



Original Article

Results of a multicentre dosimetry audit using a respiratory phantom within the EORTC LungTech trial



Marie Lara Lambrecht^a, David J. Eaton^b, Jan-Jakob Sonke^c, Ursula Nestle^d, Heike Peulen^a, Damien C. Weber^e, Marcel Verheij^c, Coen W. Hurkmans^{a,*}

^aDepartment of Radiation Oncology, Catharina Hospital, Eindhoven, The Netherlands; ^bRadiotherapy Trials QA Group, Mount Vernon Hospital, Northwood, United Kingdom; ^cDepartment of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands; ^dDepartment of Radiation Oncology, Kliniken Maria Hilf, Mönchengladbach, Germany; ^eCenter for Proton Therapy, Paul Scherrer Institute, ETH Domain, Villigen, Switzerland

ARTICLE INFO

Article history:

Received 11 January 2019

Received in revised form 6 June 2019

Accepted 7 June 2019

Available online 25 June 2019

Keywords:

Respiratory phantom

4D dosimetry

Multicentre study

SBRT

ABSTRACT

Introduction: The EORTC 22113-08113 LungTech trial assesses the safety and efficacy of SBRT for centrally located NSCLC. To insure protocol compliance an extensive RTQA procedure was implemented.

Methods: Twelve centres were audited using a CIRS008A phantom. The phantom was scanned using target inserts of 7.5 mm and 12.5 mm radius in static condition. For the 7.5 mm insert a 4DCT was acquired while moving according to a \cos^6 function. Treatment plans were measured using film and an ionization chamber. Wilcoxon's signed-rank tests were performed to compare the three plans across institutions. A Spearman correlation was calculated to evaluate the influence of factors such as PTV, slice thickness and total number of monitor units on the dosimetric results.

Results: The reference output dose median [min, max] variation was 0.5% [−1.1, +1.5]. The median deviations between chamber doses and point-planned doses were 1.8% [−0.1; 6.7] for the 7.5 mm and 1.1% [−2.8; 5.0] for the 12.5 mm sphere in static situation and 3.2% [−3.2; 15.7] for the dynamic situation. Film gamma median pass rates were 92.0% [68.0, 99.0] for 7.5 mm static, 96.2% [73.0, 99.0] for 12.5 mm static and 71.0% [40.0, 99.0] for 7.5 mm dynamic. Wilcoxon's signed-rank tests showed that the dynamic irradiations resulted in significantly lower gamma pass rates compared to the 12.5 mm static plan ($p = 0.001$). The total number of MUs per plan was correlated to both film and IC results.

Conclusion: An end-to-end audit was successfully performed, revealing important variations between institutions especially in dynamic irradiations. This shows the importance of dosimetry audits and the potentials for further technique and methodology improvements.

© 2019 Elsevier B.V. All rights reserved. Radiotherapy and Oncology 138 (2019) 106–113

In 2014 the EORTC launched the phase II 22113-08113 LungTech trial to assess safety and efficacy of SBRT for centrally located NSCLC [1]. The primary endpoint of the study is freedom from local progression at three years. In addition toxicity is scored according to the Common Terminology Criteria for Adverse Events (CTCAE V4). SBRT results in excellent local control similar to surgery in the case of peripheral lesions [2]. For central lesions, located near critical normal mediastinal structures, only few studies with clinical results have recently been published [3]. This scarcity of results regarding the efficacy of the treatment and mainly the associated toxicity for the OARs is a source of cautiousness for SBRT of centrally located lung tumours. The LungTech trial investigates a risk adapted hypo-fractionation approach with 60 Gy delivered to the tumour over a course of 8 fractions within a treatment duration of two weeks. As lung SBRT involves the precise delivery of high

fraction doses to a moving tumour in a low density surrounding, the technique poses precision challenges for both the dose calculation and dose delivery. As this is an international trial, it was also anticipated that a broad range of techniques and equipment would be used. Therefore, a comprehensive radiation therapy quality assurance (RTQA) procedure has been developed [4]. The procedure included a benchmark exercise and evaluation of 4D-CT and 4D PET-CT quality, of which the results have been recently published. This study reports on the end-to-end tests investigating the dosimetric accuracy of the IMRT and VMAT treatments using a lung anthropomorphic phantom in static and respiratory conditions.

In 2009, Vinogradskiy et al. studied the accuracy of single and multiple beam radiotherapy and IMRT, comparing 3D and 4D dose calculations with measurements in a rigid and deformable lung phantom in a single centre [5]. Ong et al. and Li et al. studied the dosimetric interplay effect of a moving tumour in combination with simultaneously moving leaves using single centre

* Corresponding author.

E-mail address: coen.hurkmans@cze.nl (C.W. Hurkmans).

measurements [6,7], focussing on flattening filter free treatments. In 2013, Nishio et al. published the results of a multicentre study using a self made water-tank like, non-moving phantom on the accuracy of lung SBRT in the JCOG 0702 phase I trial of lung SBRT [8]. Recently, a number of papers have been published on trial QA of lung radiotherapy with, mainly data from centres in the United Kingdom [9–11]. Palmer et al. used a phantom with a moving lung like insert for assessing non SBRT lung treatment [11]. They presented results from 12 centres in the United Kingdom. Only one of these, Distefano et al. [3] looked at lung SBRT doses, their volume was nonetheless larger than would be typically considered for SBRT treatments.

This study is therefore unprecedented in evaluating lung SBRT in a multicentre international setting with an anthropomorphic respiratory phantom. Our report includes data in an international clinical trial setting. Furthermore, we provide an overview of the treatment techniques and equipment used by the participating institutions.

