PHYSICS CONTRIBUTION

REPLACING PRETREATMENT VERIFICATION WITH IN VIVO EPID DOSIMETRY FOR PROSTATE IMRT

LEAH N. MCDERMOTT, B.A./B.Sc.(Hons), MARKUS WENDLING, PH.D., JAN-JAKOB SONKE, PH.D., MARCEL VAN HERK, PH.D., AND BEN J. MIJNHEER, PH.D.

Department of Radiation Oncology, Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

Purpose: To investigate the feasibility of replacing pretreatment verification with in vivo electronic portal imaging device (EPID) dosimetry for prostate intensity-modulated radiotherapy (IMRT).

Methods and Materials: Dose distributions were reconstructed from EPID images, inside a phantom (pretreatment) or the patient (five fractions in vivo) for 75 IMRT prostate plans. Planned and EPID dose values were compared at the isocenter and in two dimensions using the gamma index (3%/3 mm). The number of measured in vivo fractions required to achieve similar levels of agreement with the plan as pretreatment verification was determined. The time required to perform both methods was compared.

Results: Planned and EPID isocenter dose values agreed, on average, within ±1% (1 SD) of the total plan for both pretreatment and in vivo verification. For two-dimensional field-by-field verification, an alert was raised for 10 pretreatment checks with clear but clinically irrelevant discrepancies. Multiple in vivo fractions were combined by assessing gamma images consisting of median, minimum and low (intermediate) pixel values of one to five fractions. The “low” gamma values of three fractions rendered similar results as pretreatment verification. Additional time for verification was ~2.5 h per plan for pretreatment verification, and 15 min ± 10 min/fraction using in vivo dosimetry.

Conclusions: In vivo EPID dosimetry is a viable alternative to pretreatment verification for prostate IMRT. For our patients, combining information from three fractions in vivo is the best way to distinguish systematic errors from non-clinically relevant discrepancies, save hours of quality assurance time per patient plan, and enable verification of the actual patient treatment. © 2007 Elsevier Inc.

In vivo dosimetry, EPID dosimetry, IMRT dose verification, Prostate.

INTRODUCTION

As dose prescriptions and the complexity of radiotherapy plans increase, so too do the demands for accurate and efficient means of verifying the dose delivered to patients. An ideal verification system would enable a check of the patient’s absolute dose distribution at the time and place of treatment in two, or preferably three, dimensions. The results would be in a digital format and require no additional time for setup of additional equipment before or during treatment time. The electronic portal imaging device (EPID) is a feasible candidate for such a system.

Pretreatment dosimetry is a commonly used surrogate verification method whereby the planned dose is verified before treatment. The EPIDs have been shown to be useful for intensity modulated radiotherapy (IMRT) pretreatment dosimetry (1–7). Converting an EPID image to a dose distribution pretreatment (with or without a phantom) allows for verification of the dose calculation and plan deliverability. The drawback is, however, that errors occurring at the time of treatment would be missed and it is not clear a priori how errors detected pretreatment would translate to errors within a patient. Furthermore, the workload becomes insurmountable for most clinics to do this for every patient plan. An alternative is in vivo EPID dosimetry, which we define as determination of the dose distribution inside the patient based on EPID images acquired at the time of treatment. In vivo EPID dosimetry has several advantages: the panel is already fixed to the linac, high resolution two-dimensional (2D) digital images are available immediately after irradiation and the images contain both dose and anatomic information, providing a check and documentation of the ac-
tual patient treatment. In addition, in vivo EPID dosimetry takes very little additional clinical time, because measurements are acquired during treatment time. These advantages are only relevant, however, if an acceptable error detection accuracy can be achieved with in vivo EPID dosimetry.

