

Received:  
20 August 2020

Revised:  
15 October 2020

Accepted:  
28 October 2020

© 2021 The Authors. Published by the British Institute of Radiology under the terms of the Creative Commons Attribution 4.0 Unported License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Cite this article as:

Bedford JL, Hanson IM. Optimisation of a composite difference metric for prompt error detection in real-time portal dosimetry of simulated volumetric modulated arc therapy. *Br J Radiol* 2021; **94**: 20201014.

## FULL PAPER

# Optimisation of a composite difference metric for prompt error detection in real-time portal dosimetry of simulated volumetric modulated arc therapy

JAMES L BEDFORD, PhD and IAN M HANSON, PhD

Joint Department of Physics, The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London, UK

Address correspondence to: Dr James L Bedford  
E-mail: [james.bedford@icr.ac.uk](mailto:james.bedford@icr.ac.uk)

**Objectives:** In real-time portal dosimetry, thresholds are set for several measures of difference between predicted and measured images, and signals larger than those thresholds signify an error. The aim of this work is to investigate the use of an additional composite difference metric (CDM) for earlier detection of errors.

**Methods:** Portal images were predicted for the volumetric modulated arc therapy plans of six prostate patients. Errors in monitor units, aperture opening, aperture position and path length were deliberately introduced into all 180 segments of the treatment plans, and these plans were delivered to a water-equivalent phantom. Four different metrics, consisting of central axis signal, mean image value and two image difference measures, were used to identify errors, and a CDM was added, consisting of a weighted power sum of the individual metrics. To optimise the weights of the CDM and to evaluate the resulting timeliness of error detection, a leave-pair-out strategy was used. For each combination

of four patients, the weights of the CDM were determined by an exhaustive search, and the result was evaluated on the remaining two patients.

**Results:** The median segment index at which the errors were identified was 87 (range 40–130) when using all of the individual metrics separately. Using a CDM as well as multiple separate metrics reduced this to 73 (35–95). The median weighting factors of the four metrics constituting the composite were (0.15, 0.10, 0.15, 0.00). Due to selection of suitable threshold levels, there was only one false positive result in the six patients.

**Conclusion:** This study shows that, in conjunction with appropriate error thresholds, use of a CDM is able to identify increased image differences around 20% earlier than the separate measures.

**Advances in knowledge:** This study shows the value of combining difference metrics to allow earlier detection of errors during real-time portal dosimetry for volumetric modulated arc therapy treatment.

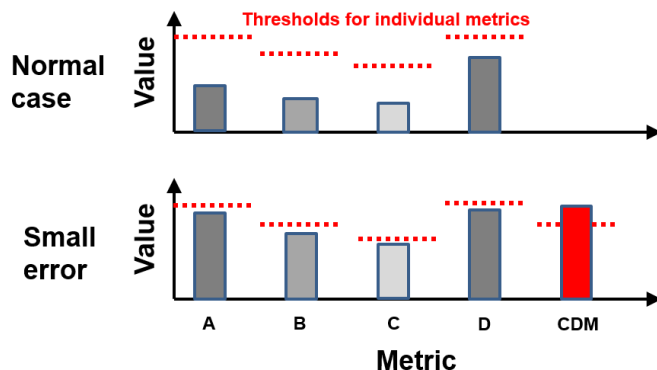
## INTRODUCTION

Portal dosimetry is a valuable means of ensuring that the dose delivered to the patient during a fraction of radiotherapy is in agreement with that predicted by the treatment planning system.<sup>1–3</sup> The two principal methods are forward projection,<sup>4,5</sup> in which portal images are predicted at the time of treatment planning, and then the delivered images are compared with these; and back projection,<sup>6–8</sup> in which the measured images are projected into the CT scan of the patient and converted into a dose distribution, which is then compared with the planned dose distribution.

The most widespread use of portal dosimetry is for verification of whole fractions of treatment, with analysis being carried out after delivery of the complete fraction.<sup>9</sup>

Often this is the first fraction or closely thereafter, with additional measurements and analyses being carried out when anatomical variations occur, subject to departmental resources.<sup>10–12</sup> However, real-time or intrafraction portal dosimetry is used to evaluate each fraction of the treatment as it is delivered.<sup>13–16</sup> This has the advantage that errors can be detected before the whole fraction of treatment has been delivered. With a constant tendency for treatments to become more hypofractionated, either in a stereotactic context<sup>17,18</sup> or otherwise,<sup>19</sup> this approach is important, as a whole fraction of treatment accounts for a large proportion of the total dose delivered. The approach also allows for a time-resolved analysis, which gives greater insight into the accuracy of the delivery than when examining the whole.<sup>20,21</sup>

Figure 1. Principle of the CDM. The individual metrics do not respond strongly enough to an error to exceed their respective thresholds, but the composite difference metric detects the error. For a large error, most or all of the metrics exceed their thresholds. CDM, composite difference metric.



The selection of appropriate metrics for optimum sensitivity and specificity of portal dosimetry is the subject of continuing research. Olaciregui-Ruiz et al<sup>22</sup> investigate a number of different treatment plan metrics and report on their success in identifying errors in a large cohort of volumetric modulated arc therapy (VMAT) patients. Monitoring and analysis software typically follows several metrics simultaneously, using appropriate thresholds for each. However, there may be value in using a single composite difference metric (CDM). This is convenient for monitoring purposes, but the principal rationale is that small but correlated changes in the individual measures may cause the composite metric to indicate an error before any of the individual measures has exceeded its own threshold (Figure 1).

This study investigates a composite difference metric consisting of a weighted power sum of several individual metrics. The method is applied to segment-resolved portal dosimetry for VMAT treatment for prostate. Various errors are deliberately introduced into the treatment plans and the resulting plans are delivered to a water-equivalent phantom. Several metrics are used individually and simultaneously to detect differences between the predicted and measured images. A composite difference metric is proposed, and an optimisation method to determine the optimum metric weights is described. The sensitivity of the method to the various errors is determined, and the number of segments taken during delivery of a treatment fraction for each error to become apparent is used as an indicator of success.

