment delay | 2 | 3 | 5 | 30 |
| | | | Toxicity and underdose | 2 | 6 | 5 | 60 |
| | Wrong or insufficient dose calculation due to wrong or inappropriate physical property assignment | Human errors | Treatment delay | 2 | 3 | 8 | 48 |
| | | | Toxicity and underdose | 2 | 6 | 8 | 96 |
| | Wrong or insufficient dose calculation due to wrong or inappropriate plan design | Human errors | Treatment delay | 1 | 6 | 8 | 48 |
| | | | Toxicity and underdose | 1 | 6 | 8 | 48 |
| | Wrong or insufficient dose calculation | Software malfunction | Treatment delay | 1 | 3 | 8 | 24 |
| | | | Toxicity and underdose | 1 | 6 | 8 | 48 |

*Table B3. Occurrence (O), severity (S) and detectability (D), scored based on TG100 recommendations, and resulting Risk Priority Numbers (RPN) for the critical steps and their related potential failure modes, causes and effects within step (C) Analysis and display, of the oART patient-specific, calculation-based QA process using Mobius3D. The RPN were calculated for each failure mode as RPN=O×S×D, with O, S and D ranging from 1 to 10.*

| Steps | Potential Failure Modes | Potential Causes of Failure | Potential Effects of Failure | *O* | *S* | *D* | *RPN* |
|---|---|---|---|---|---|---|---|
| DVH analysis and display | Failed to calculate DVH | Software malfunction | Treatment delay | 2 | 3 | 2 | 12 |
| | | | Treatment cancellation | 2 | 3 | 2 | 12 |
| | Wrong DVH calculations or display | Software malfunction | Treatment delay | 2 | 3 | 10 | 60 |
| | | | Treatment cancellation | 2 | 3 | 10 | 60 |
| | | | Toxicity and underdose | 2 | 6 | 10 | 120 |
| | DVH not reviewed or reviewed incorrectly | Human errors | Treatment delay | 2 | 3 | 3 | 18 |
| | | | Toxicity and underdose | 2 | 6 | 3 | 36 |
| Target coverage analysis and display | Failed to calculated Target coverage | Software malfunction | Treatment delay | 2 | 3 | 2 | 12 |
| | | | Treatment cancellation | 2 | 3 | 2 | 12 |
| | Wrong coverage calculations or display | Software malfunction | Treatment delay | 2 | 3 | 10 | 60 |
| | | | Treatment cancellation | 2 | 3 | 10 | 60 |
| | | | Toxicity and underdose | 2 | 6 | 10 | 120 |
| | Target Coverage not reviewed or reviewed incorrectly | Human errors | Treatment delay | 2 | 3 | 3 | 18 |
| | | | Toxicity and underdose | 2 | 6 | 3 | 36 |
| DVH limits calculation and display | Failed to calculate DVH limits | Software malfunction | Treatment delay | 2 | 3 | 2 | 12 |
| | | | Treatment cancellation | 2 | 3 | 2 | 12 |
| | Wrong DVH limits calculations or display | Software malfunction | Treatment delay | 2 | 3 | 10 | 60 |
| | | | Treatment cancellation | 2 | 3 | 10 | 60 |
| | | | Toxicity and underdose | 2 | 6 | 10 | 120 |
| | | Human errors (DVH limits set wrong) | Treatment delay | 3 | 3 | 4 | 36 |
| | | | Toxicity and underdose | 3 | 6 | 4 | 72 |
| | DVH limits not reviewed or reviewed incorrectly | Human errors | Treatment delay | 2 | 3 | 8 | 48 |
| | | | Toxicity and underdose | 2 | 6 | 8 | 96 |
| 3D gamma calculation and display | Failed to calculate 3D gamma | Software malfunction | Treatment delay | 2 | 6 | 2 | 24 |
| | | | Treatment cancellation | 2 | 2 | 2 | 8 |
| | Wrong 3D gamma calculations or display | Software malfunction | Treatment delay | 2 | 2 | 3 | 12 |
| | | | Treatment cancellation | 2 | 2 | 3 | 12 |
| | | | Toxicity and underdose | 2 | 6 | 3 | 36 |
| | | Human configuration errors (wrong tolerance, criteria) | Treatment delay | 2 | 2 | 8 | 32 |
| | | | Treatment cancellation | 2 | 2 | 8 | 32 |
| | | | Toxicity and underdose | 2 | 6 | 8 | 96 |
| | Gamma not reviewed or reviewed incorrectly | Human errors | Treatment delay | 3 | 2 | 2 | 12 |
| | | | Toxicity and underdose | 3 | 6 | 2 | 36 |
| Plan information display | Failed to grab plan information | Software malfunction | Treatment delay | 3 | 5 | 2 | 30 |
| | Wrong plan information display | Software malfunction | Treatment delay | 2 | 5 | 4 | 40 |
| | | | Treatment cancellation | 2 | 5 | 4 | 40 |
| | Plan info not reviewed or reviewed incorrectly | Human errors | Treatment delay | 2 | 3 | 4 | 24 |
| | | | Toxicity and underdose | 2 | 6 | 4 | 48 |