In vivo dosimetry in radiotherapy has traditionally relied on the use of point detectors, such as thermoluminescent dosimeters and diodes (8). MOSFETS (metal oxide semiconductor field effect transmitters) have also been used (9, 10), with reports of measured dose differences within ±5% of reference measurements. Advantages of MOSFETS over traditional point detectors are their small size, remote readout, and dose-rate independence. Even with multiple point-dose detectors (i.e., 2D matrix devices), spatial limitations make them insufficient for IMRT verification. Furthermore, measurements tend to be time-consuming, and changing anatomy makes it difficult to distinguish relevant from irrelevant errors; systematic plan errors can be masked by transient anatomic changes. Possible errors that could be easily missed include wrong leaf positions of the multileaf collimator (MLC), incorrect treatment data transfer, and wrong plan delivery. Such errors are less likely to be missed if treatment is verified in at least 2D, at a resolution comparable with the planned dose distribution. In 2D, transmission EPID images acquired during treatment have already been used for in vivo dosimetry of conformal fields (11–14); however, reports detailing clinical use for IMRT verification are lacking in the literature.

Our department has extensive experience with the use of EPID dosimetry for clinical pretreatment verification of IMRT prostate plans, whereby the dose is verified inside a homogeneous phantom (2). Replacing pretreatment with in vivo dosimetry is not straightforward, however. When acceptance criteria are used to raise an alert for potentially erroneous plans, checking one fraction is not usually sufficient, because random events during treatment (such as changes in patient anatomy and linac output) can obscure systematic errors. Random events that are not clinically relevant raise too many false positives and increase the workload unnecessarily. Waiting for too many fractions before verifying a plan, on the other hand, is potentially dangerous. One would like to combine the results from a few fractions to distinguish irrelevant, ignorable discrepancies from clinically relevant errors. Our aims, therefore, were to evaluate the verification results for a large patient group, test different protocols combining multiple fractions, find the optimal number of fractions that would render an acceptable detection rate, and investigate the potential cost/benefit of replacing pretreatment with in vivo dose verification in the clinic.

METHODS AND MATERIALS

Patient treatment plans
Seventy-five prostate cancer patients were included in this study. Plans comprised a five-field step-and-shoot IMRT technique, using 18 MV photon beams, with 15 to 40 segments per plan (average = 25) and beam angles of 0, 40, 100, 260 and 320°. Dose distributions were optimized and calculated with the treatment planning system (TPS) Pinnacle 7.4f (Philips Medical Systems, Eindhoven, The Netherlands). The prescribed dose was 78 Gy to the prostate, delivered as 2 Gy per fraction (15). Except for low-risk cases, the prostate contour also included the seminal vesicles. The dose was calculated using the adaptive convolve algorithm with a grid size of $0.4 \times 0.4 \times 0.4$ cm$^3$.

EPID dosimetry
Images were acquired with an amorphous silicon flat panel imager (iViewGT, Elekta, Crawley, UK). Details regarding the imager design, image acquisition, stability, and dosimetric characteristics have been reported extensively in previous work (16, 17). The algorithm used to determine dose images within a patient (or phantom) for the present study has also been described previously (7, 12), and will be briefly outlined here. Because the same algorithm is used for pretreatment or in vivo dosimetry, “patient” is used to refer to either the patient or the phantom. The algorithm converts segment images to an absolute 2D dose distribution in the reconstruction plane of the patient, defined as the plane perpendicular to the beam axis intersecting the isocenter (7, 18, 19). Therefore, this plane will rotate with the gantry. Pixel values of the transit dose image are processed using scatter kernels (for scatter within the EPID and scatter from the patient to the EPID), the scatter-to-primary ratio (for scattered radiation within the patient), the inverse square law, and the measured transmission to obtain the absolute dose distribution in the isocentric plane of the patient. The measured transmission of the beam through the patient is determined from images acquired for each field (or segment) both with and without the patient. The location of the reconstruction plane is arbitrary, so a correction is required to account for attenuation of the beam from the isocentric plane to the exit surface. The external contour of the patient computed tomography (CT) scan is used to obtain the ratio of geometrical path lengths, which is used to calculate the attenuation per pixel. The density of the transmission medium is assumed to be homogeneous; therefore, the dose may be incorrect for areas on the plane where the beam passed through media of non-tissue equivalent density. Reconstructed 2D dose distributions for each field are then the sum of the reconstructed dose distributions for all segments belonging to that field. We should note that the accuracy of our EPID dosimetry method ($\pm 2\%$ or 2 mm) has been published for 18 MV (7), and equivalent levels of accuracy have been achieved for verification of 6, 8, and 10 MV photon beams both at research level and in clinical practice.