Material and methods

The institutions participating in the LungTech trial were required to complete the EORTC RTQA procedure. The level 5 of this procedure, called dosimetric check was here mainly performed through a site visit of three days in each institution. The site visit was highly recommended by the trial guidelines but not mandatory. Centres that did not receive a site visit had to perform a virtual phantom procedure before they were allowed to enter patients in the trial [12]. The site visits were performed by one operator and identical the equipment.

Phantom details

The CIRS 008A phantom (CIRSinc, Norfolk, Virginia, USA) is a thoracic anthropomorphic phantom with two lung shaped regions made of lung equivalent density material, a water-equivalent mediastinum and a vertebral structure. A lung equivalent density rod is inserted in the phantom's left lung and carries an insert containing a spherical target of water equivalent density simulating a lung lesion. The rod is attached to a motion actuator moving the target according to a respiratory signal specified by the user. Two sizes of lesions were investigated with spheres of 7.5 mm and 12.5 mm radius (pictures of the phantom are displayed in annex 1).

For dosimetric purposes, each sphere size had two inserts made. One insert (sbirt rod model 008a-22) was designed to carry films of 5 × 13 cm dimensions traversing the sagittal central plane of the sphere. The film insert was carrying at the inner surface of one half three markers visible on CT, the other half of the insert had holes to allow the user to pinch the films at the place of the markers.

The other insert (micro chamber insert model 008A-06-CV) was designed to carry a Wellhofer ionization chamber (IC) of 0.04 cc volume. The IC is traceable to a secondary standard dosimetry laboratory (VSL – The Netherlands) and was always connected by the same wires to the PTW UNIDOS® E Universal Dosimeter.

Audit

Image acquisition and planning

The dosimetric part of the site visit was planned over two days, with one day dedicated to the acquisition of the planning CT and the planning itself and one day for measurements.

The CIRS phantom was scanned at the visited institution using the institutions own protocol which they would also use within the trial. Within the trial, the use of 4D-CT is mandatory to account for tumour motion due to breathing. The slice thickness of the 4D CT scan are required not to exceed 3 mm.

Multiple 3D and 4D CT scans were made for the end-to-end tests described here. A 3D static scan was made with the small sphere in place and one with the large sphere in place. A 4D scan of the small sphere with the rod moving according to a \cos^6 function with a period of 3 seconds and a 15 mm peak-to-peak amplitude was also made.

A treatment plan was made for all three scans according to the trial guidelines for PTV coverage.

Measurements

Prior to phantom measurements a beam output check was performed in water under the institutions own reference conditions. The plans were then measured twice using EBT3 films (12.7 cm × 5 cm) and the IC. After switching the film for the IC insert, the position of the phantom itself was checked using the in-room lasers and it was made sure that the insert was in the most inward position in the phantom which positioned the IC precisely in the isocentre.

The films were analysed and calibrated centrally. The film calibration was done per box using 11 dose points ranging from 0 to 25 Gy. The films were scanned using an EPSON Perfection V700 scanner. The three RGB channels were used to read the image and the resolution was set to 72dpi. The films were always positioned at the centre of the scanner bed in the same orientation according to Wen et al. [13].

The analysis of the films was performed in RIT113 version 6.3 (Radiological Imaging Technology, Inc, USA). Each film was rescaled at the centre by the result of the IC for each plan measured. Global gamma analyses were performed using film dose as the reference, a normalization to 100% at the centre of the sphere, a dose threshold at 20% of dose maximum and a 3%/3 mm deviation criteria. For each plan, two films were simultaneously placed in the insert and the results of the film giving the best gamma pass rate values were used in the analysis.

Institutions that wanted to test a new irradiation technique or a new calculation algorithm in addition to their clinical practice, were allowed to undergo additional measurements. In addition, if after unsatisfying results institutions were able to find a cause and propose a solution, new calculations could be sent to compare with the measurement or in exceptional cases a second visit could be organized.

The additional measurements or calculations that were planned by the institutions voluntarily are referred to as 'bis' in this analysis. The measurements that were performed after unsatisfying results are referred to as 'follow-up measurements'.

Statistical analysis

For the statistical analysis of the QA measurement for each plan only the twelve definitive measurements, excluding the "bis" measurements, were taken into account. In case a follow-up measurement was performed for the main technique this measurement was considered as the definitive measurement. The statistical analysis was performed in XLSTAT® (Addinsoft, inc, USA). Across all institutions a Wilcoxon signed rank test was performed between the three-measured plans for both the IC results and the gamma pass rates in order to evaluate if a significant difference existed between the three plans. A Spearman correlation was calculated to evaluate if the beam output result was correlated with the dose deviation for the phantom measurements for both IC measurements and film measurements. Spearman's correlations were also calculated to evaluate the association between the results measured by the IC and the gamma pass rate and the influence of the total number of MU, the slice thickness of the CT and the dose grid and the PTV on the IC results and the gamma pass rate. In addition, comparisons of IC and gamma pass rate medians using a mood test (due to the inequality of the populations size) was performed

between the two dominant irradiation techniques (VMAT & IMRT) used within the visited institutions.

Results

Equipment and technique used

Sixteen centres were visited. For one centre, the film measurements failed which was only discovered after the visit as the films were scanned centrally. Three centres did not send their treatment plan. As required by the trial guidelines all 16 institutions used a type B or more advanced algorithm [14]. One institution used a Monte-Carlo algorithm and three institutions (one in clinical routine, two to test an algorithm that was not yet clinically implemented) used the Acuros linear Boltzmann transport equation algorithm. One institution intended to switch from 6MV to 6MV FFF (institution 10). There was only little diversity in the treatment machines used as presented in Table 1.

