

Evaluation of automated pre-treatment and transit in-vivo dosimetry in radiotherapy using empirically determined parameters

Evy Bossuyt^{a,*}, Reinhilde Weytjens^a, Daan Nevens^{a,b}, Sarah De Vos^a, Dirk Verellen^{a,b}

^a Iridium Kankernetwerk, Radiation Oncology Department, Antwerp, Belgium

^b Faculty of Medicine and Health Sciences, Antwerp University, Antwerp, Belgium

ARTICLE INFO

Keywords:
In-vivo
Transit dosimetry

ABSTRACT

Background and purpose: First reports on clinical use of commercially automated systems for Electronic Portal Imaging Device (EPID)-based dosimetry in radiotherapy showed the capability to detect important changes in patient setup, anatomy and external device position. For this study, results for more than 3000 patients, for both pre-treatment verification and in-vivo transit dosimetry were analyzed.

Materials and methods: For all Volumetric Modulated Arc Therapy (VMAT) plans, pre-treatment quality assurance (QA) with EPID images was performed. In-vivo dosimetry using transit EPID images was analyzed, including causes and actions for failed fractions for all patients receiving photon treatment (2018–2019). In total 3136 and 32,632 fractions were analyzed with pre-treatment and transit images respectively. Parameters for gamma analysis were empirically determined, balancing the rate between detection of clinically relevant problems and the number of false positive results.

Results: Pre-treatment and in-vivo results depended on machine type. Causes for failed in-vivo analysis included deviations in patient positioning (32%) and anatomy change (28%). In addition, errors in planning, imaging, treatment delivery, simulation, breath hold and with immobilization devices were detected. Actions for failed fractions were mostly to repeat the measurement while taking extra care in positioning (54%) and to intensify imaging procedures (14%). Four percent initiated plan adjustments, showing the potential of the system as a basis for adaptive planning.

Conclusions: EPID-based pre-treatment and in-vivo transit dosimetry using a commercially available automated system efficiently revealed a wide variety of deviations and showed potential to serve as a basis for adaptive planning.

1. Introduction

In-vivo dosimetry is recommended in radiotherapy to avoid major treatment errors and to improve accuracy [1–12]. In clinical routine, placing the traditional point detectors on the patient's skin is not always feasible and requires additional setup time. Moreover, it is often not possible to use these in combination with Intensity Modulated Radiation Therapy (IMRT) or Volumetric Modulated Arc Therapy (VMAT) for which verification is limited to pre-treatment solutions. Since Electronic Portal Imaging Device (EPID) images contain dose information, many groups have investigated their use for radiotherapy dose measurement [12–22]. Pre-treatment dose verification and in-vivo dosimetry using EPIDs has become routine practice in a growing number of clinics in recent years. Mainstream acceptance, however, has been hampered by a

lack of commercially available solutions. Fully automated EPID-based systems are now emerging making it feasible to perform dosimetric quality assurance (QA) on every field for every patient. For a comprehensive literature review on electronic portal imaging for radiotherapy dosimetry the reader is referred to van Elmpt et al. [23].

Olch and colleagues [22] published the first report of the clinical use of a commercially available automated system for EPID-based dosimetry. It presents results for 855 fractions using the first treatment fraction as baseline to compare with later fractions, showing the system's capability to detect important changes in patient setup, anatomy and external device position. We present results for more than 30,000 fractions, of which most are based on absolute dose prediction (as opposed to comparing to baseline), investigating if this broadens the variety of detected deviations.

* Corresponding author.

E-mail address: Evy.Bossuyt@gza.be (E. Bossuyt).

<https://doi.org/10.1016/j.phro.2020.09.011>

Received 27 February 2020; Received in revised form 7 August 2020; Accepted 23 September 2020

Available online 9 November 2020

2405-6316/© 2020 The Authors. Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the

CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1
Summary of empirically determined parameters for gamma analysis of in-vivo transit dosimetry results.

	Normalization (Local/ Global)	Dose Difference Tolerance (%)	Distance Tolerance (mm)	Low Dose Threshold (%)	Passing Tolerance Level (%)
Breast	Local	7	6	20	90
Whole Brain RadioTherapy	Local	7	3	20	90
Palliative treatments	Local	7	5	20	93
H&N and Brain	Global	3	3	20	95
Rectum	Global	5	5	20	93
Other treatment sites with mask	Global	5	3	20	95
Other treatment sites without mask (including lung, pelvis, abdomen,...)	Global	5	5	20	95
Stereotactic 1 mm	Local	10	1	20	95
Stereotactic 2 mm	Local	10	2	20	95
Stereotactic 3 mm	Local	10	3	20	95

There is currently no consensus on which parameters for gamma analysis are most appropriate for specific treatment sites or equipment used [12,18,22,23]. A common set of parameters is not feasible for clinical practice. This study investigates the possibility to determine parameters empirically, balancing the rate between detection of clinically relevant problems and the number of false positive results.

2. Materials and methods

The study was conducted in the Iridium Kankernetwerk, Belgium with ten linear accelerators: three ‘old generation’ machines: Varian Clinacs; and seven ‘new generation’ machines: six TrueBeams and one TrueBeam STX (Varian Oncology Systems, Palo Alto, CA). Two treatment planning systems were used to create plans: RayStation 9A (RaySearch Laboratories, Stockholm, Sweden) for stereotactic plans (both intra and extra-cranial) and Eclipse v13.6 (Varian Medical Systems, Palo Alto, CA) for all other types of plans. An automated web-based system was installed early 2017, for both pre-treatment and in-vivo QA based on EPID measurements (PerFRACTION™, part of SunCHECK™, Sun Nuclear Corporation). Clinical use of the software (version 1.7) started in October 2017 on two of the machines, gradually including all other machines until in February 2018 the software was used for all machines and all patients (version 2.0 in May 2018, version 2.2 in November 2019), pushing DICOM data to the server, actively retrieving images, and automatically calculating results in the background. Failing fractions were checked daily by the responsible physicists and physicians, and appropriate actions were taken. For this study, calculated results and comments on causes and actions for failed fractions were extracted from the system using structured query language (SQL) scripts. A clinical validation of the efficiency and performance on detection of errors was carried out, along with an analysis of causes and actions taken for failed fractions.

Pre-treatment QA (‘Fraction 0’) was performed for all VMAT plans. Preparation consisted of exporting plans to PerFRACTION, assigning a ‘QA template’ containing (treatment site specific) tolerance levels, and scheduling integrated images. All images were automatically retrieved after measurement and compared with calculated predicted dose images using a Global Gamma analysis of 3%/2mm, a low dose threshold of 10% and a passing tolerance level of 98%. All treatment plans were supervised by a physicist, including the results of the ‘Fraction 0’ measurement. If the automatic verification failed, plan and results were checked in detail and re-calculated with 3%/3mm and/or 4%/2mm. If still failing, results were discussed with a physician. In total 3136 fractions from 2686 patients were analyzed.

Transit EPID dosimetry with integrated images was performed in 43% of all treatment fractions: as standard, for the first three days of treatment and weekly thereafter (or more if fraction results failed); daily in the case of short treatment schedules. All causes of failed fractions were analyzed. If failure was due to more than one reason, the cause with the greatest contribution to the failure was assigned. In the start-up period between October 2017 and August 2018, in-vivo results were calculated using ‘relative’ (image-to-image) comparison, using one of the first acquired images as a reference (baseline). After this period, a first analysis was performed in which 8621 fractions with transit EPID measurements were included [24]. In September 2018 ‘absolute’ verification was introduced allowing comparison of images to calculated predicted dose. Breasts with nodal irradiation remained being calculated relative to a baseline image, because a longitudinal panel shift was needed for analysis of the entire breast and predicted dose could not be calculated with these panel shifts. A second analysis was performed for patients treated between September 2018 and August 2019 with in total 24,011 fractions from 3671 patients using transit EPID measurements. The latter will be the main subject of this report.

During the start-up period we tried to determine ‘appropriate’

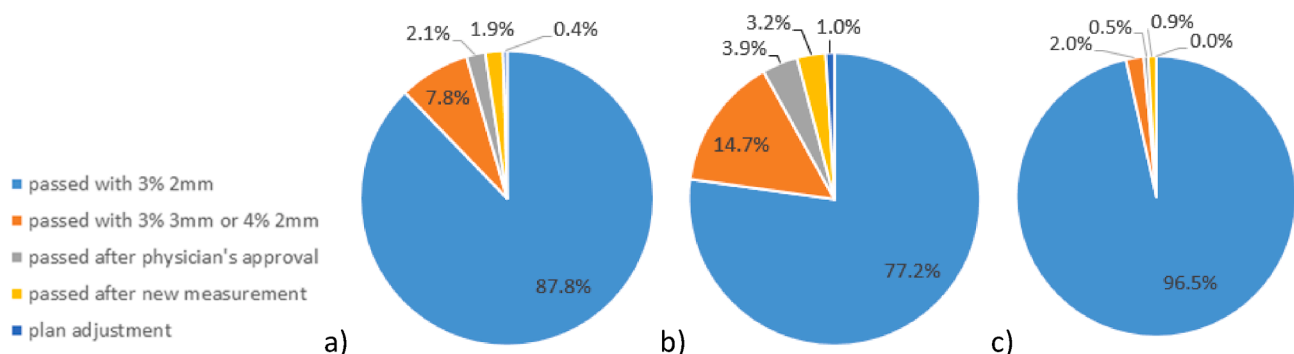


Fig. 1. Pre-treatment QA results: a) All results, b) Only old generation machine type results, c) Only new generation machine type results.

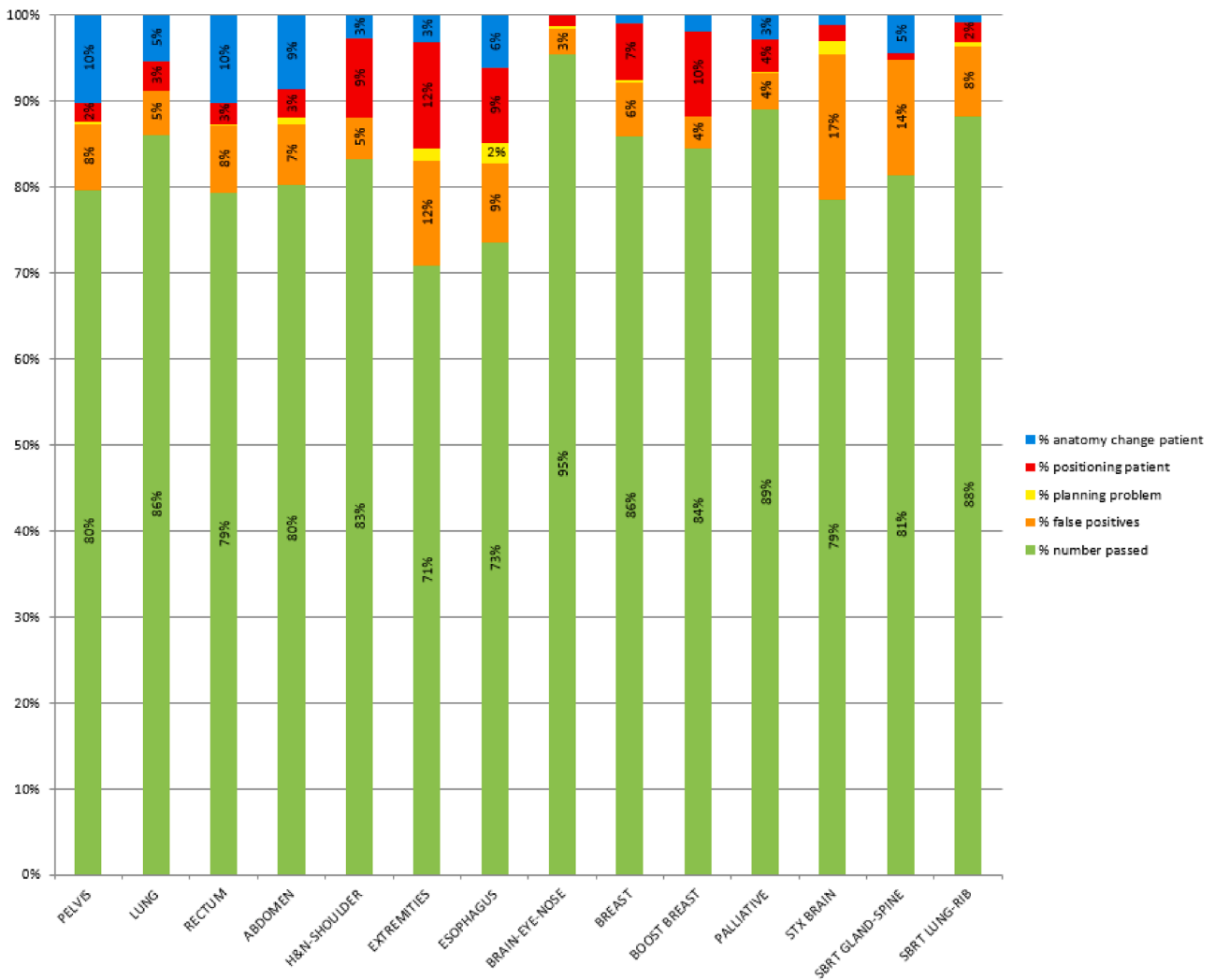


Fig. 2. Analysis of measured in-vivo fractions, classified per treatment site and/or technique, indicating the number of passed fractions, the number of false positives and the number of failed fractions due to planning problems, deviations in patient positioning and changes in patient anatomy.