Dose verification in the phantom and the patient
Dose distributions of each plan were verified in 2D with EPID images pretreatment (in a phantom) and in vivo (in a patient) during five treatment fractions.

For phantom verification, plans were recalculated with the TPS, replacing the patient planning CT scan with a phantom CT scan. The phantom was a polystyrene slab phantom of base 30 × 30 cm$^2$ (parallel with the table surface) and height 20 cm. The source-surface distance was 90 cm at a gantry angle of 0°. No plan parameters were changed, so the same MLC settings, collimator and gantry angles, energy, and numbers of monitor units were used as for the patient treatment. The dose distribution was recalculated at a higher resolution for higher accuracy using a grid size of $0.2 \times 0.2 \times 0.2$ cm$^3$. 


Table 1. Criteria levels for comparing measured (electronic portal imaging device) and planned (treatment planning system) dose distributions field-by-field

<table>
<thead>
<tr>
<th>Level</th>
<th>$\gamma_{\text{avg}}$</th>
<th>$\gamma_{\text{max}}$</th>
<th>Corresponding $\Delta D$ (%) and DTA (mm)</th>
<th>$P_{\gamma &lt; 1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strict</td>
<td>0.50</td>
<td>1.33</td>
<td>$\pm 1.5%$, 1.5 mm $\pm 4.0%$, 4.0 mm</td>
<td>99%</td>
</tr>
<tr>
<td>Medium</td>
<td>0.67</td>
<td>2.00</td>
<td>$\pm 2.0%$, 2.0 mm $\pm 6.0%$, 6.0 mm</td>
<td>95%</td>
</tr>
<tr>
<td>Easy</td>
<td>0.83</td>
<td>2.67</td>
<td>$\pm 2.5%$, 2.5 mm $\pm 8.0%$, 8.0 mm</td>
<td>90%</td>
</tr>
<tr>
<td>Very easy</td>
<td>1.00</td>
<td>3.33</td>
<td>$\pm 3.0%$, 3.0 mm $\pm 10.0%$, 10.0 mm</td>
<td>80%</td>
</tr>
</tbody>
</table>

*Abbreviation:* DTA = distance-to-agreement.

Corresponding percentage dose difference, $\Delta D$ (%) or DTA (mm) values are also shown, given that $\gamma = 1$ if $\Delta D = 3\%$ or DTA = 3 mm. The medium level is used clinically in our department for *in vivo* dosimetry of intensity-modulated radiotherapy prostate treatments.

For prostate IMRT patients treated in our department (including all patients for this study), patient setup is verified at the first two fractions and once per week thereafter. Additional fractions are included if setup corrections are required, according to shrinking action level decision rules (20). *In vivo* EPID dose verification was performed for the same fractions as setup verification. Additional EPID dosimetry images were always acquired at the third fraction to ensure five *in vivo* measurements were made at most within the first 2 weeks plus 1 day of treatment. Therefore, EPID treatment images were acquired at the first three fractions and then once per week, as well as for any fractions requiring additional setup verification. The first 5 measured fractions for each patient were included in this study, yielding data for 375 fractions measured *in vivo*, with all 75 plans verified in a phantom pretreatment.

Plans were first compared at the isocenter, by summing the planned and EPID isocenter doses for each field, respectively. Each field was then compared in 2D at the plane corresponding to the location of the reconstructed dose distribution determined with the EPID. The resolution of the EPID dose distribution was 0.1 × 0.1 cm². Evaluations were performed using the $\gamma$ index (21), with a dose difference tolerance of 3% of the maximum planned dose per field and a distance-to-agreement tolerance of 3 mm. Plans were assessed based on a combination of three $\gamma$ parameters, namely the average ($\gamma_{\text{avg}}$), maximum $\gamma$ ($\gamma_{\text{max}}$), and the percentage of points in agreement ($P_{\gamma < 1}$). These parameters were calculated from the $\gamma$ evaluation of each field within the 20% isodose line of the planned dose distribution. Discrepancies were counted according to their source for both pretreatment and *in vivo* dosimetry.