## METHODS AND MATERIALS

### Patients and treatment plans

Treatment plans for six prostate patients were retrospectively used in this study. The patients consented for their images to be used for research purposes. Three planning target volumes (PTVs) were used to create the treatment plans, two based on the prostate alone and the third, outermost volume, based on the prostate plus seminal vesicles.<sup>23,24</sup> The prescribed dose was 60 Gy in 20 fractions to the most central PTV. The treatment plans consisted of a single 358° gantry arc comprised of 180 segments at a spacing of 2°, with collimator angle 2° throughout. The plans were created using AutoBeam v5.8<sup>25</sup> using a dose grid of 4 mm ×

4 mm × 4 mm resolution and a final calculation and renormalisation was carried out with Pinnacle<sup>3</sup> (Philips Radiation Oncology Systems, Madison, WI) using a dose grid of 2.5 mm × 2.5 mm × 2.5 mm resolution. The beam energy was 6 MV and the plans were designed for delivery with a VersaHD accelerator (Elekta AB, Stockholm, Sweden).

After inverse planning, AutoBeam was used to recalculate the plan on a water-equivalent phantom of dimensions 300 mm long (G–T direction) × 300 mm wide (A–B direction) × 200 mm high, with the isocentre located at the centre of the phantom. The point dose at the isocentre and the mean dose over a volume corresponding to the 24 mm length and 6 mm diameter of a Farmer 2571 ionisation chamber (Saint Gobain Crystals and Detectors, Reading, UK) were noted for verification purposes.

### Predicted images

The AutoBeam software was used to calculate predicted portal images for each treatment plan. The prediction model projected the dose distribution at the isocentre plane onto the image plane with 512 × 512 pixels of size 0.8 mm × 0.8 mm at a source-panel distance of 1600 mm, giving a pixel size of 0.5 mm at the isocentre plane. The prediction model distinguished between the in-field region of the image panel, defined as the region within the beam's eye view of the treatment beam, and the out-of-field region, which was the remaining part of the panel. The in-field signal was calculated as the dose at the isocentre plane, corrected to give dose in a water-equivalent medium, and projected to the image plane, allowing for exponential attenuation. The out-of-field signal was calculated as a uniform signal, proportional to the field size and based on a scatter factor as a function of the mean equivalent path length from isocentre plane to detector plane. The union of both in-field and out-of-field regions was then convolved with a pair of Gaussian scatter kernels to allow for in-panel scatter.<sup>5</sup>

Previous work<sup>16,26</sup> indicated that an accurate prediction model was important. This was because a more accurate prediction led to a lower error threshold, which in turn allowed for greater sensitivity to errors. In the present study, a model of couch attenuation between the isocentre and the image panel was therefore included. This model consisted of an additional radiological path length between the isocentre plane and the image plane, for gantry angles between 300° and 60° through 0°. The additional path length was 8.0 mm, divided by the cosine of the gantry angle so as to allow for oblique passage through the couch. The path length at gantry angle 60° was therefore 16 mm.

A model of gantry and panel sag was also used to account for any movement of the radiation isocentre and image panel. Portal images for a 100 mm × 100 mm field were acquired at all gantry angles at intervals of 10° and the centre of the imaged field and the central image intensity were evaluated. Lateral displacement of the panel and variation in intensity were negligible, but the longitudinal displacement of the panel was found to be significant and predictable. This was similar to the observations of Poludniowski et al<sup>27</sup>. The longitudinal displacement was therefore included

in the prediction model as a sinusoidal shift with peak-to-peak amplitude of 4 mm.

Predicted images were produced for each control point of the VMAT arc and stored as a single binary file for each patient, which was subsequently loaded into the AutoDose v. 1.1 software to provide the predictions needed for comparison with real-time images.<sup>16</sup>

### Measured images

The treatment plans were delivered to a Solid Water phantom (Radiation Measurements, Inc., Middleton, WI), with geometry as described above. The dose at the isocentre was verified using a Farmer ionisation chamber, corrected for accelerator output. From the treatment plan for each of the patients, further plans were created by deliberately introducing errors into all 180 segments. The errors consisted of a 4% increase in monitor units, a retraction of 4 mm of all multileaf collimator (MLC) leaves, a shift of 4 mm of all MLC leaves, and introduction of an air space of 20 mm width (left-right direction)  $\times$  50 mm height (anteroposterior direction) into the phantom, this latter to simulate rectal gas.<sup>26</sup> In three of the patients, the magnitudes of the errors were varied to give a 2–10% increase in monitor units in 2% steps, a retraction of 2–10 mm in 2 mm steps of all MLC leaves, a shift of 2–10 mm in 2 mm steps of all MLC leaves, and introduction of an air space of 10–50 mm width in 10 mm steps into the phantom. These systematic variations ensured that the method was able to detect errors of various magnitudes. The correct and erroneous treatment plans were then delivered to the water-equivalent phantom.

Portal images were measured using an iViewGT portal imager (Elekta), which consisted of an amorphous silicon panel (Perkin Elmer, Santa Clara, CA), recording images in JPEG format at a size of  $512 \times 512$  pixels.<sup>28</sup> The AutoDose software used the gantry angle information stored with each image to bin the images according to treatment plan segment. In a real-time context, the measured images would be compared with the predicted images as the treatment progressed, and categorised according to a traffic light system, but in this study, the images were stored and processed retrospectively, segment by segment. No correction for accelerator output was applied in this process. Both predicted and measured images were rescaled by  $10^{-4}$  to bring them into a range of approximately 0–100, thereby allowing convenient handling.

### Difference metrics

Four metrics were used to identify differences between predicted and measured images: central axis signal (CAS), mean image value (MIV), mean absolute difference as a percentage of maximum predicted image intensity (MDM) and mean absolute difference as a percentage of local predicted image intensity (MDL). The latter three of these metrics were computed for regions where the predicted image had an intensity of greater than 10% of the maximum. A composite difference metric was then added:

$$CDM = \sum_{i=1}^4 w_i m_i^{w_i} \quad (1)$$

where  $m_i$  were the individual metrics and  $w_i$  were weighting factors, which took values of between 0 and 1. The first two metrics had dimensions of image intensity and the latter two were percentages, but the rescaling of the image intensities allowed for practical computation of the CDM. The CDM itself was considered as a measure of the presence of an error, analogous to a human operator assessing a variety of measures, with their respective dimensions, and making a decision as to whether there was an error.