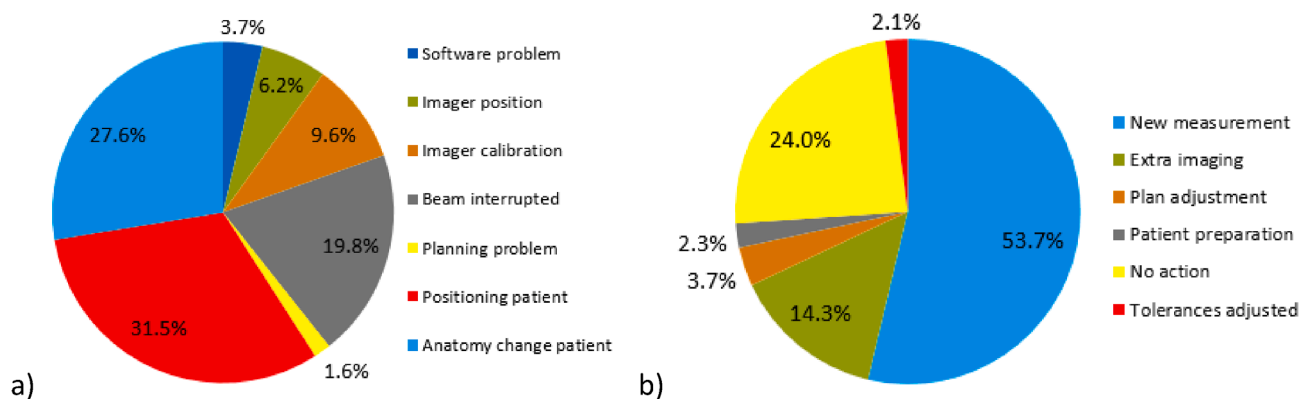


Fig. 3. Causes and actions for failed fractions in the second analysis: a) Causes: software problems including software bugs and irradiation on another (matched) machine with a different imager causing a deviation when comparing to a baseline of another machine; wrong imager position; problems with imager calibration; interrupted beams causing missing dose in the image; planning problems; deviations in patient position; changes in patient anatomy. b) Actions: taking a new measurement; adding extra pre-treatment imaging; plan adjustment; taking measures regarding patient preparation (e.g. bladder and rectum protocol); taking no action; adjusting the used tolerances.

parameters for gamma analysis, balancing the rate between detection of clinically relevant problems and the number of false positive results. Using the ‘QA templates’ in the software, ‘standard’ parameters for gamma analysis were assigned. We chose to create templates depending on the treatment site, as the treatment site largely determines the used immobilization and treatment technique, and acceptable tolerance. We ended up with seven different sets of analysis parameters for non-stereotactic plans and three for stereotactic plans, summarized in Table 1.

A Global Gamma Analysis with 3% dose difference tolerance, 3 mm distance tolerance, 10% low dose threshold and 95% passing tolerance level was used at the start, using one of the first acquired images as baseline. However, these levels turned out to be too strict for most plans with an excessive number of false positives. The challenge is to detect large discrepancies that could have clinical consequences as opposed to minor issues and therefore find a good balance between the detection of clinically relevant problems without increasing the rate of both false negatives, or false positives.

It was decided to use the Planning Target Volume (PTV) margin as distance tolerance, a patient shift within the PTV margin being considered as clinically acceptable. The dose difference tolerance was increased to 5% for most treatments except Brain and H&N, because the variation in patient thickness produced failed fractions that were clinically irrelevant. It was increased to 7% for breast, Whole Brain Radiation Therapy (WBRT) and palliative treatments and to 10% for stereotactic treatments (the very small fields caused acceptable large dose differences in the high dose regions). The low dose threshold was increased to 20%, as the 10% threshold produced fractions failing in low dose regions close to field edges (considered clinically less relevant). A Global Gamma Analysis was applied for VMAT plans and a Local Gamma Analysis for 3D plans and stereotactic plans. This was based on reports

showing that global analysis masks some problems, but also that local analysis magnifies irrelevant errors at low doses [25]. Since integrated images for VMAT plans consist of a summation of dose from all angles, they are less sensitive to geometric errors and local analysis tends to magnify irrelevant errors for large VMAT plans. The passing tolerance levels were set at 95% except for rectum and palliative treatments (93%) and for breast and WBRT treatments (90%). Due to the skin flash, a small shift in breast and WBRT treatments causes a lot of points to fail and 95% was considered too strict. In rectum treatments 93% provided a value that seemed to balance the detection of clinically relevant air bubbles against irrelevant ones.

3. Results

Pre-treatment QA proved to be time-efficient, requiring five to ten minutes per plan, including preparation, measurement and verification. Fig. 1 shows results for pre-treatment QA, with 88% of fractions passing the automatically calculated gamma analysis of 3%/2mm with a 98% tolerance level (Fig. 1a). In 8% of cases, results were scored as ‘acceptable’ by a physicist with a gamma analysis of 4%/2mm or 3%/3mm. In 2% of cases, results were determined to be acceptable after discussion with a physician. In less than 1% of cases, results were unacceptable and the plan was adjusted. A difference was observed between results for different machine types, Fig. 1b and 1c respectively. On the new generation machines, 97% of measurements passed with the standard 3%/2mm analysis, compared to 77% on the old generation machines.

Of the 8621 fractions of the first in-vivo analysis after the start-up period, 1523 failed (18%), of which 7% were false positives and 11% caused by patient related issues. Of the 24,011 fractions of the second analysis, 3766 failed (16%), of which 6% were false positives and 10% caused by patient related issues. Of these 6% false positives, more than

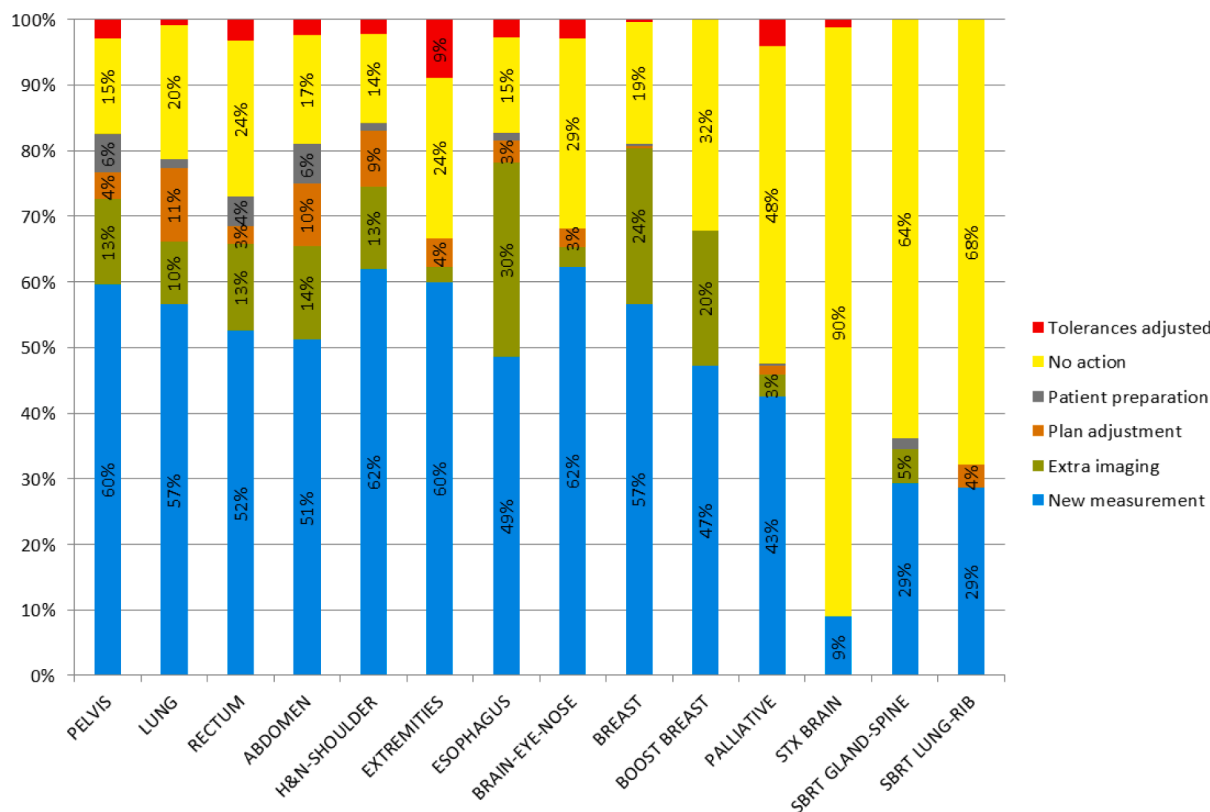


Fig. 4. Analysis of the actions for failed fractions in the second analysis classified per treatment site and/or technique: taking a new measurement; adding extra pre-treatment imaging; plan adjustment; taking measures regarding patient preparation (e.g. bladder and rectum protocol); taking no action; adjusting used tolerances.

4% were measurements performed on old generation machines, less than 2% on new generation machines. Fig. 2 shows an analysis of these measured fractions, classified per treatment site and/or technique.

Fig. 3 shows all causes and actions of the 3766 failed fractions of the second analysis in detail: false positives were mostly caused by interrupted beams (20%) and imager calibration issues (10%). Patient related failed fractions were in 32% of cases caused by a deviation in patient positioning and in 28% of cases by anatomy changes such as weight loss, deviations in bladder or rectal filling and tumor shrinkage. Actions most commonly taken were: repeating the measurement while taking extra care in positioning, breath hold technique, shoulder or arm position (54%) or adding extra imaging to obtain more information and to help differentiating between true and false positives (14%). In 24% of cases no extra actions were taken (e.g. acceptable for palliative cases or last fractions of treatment). In 4% of cases, plan adjustments were made. Fig. 4 shows all actions for failed fractions classified per treatment site and/or technique.

Examples of a wide variety of detected errors and deviations, together with the appropriate corrective actions, are illustrated in Appendix A, including decision charts for taking actions when a measurement is out of tolerance (Fig. A.1). Several of the discovered errors would have led to a dose difference of the total treatment of more than 5% if uncorrected. Some of them are also discussed in the Appendix. Examples of what could be detected when comparing to a baseline include: patient position (Fig. A.2), breast swelling (Fig. A.3), pneumonia (Fig. A.4), planning problems where the skin flash tool was not well used (Fig. A.5), unintended table shift (Fig. A.6) and tumor reduction (Figs. A.7 and A.8). More errors could be caught when comparing to predicted dose, including planning issues (Fig. A.9), problems at simulation (Figs. A.10 to A.12), problems with patient preparation (Fig. A.13), errors during imaging and treatment (Fig. A.14), weight loss at the start of treatment (Fig. A.15), issues related to the breath hold technique (Fig. A.16) and problems with immobilization devices (Fig. A.17).

