ESTRO project

An inter-centre quality assurance network for IMRT verification: Results of the ESTRO QUASIMODO project

Sofie Gillis*, Carlos De Wagter, Joerg Bohsung, Bruce Perrin, Peter Williams, Ben J. Mijnheer

*Department of Radiotherapy, Ghent University Hospital, Gent, Belgium
bDepartment of Radiotherapy, Charite Campus Mitte, Berlin, Germany
cNorth Western Medical Physics, Christie Hospital NHS Trust, Manchester, UK
dDepartment of Radiotherapy, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
eSiemens, Heidelberg, Germany

Abstract

Background and purpose: IMRT necessitates extension of existing inter-centre quality assurance programs due to its increased complexity. We assessed the feasibility of an inter-centre verification method for different IMRT techniques.

Materials and methods: Eight European radiotherapy institutions of the QUASIMODO network, have designed an IMRT plan for a horseshoe-shaped PTV surrounding a cylindrical OAR in a simplified pelvic phantom. All centres applied common plan objectives but used their own equipment for planning and delivery. They verified the delivery of this plan according to a common protocol with radiographic film and ionisation chamber measurements. The irradiated films, the results of the ionisation chamber measurements and the computed dose distributions were sent to one analysis centre that compared the measured and computed dose distributions with the gamma method and composite dose–area histograms.

Results: 4% (relative to the prescribed dose) and 3 mm (distance-to-agreement) were decided feasible gamma criteria. The composite dose–area histograms showed a maximum local deviation of 3.5% in the mean dose of the PTV and 5% in the OAR. Systematic differences could be identified, and in some cases explained.

Conclusions: This multi-centre dosimetric verification study demonstrated both the feasibility of a multi-centre quality assurance network to evaluate any IMRT planning and delivery system combination, as well as the validity of the methodology involved.

Keywords: IMRT delivery; QA; Radiographic film; Multi-centre intercomparison; Entire treatment chain

Intensity modulated radiation therapy (IMRT) allows to deliver radiation dose more conform to irregularly shaped target volumes. A high geometric and dosimetric accuracy is required for these advanced techniques and the verification of the delivery of these IMRT dose distributions is a prerequisite for a safe and efficient application.

In literature as well as in a survey amongst European centres performing IMRT, it has been demonstrated that currently each institution applies its own quality assurance (QA) methods, thus complicating intercomparisons between institutions. As the ESTRO QUASIMODO (Quality Assurance of Intensity Modulated radiation Oncology) network aimed at developing and disseminating more uniform guidelines for validation of newly designed IMRT techniques, a standardized verification test was set up that dosimetrically validates the IMRT planning and delivery chain. The proposed independent verification test for IMRT can be considered as an extension of existing external dosimetry quality audits organised by international organisations such as IAEA and ESTRO. The relevance of these audits has been clearly shown in the past.

Realising that each institution uses its own specific equipment and method for planning and delivery, with the inherent variability of features and restrictions, the dose distribution due to the complete treatment fraction was judged to be a better test item than the multiple dose distributions from the separate beams. In this test, a range of IMRT prescriptions was generated by the participating centres for a fictive clinical case, however, all aimed at the same dosimetric objectives. The IMRT dose distribution was measured in a number of planes and compared to the planned dose distribution. Radiographic film was preferred as detector material, as high spatial accuracy was required and mailed film dosimetry was already judged to be feasible in the context of external audits of photon beams.

It was the aim of this study to assess the feasibility of such an IMRT verification method for a wide range of solutions.
Hence, the method can be considered more solid and applied with more confidence. This paper reports the findings and results of IMRT delivery verification in eight participating centres.

Materials and methods

QUASIMODO network

The QUASIMODO network belongs to ESTRO’s ESQUIRE project and is supported by 15 radiotherapy institutions from nine European countries. The network involves major linac manufacturers and treatment planning systems (Table 1). The first aim of this network was to draft a list of practical commissioning tests for treatment planning systems for non-IMRT [14]. The second goal was to draft guidelines for the verification of IMRT. In this, a QA test for the complete planning and delivery chain workable in any centre is crucial. This paper reports the findings and results of IMRT delivery verification part of the project.

Verification protocol

A simplified ‘pelvis’ phantom (CarPet) was constructed that contained a virtual concave planning target volume (PTV) that surrounded a cylindrical organ at risk (OAR). The CarPet phantom (Fig. 1(a)) consists of 16 identical transverse polystyrene (Polystyrol 495F, BASF, Germany) slabs each with a thickness of 1 cm and a density of 1.04 g/cm³. These slabs are held together with two clamping screws [21]. In each slab, five holes with a diameter of 25 mm were drilled to allow ionisation chamber measurements, using a special holder, or to reference mark the films with a customized prick pin. Films can be put between the slabs and are irradiated parallel to the central axis of coplanar beams.