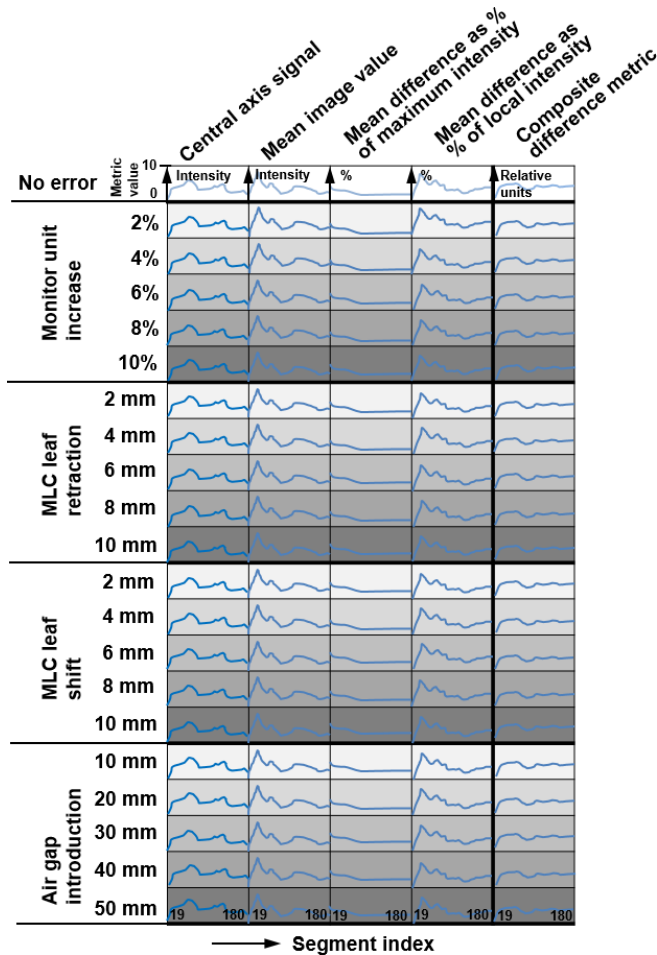
This form of metric was found by empirical investigation. Other forms, such as a simple weighted sum or product were found to be less effective, so were not pursued. However, from a theoretical perspective, the usefulness of Equation (1) can be seen from the fact that with  $w_i$  less than 1, the CDM takes the form of a sum of functions similar to  $y = \sqrt{x}$ . These functions increase rapidly at low values of  $x$  and then plateau. The result is that the CDM has the maximum impact when all of the individual metrics are raised slightly, in contrast to a single metric with a large elevation in value. The weights are equal to the powers to limit the number of combinations of parameter to be searched, while allowing a broad search space.

The composite difference metric was calculated for each segment of the treatment plan. The first 10% of segments were not considered as the image signal was unstable in this period. The segments of the plans were approximately evenly weighted, so that the first 10% of segments was expected to deliver approximately 10% of the dose. By removing these segments from the analysis, the potential existed for an error in this part of the plan to remain unnoticed, allowing incorrect dose delivery. Such an error was expected to be much less than the total dose delivered by the initial segments. After the first 10% of segments, a running sum of 10 segments, or “section value”, was used to provide information on each part of the VMAT arc while minimising image fluctuation. The section value for segment  $s$  consisted of the sum of segments  $s - 9$  to  $s$ . Figure 2 summarises the data used.

### Optimisation of the composite difference metric

Based on the above arrangement of data, an optimisation routine was used to determine the optimal values of the weighting factors in the CDM. In order to ensure that the results were generally applicable and not just optimal for the patients used in the optimisation, a leave-pair-out strategy was used. In the development of prediction models, particularly in artificial intelligence, an optimisation method is commonly used to ensure optimal prediction. When the sample used for optimisation is limited, the possibility exists that the resulting parameters give good prediction for that sample only, and rather less good results in further patients, in which case overfitting is said to occur. Cross-validation is used to avoid this, by selecting an optimisation sample for the fitting and then a validation sample for testing the results. In leave-pair-out cross-validation, each pair of data items is set aside in turn for validation, while the remaining data items are used for optimisation. The interested reader is referred to the more formal descriptions given by Kohavi<sup>29</sup> and Hastie et al.<sup>30</sup> In the context of this study, the CDM weights for four of the patients were optimised, and the results were then applied to the

Figure 2. Summary of data used for evaluation of the composite difference metric. Each cell represents a sequence of metric values for segments 19–180 of the treatment plan (shown schematically only). There are four metrics and the composite difference metric. Measurements are available for no error and increasing error of one of four types. MLC, multileaf collimator.



remaining two patients. For purposes of uniformity, the optimisation set contained two patients with errors of varying magnitude and two patients without (see Measured images), while the test set contained one patient with errors of varying magnitude and one without. This gave nine unique combinations of optimisation set and test set. The process described below was therefore carried out nine times and the median and range of the results recorded for the test patients.

The first step in the optimisation of CDM weights (Figure 3) was to calculate the thresholds for the individual metrics, based on the normal deliveries of the patients (the top row of Figure 2). In order to avoid the possibility of false-positive results occurring when transferring optimisation results from the four optimised patients to the two test patients during leave-pair-out, the threshold was set as the median plus range of the maximum metric value observed in the normal deliveries for the four patients. The metrics were evaluated at control points 19–180 in the normal deliveries, the maximum value of the metric was

Figure 3. Flow diagram showing the optimisation process for determination of weighting factors in the composite difference metric. The greyscale intensity represents the computational workload. CDM, composite difference metric.