4. Discussion

This study presents clinical results for more than 3000 patients, for EPID-based pre-treatment and in-vivo transit dosimetry. Analysis with an empirically determined set of parameters shows a wide variety of detected errors and deviations.

Our measurements showed pre-treatment QA to be time-efficient (five to ten minutes) and revealing some deviations, with results differing between machine types. Extra verifications with Arc CHECK and point measurements showed most of the deviations to be acceptable and causes for failure mostly related to technical problems with both machines and imagers (e.g. the imager position of the aS1000 imagers not being corrected mechanically during VMAT and hence less stable than the aS1200 imagers). In addition, the more complex plans showed better results at new generation machines. It is known from log file analysis that MLC positioning errors substantially differ between machine types [26,27]. Omitting pre-treatment verification for non-hypofractionated plans seems feasible, since in-vivo transit dosimetry comparing to predicted dose is performed, which would also indicate problems with patient plans or machine failures. However, if the patient has failing in-vivo fractions, it would require an extra ‘Fraction 0’ measurement to eliminate plan issues as the cause of failure. This would introduce an extra workflow and more manual work compared to performing the measurement for every VMAT plan before treatment as a standard procedure. In the future it might be possible to predict ‘difficult’ plans using complexity thresholds or Machine Learning. Properly generating and cleaning the large amount of current data will open the door for Machine Learning tools to filter out machine and specific issues before the actual QA is needed.

Tolerance levels for in-vivo dosimetry were empirically determined in the start-up period, using an AMARA-principle: detecting errors, but only As Many As Reasonably Achievable, taking into account economic and societal factors. Economic factors include costs of in-vivo systems and time spent on measuring and analyzing results, including false positives. Societal factors include patient comfort, e.g. implanting detectors or placing detectors on the skin reduce patient comfort compared to deploying an imager during treatment. Finding a good balance will be different for every in-vivo system and every department. Comparing clinical results from multi-center collaborations might help to determine a set of generally accepted recommendations.

An AMARA-principle can be based on a few pillars. The first is knowing the sensitivity, strengths and weaknesses of the system. This can influence decisions on the tolerance levels to be used, but also on decisions such as the level of image guidance to be used; e.g. in-vivo EPID dosimetry is quite sensitive to a change in patient thickness [28], but the sensitivity to patient position variations is dependent on treatment site [18]: in anatomical regions with few density differences like abdominal regions, a shift in patient’s position will have less effect on the transmitted radiation than in Head & Neck (H&N) regions. Also, sensitivity can depend on technique: e.g. VMAT could be less sensitive to detect patient position variations compared to IMRT, which in turn could be less sensitive than 3D conformal radiotherapy [19].

Secondly, reducing the number of false positives and introducing an automated probability score for being false positive, might make it less likely a real error will be classified as a false positive due to the observer becoming ‘error fatigued’. As a first step, a script has been introduced to identify false positives related to interrupted beams or incorrect positioning of the imager: a report is generated showing the number of detected Monitor Units in the image and the position of the imager compared with the planned values. The latter will be extended to produce suggestions for possible causes and actions in the near future.

Finally, re-evaluating tolerance levels on a regular base, especially after introducing software or hardware changes or introducing new immobilization devices, imaging or treatment strategies, is necessary to re-assess if parameters can be made tighter to diminish the number of false negatives without adding false positives or clinically irrelevant errors. Software upgrades in which shifted imager positions can be taken into account or replacement of equipment, might also influence decisions on used tolerance levels.

The latter is illustrated in the adjustments made to tolerance levels during the startup period to decrease the number of false positives (Table 1). Using these tolerance levels, a level of 16% failed fractions was reached, of which 6% were false positives and 10% were caused by patient related issues. This was judged to be an acceptable balance between the detection of clinically relevant problems and the number of false positive results. In the report of Olch et al. [22], tighter tolerance levels were used and the 3%/3mm analysis only produced 8% failed fractions. However, the study reports only 57 patients, all being treated on a new generation machine (of which we also reported less failed fractions) and strongly immobilized, probably producing a higher degree of positioning certainty (close to 3 mm). In our study, results of 3671 patients were analyzed, receiving treatments using various immobilization devices and imaging strategies, supporting the suggestion of determining tolerance levels according to an AMARA-principle allowing variation based on the changes in treatment strategies, desired sensitivity levels and/or machine type.

Evaluating results on a regular basis offers important insights in the quality of treatments and indicates possible items for improvement, or helps departments to decide on future projects. An illustration is given in Fig. 2 where the largest number of positioning issues was related to extremities (12%), breast boost (10%), esophagus (9%) and Head & Neck (H&N) patients (9%). As a result, extra pre-treatment imaging was introduced for breast boosts and more attention was given to proper

positioning and avoiding shoulders in treatment planning for H&N. The largest number of anatomy changes were seen for pelvis (10%), rectum (10%) and abdomen (9%). Earlier reports have shown that these regions might have a lower sensitivity to patient position [18,28]. From Fig. 4 it can be observed that most plan adjustments occurred in lung (11%), abdominal (10%) and H&N treatments (9%), showing the software's potential to serve as a basis for adaptive planning. However, further investigation is necessary to determine the number of false negatives and the limitations of the system. Currently, a study is being conducted to investigate the sensitivity in detecting H&N patients that require an adaptive plan during the course of treatment and to estimate the number of false negatives using current tolerances.

In conclusion, a commercially available automated pre-treatment and in-vivo transit dosimetry system has been clinically implemented for all patients, efficiently revealing a wide variety of deviations using an empirically determined set of parameters for gamma analysis. Results show its potential to serve as a basis for adaptive planning and the number of false positive results dependent on machine type. Tolerance levels should be wisely chosen according to an AMARA-principle, balancing the rate between detection of clinically relevant problems and the number of false positive results.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Iridium Kankernetwerk is a member of the SunCHECK Customer Advisory Board and a Reference center for Sun Nuclear Corporation, but there has been no significant financial support for this work that could have influenced its outcome.

Appendix A

In this appendix examples of detected errors and deviations are illustrated, together with the appropriate corrective actions. In Fig. A.1 decision charts are shown for taking actions when a measurement is out of tolerance. The screenshots from the software in the examples show the delivered dose on the left and the expected dose (baseline or predicted dose) on the right with a blue color for low doses and a red color for high doses. In the middle the gamma comparison is shown. A green color means the gamma comparison is within tolerance levels, an orange color means it is out of tolerance levels. In the profiles a more detailed comparison between delivered and expected dose can be seen.

Fig. A.2 shows a problem with patient positioning. When comparing delivered with expected dose, a shift in breast position can be seen. The best way to distinguish between breast positioning and breast swelling, is to look at the position of the thoracic wall. In this case a shift of the thoracic wall can be seen in the software. Images were matched in the offline review application of the planning system and a shift of more than 1 cm was detected (being more than the distance tolerance level of 6 mm for a breast case). The action in this case was to schedule pre-treatment setup images the next fraction. If the deviation is still present the next day, then 3 days of imaging are performed to define a better standard position.

Fig. A.3 shows a case of a patient with breast swelling. Here the position of the thoracic wall was fine. No direct action was considered necessary as there was sufficient covering and skin flash. Follow-up and a new simulation for the boost was needed.

Fig. A.4 shows the results of a patient who developed a pneumonia. The transit dose was less than the expected dose, because of fluids in the lung. The action was a treatment with antibiotics, after which results normalized again.

Fig. A.5 shows a detected planning problem where the skin flash tool was not used well. The expected dose in this case was the baseline dose from the first fraction. Because the skin flash tool was not properly applied, a small shift in patient position induced a large dose error in

transit dose. The action here was to make a new plan.

Fig. A.6 shows a case where the last field failed. A table shift was applied here to avoid table-gantry collision. Unfortunately, it was not shifted back, causing the last field to be given completely next to the target. Since this was a plan with 2 opposed fields and 10 fractions, this meant half of the dose was given next to the target and this would have meant a 5% under dosage to the target if left undetected. The action was to deliver this field one time extra in the next fraction.

Fig. A.7 shows a patient having a large tumor reduction after only 6 fractions. The action was to make a new plan on a new planning CT.

Fig. A.8 also shows tumor reduction combined with some weight loss. Originally a dose maximum of 106% is planned. After detecting failed fractions and seeing limited tumor reduction and weight loss on CBCT, an evaluation CT was made on which the original plan showed an increase of the dose maximum to 116% and a high increase on the target doses and trachea and spinal cord doses.

Fig. A.9 shows a planning problem where a hip implant was not avoided and showed serious blockage of the dose. A new plan was made with VMAT avoiding the hip implant. Without in-vivo measurements, this would have stayed undetected.

Fig. A.10 shows a case where a very slow breathing induced a CT with artefacts. However, it remained unnoticed that these artefacts actually masked very large movement of the tumor. For this patient a new plan was made with a larger PTV.

Fig. A.11 shows a case where the patient was simulated with the arms down. At planning it was decided to plan with the arms removed from the body contour. However, placing a note in the patient file indicating arms should be crossed on the chest instead of next to the body, was missed. At the machine, patient was installed as seen on the patient photographs and setup instructions from the simulation, resulting in dose difference in the lateral fields. For this first fraction, this resulted in an under dosage of 6%. The action here was to notify the therapists and to adjust the patient file and setup instructions.

Fig. A.12 shows a case where an under dosage is detected. Careful examination showed a shift of 1 cm was necessary for good tumor matching. However, this in turn induced a bad chin and shoulder position, having an effect on the dose. The patient probably swallowed during simulation and a new plan was made on a new CT.

Fig. A.13 shows a patient with air in the rectum. The action here was to instruct the patient for better bowel preparation and provide dietary guidance.

Fig. A.14 shows a case with a perfect online match, so in offline review nothing irregular was observed, but the table was not shifted to the correct values after the match. In this case the patient received an extra fraction with a new plan compensating for the missed area.

Fig. A.15 shows weight loss at the first fraction. The action was to make a new plan on a new planning CT. Analysis of the old plan on the new CT showed an increase of the mean dose on the bowel from 24 Gy to 30 Gy and mean dose on the bladder from 42 Gy to 46 Gy if the plan had not been adjusted.

Fig. A.16 shows a patient using the breath hold technique for a breast irradiation. The matching on the pre-treatment image looked fine, although with a rather large table shift. The in-vivo software however showed a quite large deviation in position while treated. A closer look in the next fraction revealed patient had a bad breath hold technique, sometimes arching the back. The patient received extra instructions and training.

Fig. A.17 shows the discovery of a systematic difference between the belly board used at simulation and those used at some of the treatment machines. This difference had been present for years and had never been detected. After seeing multiple belly board patients with in-vivo deviations, belly boards were thoroughly checked and it was discovered that the white position indicators were not at the same height relative to the opening for all systems. These position indicators are used to position the patient at the correct height on the belly board.