Assessing multiple fractions

After 2D acceptance criteria are defined, it is straightforward to apply them to EPID *in vivo* measurements of a single fraction. Because of random events, however, it is better to combine multiple fractions. The objective is to automatically accept/reject dosimetry images, avoid detecting irrelevant discrepancies and, more importantly, avoid missing real errors. Taking one field at a time, corresponding $\gamma$ pixel values from 1 to $n$ fractions were sorted in ascending order. Composite $\gamma$ images were created based on three different methods combining multiple fractions for assessment (per field). The methods included the median $\gamma$ value per pixel (median-$\gamma$-image), the minimum $\gamma$ value per pixel (min-$\gamma$-image), and the “low” $\gamma$ value per pixel (low-$\gamma$-image), half-way between the minimum and the median. For example, if nine fractions were assessed, the $\gamma$ values would be arranged in ascending order per pixel and numbered one to nine. The median-$\gamma$-image would be a composite $\gamma$ image comprising all the pixel values at the fifth (middle) position, the min-$\gamma$-image would comprise all the first (lowest) values and the low-$\gamma$-image would comprise all the pixel values in the third position.

Alert criteria

Four criteria levels were applied to each median-, min-, and low-$\gamma$-image to raise an alert to potentially problematic plans, based on $\gamma_{\text{avg}}$, $\gamma_{\text{max}}$, and $P_{\gamma < 1}$. The four levels tested were labeled “strict”, “medium”, “easy”, and “very easy” (Table 1). The medium level corresponded to the clinical criteria used in our department for both pretreatment and *in vivo* dosimetry. Additional levels were designed to compare the sensitivity of the results. The number of plans for which an alert was raised was counted for pretreatment ($f_{\text{pr}}$) and for each of $n$ fractions in *in vivo* ($i_{\text{in}},(n)$). A flow chart describing the process is given in Fig. 1.

Furthermore, the correlations between pretreatment and *in vivo* results of the three $\gamma$ parameters ($\gamma_{\text{avg}}$, $\gamma_{\text{max}}$, and $P_{\gamma < 1}$) were compared pairwise for each field, for 1 to $n$ fractions, using the three combination methods.

Developing a clinical protocol

After the optimal number of treatment fractions required to replace pretreatment dosimetry was found, a plan of action was devised for the clinic based on *in vivo* dosimetry. To this end, discrepancies encountered from verification of 75 patient plans were divided into three groups. If the influence of a discrepancy was isolated to a single field, if it were due to an easily identifiable random cause, and if it would not be considered clinically relevant, the discrepancy was classified as group A. If the influence of a discrepancy was isolated to individual fractions (all fields), but not more than two of the five fractions investigated, and could not be easily identified from *in vivo* dosimetry, the discrepancy was classified as group B. Otherwise the discrepancy was considered plan dependent, systematic, and would affect one or several fields in every treatment fraction. This type of error was classified as group C.

Time

For both pretreatment checks and *in vivo* dosimetry, the additional time outside routine patient planning and treatment time was assessed and compared. This included additional time to plan, measure, and analyze the dose distributions in 2D for each field.
RESULTS

Verification at the isocenter

Comparison of point dose values at the isocenter showed that EPID and planned values agreed well. For pretreatment verification of 75 plans, the average ratio of the EPID and planned isocenter dose values in the phantom was 0.99 ± 0.01 (1 SD), ranging from 0.96 to 1.01. For the fields measured in vivo, the ratio of EPID and planned values over 1,860 fields was 0.99 ± 0.01 (1 SD), ranging from 0.96 to 1.02. Outliers resulting from image acquisition errors were excluded (15 fields). The correlation between in vivo and pretreatment isocenter values for the total plan was low, with a correlation coefficient $r = 0.023$ (i.e., the magnitude of the discrepancies detected was much smaller than the measurement uncertainty).