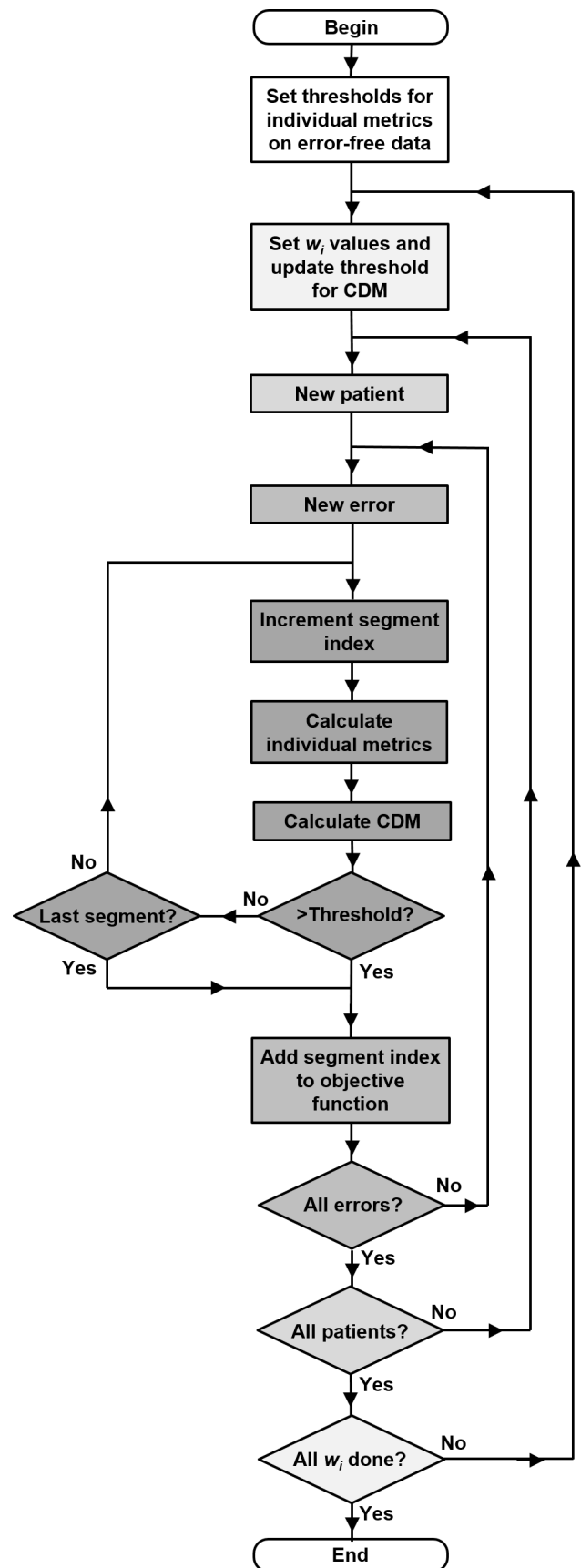
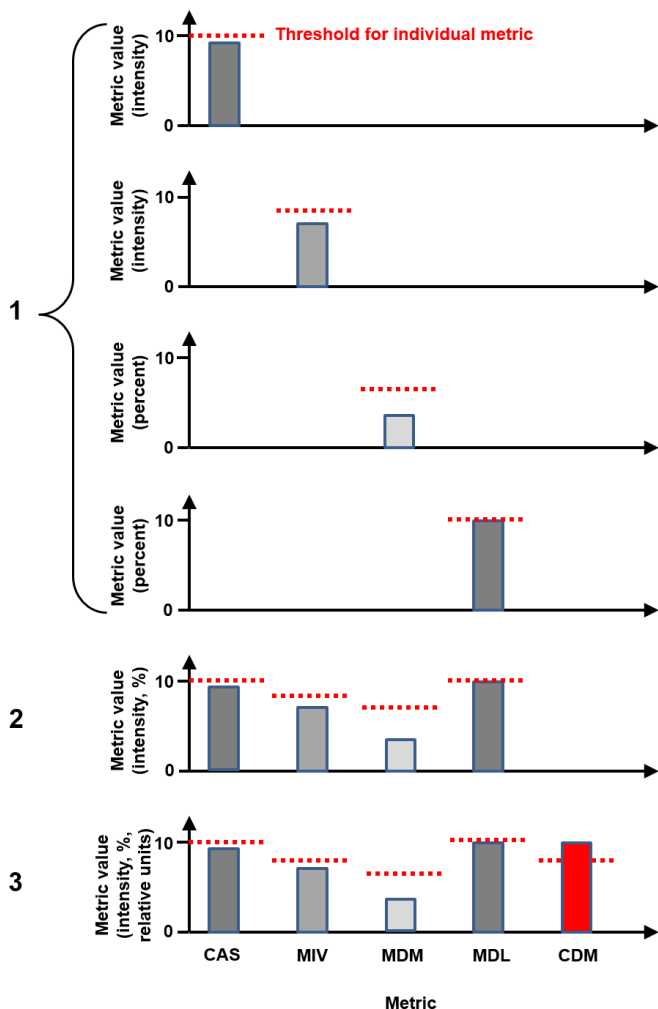


Figure 4. Three paradigms considered in this study. Each of the metrics is considered individually, then together, then in combination with the composite difference metric. The scales shown on the vertical axes are realistic, but the metrics and their thresholds are schematic only. CAS, central axis signal; CDM, composite difference metric; MDL, mean difference as a percentage of local intensity; MDM, mean difference as a percentage of maximum intensity; MIV, mean image value.



found, and the median and range of these maxima over the four patients were found. The value of the median plus range was taken to be the threshold. A distribution of maximum values was therefore implicitly assumed for the error-free cases, and the median plus range was taken to be the upper extreme of that distribution.

The optimisation process itself consisted of an exhaustive search over the combinations of  $w_b$ , with allowed values ranging from 0.00 to 1.00 in steps of 0.05. In contrast with the individual metrics, the thresholds for the CDM varied according to the weighting factors used to calculate the CDM. This was simply because the value of the CDM changed according to the weighting factors used. The threshold value was therefore recalculated every time that the weighting factors were updated, by taking the median plus range of the maximum error value in the

normal deliveries for the four patients. The search method investigated each of the four patients in turn, and within each patient, the error cases in turn. The method examined each running sum of 10 control points in turn and evaluated both the individual metrics and composite metric. The index of the first segment at which an error was detected by any of these five metrics passing the threshold was noted and the objective value was calculated as the mean of these indices over the four patients and various error cases. The final objective value was therefore the minimum value of the mean segment index for all patients and error cases at which the errors could be detected.