Table 1

<table>
<thead>
<tr>
<th>No. of treatment verification</th>
<th>Linac manufacturer</th>
<th>Photon beam quality</th>
<th>IMRT method</th>
<th>TPS</th>
<th>Computational grid size X×Y×Z (mm×mm×mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Varian</td>
<td>6 MV</td>
<td>Sliding window</td>
<td>Cadplan/Helios V6.3.6</td>
<td>2.5×2.5×2.5</td>
</tr>
<tr>
<td>2</td>
<td>Siemens</td>
<td>6 MV</td>
<td>Step and shoot</td>
<td>Helix-TMS V6.1.a 1</td>
<td>1.5×1.5×2</td>
</tr>
<tr>
<td>3</td>
<td>Elekta</td>
<td>18 MV</td>
<td>Step and shoot</td>
<td>Pinnacle V6.2b</td>
<td>4×4×4</td>
</tr>
<tr>
<td>4</td>
<td>Elekta</td>
<td>15 MV</td>
<td>Step and shoot</td>
<td>Hyperion</td>
<td>4×4×4</td>
</tr>
<tr>
<td>5</td>
<td>Varian</td>
<td>20 MV</td>
<td>Sliding window</td>
<td>Eclipse/Helios V7.1.35</td>
<td>2.5×2.5×2</td>
</tr>
<tr>
<td>6</td>
<td>Elekta</td>
<td>8 MV</td>
<td>Step and shoot</td>
<td>Pinnacle V6.2b</td>
<td>3×3×3</td>
</tr>
<tr>
<td>7</td>
<td>Elekta</td>
<td>10 MV</td>
<td>Step and shoot</td>
<td>Plato ITP</td>
<td>1.95×1.95×2</td>
</tr>
<tr>
<td>8</td>
<td>Elekta</td>
<td>18 MV</td>
<td>IMAT</td>
<td>Pinnacle + in-house optimization</td>
<td>4×4×4</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>Step and shoot</td>
<td>Helix-TMS V6.1.a 1</td>
<td>1×1×2</td>
</tr>
</tbody>
</table>

Additional information of the different treatment verifications: linac manufacturer, beam quality, delivery method, TPS and computational grid size

Planning procedure

A common CT dataset of the CarPet phantom was acquired with pixel size of 0.98 mm and slice thickness of 2 mm. An anterior-posterior shift in the longitudinal direction of 0.25 cm/cm distance was applied to the horseshoe-shaped PTV around the circular OAR contours (Fig. 1(a)). The CT dataset and the contours of the volumes of interest were made available to the participating centres in DICOM-CT format and DICOM-RT-Structure format, respectively. A set of dosimetric planning objectives was defined. The mean dose to the PTV had to be 2.00 Gy, the prescribed dose. The fractional volume of the PTV that received more than 95% of the prescribed dose (V95) had to be more than 99% but the percentage of the volume that exceeded 105% of the prescribed dose (V105) had to be lower than 5%. For the OAR it was stipulated that the percentage of the volume that received at least 70% of the prescribed dose (V70) had to be lower than 1%. For the partial volume of the body without the PTV, which received more than 80% of the prescribed dose (V80), a maximum of 15% was allowed. The maximum dose (Dmax) in this volume had to be lower than 105% and the partial volume that received more than the prescribed dose had to be less than 2%. The incident beams had to be all of the same beam quality but no recommendations were made about the beam arrangements and the beam quality itself.

1 CarPet is formed by the concatenation of the first three letters of the forenames of two members of the QUASIMODO group, CDW and PW, as the test tool design was based on their suggestions.
The dose calculation grid had to be less than 3 mm in all directions. Finally, it was suggested to keep the treatment delivery time within 20 min. All participating centres were asked to implement this information into their TPS and to attempt to reach the common dosimetric objectives using their own specific planning software. The resulting computed dose distributions were sent in DICOM-RT-Dose or RTOG format to the processing centre. In a forthcoming paper, a detailed comparative analysis of the treatment plans will be presented.

**Experimental dosimetry**

For the verification, EDR2 (Eastman Kodak Co., Rochester, NY, USA) was chosen as this radiographic film can handle a dose of more than 2.00 Gy without saturation [2]. All films used within one verification experiment were of the same emulsion batch and all jackets were punctured at the four corners to avoid air pockets [23].

In each participating centre, the calibration films, in their jacket, were irradiated perpendicularly to the central axis with a field of 5 cm × 5 cm at 10 cm depth in a layered water-equivalent phantom. The source-surface distance (SSD) was 90 cm. Four fields were given per film in a predefined position so that successive dose points in the calibration curve came from the same film. Five films were needed to construct a calibration curve. Film 1 was a blank film. Film 2 was irradiated to 10, 20, 30, and 40 cGy. Film 3 was exposed to 50, 60, 70, and 80 cGy. Film 4 received 100, 120, 140, and 160 cGy. Film 5 was exposed to 180, 200, 220, and 240 cGy.

At the processing centre, the seven treatment films were resized (Fig. 1(a)) in the dark room and the jackets thermally resealed in the processing centre. This procedure ensured that the films’ positions in the jacket were well defined in the anterior-posterior direction of the phantom and limited attenuation by pieces of film sticking out of the phantom as much as possible (Validity tests). At the time of treatment in the participating centre, the phantom was filled with seven films every 2 cm. Every film was carefully aligned to the edge of the phantom in the anterior-posterior direction through feeling the edge of the film through the envelope. Subsequently, the next two slabs, filled with dummy disks except for the left- and right-hand holes (Fig. 1), were repeatedly

Fig. 1. (a) Transverse and sagittal view of the polystyrene slab (CarPet) phantom conceived for the dosimetric verification of IMRT of prostate cancer. Indicated is the planning target volume (in black) surrounding an organ at risk. The sagittal view denotes the tilt in the (longitudinal) Z-direction. In the experiments described, the seven films are at Z = 60 mm, Z = 40 mm, Z = 20 cm, Z = 0 mm, Z = 20 mm, Z = −40 mm, Z = −60 mm. (b) Central CT-slice with contours of OAR and PTV and applied co-ordinate system.