Two institutions voluntarily tested two irradiation techniques (9 and 10). They used IMRT in clinical practice and wanted to switch to a VMAT technique. Two institutions (1 and 2) made a double calculation of their plan using two different algorithms, the AAA algorithm, clinically used, and Acuros. The non-clinically used techniques are referred to as bis measurements in the figures and are excluded from the statistics.

Beam output in reference condition

Among the participating centres the median [min, max] dose deviation from the expected value for the output in water in reference condition was 0.5% [−1.1%, +1.5%].

Point dose measurements

The median [min, max] dose deviation of the measured dose corrected for daily output fluctuations with respect to the planned dose can be found in Table 2. The median results for the 7.5 mm radius sphere were 1.8% and 3.2% in static and dynamic condition, respectively and for the 12.5 mm radius sphere in static condition the observed deviation was 1.1%. The dispersion of the dose deviation (%) across institutions is represented in Fig. 2.

Individuals point dose measurement results are displayed in Fig. 3a. The two institutions (1 and 2) that made a double calculation with AAA and Acuros had both a reduction in their dose deviation using Acuros. The decrease in the error was for measurement 1 and 1bis, and measurement 2 and 2bis −0.02% [from −0.12% to 0.10%] and −0.2% [from 2.0% to 1.8%] for the static 7.5 mm sphere. For the dynamic 7.5 mm sphere and for the 12.5 mm static this decrease was −0.4% [from −2.2% to −1.8%] and −0.6% [from −3.2% to −2.6%] and −0.2% [from 1.7% to 1.5%] and −1.2% [from 5.0% to 3.8%], respectively. The two institutions (9 and 10) that irradiated two plans and wanted to switch from IMRT to VMAT had worse results using their latest techniques. Four institutions were unsatisfied with the extend of the deviations measured. Two institutions, 9 and 10, provided a new calculation for their bis technique for the plan made with the small sphere moving. Institution 9 recalculated the plan based on the exhale phases instead of the global average CT image as they discovered density issues with the HU profile of the sphere for their averaged image. This institution was using a free-breathing approach in their clinical practice and had no precedent with the 4DCT technique required by the trial guidelines which they now tested for the first time and planned to use within the trial. Institution 10 that was using an ITV contour based on the average image calculated on

Table 1
Institutions treatment technique and equipment.

Institution	Energy (MV)	TPS (algorithm)	Delivery technique*	Linac model	Number of arcs (full/partial)/ Number of beams	Control points spacing (°)	CT slice spacing (mm)	Dose grid (mm)
1	6	Eclipse (AAA)	ss IMRT	Varian Truebeam	8	2	3	1
1 bis	6	Eclipse (Acuros)	ss IMRT	Varian Truebeam	8	2	3	1
2	6	Eclipse (AAA)	VMAT	Varian Truebeam	6 (partial)	2	3	1
2 bis	6	Eclipse (Acuros)	VMAT	Varian Truebeam	6 (partial)	2	3	1
3	6	Iplan (MC)	DCAT	Varian Truebeam STX	1 (partial)	6	3	1
4	8	Eclipse (AAA)	DCAT	Varian Clinac	2 (partial)	2	3	2.5
5	6	Eclipse (AAA)	VMAT	Varian Clinac	2 (partial)	1.5	3	1
6	6	Eclipse (AAA)	VMAT	Varian Truebeam	2 (partial)	2	4	2.5
7	18	Eclipse (AAA)	sw IMRT	Varian Clinac	9 [#]	-	3	2.5
8	6FFF	Eclipse (Acuros)	VMAT	Truebeam HD	3 (partial)	2	2	3
9	6	Eclipse (AAA)	ss IMRT	Varian Truebeam HD	13	-	2	2.5
9 bis	6	Eclipse (AAA)	VMAT	Varian Truebeam	2 (full)	2	2	2.5
10	6	Pinnacle (CC)	ss IMRT	Elekta Synergy	9	-	3	2.5
10 bis	6FFF	Pinnacle (CC)	VMAT	Elekta Synergy	1 (full)	2	3	2.5
11	6	Eclipse (AAA)	VMAT	Varian Novalis Tx	2 (full)	2	3	2.5
12	6	Pinnacle (CC)	VMAT	Elekta Precise	1 (full)	4 and 2	3	2.5

* ss = step and shoot, DCAT = Dynamic Conformal Arc Therapy, sw = sliding window.

[#] Non-coplanar.

Table 2

Comparison of results between the dynamic irradiation and the static irradiation, for the film dosimetry and for the ionization chamber results. Minus sign represents a lower measured than calculated dose.

	Plan 1	Plan 2	Plan 3
Point dose deviation (Gy)	1.8	3.2	1.1
Median (range)	(−0.1, 6.7)	(−3.2, 15.7)	(−2.8, 5.0)
Film gamma pass rate (%)	92.0	71.0	96.2
Median (range)	(68.0, 99.0)	(40.0, 95.0)	(73.0, 99.0)

Plan 1 = 7.5r sphere static, Plan 2 = 7.5r sphere dynamic, Plan 3 = 12.5r sphere static

the average CT, they recalculated on a mid-position image. These new calculations decreased the deviation with the measured dose from 6.2% to 2.5% and from 5.2% to 1.5% for measurement 9bis and measurement 10bis, respectively. Institution 12 had deviations higher than 3% with their clinically used technique for the three plans tested. They retested the plans with their delta 4 phantom and could not reproduce the deviations observed during the visit with the anthropomorphic phantom. After a discussion with neighbouring centres using the same combination of TPS and treatment device, they were encouraged to reduce the spacing between the MLC control points from 4 degrees to 2 degrees and requested a second visit. The visit measured a decrease in the deviations from 3.8% to 2.9%, from 8.6% to 3.2% and from 4.6% to 0.0% for respectively the static 7.5 mm, the dynamic 7.5 mm and the 12.5 mm static plan. Institution 6 had a deviation up to 15.7% with the dynamic plan of which the cause could not be found. The institution changed its technique from VMAT to IMRT and its treatment device and beam energy from Truebeam 6 FFF to a Clinac 6 MV and later provided the results of a virtual phantom QA that met the trial criteria.