### Evaluation of detection ability

The thresholds for the individual metrics and the CDM, together with the weights of the CDM, were then applied to the two evaluation patients. For comparison purposes, the maximum values of the metrics, based on the section images (*i.e.* running sums of 10 control points), reached over segments 19–180, were evaluated. The mean index for the two patients at which an error was detected was also taken to represent the effectiveness of the metric in identifying an error. Specifically, three different error detection scenarios were evaluated (Figure 4):

- (1) Each of the four individual metrics was considered in isolation. For each metric in turn, the threshold was set and the mean index at which the errors were detected was noted.
- (2) The four individual metrics were considered separately but simultaneously. This corresponded most closely to what is currently carried out clinically. The threshold was determined for each individual metric, and then the first index at which *any* of the four metrics exceeded its respective threshold was noted.
- (3) The CDM was used in addition to the four individual metrics separately, *i.e.* in addition to Scenario 2. This scenario represented the results of the optimisation process.

## RESULTS

For all error-free treatment plans, the dose measured using a  $0.6\text{ cm}^3$  ionisation chamber at the centre of the water-equivalent phantom is within 1% of the volume-averaged dose predicted by the AutoBeam inverse treatment planning system. Examples of the section values (*i.e.* running average of 10 segments) of the four individual metrics are shown in Figure 5 for the error-free case.

The complete set of weighting factors searched by the optimisation method for the CDM constitutes a four-dimensional space, which is difficult to visualise in its entirety, but several planes through this space are shown in Figure 6. Each plot shows the variation of the mean segment index at which errors are first identified, which is the objective function value used by the weight optimisation. In each plot, the mean segment index is shown as a function of  $w_1$ , the weight given to the central axis signal, and  $w_2$ , the weight given to the mean image value. The four plots show this region for four different combinations of  $w_3$ , mean difference relative to maximum image intensity, and  $w_4$ , mean difference relative to local image intensity. For simplicity, this result does not use the leave-pair-out strategy, so all six

Figure 5. Section values of the four difference metrics for error-free delivery in one patient. The green background indicates that the error threshold has not been exceeded, and in the real-time implementation, this background changes to orange (near miss) or red (error) for those segments that exceed the chosen thresholds. Note that the scales differ on the vertical axes.

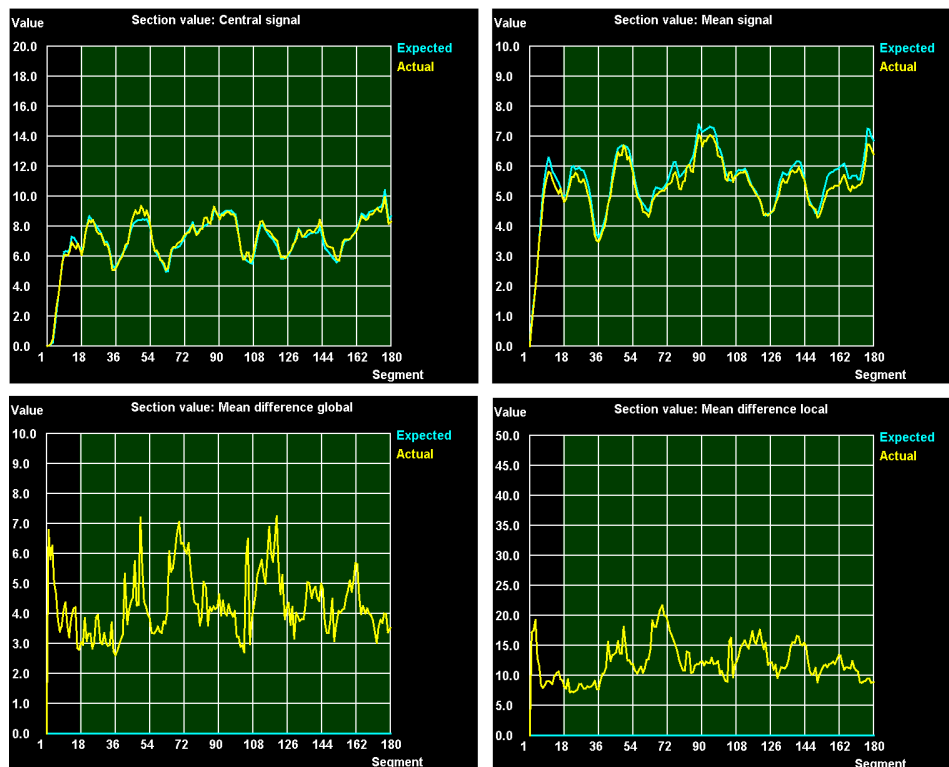


Figure 6. Mean segment index at which errors are first identified, as a function of the weighting factors,  $w$ .

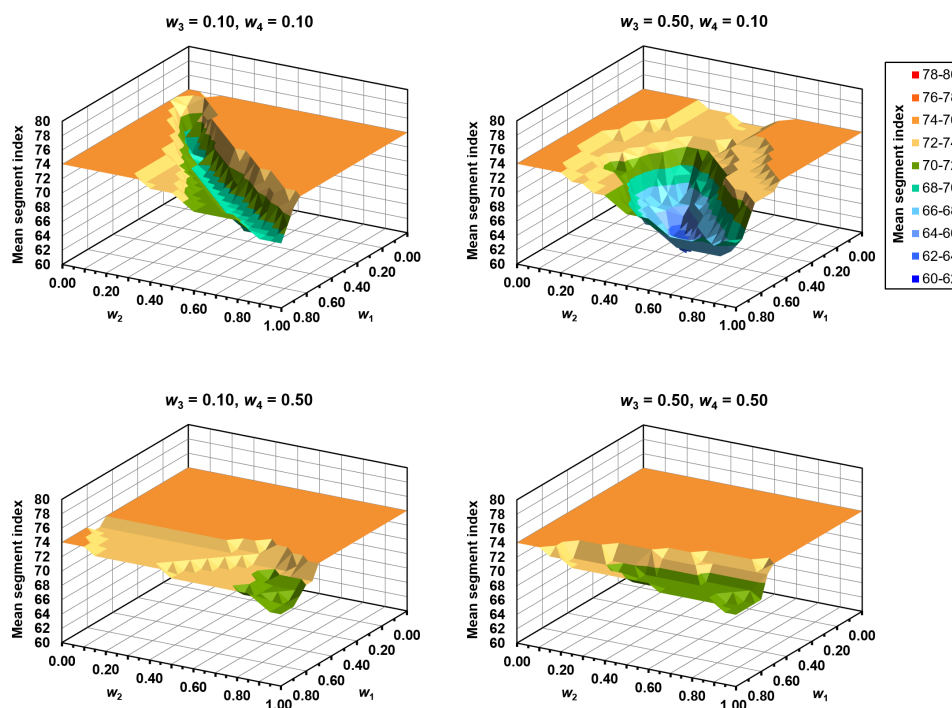


Table 1. Median and range values of the optimal weightings for the composite difference metric, obtained using the leave-pair-out method

Metric	Weight
CAS	0.15 (0.00–0.45)
MIV	0.10 (0.10–0.35)
MDM	0.15 (0.10–0.25)
MDL	0.00 (0.00–0.20)

CAS, central axis signal; MDL, mean absolute difference relative to local image intensity; MDM, mean absolute difference relative to image maximum; MIV, mean image value.

patients are included. The value of 74 at which the segment index plateaus is the value attained by the individual metrics, so the combination of the individual metrics and the CDM always achieves a lower value than this.