![Action in participating centre](image1.png)

![Action in processing centre](image2.png)

Fig. 2. Flow chart of the verification procedure. The steps included in the two reproducibility tests are also indicated.
piled on top of each film. To register the exact position of the underlying film against the phantom, the film was pin-pricked through the remaining holes using a dedicated pricking rod. Afterwards the holes were filled with dummy disks. As a result, seven films were positioned between slabs 2–3, slabs 4–5, slabs 6–7, slabs 8–9, slabs 10–11, slabs 12–13, and slabs 14–15, respectively. Every film was pin-pricked in a specified corner for identification. The phantom was screwed tightly together to adequately squeeze the films within the phantom. On the treatment couch, the phantom was positioned using the isocentric lasers and the cross lines that indicated the centres of the outer slices. The exposed but still undeveloped calibration and treatment films were sent back to the processing centre by postal mail.

The aimed accuracy of the verification method required ionisation chamber measurements in addition to the film calibration, as is described in Data processing at the processing centre. Each participating centre performed two measurements with their own 0.6 cm² Farmer-type ionisation chamber (applying their own beam calibration protocol) in the central slice, a first one in the PTV and a second one in the OAR. The CarPet phantom was positioned on the treatment couch in the same way as mentioned above. A polystyrene cylindrical adaptor to hold the ionisation chamber was slid in the hole that was created by removing the dummy disks in the PTV and OAR volume, respectively. The cylinder was positioned in such a way that the effective measuring point of the ionisation chamber was situated just in the central plane.

Data processing at the processing centre

At the processing centre, all computed dose distributions were specified in a common co-ordinate system as shown in Fig. 1. As the computation grid was coarser than the film scanning grid in X- and Y-direction, the three-dimensional computed dose distribution was interpolated trilinearly into a grid of 1 mm × 1 mm × 2 mm, i.e. between the scanning grid size and the computational grid size.

For any delivery verification, all irradiated films were developed in the processing centre in one session and in arbitrary order to randomise the effects of possible drifts in temperature or composition of the developer solution. An automatic film processor (KODAK RP X-Omat processor model M7B) was used with standard RP X-Omat chemicals. The temperature of the developer solution was 32 °C. Optical density (OD) data were obtained with a Vidar VX-R12 digitiser (Vidar Systems Corporation, Herndon, VA, USA) equipped with the standard broadband fluorescent tube (Philips F17T8/TLB41). The features of this 12-bit CCD based film digitiser have been described extensively by Mersseman and De Wagter [13]. All films were digitised with a speed of 14 ms/line and at a resolution of 0.34 mm (75 dpi). The resulting 16-bit TIFF images comprehended all information of the 12-bit CCD. The procedure of achieving OD from the digitiser raw data can be found in [12].

The calibration TIFF images were divided into quadrants. In each quadrant, the contour of the field was searched automatically and the geometric centre was determined. Subsequently, the readings were averaged over the central 5 × 5 pixels (1.7 mm × 1.7 mm) and converted to OD. A third-order polynomial curve was fit through the OD-dose points, yielding the calibration curve and its analytical expression. Next, the OD-data of the treatment films were converted to film dose through this calibration curve.

Although the calibration of films allowed for the determination of absolute doses, there was a residual uncertainty caused by other factors including irradiation geometry. The calibration films were irradiated normal to the radiation beams whereas the treatment films were irradiated parallel to the beams [3]. These uncertainties were minimised by rescaling of doses determined from the films by ionisation chamber measurements as outlined in the next paragraph. However, the film calibration process was still necessary to correct for the non-linear response of the films.

In order to scale the film doses to the two absolute dose values determined with the ionisation chamber, a two-parameter linear conversion was conceived resulting in a linear rescaling that was applied to all treatment films

\[ D_m = a \times D_{\text{film}} + b \]  \hspace{1cm} (1)

with

- \( D_m \) dose rescaled to ionisation chamber measurements
- \( D_{\text{film}} \) film dose, i.e. dose derived from film OD through calibration curve

The parameters \( a \) and \( b \) were derived by identifying the two chamber-measured dose values to the film doses averaged over 5 mm × 5 mm squares in the longitudinally central film where the two chamber measurements had been performed. As each centre carried out the ionisation chamber measurements with their own chamber using their own beam calibration protocol, errors in beam calibration were not taken into account by this procedure.

The two pin-prick marks on the film allowed to correct for possible rotational misalignment. Subsequently, a co-ordinate system was attached to each film. The position \( X = Y = 0 \) mm was situated in the middle of the two prick marks. The Z-value was determined according to the position of the film in the phantom. \( X \) and \( Y \) were directed as for the computed dose distribution. In a next step, a bilinear interpolation was made to a grid of 1 mm × 1 mm resulting in the same grid size as for the interpolated computed dose distribution.