Film dose measurement

Individual γ passing rate results are displayed in Fig. 3b. The median and range [min, max] γ pass rate for 3%/3 mm was 92.0% [68.0, 99.0], 71.0% [40.0, 95.0] and 96.2% [73.0, 99.0] for the static 7.5 mm, the dynamic 7.5 mm and the static 12.5 mm, respectively. For institution 1 that provided plans recalculated with Acuros the gamma pass rates improved from 92% to 94%, from 67% to 72% and from 98% to 99%, for respectively the 7.5 mm plan static, the 7.5 mm plan dynamic and the 12.5 mm plan static when using Acuros. For institution 2 the film measurements of the 7.5 mm plan static were inexploitable due to a phantom set-up error, nonetheless similarly to institution 1, the Acuros algorithm showed better agreement with the planned dose; with gamma pass rate improvements from 67% to 70% for the 7.5 mm dynamic plan and from 97%

to 98% for the 12.5 mm static plan. No films were measured during the first site visit at institution 12 at the light of the deviations found and reproduced with the IC, the films were irradiated only during the second visit (additional measurements) where the cause of the deviations had been identified and corrected.

Statistical analysis

The results of the Wilcoxon sign rank tests comparing the QA results of the three plans did not find any significant difference between the three plans results of the IC measurements. Regarding the film γ passing rate the test highlighted a significant difference solely for the film QA results between the 12.5 mm sphere static and the 7.5 mm sphere in dynamic condition with $p < 0.001$. The correlation (Spearman's) between the absolute IC and the film QA results was significant only in the case of the 7.5 mm sphere in static condition (Table 3).

Additionally a Spearman correlation test failed to find any correlation between each institution beam output result and the institution films and IC phantom QA performance.

The comparison of median IC deviation between the institution using VMAT and the institution using IMRT as irradiation technique was significant with $p = 0.024$ in favour of IMRT. The comparison was also significant for the median films gamma pass rate with $p = 0.001$, also in favour of IMRT.

The Spearman correlation matrix (Table 4) highlighted a correlation for both the IC results and the film gamma pass rates with the total number of MU with $p = 0.003$ and 0.020, respectively, where a lower number of MU was correlated with a better pass rate.

Discussion

The main objective of this study was to investigate the treatment planning and delivery accuracy of centres participating in the LungTech trial. Although accuracy was high in some centres we noticed a wide variation across institutions.

The beam output in reference condition for the energy used by the institutions showed a high level of agreement between expected and measured dose across all institutions (median deviation +0.5%). However this high accuracy was not correlated with the phantom measurements that showed important inter-institution performance variations. This demonstrates that a beam output is not a sufficient minimum requirement for insuring dosimetric accuracy within a trial and could actually be discarded if a treatment plan is measured on a phantom. The results of the IC chamber placed at the centre of the sphere were also poorly

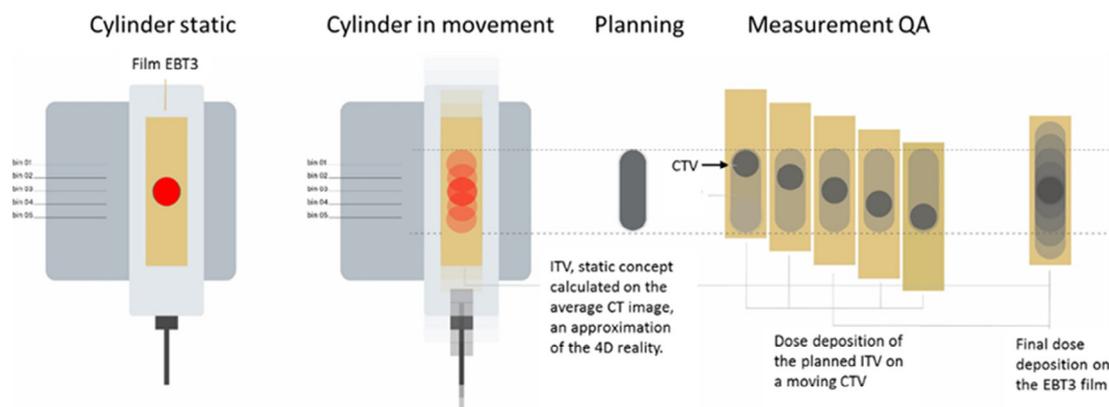


Fig. 1. Schematization of the experiment and the differences between 3D and 4D calculations in matching the measurement.