At low values of  $w_3$  and  $w_4$ , there is a clear minimum in the value of the segment index, reached with approximately equal values of  $w_1$  and  $w_2$ . This minimum region rotates somewhat at higher values of  $w_3$  and  $w_4$ , and is less marked. At values of  $w_3$  and  $w_4$  approaching unity, the segment index levels out at the plateau value of 74.

The median and range of the CDM weights chosen by the optimisation process are shown in Table 1. The constituent metrics are similarly weighted, with the exception of the mean difference relative to local image intensity. The form of the optimisation space for this exact combination of  $w_3$  and  $w_4$  is shown in Figure 7.

The segment index at which the errors are identified is summarised in Table 2. Each of the nine combinations of patient in the leave-pair-out approach yields a mean segment index for error identification, and the table shows the median and range of these values. Considering the metrics individually, the central axis signal is moderately sensitive to errors and allows timely detection of errors. Mean image value is not responsive to errors,

Figure 7. Mean segment index at which errors are first identified, as a function of the weighting factors,  $w$ , for the optimal combination of  $w_3$  and  $w_4$ .

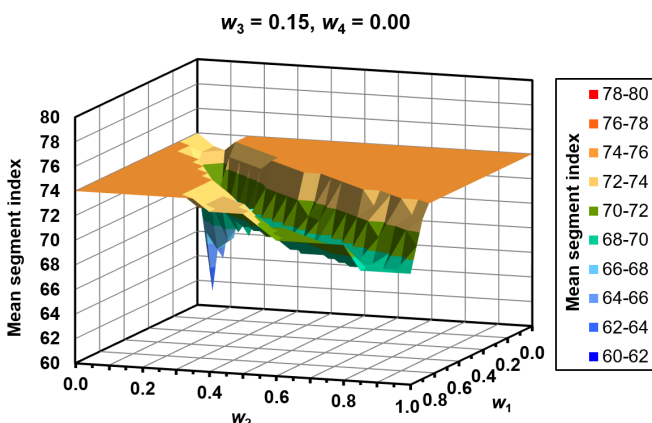


Table 2. Median and range values of the mean segment index at which errors are detected, using individual metrics, multiple separate metrics and the composite difference metric

Metric	Segment index
CAS	136 (58–170)
MIV	156 (138–181)
MDM	101 (82–146)
MDL	99 (85–141)
All of the above with respect to their own respective thresholds	87 (40–130)
All of the above plus CDM	73 (35–95)

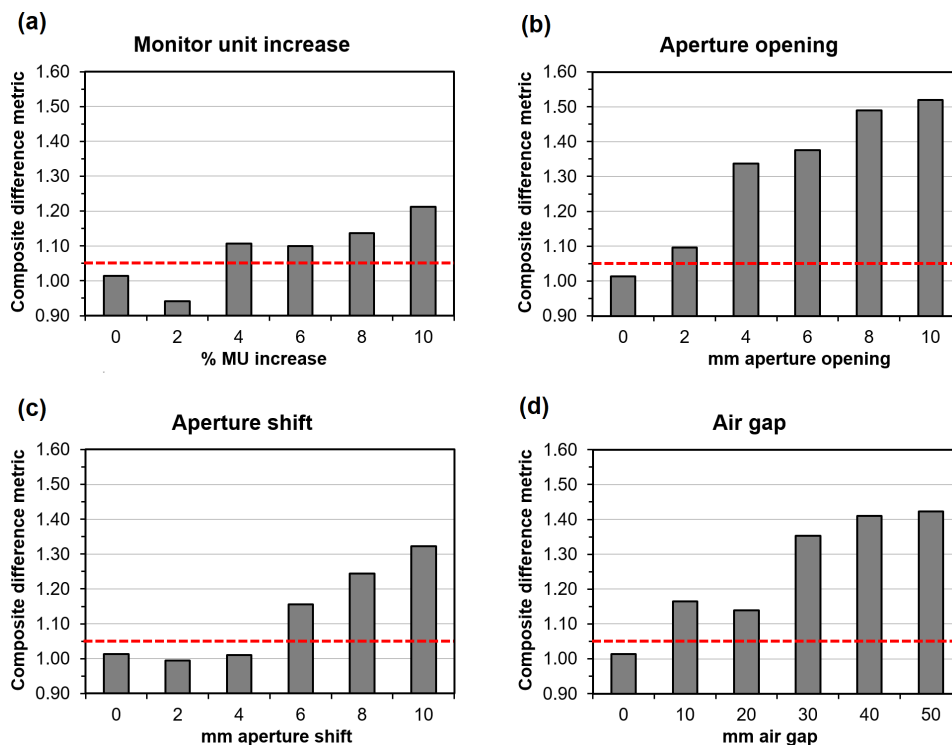
CAS, central axis signal; CDM, composite difference metric; MDL, mean absolute difference relative to local image intensity; MDM, mean absolute difference relative to image maximum; MIV, mean image value.

because an increase in image intensity in one part of the image can be offset by a decrease in image intensity elsewhere. The mean absolute differences are the most effective ways of detecting errors in a timely manner, if a single metric is to be used. One patient of the six exhibits a false-positive result for the central axis signal and mean image value in all of the three cases where that patient is used in the test set. In other words, the thresholds for those image metrics are exceeded for the normal delivery.