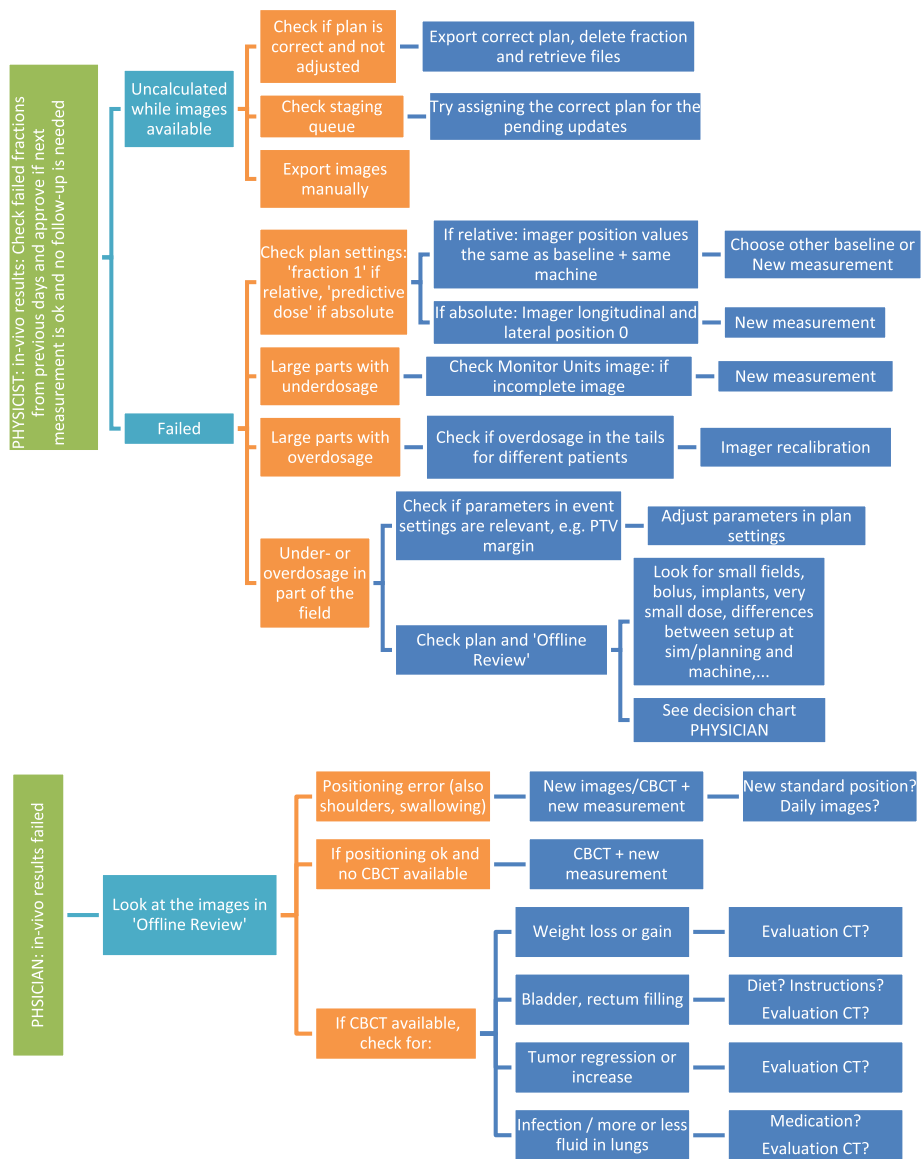


Fig. A1. Decision charts for when a measurement is out of tolerance. The first decision chart is meant for physicists including detecting false positives, the second decision chart is meant for physicians to take actions for patient related errors. Results are always compared with the available imaging. Taking Cone Beam Computed Tomography (CBCT) is often one of the first actions.

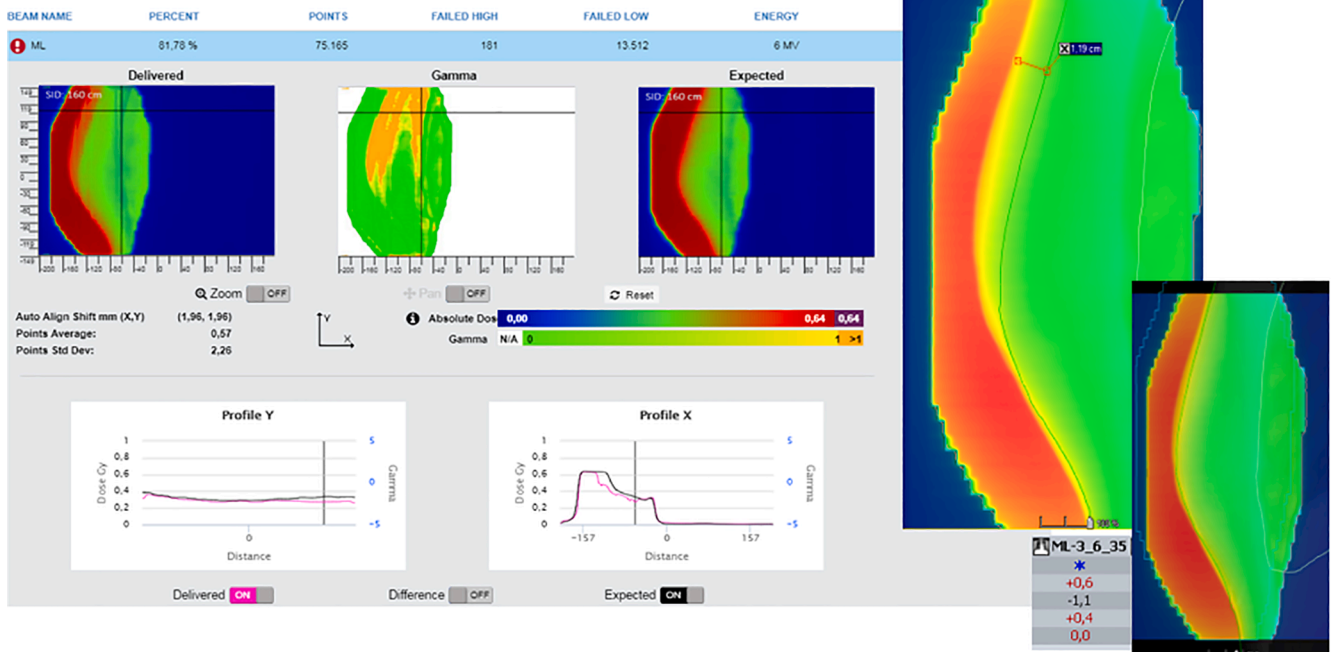


Fig. A2. Deviation in breast position. On the left, in-vivo software results. On the right, offline review images where the integrated images could be matched with the Digitally Reconstructed Radiograph (DRR).

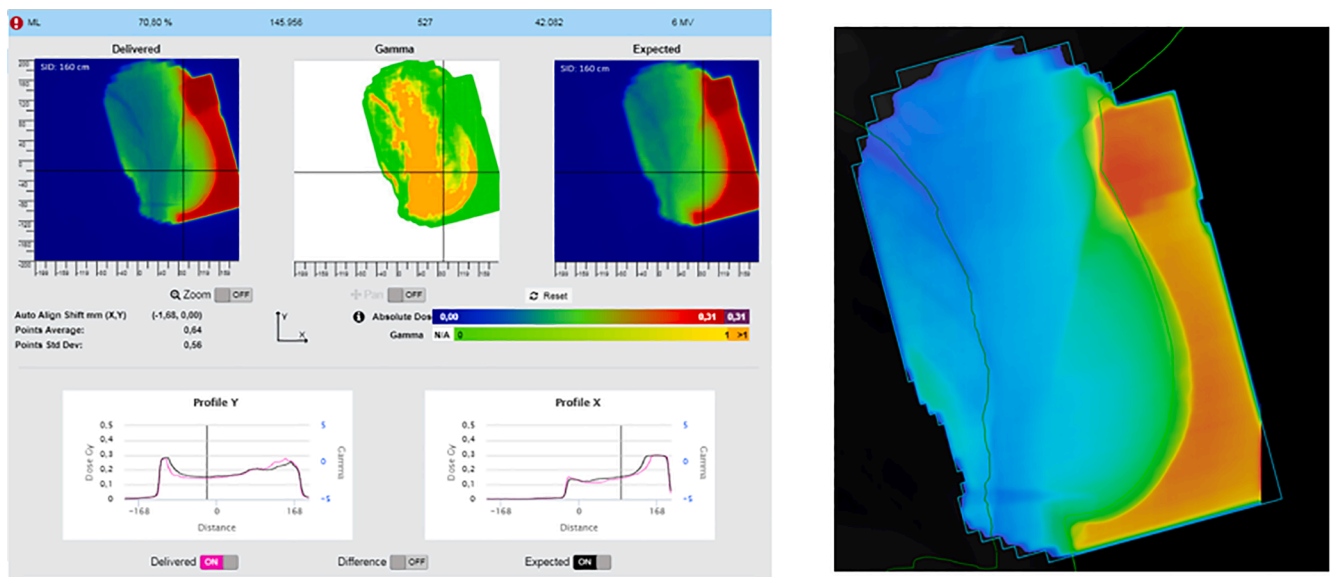


Fig. A3. Breast swelling. On the left, in-vivo software results. On the right, offline review images where the integrated images could be matched with the DRR.

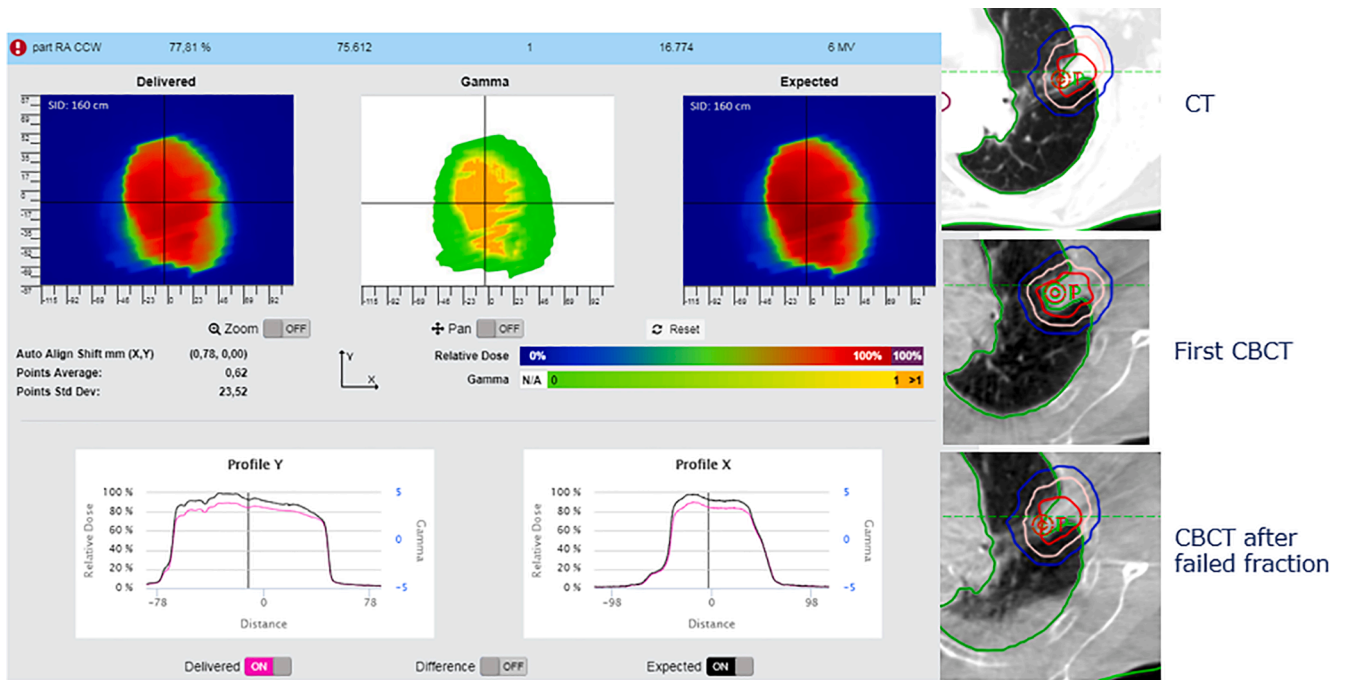


Fig. A4. Pneumonia. On the left, in-vivo software results. On the right, transversal views of planning CT and CBCT's with the structures projected from the planning CT.