Individual and inter-centre comparison

As both geometric and dosimetric accuracy are important in IMRT, we decided to use the gamma (γ) index approach [10,11]. The dosimetric criterion is a dose-difference presented as a percentage of the prescribed dose. In the terminology of Low and Dempsey [11], the measured dose distribution was taken as reference and the computed dose distribution was appointed as to be evaluated. If \( \gamma(i) < 1 \) the dose delivered at point \( i \) is considered to be within the tolerance criteria and is hence accepted with regard to the computed, i.e. intended, dose. It should be noted that lower γ-values are obtained by considering the full three-dimensionality of the calculated dose matrix and hence by incorporating dose variation in the direction perpendicular to the film, making the verification more realistic in case of longitudinal dose gradients. Adequate
values for the tolerance criteria, will be estimated based on the experiments outlined in Sections Reproducitivity tests and Evaluation of the gamma tolerance criteria.

In addition, dose-volume histograms are of clinical interest. However, as film provides only two-dimensional information, it was not possible to derive identical dose-volume histograms in the measured and computed dose distribution. To get an idea of the correspondence between computed and measured dose distributions in the PTV and OAR, a dose-area histogram (DAH) was made per Z-position. The common 1 mm × 1 mm grid per Z-position was used for the calculated as well as for the measured DAH. In this way, both DAHs were derived from identical spatial information. In order to represent those extensive data compactly, also a composite DAH was constructed containing all information of the five evaluated Z-positions (Z = −40 mm, Z = −20 mm, Z = 0 mm, Z = 20 mm, Z = 40 mm). Therefore, all dose points from the different Z-positions were considered, each with equal weights.

Validity tests

As four calibration fields were registered on one film, the OD value of one field could be influenced by the three other fields (Section Experimental dosimetry), and in this way a systematic error could be introduced in the calibration curve. Therefore, a calibration curve, assembled according to the protocol, was compared to a calibration curve obtained by a one-field-per-film approach. All calibration films were irradiated on the same day and were developed and scanned in a randomised order.

As reported by many groups [2,5,7], air surrounding radiographic film, irradiated parallel to the central axis, will influence its response strongly. As films in the CarPet phantom were irradiated parallel to the axis of coplanar beams, it was worthwhile to assess the magnitude of this effect as a function of depth. A calibration curve was constructed for perpendicular exposure at a depth of 5 cm in a layered polystyrene phantom. The SSD was 95 cm and 15 dose points between 0 and 180 cGy were used. Each field had a size of 10 cm × 10 cm and the beam quality was 6 MV (Elekta, Crawley, UK). For the parallel exposure, a film was sandwiched between two polystyrene blocks of 50 cm × 33 cm × 9 cm. The edge of the film was positioned at the surface of the phantom and marks were made at 5 cm depth to record the orientation of the film and its relative position to the phantom edge. The gantry was rotated to 90°, a field size of 2 cm × 15 cm was set and 100 MUs were delivered. The phantom was positioned so that the film was parallel to the central axis and the SSD was 95 cm. Three parallel-film positions were investigated: (i) containing the central axis, (ii) 2 cm, and (iii) 5 cm offset to the central axis. For each position, three films were irradiated to assess the reproducibility of the measurements. Nine other films were irradiated in the same way, but with one difference: these films were squeezed in the phantom to remove possible air pockets using two carpenter clamps near the anterior surface of the phantom and a third carpenter clamp near the posterior surface. After processing and scanning the films, the position of each film relative to the phantom and the irradiation field was reconstructed and depth-dose curves were acquired along the central lines of the beam’s film images using the calibration curve. The associated depth-dose curves were averaged and the two resulting curves per parallel-film position were compared to each other.

For each parallel-film position, the depth-dose curve measured with a diamond detector (PTW, Freiburg, Germany) was taken as reference. The validity of the dose-rate corrected depth-dose curve measured with a diamond detector was checked with a 0.125 cm³ ionisation chamber (PTW, Freiburg, Germany) for a 6-MV beam with field size of 2 cm × 15 cm and SSD of 95 cm on the central axis resulting in an average difference between the two curves of −0.04% and standard deviation of 0.23% for the depth range of 1–20 cm.

In addition, we ensured that for the film position containing the isocentre, the squeezed film showed the same response for parallel irradiation as for extra 1° gantry tilt in this rectangular phantom and by consequence a slightly different incidence on the film [20]. This justified the same handling for films positioned in the central transverse slice in the CarPet phantom as for films inserted at off-axis transversal positions.

As the original films do not conform to the phantom shape, it was worthwhile to investigate the perturbing effect caused by pieces of film that stick out of the phantom. Following the procedure described in the former paragraph, a calibration curve was constructed. The same phantom configuration and settings of the linear accelerator were taken. The orientations of the film were identical as described in the former paragraph except for the position of the film edge relative to the phantom surface. In a first situation, the edges of the films were positioned at the surface of the phantom. In the second situation, the films stuck out of the phantom by 5 cm. For each film-stick-out condition, the three positions of the film relative to the beam were considered (0, 2, 5 cm off-axis). During all irradiations, the films were squeezed with carpenter clamps and each film exposure condition was repeated three times.

Reproducibility tests

A complete set calibration and treatment films, were scanned twice with a time interval of 6 months to assess the reproducibility of the film scanning and data processing loop (Fig. 2). The complete procedure was followed to handle both sets. The same ionisation chamber measurements were used to rescale both sets. Instead of comparing the two resulting measured dose distributions to the computed one, they were compared to each other in two dimensions with the gamma evaluation method. A pair of tolerance criteria was searched for which the 95th percentile of gamma in the PTV and OAR was around 1 in order to estimate the contribution of the scanning error to the total uncertainty of the verification chain.