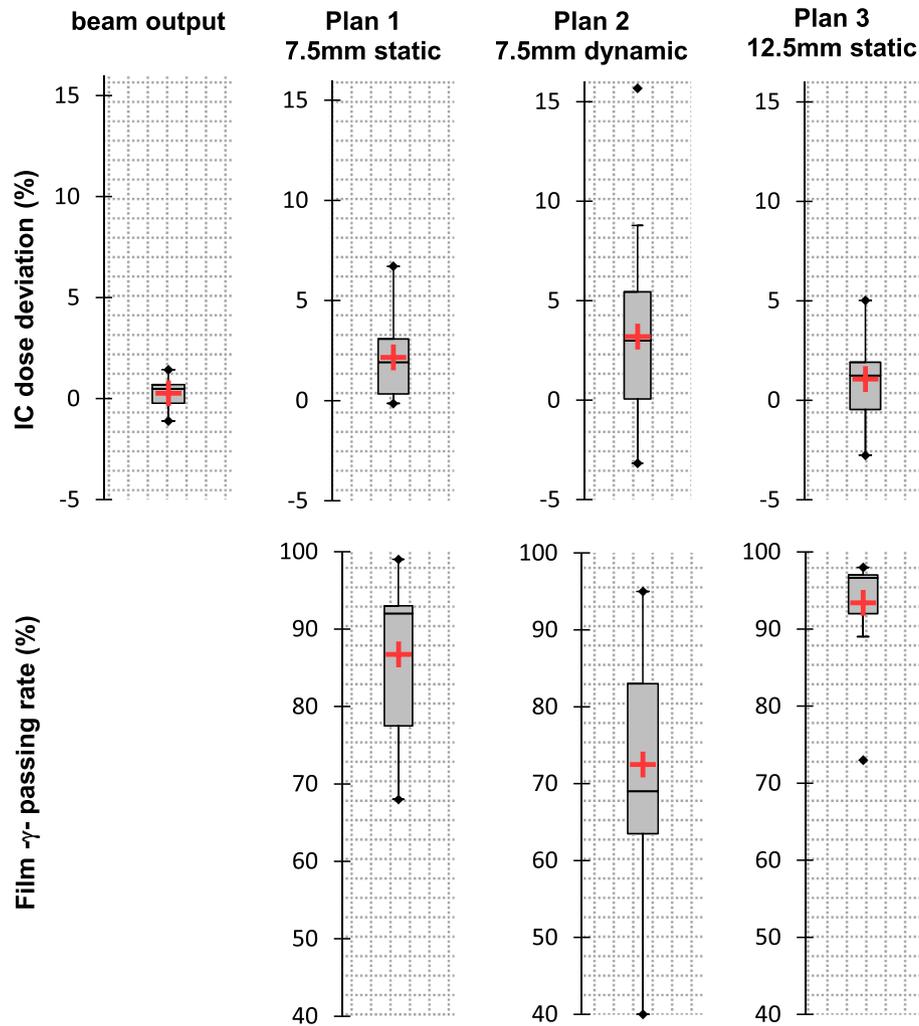


Fig. 2. Boxplots of the measured dose (IC) deviation (%) and the film γ passing rate (%) across the institutions tested for the beam output in water (IC) and the three measured plans. The red cross indicates the mean value.

correlated with the film γ pass rates, highlighting the necessity of at least 2D plane measurements to ensure plan quality.

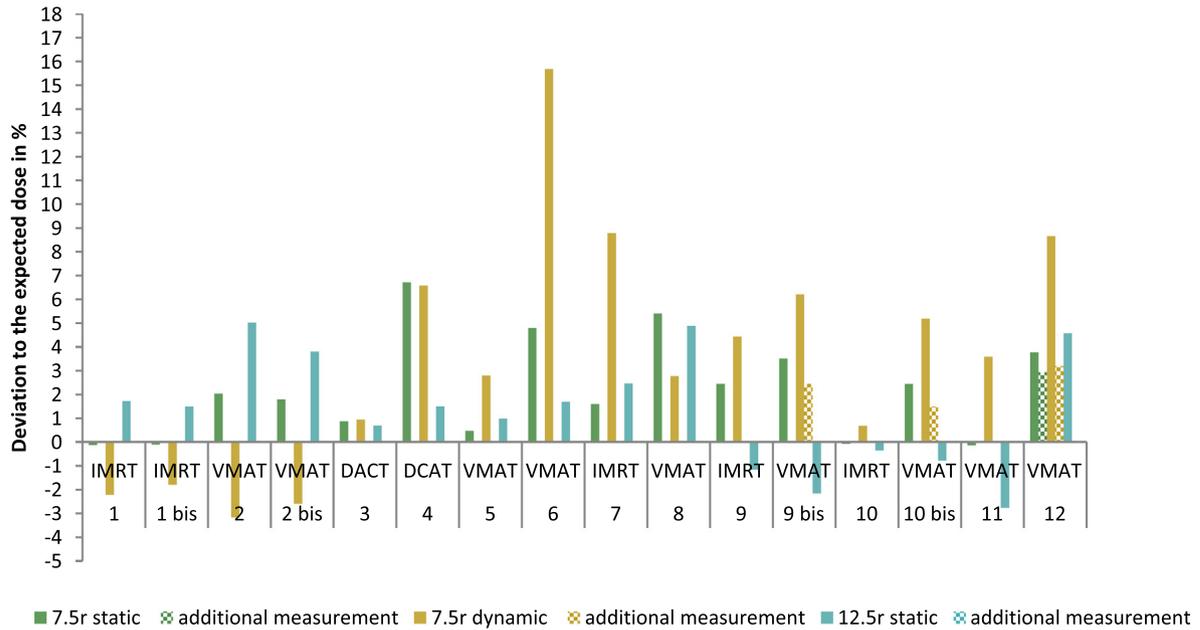
When using dynamic phantoms, there is an intrinsic difference between the dynamic measurement and a plan based on an average CT image of a moving phantom, and a deviation to the planned dose is thus expected (Fig. 1).