Verification in 2D

Individual measured fields compared very well with planned dose distributions. An example of dose profiles for one IMRT field measured in a phantom pretreatment and for five fractions in vivo is shown in Fig. 2. This was a patient with gas pockets in the rectum on two of the five fractions measured in vivo. Because our EPID dosimetry algorithm does not account for inhomogeneities, an accurate determination of dose in the ray-path intersecting the gas pocket is not possible. It does, however, indicate where potential dose discrepancies would occur from different patient anatomy between treatment and planning.

Figure 3 shows an example of $\gamma$ evaluations for one patient plan: (a) pretreatment, (b) $\gamma$ images, and (c) low-$\gamma$ images for five in vivo fractions. For this patient, both a systematic dose calculation error and random discrepancies can be seen by comparing pretreatment verification in a phantom with in vivo measurements. By considering more fractions, the low-$\gamma$ images “improve” as large random errors are suppressed, allowing optimal detection of systematic errors after three fractions. The systematic error for this plan (12% local underdosage) was not considered clinically relevant because the error was in one field, covered a small

---

Fig. 1. Flow chart to compare the number of plans alerted with pretreatment and in vivo dosimetry. The composite in vivo $\gamma$ image of $n$ fractions was made using the median-, min-, and low-$\gamma$-image methods. The values $i_{pt}$ and $i_{iv}(n)$ are the number of plans for which an alert was raised for pretreatment and in vivo verification, respectively. These were counted for all four criteria levels (Table 1).

Fig. 2. Planned and measured absolute dose profiles for one field of an intensity-modulated radiotherapy plan. The location of the profile is indicated in the EPID image. (a) Calculated and measured in a phantom, (b) calculated in a patient and measured in vivo over five treatment fractions. Random discrepancies occurred at two fractions from gas pockets in the rectum.
Comparing pretreatment and in vivo dosimetry

The results for the combination of multiple fractions using three methods (median-, min-, and low-\(\gamma\)-image) are shown in Fig. 4 (process outlined in Fig. 1), alongside the number of plans alerted after a pretreatment check. Each histogram illustrates \(i_{\text{pt}}\) and \(i_{\text{iv}}(n)\), the number of plans for which an alert would have been raised for the four levels of alert criteria. To ensure a similar error-detection rate for in vivo dosimetry as pretreatment, a similar group of plans should be alerted. Therefore \(i_{\text{iv}}\) should be similar to \(i_{\text{pt}}\) for a given number of fractions, method, and criteria level. Overall, the median-\(\gamma\)-image renders much higher values of \(i_{\text{iv}}(n)\) than \(i_{\text{pt}}\), indicating unnecessary workload required to check error free plans with a phantom (Fig. 4a). With the min- and low-\(\gamma\)-image methods (Fig. 4b and 4c), checking one or two
Fractions would render high $i_{\nu}(n)$ values; however, three or more would provide an appropriate rate of detection for all levels of criteria. For the min-$H9253$-image method, however, the number of alerts raised continues to fall after three fractions, as random events or measurement uncertainty compensate potentially erroneous plans. Considering more fractions, zero errors would be detected much sooner with the min-$H9253$-image than the low-$H9253$-image method, increasing the chance of false negatives. Therefore, the low-$H9253$-image with the medium criteria was considered the best option.

Assuming that pretreatment dosimetry gives an "actual" measure of agreement between planned and measured dose, the decision to replace pretreatment with in vivo dosimetry was justified by assessing the relative correlation between both methods for the range of evaluation parameters. Figure 5 shows an example of the correlation between pretreatment and in vivo dosimetry for $\gamma_{\text{max}}$ with multiple fractions combined using low-$\gamma$-images. The medium criterion for $\gamma_{\text{max}} (\gamma = 2.0)$ is also plotted, defining the following quadrants:

- lower left = both pretreatment and in vivo pass,
- upper right = both pretreatment and in vivo raise an alert,
- upper left = false positives (work load too high), and
- lower right = false negatives (detection accuracy too low).

Assessing the in vivo results over increasing number of fractions improves the correlation, rendering fewer false positives. As random outliers are suppressed, the low-$\gamma$-image approaches a measure of the “actual” agreement between the plan and the delivered dose distribution.