Taking all of these metrics and examining them with respect to their respective thresholds gives a more timely response to errors than using any single metric. This is also shown in Table 2. A further reduction in mean value of segment index at which errors are detected is seen when the composite difference metric is used in addition to the individual metrics. The benefit of the CDM, expressed as a ratio of the segment index for error detection using multiple separate metrics plus CDM vs that using only multiple separate metrics, has a median value of 0.83 (range 0.73–0.98). The one patient that exhibits a false-positive result for the individual metrics also shows a false-positive result for multiple separate metrics and for the CDM, since these metrics are based on the individual metrics (see paradigms 2 and 3 of Figure 4).

These results are based on multiple patients, various types of error and various magnitudes of error taken together, but the magnitudes of the composite difference metric for the various errors separately are shown for a single patient in Figure 8. This figure shows the maximum value of the CDM attained in segments 19–180. The indices of the segments at which the errors are first detected, by passing the thresholds, are shown for a single error magnitude in all patients in Figure 9. Note that the values of the individual metrics are not explicitly included in Figure 8, but they are used both individually and in the CDM in Figure 9. From Figure 9, it can be seen that an aperture opening is the earliest to be detected of the four types of error, followed by an aperture shift, with a monitor unit error and air gap error being less easily detected. It is clear that the inclusion of the CDM increases the responsiveness of the system to errors. For the 4 mm aperture opening, there is

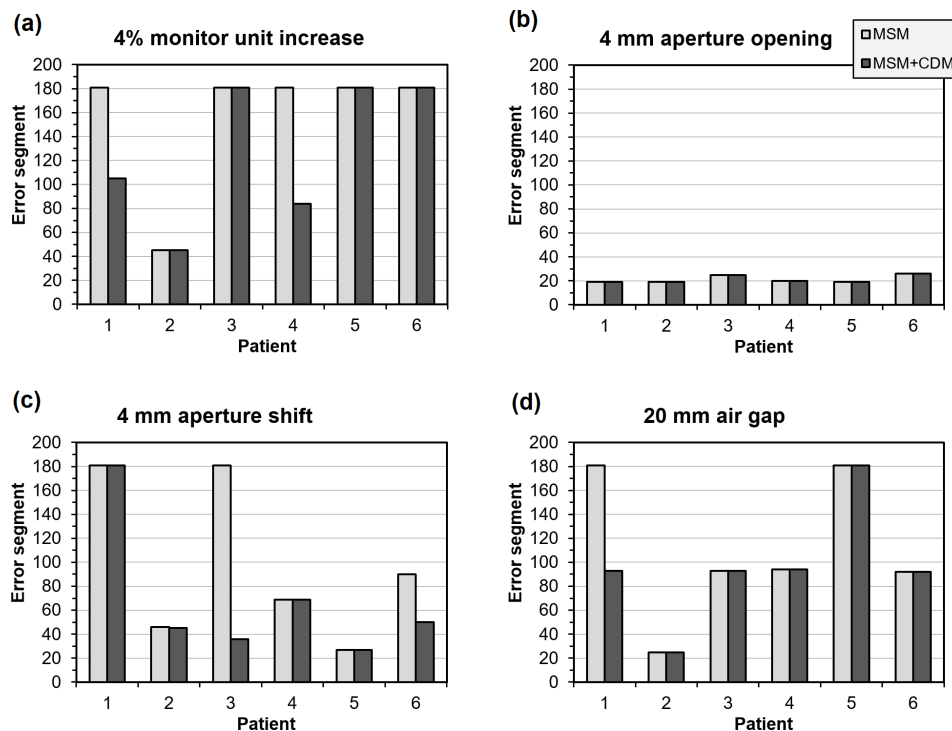
Figure 8. Maximum value attained by the composite difference metric in a single patient over segments 19-180 for various error types and magnitudes. The dashed line shows the threshold, based on the case with no error.



little difference between multiple separate metrics and use of the CDM due to the immediate detection of errors at Segment 19 in both instances. There is also very little difference in the

case of a 4 mm aperture shift, for a similar reason. However, in certain cases, there are very marked improvements in error detection.

Figure 9. First segment at which an error is detected in the six patients, for various error types. CDM, composite difference metric; MSM, multiple separate metrics.





## DISCUSSION

The composite difference metric used in this study shows improved timeliness of error detection compared to a single difference metric or multiple difference metrics compared against their own respective thresholds. The mathematical form of the CDM itself gives maximum benefit when several individual metrics show slightly elevated values. Other forms may be possible, but empirical experimentation has shown that forms such as a simple-weighted sum or weighted product are not as effective as the form given by Equation (1), and in some cases fail to produce any benefit at all. Nevertheless, the metrics used in this study are presented as a proof of concept, and there is scope for further investigation of different forms of metric. With this in view, the optimisation method presented in this study (Figure 3) is not dependent on the exact form of the CDM used, as the CDM is merely evaluated at each iteration of the process. The magnitude of the improvement with the CDM is around 15–20%, which is expected to translate into a corresponding sparing of incorrectly delivered dose, should an error occur.

Unlike the specific tests used by Passarge et al,<sup>31</sup> each of which is designed to identify a particular error, the metrics used in the present study are general image comparison metrics. This is to enable as broad as possible a range of errors to be identified. Furthermore, it is likely that the same approach can be used in the context of back-projection portal dosimetry, but with the metrics chosen as structure-based, dosimetric metrics. For example, Olaciregui-Ruiz et al<sup>22</sup> use a variety of dosimetric statistics in conjunction with the back-projection method to identify anatomical changes during treatment. Similarly, the optimal weightings of the composite difference metric may vary for different treatment sites, in the same way that alert criteria may be adapted specifically for each treatment site.<sup>32</sup>  $\gamma$  analysis has not been used in this study as allowing a small spatial error in the segments of a VMAT plan may cause too large an error in the overall delivered dose.