To evaluate the reproducibility of the complete verification loop (Fig. 2), one treatment delivery was repeated to a different CarPet phantom after 4 weeks. So for each verification experiment, a set of calibration films and a set of treatment films were irradiated and the aforementioned ionisation chamber measurements were conducted.
(Experimental dosimetry). Next, the sets were developed and scanned at different dates. Instead of comparing each set of measured data to the computed dose distribution, they were compared to each other using the gamma evaluation method in two dimensions. A set of tolerance criteria was searched for which the 95th percentile of the gamma in the PTV and the OAR was around 1.

The accuracy of the ionisation chamber measurements may have been affected during the IMRT treatments, as there were moments that the ionisation chamber was irradiated by narrow beam segments or was partially irradiated. The possible impact of this on absolute dosimetry has been investigated in terms of Monte Carlo computed stopping-power ratios [18] and ion chamber perturbation [1] at 6 MV. Based on these studies we neglected the possible effects in the context of this paper. In addition, centre of treatment verification 6 conducted the ionisation chamber measurements described in Section Experimental dosimetry for their treatment plan five times at different dates. The reproducibility of these measurements will be discussed.

Evaluation of gamma tolerance criteria

To obtain a reasonable value for the spatial gamma tolerance criterion, the spatial uncertainty, emanating from the verification procedure (phantom set-up) itself, rather than from the dose delivery, was assessed by repeating a complete verification experiment for the case of a simple dose distribution with sufficient geometric detail. The isocentre was defined in the centre of the CarPet phantom and a simple ‘box’ configuration consisting of four beams, with gantry positions of 0, 90, 180, and 270°, delivering each 50 cGy to the isocentre, was established. Each beam had a field size of 10 cm × 10 cm and the photon beam quality for all beams was 18 MV (Elekta, Crawley, UK). One resized film was positioned between the two central slabs. This ‘box’-treatment delivery verification was repeated once. The developed films were scanned and the resulting distributions were converted to dose as described in Section Data processing at the processing centre but without rescaling, as no ionisation chamber measurements were performed, because our main interest was the spatial accuracy. The resulting dose distributions were compared to each other in two dimensions for different values of the spatial gamma tolerance criterion in order to find a reasonable estimate.

In order to obtain a value for the dosimetric gamma tolerance criterion, the dosimetric accuracy related to the verification procedure was determined. Therefore, the whole verification procedure was conducted for one anterior 6-MV 200-MU beam with a field size of 2 cm × 15 cm delivered to the CarPet phantom filled with five resized films (between slabs 4-5, slabs 6-7, slabs 8-9, slabs 10-11, and slabs 12-13). The length of the field was in the longitudinal direction. The isocentre was situated between slabs 8 and 9, at a height of 5 cm above the phantom centre. A calibration curve was constructed as described in the protocol. The developed films were scanned, the resulting images were converted to dose and from each film, a depth-dose curve was derived along the central line of the beam’s film image. These film depth-dose curves were rescaled according to the doses measured in the centre of the field at a depth of 5 and 10 cm. For these curves, the root mean square deviations (RMSD) from the depth-dose curves measured with a diamond detector in water for the same configuration, were computed over the depth range of 1-20 cm with a step size of 0.5 mm.

Results and discussion

Validity tests

The comparison of the one-field-per-film and 4-fields-per-film calibration procedures indicates that the maximal local deviation in OD between the corresponding calibration points is less than 1% above an OD-value of 0.4. In terms of dose this means a maximal deviation of 0.6% relative to 2.00 Gy. Hence, the influence of irradiating the calibration films with four fields induces no significant uncertainty in the whole verification chain. The effect of clamping the films in the rectangular phantom on the derived depth-dose curves is presented in Fig. 3. The results reveal that air around the film enhances the film response with increasing importance at increasing depths and that squeezing is most effective for the film that contains the central axis. The under-response of the film at shallow depth is observed in all slices and is a major motivation for the two-parameter conversion as explained in Section Data processing at the processing centre. The effect of film sticking 5 cm out of the rectangular phantom is illustrated in Fig. 4. For the film that contains the central axis, the 5 cm film out of the phantom causes an overall attenuation of approximately 5% in the central axis depth-dose curve. This effect of attenuation diminishes as the distance of the film to the central axis increases. This observation would mean for the CarPet experiment that pieces of film sticking out of the phantom cause a differential weighting of the beams, as different beam incidences would imply different radiological path lengths in the pieces of film that stick out. Therefore, resizing the films to the shape of the CarPet phantom minimized the attenuation effect in the CarPet dose verification experiment.

Reproducibility tests

The reproducibility test of the film scanning and data processing loop included the fitting of the calibration curve, the scanning process, and the realignment of the films in the existing co-ordinate system (Fig. 2). The 95th percentile and the median of the gamma values with tolerance criteria of 1 mm and 1% in the OAR and PTV are shown in Table 2. The critical value of 1 was reached nowhere, ensuring the reliability of this loop and its small contribution to the overall uncertainty.