Figure 6 summarizes the correlation for all three methods used to combine in vivo fraction measurements (median-, min-, and low-$\gamma$-image), for each of the three $\gamma$ parameters ($\gamma_{\text{avg}}$, $\gamma_{\text{max}}$, and $P_{\gamma < i}$). The correlation improves from one to three fractions in all cases, with little improvement after

---

Fig. 4. The number of plans for which an alert was raised after a pretreatment phantom check ($i_{\nu}$) for 75 intensity-modulated radiotherapy prostate plans is given in panels a-c. Because of small dose calculation errors, more plans were alerted with strict criteria, decreasing as the criteria were relaxed. Corresponding data are also shown considering one to five in vivo fractions ($i_{v_{\nu}}$). For in vivo measurements, the criteria were applied to the (a) median-, (b) min-, and (c) low-$\gamma$-images. Ideally, $i_{\nu}$ and $i_{v_{\nu}}$ should alert and pass the same plans at each level. Compared with pretreatment, too many plans are alerted using the median-$\gamma$-image because of ignorable outliers (e.g., gas pockets). The number of plans alerted converges to zero using the min-$\gamma$-image. Three fractions is sufficient using the low-$\gamma$-image and the medium criteria to achieve a similar accuracy as a check of every plan in a phantom.

---

Fig. 5. An example of the correlation ($r$) between pretreatment and in vivo EPID dosimetry. Each dot represents the $\gamma_{\text{max}}$ value of one field. The $\gamma_{\text{max}}$ values for all pretreatment fields are compared with the $\gamma_{\text{max}}$ of the low-$\gamma$-image calculated for $n = 1$ to 5 fractions. The limits of the medium-level alert criteria are also shown. Assessing the in vivo results over increasing number of fractions improves the correlation with the pretreatment results. As random outliers are suppressed, the low-$\gamma$-image approaches a better measure of the agreement between the plan and the delivered dose distribution.
four or five fractions. The min- and low-$\gamma$-images render better correlation than median-$\gamma$-images. For the min-$\gamma$-image, however, the $\gamma_{\text{avg}}$ and $P_{\gamma < 1}$ correlations decrease at five fractions, with little improvement after four or five fractions. The $p$ value shows the correlation is significant ($p < 0.05$) after two fractions in all cases.

For these plans, although the dose discrepancies were relatively large for the field in which they occurred, only one field of the five was alerted and the absolute discrepancy was considered to be a small proportion of the total planned dose at this location ($< 5\%$). Therefore, replanning was not considered necessary for these patient plans. For this patient group, no clinically relevant false negatives would have been missed by checking three fractions with the low-$\gamma$-image method and using the medium criteria level.

Types of errors detected

A summary of the type of errors, as alerted with the medium criteria, can be found in Table 2 for pretreatment and in vivo measurements. The errors are grouped according to groups A, B, and C. In vivo fractions were assessed separately.

Group A errors were easily determined by checking raw EPID treatment images. Gas pockets were detected in 7% of fields. An obstructing table arm (behind the patient, 4%) was detected if the patient was not positioned centrally on the treatment couch, and was, therefore, a random error. Image acquisition errors (1%) occurred when the operator was too late to start image acquisition. Additional checks are in place at the linac to ensure all segments were delivered to the patient. Additional fractions should be checked to verify these fields.

Group B errors could not be identified with certainty by checking treatment fields (e.g., variation in linac output (outside tolerance) or dose differences from errors in patient setup). These plans would require a phantom check to verify any systematic errors. The number of fractions affected by a gas pocket was also included (17% of fractions). For pretreatment verification, EPID measurements were corrected for linac output variation.