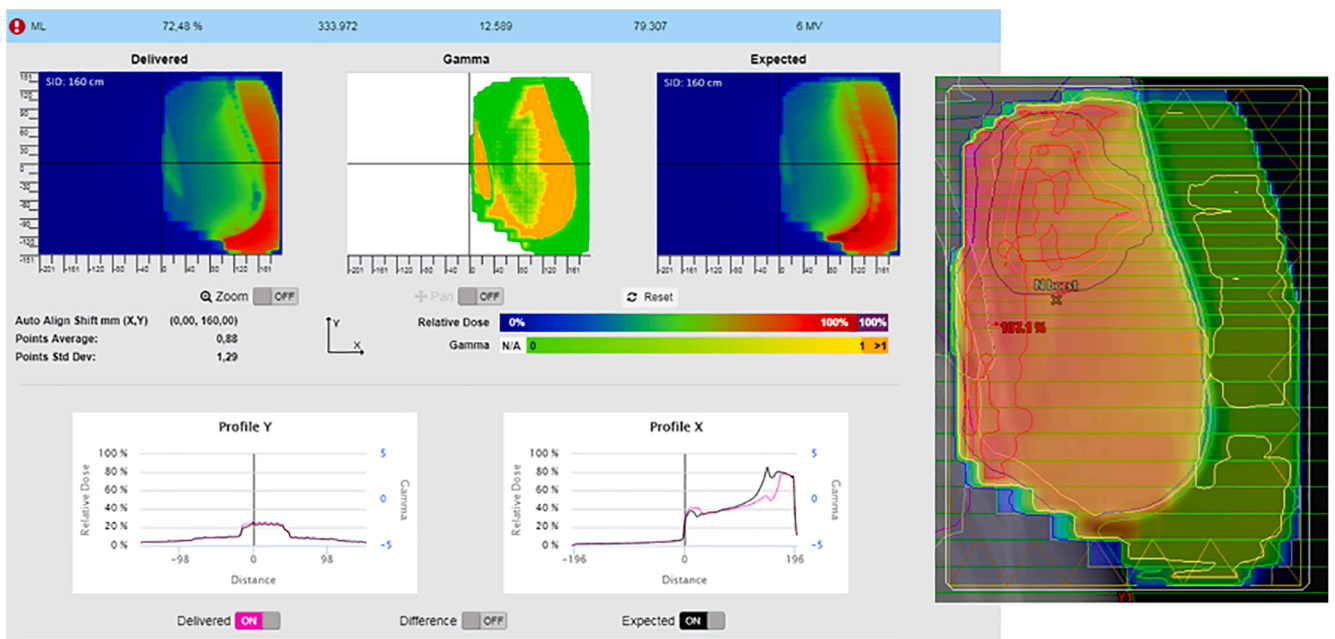


Fig. A5. Planning problem: skin flash tool not well used. On the left, in-vivo software results. On the right, Beam's Eye View (BEV) of the planned dose in the planning system.

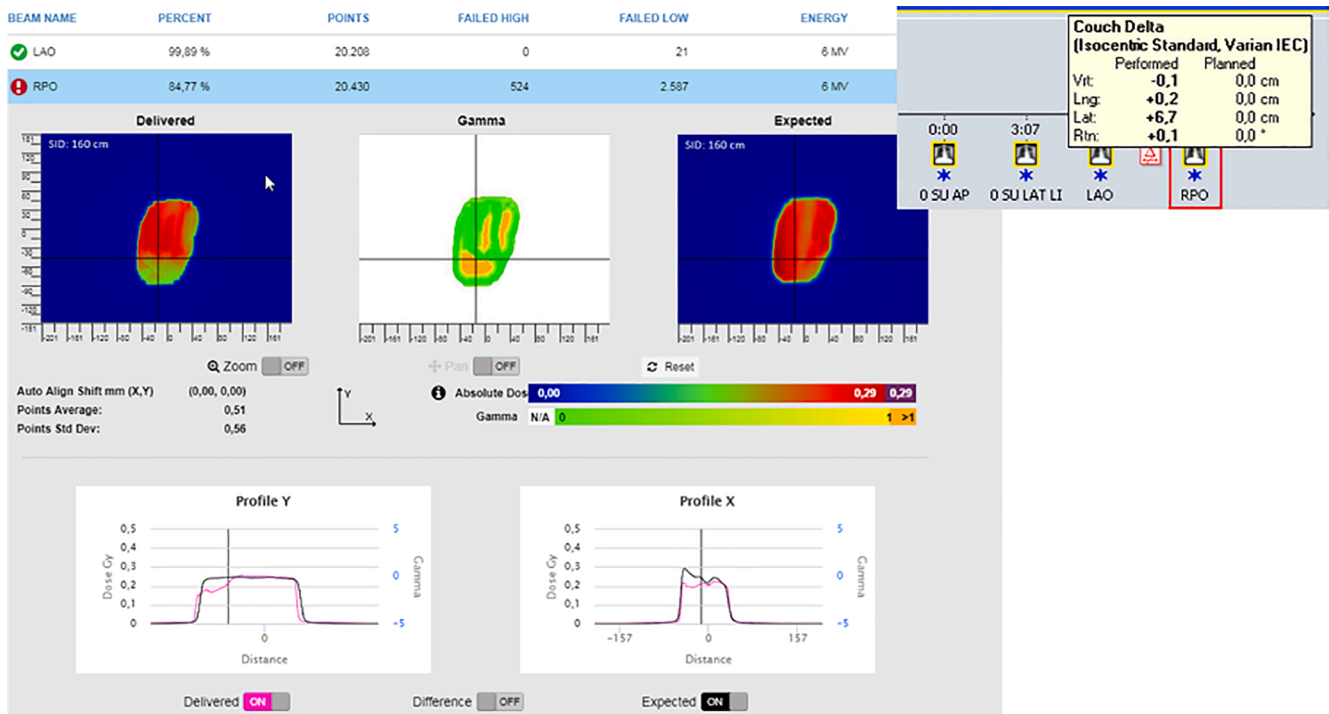


Fig. A6. Wrong table shift. On the left, in-vivo software results. On the right, offline review results showing a table shift of almost 7 cm before the last field.

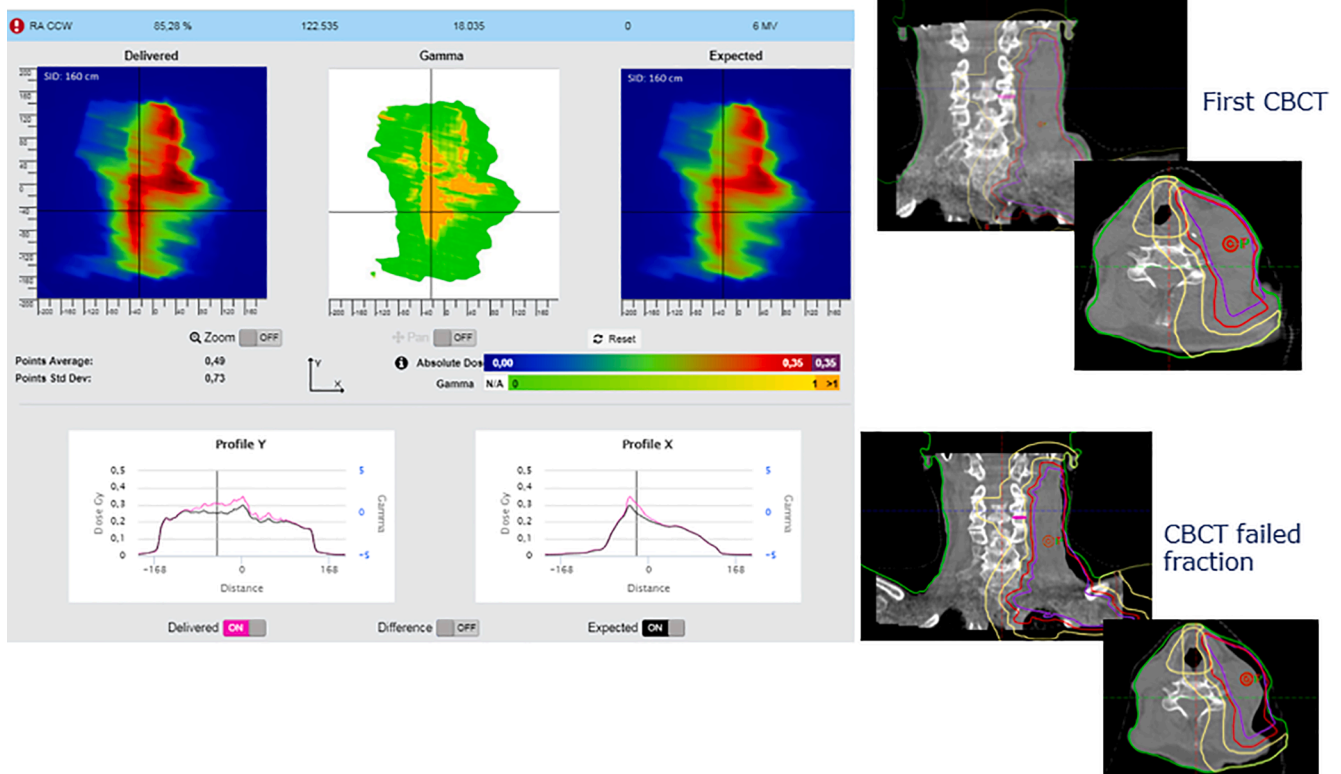


Fig. A7. Tumor reduction. On the left, in-vivo software results. On the right, frontal and transversal views of the CBCT's, with the structures projected from the planning CT.

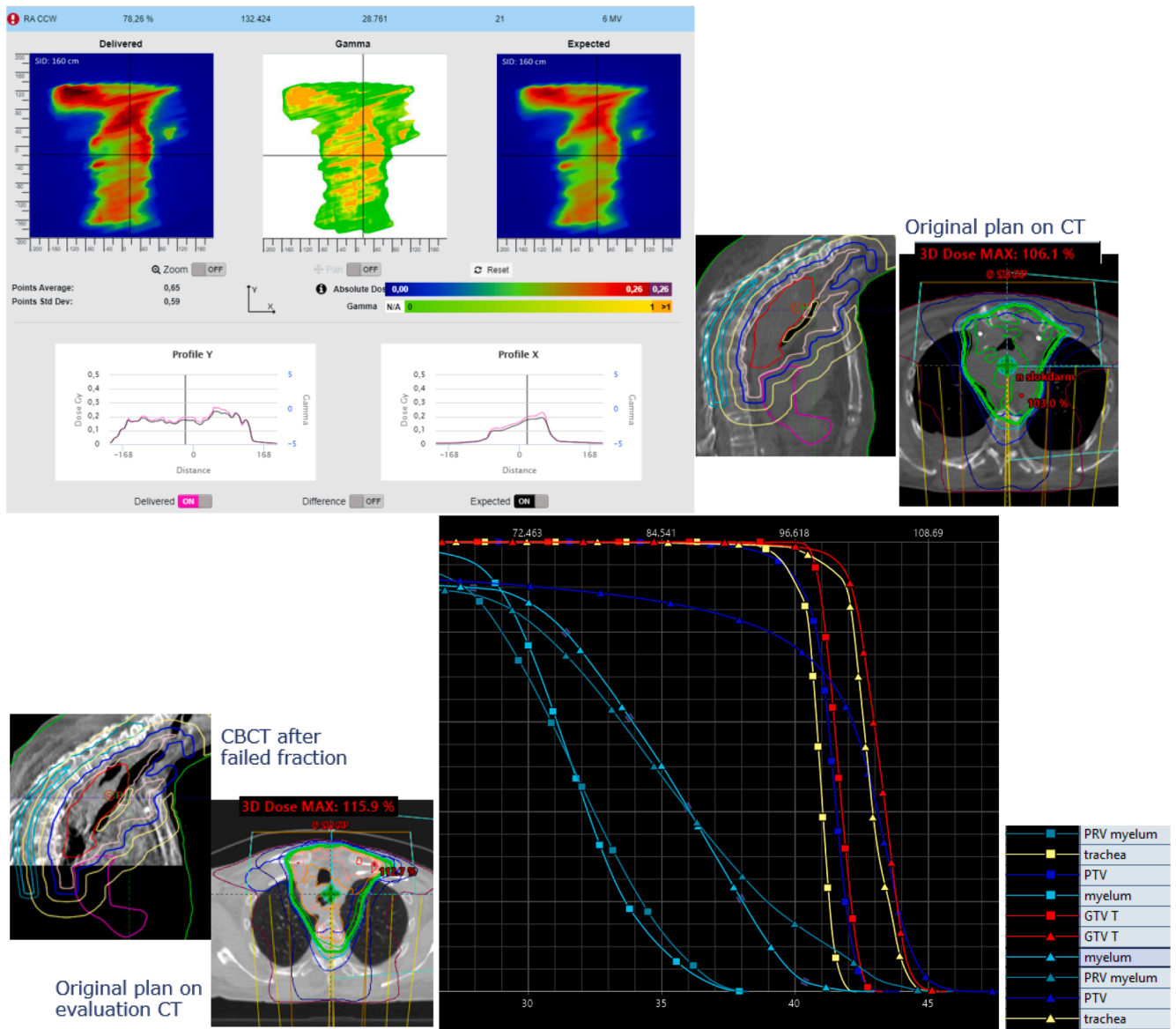


Fig. A8. Tumor reduction and weight loss. On the top left, in-vivo software results. On the top right, sagittal and transversal views of the original plan. On the bottom left, sagittal view of the CBCT and transversal view of the original plan on the evaluation CT. On the bottom right, the change in Dose Volume Histogram for several structures (squares representing the original plan on the original CT and triangles the original plan on the evaluation CT).