In order to evaluate the reproducibility of the complete verification loop, gamma evaluation plots for two different sets of tolerance criteria, {1 mm, 2%} and {2 mm, 2%} are displayed in Fig. 5. It can be observed that the spatial tolerance criterion of 1 mm cannot be maintained without unreasonably high dose tolerances. However, as can be observed from Table 2, tolerance criteria of 2 mm and 2% are feasible for the complete verification loop. The 95th percentile of the gamma values in the PTV exceeds 1 only in two of the five slices but the
overall 95th percentile stays below 1. For the OAR, no value of gamma above 1 in 95th percentile is recorded. It should be realized that in this experiment (Fig. 2), not only the verification loop itself contributes to the overall uncertainty, but also the reproducibility of the IMRT treatment delivery—a quantity to assess—has contributed, including differences in linac performance, for example beam symmetry, between the two deliveries.

The ionisation chamber measurements, carried out for reproducibility purposes, were corrected for deviations in output calibration of the linear accelerator. The standard deviations relative to the prescribed mean dose in the PTV were 0.98 and 0.74% for the OAR and PTV, respectively.

**Estimation of gamma tolerance criteria**

The spatial uncertainty emanating from the verification procedure itself was assessed by the comparison of the two film-measured box-treatment dose distributions with each other. When using gamma criteria of 1 mm and 1%, the high dose gradient regions are still accepted ($\gamma \leq 1$). Therefore,

![Central depth-dose curves showing the effect of clamping on the dose distribution along the central axis (a), at 2 cm off-axis (b), and 5 cm off-axis (c).](image-url)
we have to take a spatial uncertainty, inherent to the verification method itself, of 1 mm into account.

From the reproducibility test of the whole verification loop one can estimate that the spatial reproducibility of the whole verification procedure is around 2 mm. The increased spatial uncertainty with regard to the results of previous paragraph emanates mainly from the use of a different phantom for both verifications. An additional source of uncertainty is the geometric correlation of the films to the computed dose distribution, as the geometric centre of the phantom was not indicated in the CT-image set. Therefore, each participating centre had to define the isocentre by themselves resulting in some additional geometric uncertainty between film and computed-dose co-ordinate systems. Finally, in the estimation of the gamma spatial tolerance criterion also a tolerance for possible computational dosimetric and spatial errors should be incorporated. Considering all these uncertainties, we propose 3 mm as a realistic spatial tolerance criterion for the gamma index.

Fig. 4. Central depth-dose curves showing the effect of attenuation by parts of the film sticking out of the phantom on the central axis (a), 2 cm off-axis (b), and 5 cm off-axis (c).
To examine the dosimetric accuracy, i.e. the dose errors inherent to the film measuring method in combination with the ionisation chamber measurements, the diamond detector was used as gold standard. The root mean square deviation (RMSD), relative to the maximum dose for the central depth–dose curve was evaluated in each measured slice for the 2 cm\(^{15}\) cm field. The RMSD did not change with slice offset to the central axis if the film jackets were under mechanical pressure and the films are cut to the dimension of the phantom. The overall RMSD for all five slices was 1.1% and the maximum value per slice was 1.3%. In addition to this result, the reproducibility in dose of 2% has to be taken into account. Including a further tolerance for computational uncertainty, we consider 4% (relative to the prescribed dose) as a realistic dosimetric tolerance criterion for the gamma evaluation.

### Verification

#### Calibration curves

All calibration curves, constructed for the eight different centres, are shown in Fig. 6. All measurements were done with films from the same emulsion batch. Differences in calibration curves between different centres could be caused by different fading times between irradiation and development, by different energy response or by different composition of the developer solution at the time of development. Time elapsed between the first and last development session amounted to 8 months.

#### Typical evaluation results

Table 1 lists the beam quality, the IMRT method, and computational grid size applied in the nine treatment verifications analysed in this study. Despite the aimed computational grid size less than or equal to 3 mm in all directions, some centres used a coarser grid. For treatment verification 9, the gamma evaluation was performed originating from a computed dose grid of 2 mm\(\times\)2 mm\(\times\)2 mm and 4 mm\(\times\)4 mm\(\times\)2 mm. The results were compared and the differences were insignificant.

The film thickness of about 0.7 mm, jacket included, introduced a cumulative spatial film shift in the longitudinal direction of the CarPet phantom. Therefore, it was decided to exclude the outer films (\(Z = \pm 60\) mm), where the longitudinal dose gradient was maximal on top of that, from this analysis.

In Fig. 7(a–e), a gamma evaluation for the films of treatment verification 5 is shown for the criteria of 4% of the prescribed dose and 3 mm. The computed absolute dose distribution for the respective slice was superimposed in white on the coloured gamma distribution. In Fig. 7(f), a representative example of the composite DAHs for the PTV and OAR are shown for the same centre. Because the verification and computation composite DAHs for PTV and OAR are derived from dose distributions of the same spatial resolution, the curves are directly comparable to each other. For this particular treatment verification, one can observe in Fig. 7(f) that the PTV DAH derived from the computed dose distribution shows the same shape as the corresponding DAH derived from the measurements but is shifted towards lower