Plans based on the static ITV concept only approximate the influence of shape and density changes occurring during motion on the dose distribution. An ideal plan would require a continuous set of calculations along the tumour trajectory. None of the centres visited in this study were clinically performing a 4D dose calculation which could have theoretically reduced the deviation found between planned and measured doses. Nonetheless studies with ITV 4D dose calculations differ in conclusion on the significance of 4D calculation advantage in comparison to 3D dose calculations. Ehrbar et al. [15] have found that the dosimetric difference between 3D and 4D calculations for lung using a rigid translation approach was within -2.1% to $+1.4\%$ which therefore could only partially explain the larger deviations observed here. Ehrbar et al. also found that 3D calculations always underestimated the dose compared to 4D calculations. In our study we also found an underestimation (of the median deviation), since all the institutions were performing 3D calculations. On the other hand Vinogradskiy et al. [5] compared in their study 3D and 4D dose calculations to TLD and gafchromic film measurements on a semi-anthropomorphic

breathing phantom in rigid and deformable motion cases. They found a significant ($p < 0.001$) gamma pass rate (5%/3 mm) improvement with 4D calculation over 3D calculation for different breathing patterns (regular and irregular). The significantly different results might be explained by the varying quality of the 4D dose accumulation algorithms as described in Li et al. [16]. They found differences in the PTV summed minimum dose up to 11% when comparing various dose accumulation algorithms.

Another known cause of deviation between planned and delivered dose in dynamic irradiations is the (de)synchronization between MLC motion and target motion also known as interplay effect. Ong et al. [17] evaluated this effect in RapidArc flattened beam SBRT lung treatments and found this to be negligible; $< 3\%$ gamma pass rate (3%/1 mm) for 1 fraction delivered using two arcs. They also found that this deviation diminished when increasing the number of fractions. However, the interplay effect is more important when the irradiation time is shortened as it is for Flattening-filter-free irradiations. The decrease in breathing cycle repetition within a fraction might increase the dosimetric inaccuracy due to the uncertainty of the start breathing phase. Li et al. [6] found the combination of motion and interplay effect within SBRT FFF irradiations to be significant except in the case where margins from ITV to PTV used were equal or superior to 5 mm. Ong et al. [7] found significant deviations for FFF SBRT irradiation at 2400 MU/min in case of single arc irradiation. They

a



b

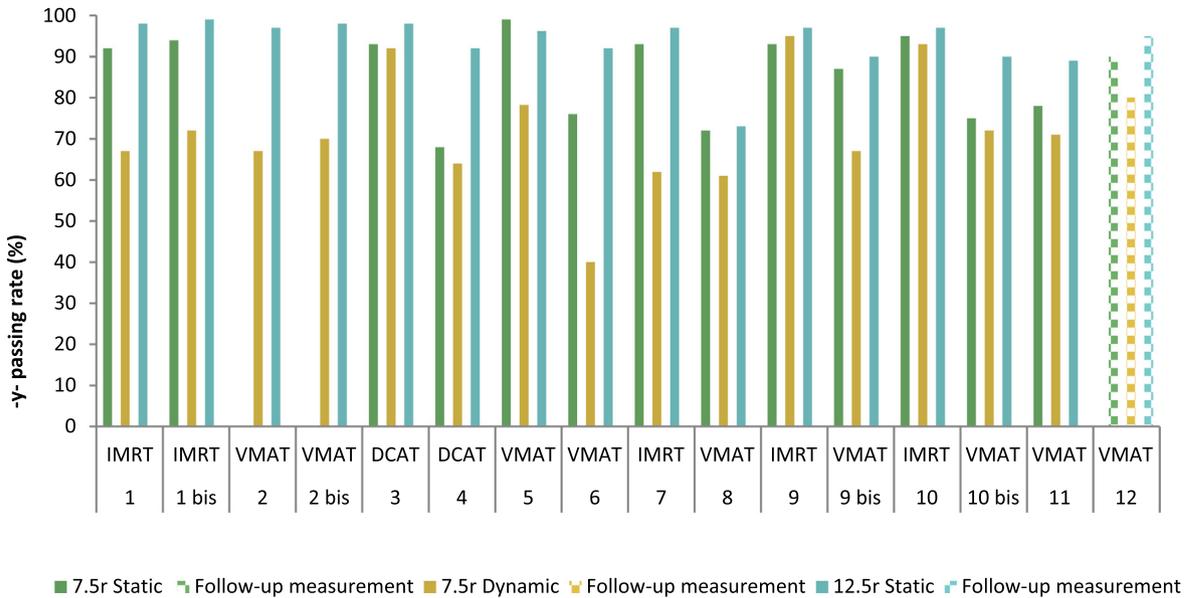


Fig. 3. (a) Results of the measured dose (IC) deviation (%), (b) results of the film γ passing rate (%).

Table 3

Spearman's correlation coefficients (*P*-values) calculated between the IC results and the γ pass rate for each plan.

	Plan 1 IC/ γ	Plan 2 IC/ γ	Plan 3 IC/ γ
Coefficients (<i>p</i> -values)	-0.78 (0.015)	-0.37 (0.130)	-0.48 (0.407)

Significant values are given in bold.

found a gamma pass rate ranging from 83.1% to 96.9%. The gamma pass rate when making use of a second arc was then improved to a range of 92.7–99.2% and was systematically superior to 97% when

summing several fractions. Our study is based on a single fraction measurement, within the participation institutions, two used FFF 6 MV, institution 8 used three partial arcs and had IC pass rates for the dynamic plan of 2.8%, and γ pass rates of 61%. Institution 10 in their bis measurement used one FFF double arc and had an IC dose deviation for the dynamic plan of 5.2% and a γ pass rate of 72%. These results seem to agree that interplay is worse for FFF treatments.

In our study we observed that VMAT delivery discretization (spacing between control points) was responsible for the deviations observed in institution 12. This importance of VMAT discretization was also highlighted as a prime accuracy factor in the

Table 4
Spearman's correlation coefficient (*p* values) between planning parameters and IC dose deviation and γ pass rates.