*Table B4. Occurrence (O), severity (S) and detectability (D), scored based on TG100 recommendations, and resulting Risk Priority Numbers (RPN) for the critical steps and their related potential failure modes, causes and effects within step (D) Data storage and record, of the oART patient-specific, calculation-based QA process using Mobius3D. The RPN were calculated for each failure mode as RPN=O×S×D, with O, S and D ranging from 1 to 10.*

| Steps | Potential Failure Modes | Potential Causes of Failure | Potential Effects of Failure | *O* | *S* | *D* | *RPN* |
|---|---|---|---|---|---|---|---|
| Data Storage and Treatment record | Stored data cannot be retrieved completely or correctly | Software malfunction | Treatment delay | 1 | 2 | 3 | 6 |
| | | | Treatment based on incomplete information | 1 | 6 | 3 | 18 |
| | | | Treatment cancellation | 1 | 2 | 3 | 6 |
| | Data cannot be stored | Software malfunction | Treatment delay | 1 | 2 | 3 | 6 |
| | | | Treatment based on outdated information or incomplete information | 1 | 2 | 3 | 6 |
| | | | Treatment cancellation | 1 | 2 | 3 | 6 |
| | | | Treatment delay | 1 | 2 | 3 | 6 |
| | Data corrupted or lost | Software malfunction | Treatment delay | 1 | 2 | 3 | 6 |
| | | | QA with incomplete data - suboptimal analysis | 1 | 2 | 3 | 6 |
| | | | Toxicity and underdose | 1 | 6 | 3 | 18 |
| | | Human Errors | Treatment delay | 1 | 2 | 8 | 16 |
| | | | QA with incomplete data - suboptimal analysis | 1 | 2 | 8 | 16 |
| | | | Toxicity and underdose | 1 | 6 | 8 | 48 |

*Table B5. Occurrence (O), severity (S) and detectability (D), scored based on TG100 recommendations, and resulting Risk Priority Numbers (RPN) for the critical steps and their related potential failure modes, causes and effects within step (E) Delivery QA, of the oART patient-specific, calculation-based QA process using Mobius3D. The RPN were calculated for each failure mode as RPN=O×S×D, with O, S and D ranging from 1 to 10.*

| Steps | Potential Failure Modes | Potential Causes of Failure | Potential Effects of Failure | *O* | *S* | *D* | *RPN* |
|---|---|---|---|---|---|---|---|
| Log file and treatment record transfer to Mobius3D | Treatment record failed to transfer | Network issues | Treatment delay | 2 | 3 | 2 | 12 |
| | | | Treatment cancellation | 2 | 4 | 2 | 16 |
| | | Human configuration errors | Treatment delay | 1 | 3 | 3 | 9 |
| | | | Treatment cancellation | 1 | 4 | 3 | 12 |
| | | Software malfunction | Treatment delay | 2 | 3 | 2 | 12 |
| | | | Treatment cancellation | 2 | 4 | 2 | 16 |
| Log file associated to patient plan | Wrong or incomplete treatment record are being associate with patient plan | Software malfunction | Treatment delay | 1 | 4 | 7 | 28 |
| | | | QA with incomplete data--suboptimal analysis | 1 | 4 | 6 | 24 |
| | | | Toxicity and underdose | 1 | 8 | 6 | 48 |
| | | Network issues | Treatment delay | 1 | 2 | 6 | 12 |
| | | | QA with incomplete data--suboptimal analysis | 1 | 4 | 6 | 24 |
| | | | Toxicity and underdose | 1 | 8 | 6 | 48 |
| | | Data corrupted or lost | Treatment delay | 1 | 2 | 6 | 12 |
| | | | QA with incomplete data--suboptimal analysis | 1 | 4 | 6 | 24 |
| | | | Toxicity and underdose | 1 | 8 | 6 | 48 |
| Dose calculation based on log file | Failed to calculate dose | Software malfunction | Treatment delay | 1 | 3 | 1 | 3 |
| | | | Treatment cancellation | 1 | 3 | 1 | 3 |
| | | | Toxicity and underdose | 1 | 6 | 8 | 48 |
| | Wrong or insufficient dose calculation | Software malfunction | Treatment delay | 1 | 3 | 8 | 24 |
| | | | Toxicity and underdose | 1 | 6 | 8 | 48 |
| DVH, Gamma, RMS calculations and display | Failed to calculate DVH, Gamma, RMS, D&V limits | Software malfunction | Treatment delay | 1 | 3 | 2 | 6 |
| | | | Treatment cancellation | 1 | 3 | 2 | 6 |
| | Wrong DVH, Gamma, RMS, D&V limits calculations or display | Software malfunction | Treatment delay | 1 | 3 | 10 | 30 |
| | | | Treatment cancellation | 1 | 3 | 10 | 30 |
| | | | Toxicity and underdose | 1 | 6 | 10 | 60 |