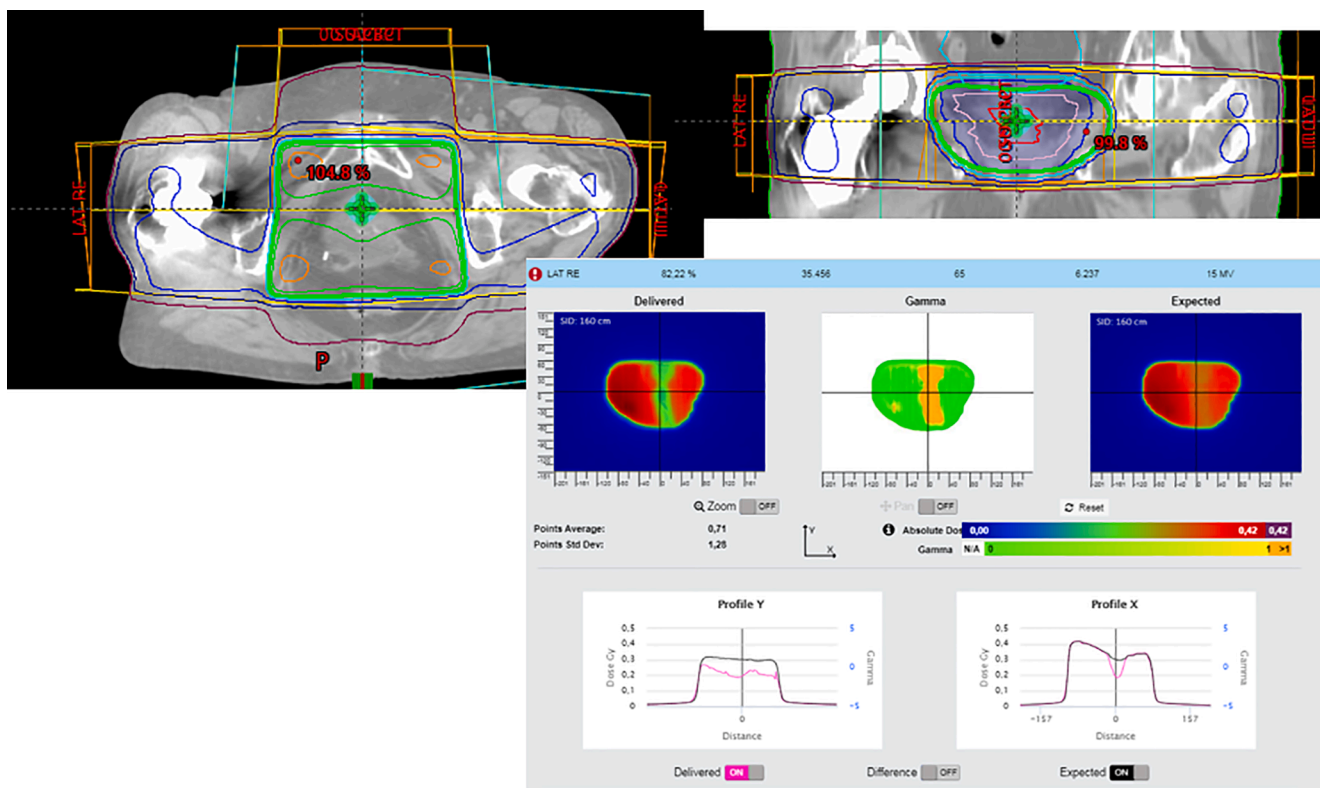


Fig. A9. Planning problem not avoiding hip implant. On the top, transversal and frontal view of the plan on the planning CT. On the bottom, in-vivo software results.

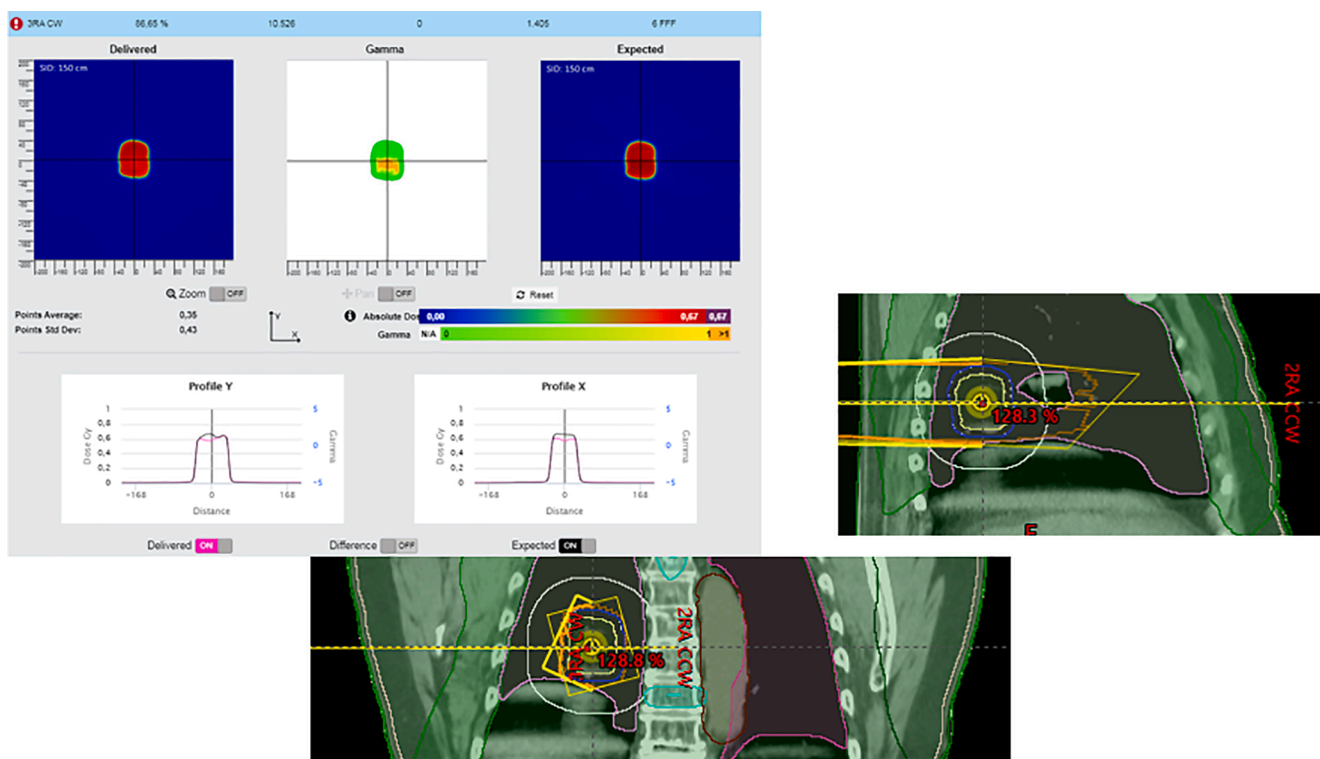


Fig. A10. Problem where slow breathing induced artefacts masking large tumor movement. On the top left, in-vivo software results. On the bottom right, sagittal and frontal view of the planning CT.

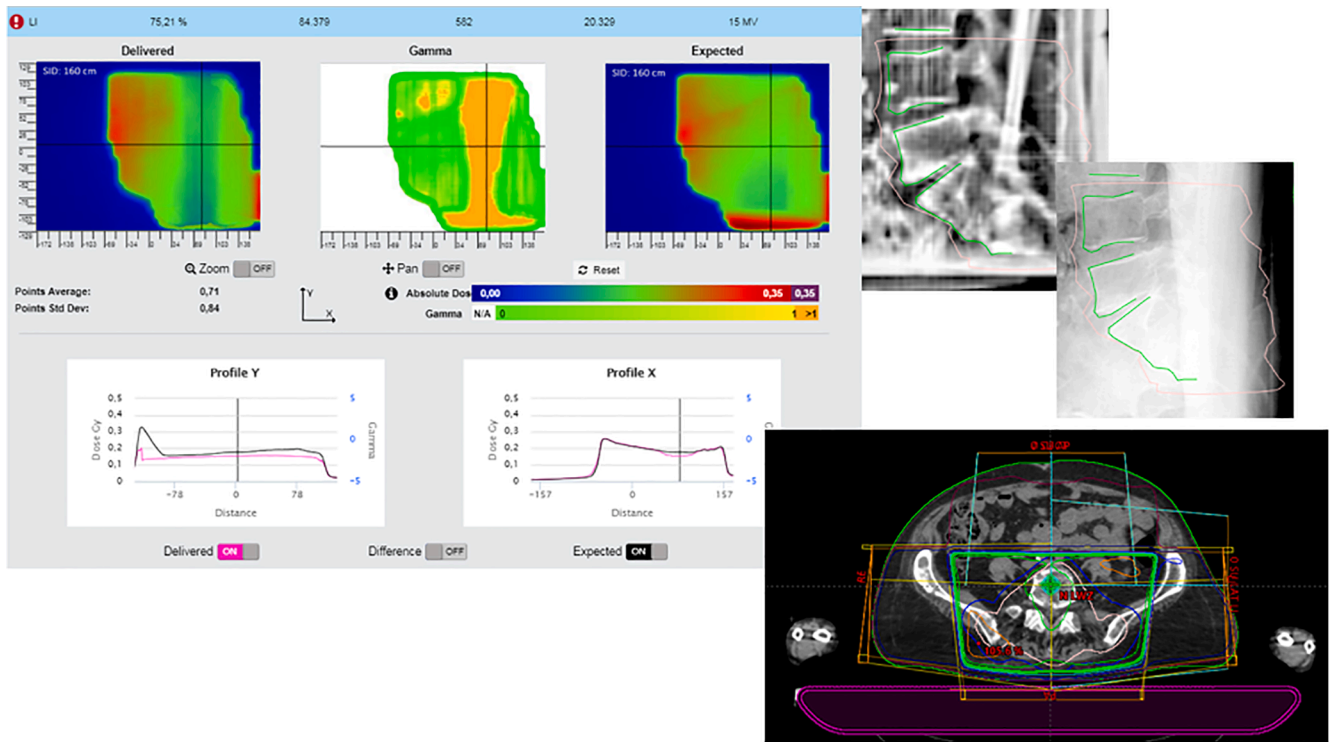


Fig. A11. Simulated with the arms down, but planned with the arms up without notification. On the left, in-vivo software results. On the right, offline review images of setup fields and transversal view of the plan on the planning CT.