The number of group C errors was higher for in vivo dosimetry (23%) than for pretreatment checks (13%). This is because there are five in vivo measurements for each plan, compared with one pretreatment. Random events and measurement uncertainty will result in variation in the measured dose, and the same plan will have a different dose distribution when calculated on a phantom, so errors will be a different proportion of the maximum dose (per field). Many will fall on the border of the alert criteria, so more in vivo

<table>
<thead>
<tr>
<th>Group</th>
<th>Source</th>
<th>Pretreatment</th>
<th>In vivo</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, number of fields</td>
<td>Gas pocket</td>
<td>375</td>
<td>1875</td>
</tr>
<tr>
<td></td>
<td>Table arm</td>
<td>2 (1%)</td>
<td>77 (4%)</td>
</tr>
<tr>
<td></td>
<td>Image acquisition</td>
<td>0 (0%)</td>
<td>21 (1%)</td>
</tr>
<tr>
<td>B, number of</td>
<td>Gas pocket</td>
<td>N/A</td>
<td>63 (17%)</td>
</tr>
<tr>
<td></td>
<td>Patient setup or linac output</td>
<td>corrected</td>
<td>42 (11%)</td>
</tr>
<tr>
<td>C, number of plans</td>
<td>Dose calculation</td>
<td>10 (13%)</td>
<td>17 (23%)</td>
</tr>
<tr>
<td></td>
<td>No discrepancy</td>
<td>64 (85%)</td>
<td>9 (12%)</td>
</tr>
</tbody>
</table>

Plans were verified both pretreatment and in vivo for five treatment fractions with electronic portal imaging device dosimetry.
fields are likely to raise an alert. Typical action for a group C error in a clinical protocol would require fixing the error (such as a systematic MLC leaf position or linac output calibration error), or replanning the intended treatment. Data transfer errors and systematic linac output errors would also be included in this group; however, no instances were found in this study.

**Designing a clinical protocol**

The chart in Fig. 7 shows the proposed clinical protocol based on different types of errors encountered for this patient group. Results from the 75 patient plans of this study are also included.

An initial check for large errors is first made for each field individually, using the “very easy” criteria. After applying the medium criteria to low-\(\gamma\)-images of initially three fractions, 64 plans passed. Four plans were alerted because of gas pockets or the table arm in two or three fractions. By checking an additional fraction, each of these passed. It should be noted that in clinical practice the additional measurement should be made at the next treatment fraction. Three plans had group B errors whereby both the pretreatment check and subsequent fractions revealed no systematic errors. The remaining four plans with systematic dose calculation errors were assessed individually and the errors were not considered clinically relevant. These four plans were also alerted pretreatment, and the errors were found to be not clinically relevant. Ten plans were alerted pretreatment (Fig. 4); six of these passed the *in vivo* check. These discrepancies were close to the criteria, and after three
fractions, the \textit{in vivo} dose passed. Therefore, only seven plans of this patient group (three from group B, four from group C) would have needed to be checked with a phantom, with such an \textit{in vivo} dosimetry protocol in place.

Over 1.5 years, 75 IMRT prostate patient plans from this study were clinically verified both pretreatment and \textit{in vivo}, as well as an additional 85 patient plans with \textit{in vivo} dosimetry alone. These 85 patients were not included in the study because routine pretreatment dosimetry was discontinued. For 83 cases, there were no phantom measurements to compare with \textit{in vivo} data. Two cases, however, did require an additional phantom check and were found to have dose calculation errors. These were corrected (by changing a parameter in the TPS) and the plans were reoptimized.

\textbf{Time}

The time required to recalculate a five-field patient plan with the phantom CT scan, and then send the separate plan to the linac, was approximately 60 min. To set up the phantom, acquire transit and open EPID images before patient treatment (usually performed outside daily linac treatment times) and measure the linac output required an additional 60 min. Analysis involved reading in plan data and EPID images for each segment, calculating the 2D dose distribution and performing the \( \gamma \) evaluation; on average, this required 15 min. Additional time to replan, measure, and analyze for pretreatment verification was 2 h, 15 min per plan.

To perform \textit{in vivo} dosimetry, no additional time is required at the planning stage. An additional 10 min is required per plan to acquire one open image (no attenuating medium) per segment. Analysis of each fraction typically takes 10 min to read in EPID images and run the \( \gamma \) evaluation. An additional 5 min is required at the first fraction to read in plan information, and link measured and calculated fields. For \textit{in vivo} dosimetry, only an additional 15 min \( \pm \) 10 min per fraction was required for analysis of each plan.