![Fig. 5. Distribution of the gamma index for the reproducibility study of the complete verification loop in film plane \(Z = 0\) mm using the \(1\) mm, \(1\)% criteria (left) and the \(2\) mm, \(2\)% criteria (right). The isodoses of the film-measured dose distribution have been superimposed. Gamma exceeds 1 at the holes that served as reference marks.](image)
doses. This was not confirmed by comparing the ionisation chamber measurement with the computed doses, but higher gamma values were reported for the outer slices (Table 4). As described in more detail in Section Inter-centre comparisons, the most plausible reason is probably the attenuation from a wrongly positioned moveable bar in the couch top. The OAR DAHs show slightly different forms more explicitly at the higher doses. The measured OAR DAH is globally shifted

![Fig. 6. Calibration curves for the nine treatment verifications. The number in brackets is the number of verification.](image)

![Fig. 7. Distribution of the gamma index in film planes Z=40 mm (a), Z=20 mm (b), Z=0 mm (c), Z=-20 mm (d), and Z=-40 mm (e) for {3 mm, 4%} criteria. The isodoses of the computed dose distribution have been superimposed in white. Panel (f) shows the corresponding composite dose–area histogram. (Results for treatment verification 5).](image)
Inter-centre comparisons

In Table 3, the mean dose values found in the composite DAHs for PTV and OAR are listed for the different treatment verifications. Generally, the agreement in these data for the PTV is very good, with a maximum local deviation, defined as the ratio of the difference between plan and verification mean dose value and the average value of plan and verification mean dose, of less than 3.5%, and a mean local deviation across the verifications of 1.4%. Local deviations in the OAR of up to 5.8% were seen. However, the OAR is likely to be shielded significantly in the majority of the calculated treatment fields. Thus limitations in the accuracy of planning system calculations at field edges and out of field are likely to add up causing such a reduction in agreement in the OAR.

In Table 4, the median and the 95th percentile of the gamma index characteristics (mean, median, 90th percentile, 95th percentile, and 98th percentile) between the five considered slices. However, the differences in all gamma index characteristics (mean, median, 90th percentile, 95th percentile, and 98th percentile) between the individual IMRT treatment verifications, including the two in the same centre, were statistically significant in the PTV. The gamma difference between the two dose distributions measured from the same treatment plan and used to test the reliability of the scanning data processing loop, was found to be non-significant (P=0.05). Interestingly, the two treatment verifications (8 and 9) that were performed in the same centre were also significantly different (P<0.05) in terms of gamma. In the OAR, on the other hand, the main effect of the factor ‘treatment verification’, i.e. the effect irrespective of the factor ‘film slice position’, was also significant (P<0.05), but not all gamma differences between the individual treatment verifications were found to be statistically significant. A plausible interpretation of this statistical outcome is that gamma is not affected by the offset distance of the film to the central plane.

As the purpose of verification is to detect problem areas in the intended dose distribution and its causes, the involved methods would show extra validity, if one could identify the reason for systematic unaccepted gamma values in the involved treatment verifications. Some examples of systematic deviations traced by this study are:

Table 3
Mean dose values and deviations for PTV and OAR derived from a composite DAH

<table>
<thead>
<tr>
<th>No. of treatment verification</th>
<th>PTV dose</th>
<th>OAR dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean plan (cGy)</td>
<td>Mean verification (cGy)</td>
</tr>
<tr>
<td>1</td>
<td>200.11</td>
<td>196.59</td>
</tr>
<tr>
<td>2</td>
<td>196.95</td>
<td>190.16</td>
</tr>
<tr>
<td>3</td>
<td>202.75</td>
<td>203.80</td>
</tr>
<tr>
<td>4</td>
<td>200.56</td>
<td>203.42</td>
</tr>
<tr>
<td>5</td>
<td>200.22</td>
<td>195.86</td>
</tr>
<tr>
<td>6</td>
<td>200.16</td>
<td>197.47</td>
</tr>
<tr>
<td>7</td>
<td>201.89</td>
<td>197.20</td>
</tr>
<tr>
<td>8a*</td>
<td>199.84</td>
<td>199.88</td>
</tr>
<tr>
<td>8b*</td>
<td>199.84</td>
<td>197.95</td>
</tr>
<tr>
<td>9*</td>
<td>200.77</td>
<td>194.16</td>
</tr>
</tbody>
</table>

*8 and 9 are two verifications performed at the same centre but with a different IMRT method. a and b refer to different dates of verification of the same plan.
All but two of the PTV and OAR slices of verification number 2 fall outside the criteria. The major reason was a higher response of the central film, which by the normalisation process introduced a systematic error in the other films.

The 40 mm slices of verification number 5 (Fig. 7(a) and (e)) show a reduction in the dose delivered at both ends of the PTV as well as a cool strip within the distribution that was seen on all slices of the delivered plan. The most plausible reason for this was found to be attenuation from a wrongly positioned moveable bar in the couch top which was not taken into account during dose calculation.

In the slices at Z = 20 mm and Z = 40 mm of verification number 7, the gamma values in the PTV fall outside the criteria and the gamma index in the OAR is close to or slightly above 1 in all slices. Analysis of the DAHs and profiles show a 4-5% systematic overestimate of the delivered dose by the TPS (Fig. 8). This was found to be due to an accumulation of errors: a 1.5% difference in linac calibration on the day of delivery and a less optimal kernel in the treatment planning system.

The differences highlighted for verifications 2, 7, and 9 also correspond to the larger differences seen in Table 3 for the DAH in the PTV and OAR.