Coefficients (<i>p</i> -values) for the tested variables	Total MU	PTV volume	Slice thickness	Dose grid
IC dose deviations	0.72 (0.003)	0.34 (0.196)	0.27 (0.319)	0.46 (0.079)
γ pass rate	-0.59 (0.020)	0.21 (0.423)	0.15 (0.574)	-0.43 (0.094)

Significant values are given in bold.

simulation studies of by Sothmann et al. [18] where they advised the use of 2.3 degree or less which was the minimum they could use. This seems to be an important factor to consider when generating such plans and more research on this topic is warranted. Control point spacing seems to be less critical when a dynamic conformal technique is used, as the results for institution 3 using a spacing of 6 degrees were very good.

In addition, calculation inaccuracies have been reported with type B, while being of course of smaller magnitude than those with type A algorithms; these deviations have been reported to be around 3% in inhomogeneous fields. Kry et al. [19] studied CC and AAA algorithms and showed a systematic overestimation of the dose at the centre of the target, by 3.7% on average. Dunn et al. [20] similarly presented results that demonstrate a systematic deviation of $2.9 \pm 1.2\%$ in regions distal from lung-tissue interfaces when the AAA algorithm was used. However, these results were all based on static measurements.

In our study we also found a significant negative correlation between the number of MUs delivered and both the IC deviation and the γ pass rate. This correlation might theoretically be explained by the fact that a high number of MUs generally indicates the use of smaller field sizes which are known in heterogeneous low density media to be difficult to model with type B algorithms. A similar correlation was also found to be significant by Shen et al [21] for γ pass rates and delivered MUs.

Recently several multicentre studies of related topics have been published by groups mainly located in the UK, such as Distefano et al. [9], Tsang et al. [22], Palmer et al. [11] and Eaton et al. [10]. Distefano et al. performed a national audit that showed deviations for a static anthropomorphic phantom of $0.4 \pm 1.4\%$ for alanine measurements and $98.6\% \pm 3.4\%$ for EBT3 films. The results of the static irradiations presented here are slightly inferior to Distefano's results. However differences in methodology might explain the discrepancy; such as the fact that in the Distefano audit images of the phantom and delineation of the target were provided by the auditing team to the visited centres. Their CT acquisition had a slice thickness of 1.25 mm. In contrast, in the study presented here, the phantom was scanned at each institution integrating in our results the inter-institution variability in imaging quality. For comparison, in this study, only one institution reconstructed CT slices with a thickness <2 mm. The slice thickness could not be proven to be a significant parameter to influence the measurements results in our study, however, the slice thickness variation might be too small in our dataset to detect such influence. Also in this study the GTV volume was delineated at each institution by a physician and the PTV volume was generated according to the institution own protocol. The volume of the static ITV in the UK study was 24.9 cm^3 which is larger than that in our study where the average PTV volume was 8.1 cm^3 for the small sphere and 23.3 cm^3 for the large sphere. Therefore if we only include the data from the larger sphere, our results (IC median: $1.1\% [-2.8\%; +5.0\%]$, Film median: $96.2 [73.0, 99.0]$) will be in line with the Distefano study.

The study presented by Tsang et al. was performed with the ArcCheck™ phantom where in the centre a Multiplug™ insert of 15 cm diameter of lung equivalent density was placed. All institutions used a preset benchmark patient plan that they irradiated on

this phantom. Their results using the ArcCheck™ and a 3%/3 mm γ analysis criterion were 99.0% [98.3–99.7] and with EBT3 films 94.7% [92.5–96.8]. It is likely that the absence of a lung like density inhomogeneity partly explains their better results. Palmer et al. presented the results of a multicentre QA study for non SBRT treatments with a non-spherical volume of 14 cm^3 . The amplitude of motion used in their study was half of the amplitude used in the present study. From their paper, it is not clear if calculation and measurement were scaled to the prescription dose. These differences might also explain their apparent higher accuracy with γ pass rate of 99.4 (3%/3 mm) in static condition compared to 96.2% (3%/3 mm) for the large sphere in static condition in our study. Palmer et al. also found a significant decrease in the gamma pass rate between static and dynamic irradiations.

The UK studies presented above resulted in comparable accuracy within the visited institutions. In our study the visited institutions are spread over six different countries, which could have contributed to the larger inter-institution variations found.

This study is unprecedented in evaluating lung SBRT in a multicentre European setting with an anthropomorphic respiratory phantom. In term of limitations of our study, the scanning of the films on the scanner used here has been reported to carry an optical density inaccuracy of (+3 to +5%) which might lead to that an overdosage error of +5% [23]. We could not trace why the film measurement failed for one centre. Another limitation is the relatively small amount of datasets compared to the wide range of parameters influencing the results. However, it is important to stress that testing the correlation between planning and irradiation parameters was out of the scope of the clinical trial RTQA procedure. We also only tested one motion with an asymmetric regular motion. Finally no data on the history of each centre with lung SBRT were collected, which could be interesting, as self-improvement of institutions with a specific technique over time is likely. As an example of self-improvement, Xu et al. tested different collimator angles and start breathing phases for their tracking conformal arc technique. They concluded that their EBT3 film gamma pass rate improved by using optimal sets of collimator angle, couch angle and start tracking phase [24].

In conclusion the results of our study not only demonstrate the feasibility of high planning accuracy for static and dynamic irradiations as seen in some institutions but also highlight the need for more QA in others. This exercise did lead to the discovery of deviations and, maybe more importantly, lead to improvements in clinical practice. The study overall calls for an increase in inter-institution exchange of practices and an increase in external audits to uncover deviations that might not have been detected using the institutions own QA equipment. As discussed, further detailed investigations are warranted to determine exactly which plan and delivery parameters are essential to improve the overall SBRT delivery quality. Consensus and standardization of lung SBRT QA would help to set comprehensive acceptance criteria within and also outside multicentre trials.

Declaration of Competing Interest

The authors have no conflict of interests to declare.