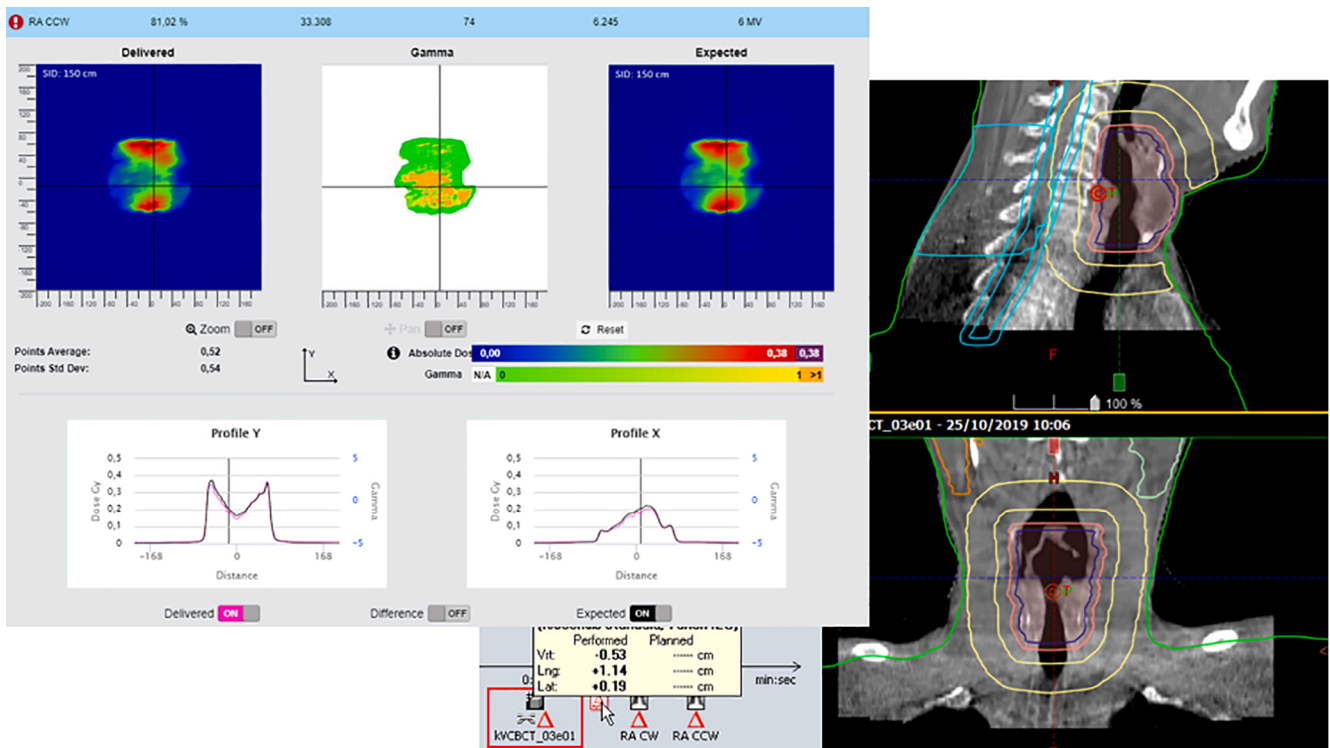
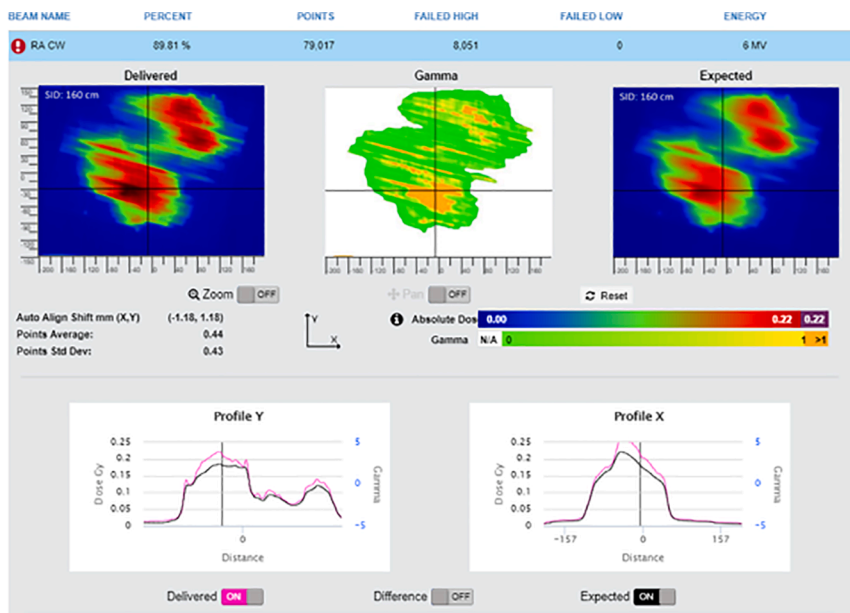


Fig. A12. Patient swallowed during simulation. On the left, in-vivo software results. On the right, offline review images of the CBCT, with the structures projected from the planning CT.



CBCT after failed fraction

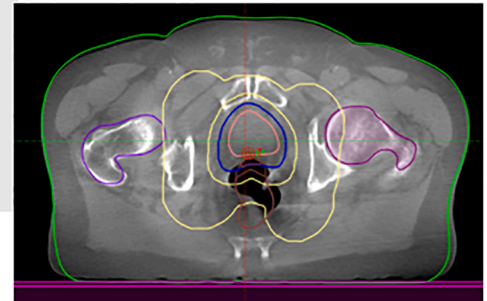


Fig. A13. Air in the rectum. On the left, in-vivo software results. On the right, transversal view of the CBCT, with the structures projected from the planning CT.

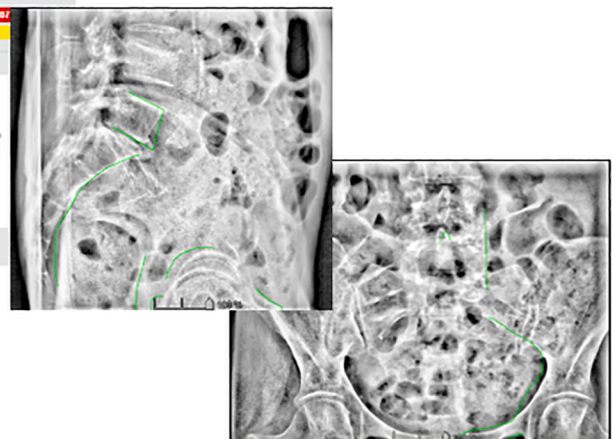
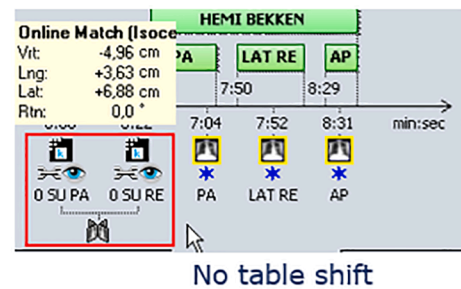
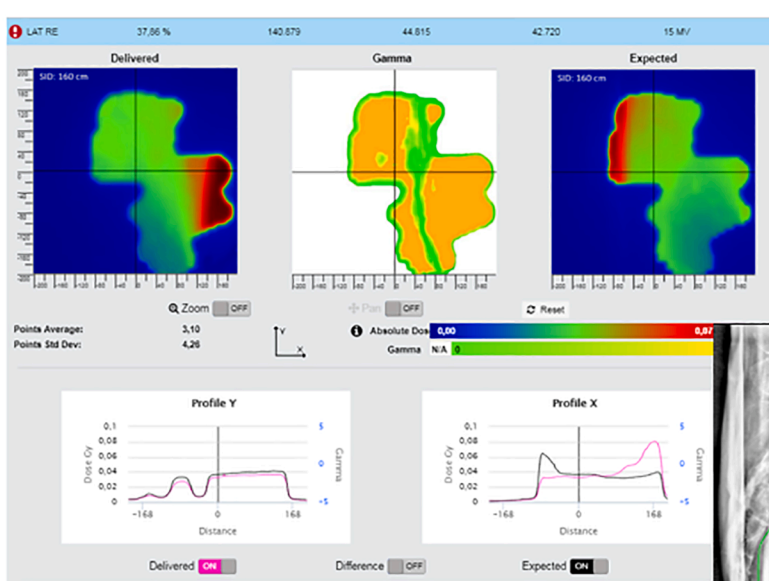
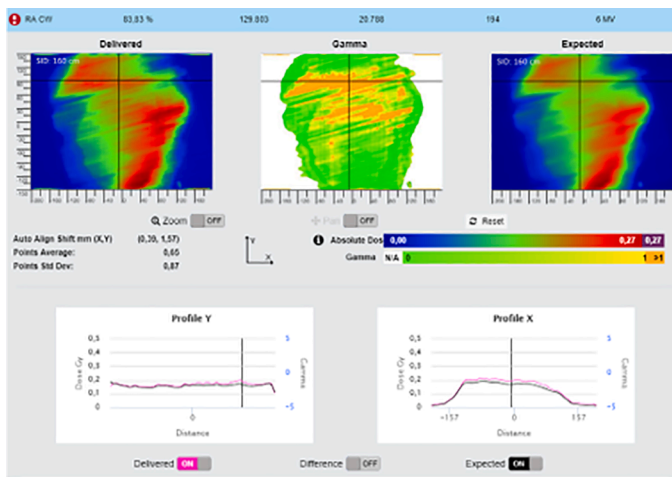


Fig. A14. Large table shift not applied after online match. On the left, in-vivo software results. On the right, offline review images of setup fields.



CBCT after failed fraction

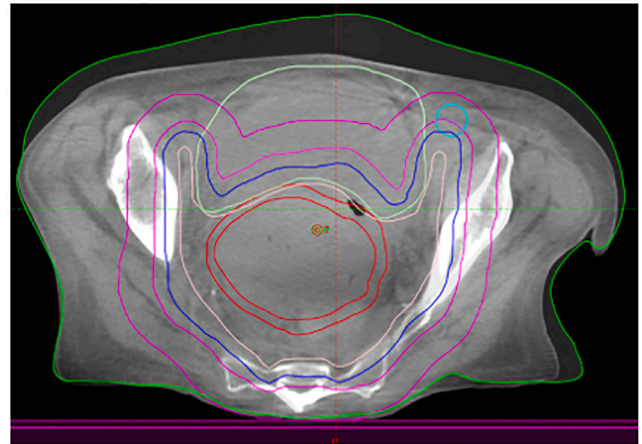


Fig. A15. Weight loss at first fraction. On the left, in-vivo software results. On the right, structures of the planning CT projected on a transversal view of the CBCT.

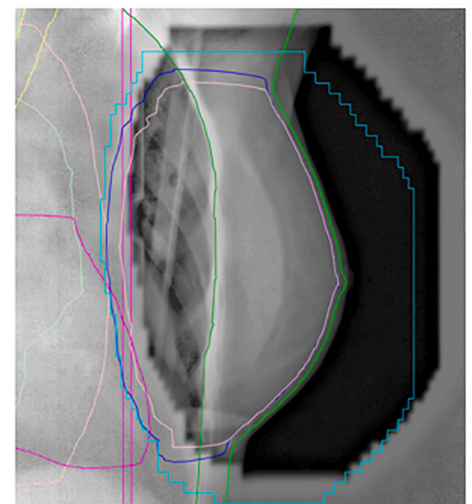
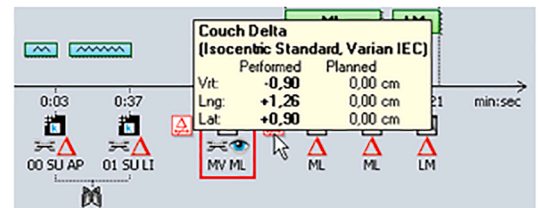
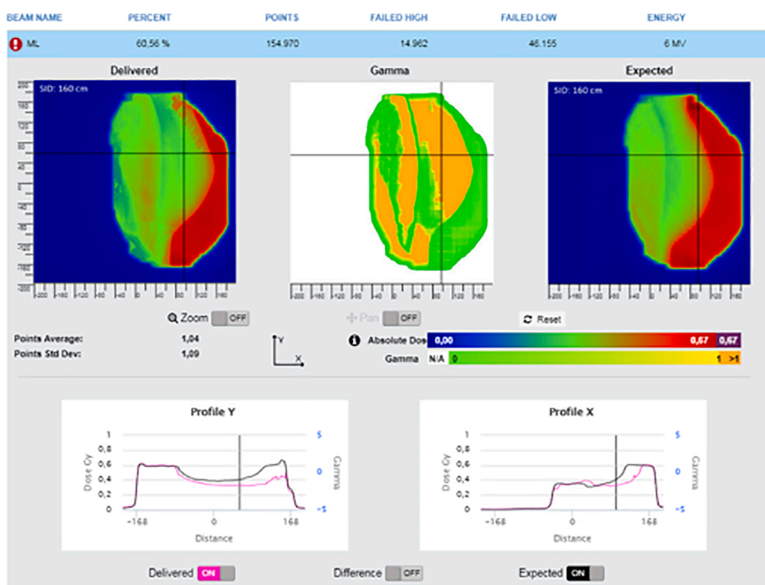


Fig. A16. Bad breathing technique. On the left, in-vivo software results. On the right, offline review setup image matched with the DRR.