Discussion of practical problems

The removal of air gaps in the film jackets is critically important since the susceptible central film that contains the central beam axes, serves as an intermediate step in rescaling the film doses. An additional issue is that dummy disks flank the PTV and OAR zones where film dose is

Table 4
Statistics of the gamma index for [3 mm, 4%] criteria for the treatment verifications in the eight participating centres

<table>
<thead>
<tr>
<th>no of treatment verification</th>
<th>PTV</th>
<th>OAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>95th percentile</td>
</tr>
<tr>
<td></td>
<td>Z = 40 mm</td>
<td>Z = 20 mm</td>
</tr>
<tr>
<td>1</td>
<td>0.36</td>
<td>0.44</td>
</tr>
<tr>
<td>2</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>4</td>
<td>0.24</td>
<td>0.32</td>
</tr>
<tr>
<td>5</td>
<td>0.28</td>
<td>0.35</td>
</tr>
<tr>
<td>6</td>
<td>0.37</td>
<td>0.50</td>
</tr>
<tr>
<td>7</td>
<td>0.89</td>
<td>0.71</td>
</tr>
<tr>
<td>8a*</td>
<td>0.35</td>
<td>0.25</td>
</tr>
<tr>
<td>8b*</td>
<td>0.21</td>
<td>0.28</td>
</tr>
<tr>
<td>9*</td>
<td>0.41</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Grey background shows gamma values larger than one. *8 and 9 are two verifications performed at the same centre but for with different IMRT method. a and b stand for different dates of verification of the same plan.
retrieved. In the first CarPet phantoms, the dummy disks did not fit optimally into the holes. This may have given rise to local over- or underpressure on the film perturbing the dose response locally. Both problems have been solved with the more recent CarPet phantoms, which were manufactured using a computer-controlled milling machine.

Evaluation of the different verification exercises also revealed a variation in the determination of the centre of the phantom when designing the treatment plan. This point plays an important role in the geometric correlation of the computed and measured dose distribution. However, this added geometric uncertainty in the procedure could be avoided in the future by electronically providing an additional volume of interest that contains only one point, i.e. the centre of the phantom as indicated by the reference lines drawn on the phantom.

If the central film shows an erratic response, this uncertainty is propagated through the rest of the evaluation. Measuring with the ionisation chamber in different slices could identify this problem. Alternatively, as the central film is more sensitive to dosimetric errors caused by air or attenuation, one might consider conducting the ionisation chamber measurements in another slice to make the method more robust.

Conclusion

This work demonstrates that the dosimetric verification of a complete IMRT treatment is a workable quality assurance tool. Film dosimetry, combined with the gamma method, allows the user to pinpoint unacceptable deviations between delivery and planning. The approach of a central service centre that prepares/develops all films, and performs the numerical comparisons to the respective computed dose distributions, enhances the reliability and validity of the intercomparison.

The agreement between measured and computed dose distributions was better than might have been expected for such a wide range of planning and delivery system combinations. Systematic differences between planned and measured dose distributions could be identified, and in some cases the cause revealed. The reliability analysis of the verification method is an important topic in these quality assurance networking activities. It is clear that for the verification of whole treatments, still more effort should be made in designing reliable dosimetry methods with intrinsically higher spatial and dosimetric accuracy. This multi-centre dosimetric verification trial demonstrated the feasibility of both a multi-centre quality assurance network to evaluate any IMRT planning and delivery system combination and the methodology involved in this verification study.

Acknowledgements

The authors appreciate the enthusiastic co-operation in this project from the following persons: R. Arrans, Hospital Virgen Macarena, Seville, Spain; A. Bakai, Universitaetsklinik fuer Radioonkologie, Tuebingen, Germany; W. Bulski, Centre of Oncology Warsaw, Warsaw, Poland; C. Fiorino, H.S. Raffaele, Milan, Italy; G. Hartmann, Deutsches Krebsforschungszentrum, Heidelberg, Germany; D. Huyskens, UZ Gasthuisberg, Leuven, Belgium; T. Knoos, Lund University Hospital, Lund, Sweden; A. Olszewska, Nederlands Kanker Instituut, Amsterdam, The Netherlands; M. Paiusco, S. Maria Nuova Hospital, Reggio Emilia, Italy; J. Rosello, ERESA, Valencia, Spain; J.-C. Rosenwald, Institut Curie, Paris, France; F. Sanchez-Doblado, Hospital Virgen Macarena, Seville, Spain; D. Thwaites, Edinburgh Cancer Centre, Edinburgh, Scotland; H. Welleweerd, UMC Utrecht, Utrecht, The Netherlands; G. Heeren, ESTRO, Brussels, Belgium.

QUASIMODO is funded by the EUROPEAN COMMISSION, Directorate General Health and Consumer Protection—Europe Against Cancer Programme, and is part of the ESQUIRE Project: Education, Science and QUality assurance In Radiotherapy in Europe, Grant Agreements 2001CVG2-005 and SPC.2002480.

This work was also supported by Grant Nos 12050401 GOA and 0111V1702 BOF of Ghent University.

* Corresponding author. S. Gillis, Department of Radiotherapy, Ghent University Hospital, De Pintelaan 185, B-9000 Gent, Belgium E-mail address: sofie.gillis@skynet.be

Received 27 July 2004; received in revised form 10 May 2005; accepted 5 June 2005; available online 15 August 2005

References