\textbf{DISCUSSION}

The advantages of \textit{in vivo} dosimetry are that a check and record of the actual treatment is gained without additional cost in measurement time. This is only an advantage if the method can be shown to be as accurate, and provide as much information as necessary, as pretreatment dosimetry. The number of measured \textit{in vivo} fractions required to replace pretreatment verification is a balance between early detection and workload. There are no precedents or recommendations to follow regarding 2D \textit{in vivo} dosimetry, so our protocols have arisen from clinical experience. Checking three fractions with \textit{in vivo} EPID dosimetry is sufficient to replace pretreatment verification of every plan for this patient group, allowing average discrepancies up to \( \gamma = 0.67 \) (corresponding to \( \pm 2.0 \%/2.0 \text{ mm} \) with \( \gamma \) criteria 3.0%/3.0 mm), maximum discrepancies up to \( \gamma = 2.00 \) (corresponding to \( \pm 6.0 \%/6.0 \text{ mm} \)), and permitting 5\% of the field within the 20\% isodose line to have \( \gamma > 1 \). When one or more of these criteria fail, the plan is verified in a phantom. It should be noted that both the acceptance levels applied and the probability that a phantom check is required depend on a number of factors. These include the treatment site, complexity of the treatment delivery, patient positioning, dose delivery, and accuracy of the TPS. We would also like to stress that for this study and in our department, \textit{in vivo} dosimetry has only replaced patient-specific phantom checks. It should not be used to replace routine MLC, linac, or administration (data transfer) quality assurance, nor provide the only dosimetric verification for the introduction of new techniques or equipment.

Statistics on a small number of fractions is complicated by the probability of large, ignorable errors (group A). An obvious choice would be to take the median. However, ignorable outliers (e.g., gas pockets, obstructing table arm) can be large, give too many false positives, and potentially obscure systematic errors. For example, considering three fractions for which discrepancies due to gas pockets occurred during two of the fractions, one would want to weight the assessment on the “good fraction”. For the same reason, we did not want to average the reconstructed dose images before comparing them with the plan. Given the \( \gamma \) value is one-sided, the \( \gamma \) values of combined fractions \textit{after} comparison with the plan can be weighted toward the values of better agreement; this is not possible if dose values are first averaged over multiple fractions. With the aim to suppress clinically irrelevant random errors and still detect any systematic errors occurring at every fraction, a solution would be to take the minimum \( \gamma \) value per pixel over multiple fractions. A problem with this method is that measurement uncertainty and compensating random events, after enough fractions, will result in a high probability of all \( \gamma \) values converging to zero. Therefore, we have chosen to use the low-\( \gamma \)-image, a compromise between the minimum and the median. In principle, the precise level of compromise between the minimum and median could be optimized, incorporating measurement uncertainty and probability of outliers. This would be overkill, however, because we are dealing with so few fractions. A halfway approximation, therefore, was considered appropriate.

The strategies proposed in this article are specifically suited to prostate treatments, because the main changes in homogeneity are due to randomly occurring gas pockets, and these are easy to detect in EPID images. Future investigations will involve incorporating inhomogeneity corrections in our algorithm, as well as addressing other treatment sites. In the end, \textit{in vivo} dosimetry is intended as a safety net, to be the last check in a series of routine quality assurance procedures. It is intended to catch large, clinically relevant errors, so calculation accuracy does not need to match that of the TPS. Our tolerance levels are based on 3%/3 mm for prostate treatments; therefore, we consider
the ±2%/2 mm accuracy of our model more than sufficient for routine use of in vivo dosimetry.

CONCLUSION

A technique has been presented for replacing pretreatment verification with EPID in vivo dosimetry. The number of measured in vivo fractions required to replace pretreatment verification is a balance between early detection and workload. Allowing 5% of the field outside 3.0%/3.0 mm, average discrepancies of ∼2.0%/2.0 mm and maximum discrepancies of ∼6.0%/6.0 mm, checking three fractions is optimal to detect similar number of systematic errors with in vivo dosimetry as pretreatment verification. The advantage of EPID in vivo dosimetry is that data are acquired during treatment, with little additional time required for measurement and analysis.

REFERENCES