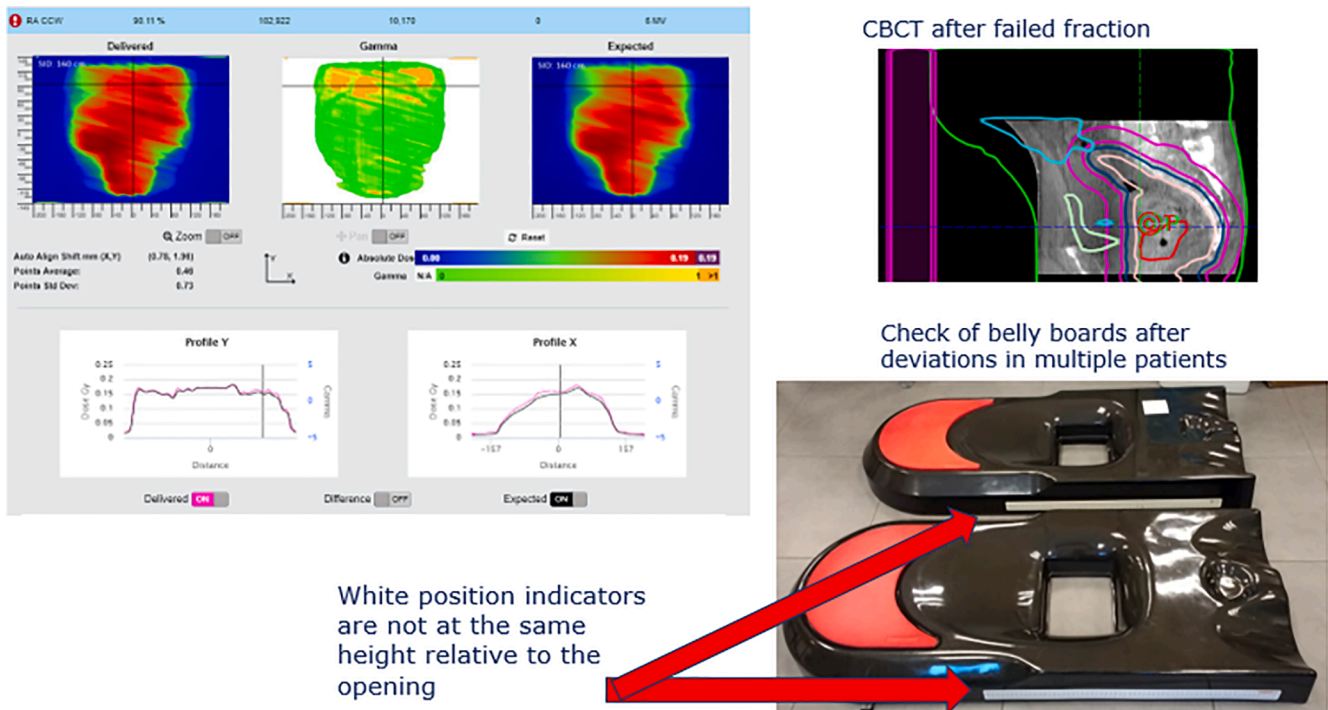


Fig. A17. Belly board problem. On the left, in-vivo software results. On the top right, structures of the planning CT projected on a sagittal view of the CBCT.

References

- Yorke E, Alecu R, Ding L, Fontenla D, Kalend A, Kaurin D, et al. AAPM report No. 87. Diode in vivo dosimetry for patients receiving external beam radiation therapy. AAPM Radiation Therapy Committee Task Group 62. Medical Physics Publishing; 2005.
- Kutcher GJ, Coia L, Gillin M, Hanson WF, Leibel S, Morton RJ, et al. Comprehensive QA for radiation oncology: report of AAPM Radiation Therapy Committee Task Group 40. Med Phys 1994;21:581–618. <https://doi.org/10.1118/1.597316>.
- Derreumaux S, Etard C, Huet C, Tromprier F, Clairand I, Bottollier-Depois J-F, et al. Lessons from recent accidents in radiation therapy in France. Radiat Prot Dosimetry 2008;131:130–5. <https://doi.org/10.1093/rpd/ncn235>.
- Ortiz López P, Cosset JM, Dunscombe P, Holmberg O, Rosenwald JC, Pinillos Ashton L, et al. ICRP publication 112. A report of preventing accidental exposures from new external beam radiation therapy technologies. Ann ICRP 2009;39:1–86.
- International Atomic Energy Agency. Development of procedures for in vivo dosimetry in radiotherapy. Human Health Report No. 8, IAEA, Vienna, 2013. https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1606_web.pdf.
- International Atomic Energy Agency. Accuracy Requirements and Uncertainties in Radiotherapy, Human Health Series No. 31, IAEA, Vienna, 2016. https://www-pub.iaea.org/MTCD/Publications/PDF/P1679_HH31_web.pdf.
- Noel A, Aletti P, Bey P, Malissard L. Detection of errors in individual patients in radiotherapy by systematic in vivo dosimetry. Radiother Oncol 1995;34:144–51. [https://doi.org/10.1016/0167-8140\(94\)01503-U](https://doi.org/10.1016/0167-8140(94)01503-U).
- Lanson JH, Essers M, Meijer GJ, Minken AWH, Uiterwaal GJ, Mijnheer BJ. In vivo dosimetry during conformal radiotherapy: requirements for and findings of a routine procedure. Radiother Oncol 1999;52:51–9. [https://doi.org/10.1016/S0167-8140\(99\)00074-2](https://doi.org/10.1016/S0167-8140(99)00074-2).
- Mans A, Wendling M, McDermott LN, Sonke JJ, Tielenburg R, Vijlbrief R, et al. Catching errors with in vivo EPID dosimetry. Med Phys 2010;37:2638–44. <https://doi.org/10.1118/1.3397807>.
- Mijnheer B, González P, Olaciregui-Ruiz I, Rozendaal R, van Herk M, Mans A. Overview of 3-year experience with large-scale electronic portal imaging device-based 3-dimensional transit dosimetry Pract Radiat Oncol. 2015;5:e679-687. <https://doi.org/10.1016/j.prro.2015.07.001>.
- Van Dam J, Marinello G. Methods for in vivo dosimetry in external radiotherapy. Physics for clinical radiotherapy: ESTRO Booklet n°1, second edition. Brussels, ESTRO, 2006. [https://doi.org/10.1016/S0936-6555\(96\)80078-1](https://doi.org/10.1016/S0936-6555(96)80078-1).
- Mijnheer B, Beddar S, Izewska J, Reft C. In vivo dosimetry in external beam radiotherapy. Med Phys 2013;40:070903. <https://doi.org/10.1118/1.4811216>.
- Olaciregui-Ruiz I, Vivas-Maiques B, Kaas J, Perik T, Wittkamper F, Mijnheer B, et al. Transit and non-transit 3D EPID dosimetry versus detector arrays for patient specific QA. J Appl Clin Med Phys 2019;20:79–90. <https://doi.org/10.1002/acm2.12610>.
- Mijnheer B, Jomehzadeh A, González P, Olaciregui-Ruiz I, Rozendaal R, Shokrani P, et al. Error detection during VMAT delivery using EPID-based 3D transit dosimetry. Phys Med 2018;54:137–45. <https://doi.org/10.1016/j.ejmp.2018.10.005>.
- van Elmpt W, Ezzell GA, Orton CG. Point/Counterpoint. EPID dosimetry must soon become an essential component of IMRT quality assurance. Med Phys. 2009; 36: 4325-7. <https://doi.org/10.1118/1.3213082>.
- Wolfs CJA, Brás MG, Schyns LEJR, Nijsten SMJJG, van Elmpt W, Scheib SG, et al. Detection of anatomical changes in lung cancer patients with 2D time-integrated, 2D time-resolved and 3D time-integrated portal dosimetry: a simulation study. Phys Med Biol. 2017;62(6044–6061). <https://doi.org/10.1088/1361-6560/aa7730>.
- Persoon LC, Nijsten SM, Wilbrink FJ, Podesta M, Snaith JA, Lustberg T, et al. Interfractional trend analysis of dose differences based on 2D transit portal dosimetry. Phys Med Biol. 2012;57:6445–58. <https://doi.org/10.1088/0031-9155/57/20/6445>.
- Bojcheko C, Ford EC. Quantifying the performance of in vivo portal dosimetry in detecting four types of treatment parameter variations. Med Phys. 2015;42: 6912–8. <https://doi.org/10.1118/1.4935093>.
- Bawazeer O, Sarasanandarajah S, Herath S, Kron T, Deb P. Sensitivity of electronic portal imaging device (EPID) based transit dosimetry to detect inter-fraction patient variations. IFMBE Proc 2018;68(3):477–80. https://doi.org/10.1007/978-981-10-9023-3_86.
- Celi S, Costa E, Wessels C, Mazal A, Fourquet A, Francois P. EPID based in vivo dosimetry system: clinical experience and results. J Appl Clin Med Phys 2016;17: 262–76. <https://doi.org/10.1120/jacmp.v17i3.6070>.
- Mijnheer B, Olaciregui-Ruiz I, Rozendaal R, Spreeuw H, van Herk M, Mans A. Current status of 3D EPID-based in vivo dosimetry in The Netherlands Cancer Institute. J Phys Conf Ser 2015;573. <https://doi.org/10.1088/1742-6596/573/1/012014>.
- Olch AJ, O'Meara K, Wong KK. First report of the clinical use of a commercial automated system for daily patient QA using EPID exit images. Adv Radiat Oncol 2019;4:722–8. <https://doi.org/10.1016/j.adro.2019.04.001>.
- Van Elmpt W, McDermott L, Nijsten S, Wendling M, Lambin P, Mijnheer B. A literature review of electronic portal imaging for radiotherapy dosimetry. Radiother Oncol. 2008;88:289–309. <https://doi.org/10.1016/j.radonc.2008.07.008>.

- [24] Bossuyt E, Weytjens R, De Vos S, Gysemans R, Verellen D. 1-year experience with automated transit in vivo dosimetry in a busy multicenter department (abstr). *Radiother Oncol*. 2019;133:S931–2. [https://doi.org/10.1016/S0167-8140\(19\)32148-6](https://doi.org/10.1016/S0167-8140(19)32148-6).
- [25] Carrasco P, Jornet N, Latorre A, Eudaldo T, Ruiz A, Ribas M. 3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification. *Med Phys*. 2012;39:5040–9. <https://doi.org/10.1118/1.4736949>.
- [26] Olasolo-Alonso J, Vázquez-Galiñanes A, Pellejero-Pellejero S, Fernando Pérez-Azorín J. Evaluation of MLC performance in VMAT and dynamic IMRT by log file analysis. *Phys Med* 2017;33:87–94. <https://doi.org/10.1016/j.ejmp.2016.12.013>.
- [27] McGarry C, Agnew C, Hussein M, Tsang Y, Hounsell A, Clark C. The use of log file analysis within VMAT audits. *Br J Radiol* 2016;89:20150489. <https://doi.org/10.1259/bjr.20150489>.
- [28] Zhuang A, Olch A. Sensitivity study of an automated system for daily patient QA using EPID exit dose images. *J Appl Clin Med Phys*. 2018;19(3):114–24. <https://doi.org/10.1002/acm2.12303>.