Acknowledgements

All the people from the centres visited which helped in the preparation and the execution of the RTQA visit are gratefully acknowledged for all their time and effort.

During the conduct of the study a PhD grant from Elekta was received.

References

- [1] Adebahr S, Collette S, Shash E, et al. LungTech, an EORTC Phase II trial of stereotactic body radiotherapy for centrally located lung tumours: a clinical perspective. *Br J Radiol* 2015;88:20150036.
- [2] Chang JY, Senan S, Paul MA, et al. Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a pooled analysis of two randomised trials. *Lancet Oncol* 2015;16:630–7.
- [3] Horne ZD, Richman AH, Dohopolski MJ, et al. Stereotactic body radiation therapy for isolated hilar and mediastinal non-small cell lung cancers. *Lung Cancer* 2018;115:1–4.
- [4] Lambrecht M, Melidis C, Sonke JJ, et al. Lungtech, a phase II EORTC trial of SBRT for centrally located lung tumours – a clinical physics perspective. *Radiat Oncol* 2016;11:7.
- [5] Vinogradskiy YY, Balter P, Followill DS, et al. Comparing the accuracy of four-dimensional photon dose calculations with three-dimensional calculations using moving and deforming phantoms. *Med Phys* 2009;36:5000–6.
- [6] Li X, Yang Y, Li T, et al. Dosimetric effect of respiratory motion on volumetric-modulated arc therapy-based lung SBRT treatment delivered by TrueBeam machine with flattening filter-free beam. *J Appl Clin Med Phys* 2013;14:4370.
- [7] Ong CL, Dahele M, Slotman BJ, Verbakel WF. Dosimetric impact of the interplay effect during stereotactic lung radiation therapy delivery using flattening filter-free beams and volumetric modulated arc therapy. *Int J Radiat Oncol Biol Phys* 2013;86:743–8.
- [8] Nishio T, Shirato H, Ishikawa M, et al. Design, development of water tank-type lung phantom and dosimetric verification in institutions participating in a phase I study of stereotactic body radiation therapy in patients with T2N0M0 non-small cell lung cancer: Japan Clinical Oncology Group trial (JCOG0702). *J Radiat Res* 2014;55:600–7.
- [9] Distefano G, Lee J, Jafari S, et al. A national dosimetry audit for stereotactic ablative radiotherapy in lung. *Radiother Oncol* 2017;122:406–10.
- [10] Eaton DJ, Tyler J, Backshall A, et al. An external dosimetry audit programme to credential static and rotational IMRT delivery for clinical trials quality assurance. *Phys Med* 2017;35:25–30.
- [11] Palmer AL, Nash D, Kearton JR, Jafari SM, Muscat S. A multicentre 'end to end' dosimetry audit of motion management (4DCT-defined motion envelope) in radiotherapy. *Radiother Oncol* 2017;125:453–8.
- [12] Weber DC, Vallet V, Molineu A, et al. IMRT credentialing for prospective trials using institutional virtual phantoms: results of a joint European Organization for the Research and Treatment of Cancer and Radiological Physics Center project. *Radiat Oncol* 2014;9:123.
- [13] Wen N, Lu S, Kim J, et al. Precise film dosimetry for stereotactic radiosurgery and stereotactic body radiotherapy quality assurance using Gafchromic EBT3 films. *Radiat Oncol* 2016;11:132.
- [14] Knoos T, Wieslander E, Cozzi L, et al. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. *Phys Med Biol* 2006;51:5785–807.
- [15] Ehrbar S, Perrin R, Peroni M, et al. Respiratory motion-management in stereotactic body radiation therapy for lung cancer – A dosimetric comparison in an anthropomorphic lung phantom (LuCa). *Radiother Oncol* 2016;121:328–34.
- [16] Li HS, Zhong H, Kim J, et al. Direct dose mapping versus energy/mass transfer mapping for 4D dose accumulation: fundamental differences and dosimetric consequences. *Phys Med Biol* 2014;59:173–88.
- [17] Ong C, Verbakel WF, Cuijpers JP, Slotman BJ, Senan S. Dosimetric impact of interplay effect on RapidArc lung stereotactic treatment delivery. *Int J Radiat Oncol Biol Phys* 2011;79:305–11.
- [18] Sothmann T, Gauer T, Werner R. 4D dose simulation in volumetric arc therapy: accuracy and affecting parameters. *PLoS One* 2017;12:e0172810.
- [19] Kry SF, Alvarez P, Molineu A, et al. Algorithms used in heterogeneous dose calculations show systematic differences as measured with the Radiological Physics Center's anthropomorphic thorax phantom used for RTOG credentialing. *Int J Radiat Oncol Biol Phys* 2013;85:e95–100.
- [20] Dunn L, Lehmann J, Lye J, et al. National dosimetric audit network finds discrepancies in AAA lung inhomogeneity corrections. *Phys Med* 2015;31:435–41.
- [21] Shen L, Chen S, Zhu X, et al. Multidimensional correlation among plan complexity, quality and deliverability parameters for volumetric-modulated arc therapy using canonical correlation analysis. *J Radiat Res* 2018.
- [22] Tsang Y, Carver A, Groom N, Harris C, Faivre-Finn C. A multi-centre dosimetry audit on advanced radiotherapy in lung as part of the Isotoxic IMRT study. *PhiRo* 2017;4:17–21.
- [23] Lewis D, Chan MF. Correcting lateral response artifacts from flatbed scanners for radiochromic film dosimetry. *Med Phys* 2015;42:416–29.
- [24] Xu Z, Yao R, Podgorsak MB, Wang LZ. Effects of collimator angle, couch angle, and starting phase on motion-tracking dynamic conformal arc therapy (4D DCAT). *J Appl Clin Med Phys* 2017;18:54–63.