As Mijnheer et al<sup>33</sup> point out, the power to detect errors depends on the baseline accuracy of the portal dosimetry prediction method. If there is a large difference between predicted and measured images in the absence of errors, the thresholds are high, so that errors are less quickly detected. With this in mind, care has been taken in this study to ensure as accurate a prediction model as possible, with the inclusion of couch attenuation and gantry sag in the model. The images are initially collated according to segment of the treatment plan and then a running sum of 10 segments is used to give an optimum balance between agreement with predictions and useful information on sections of the arc. Similar considerations are also used by other authors.<sup>34,35</sup>

Even with careful attention to the quality of the prediction and the handling of the images, there is still some dependence on the thresholds used to signify an error. In clinical practice, the thresholds are set from a sample of normal patient results, and then applied to further patients. This has been modelled in this study by calculating the thresholds on the optimisation patients and then applying them to the evaluation patients. By estimating

the maximum metric value likely to be encountered in the normal deliveries using the distribution of maxima in the optimisation patients, the number of false positives is minimised. However, one patient in this study, with higher metric values for the normal delivery, gives false-positive results. False positives are particularly unwanted in the real-time context as they are disruptive to the treatment of the patient. It is clear that a much larger collection of normal data are needed for the establishment of optimal thresholds, so that errors can be detected quickly, but false positives do not occur frequently.

This observation suggests that the use of artificial neural networks and deep learning may be beneficial for combining the metrics without requiring explicit thresholds.<sup>36–38</sup> In this case, the thresholds are implicitly built into the neural network in the various nodes and layers from which the network is composed. The connection together of the initial nodes of the network by the deeper layers may provide a similar effect to the composite difference metric examined in the present study.

As with other studies,<sup>26,33,39</sup> the method more readily identifies errors in MLC leaf positioning and is less accurate in identifying material changes such as air gap errors (Figure 9). The sensitivity to MLC opening errors occurs because there is an increased output factor as well as a shift in position. Consequently, this type of error can be identified quickly in the sequence of segments. This study does not investigate patient positioning errors as portal dosimetry is not very sensitive to positioning errors,<sup>5,33</sup> and is therefore best implemented in conjunction with some means of position verification such as cone beam CT. There are indications, however, that the segment-resolved nature of real-time portal dosimetry does allow for some detection of these errors.<sup>21</sup>

## CONCLUSIONS

A composite difference metric constructed from a weighted power sum of individual metrics is a valuable addition to the process of error detection in the context of intrafraction segment-resolved portal dosimetry. The metric is able to identify variations in measured portal images with respect to predicted images before the separate measures exceed their respective tolerances. Optimal values of the weighting factors for the metric can be found by using an exhaustive search in conjunction with appropriate thresholds based on normal deliveries. Application of the method to a series of prostate treatment images shows that simulated errors can be detected approximately 20% earlier than when using several measures separately.

## FUNDING

The authors acknowledge funding from the National Institute for Health Research (NIHR) Biomedical Research Centre at the Royal Marsden NHS Foundation Trust and the Institute of Cancer Research. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. We are also supported by a Cancer Research UK Centres Network Accelerator Award Grant (A21993) to the ART-NET consortium.

## REFERENCES

- van Elmpt W, McDermott L, Nijsten S, Wendling M, Lambin P, Mijnheer B. A literature review of electronic portal imaging for radiotherapy dosimetry. *Radiother Oncol* 2008; **88**: 289–309. doi: <https://doi.org/10.1016/j.radonc.2008.07.008>
- Mijnheer B, Beddar S, Izewska J, Reft C. In vivo dosimetry in external beam radiotherapy. *Med Phys* 2013; **40**: 070903. doi: <https://doi.org/10.1118/1.4811216>
- McCurdy B, Greer P, Bedford J. Electronic portal imaging device dosimetry. In: Mijnheer B, ed. *Clinical 3D dosimetry in modern radiation therapy*. Boca Raton, FL: Taylor and Francis; 2018. pp. 169–98.
- Chytky P, Praznik K, Van Uytven E, van Beek TA, Greer PB, McCurdy BMC. Model-based prediction of portal dose images during patient treatment. *Med Phys* 2013; **40**: 031713. doi: <https://doi.org/10.1118/1.4792203>
- Bedford JL, Hanson IM, Hansen VN. Portal dosimetry for VMAT using integrated images obtained during treatment. *Med Phys* 2014; **41**: 021725. doi: <https://doi.org/10.1118/1.4862515>
- Wendling M, McDermott LN, Mans A, Sonke J-J, van Herk M, Mijnheer BJ. A simple backprojection algorithm for 3D in vivo EPID dosimetry of IMRT treatments. *Med Phys* 2009; **36**: 3310–21. doi: <https://doi.org/10.1118/1.3148482>
- Mans A, Remeijer P, Olaciregui-Ruiz I, Wendling M, Sonke J-J, Mijnheer B, et al. 3D Dosimetric verification of volumetric-modulated arc therapy by portal dosimetry. *Radiother Oncol* 2010; **94**: 181–7. doi: <https://doi.org/10.1016/j.radonc.2009.12.020>
- Van Uytven E, Van Beek T, McCowan PM, Chytky P, Praznik K, Greer PB, McCurdy BMC. Validation of a method for in vivo 3D dose reconstruction for IMRT and VMAT treatments using on-treatment EPID images and a model-based forward-calculation algorithm. *Med Phys* 2015; **42**: 6945–54. doi: <https://doi.org/10.1118/1.4935199>
- Esposito M, Villaggi E, Bresciani S, Cilla S, Falco MD, Garibaldi C, et al. Estimating dose delivery accuracy in stereotactic body radiation therapy: a review of in-vivo measurement methods. *Radiother Oncol* 2020; **149**: 158–67. doi: <https://doi.org/10.1016/j.radonc.2020.05.014>
- Mijnheer BJ, González P, Olaciregui-Ruiz I, Rozendaal RA, van Herk M, Mans A. Overview of 3-year experience with large-scale electronic portal imaging device-based 3-dimensional transit dosimetry. *Pract Radiat Oncol* 2015; **5**: e679–87. doi: <https://doi.org/10.1016/j.prro.2015.07.001>
- Nailon WH, Welsh D, McDonald K, Burns D, Forsyth J, Cooke G, et al. EPID-based in vivo dosimetry using dosimetry check™: overview and clinical experience in a 5-yr study including breast, lung, prostate, and head and neck cancer patients. *J Appl Clin Med Phys* 2019; **20**: 6–16. doi: <https://doi.org/10.1002/acm2.12441>
- Matsushita N, Nakamura M, Sasaki M, Yano S, Yoshimura M, Mizowaki T. Analyses of integrated EPID images for on-treatment quality assurance to account for interfractional variations in volumetric modulated arc therapy. *J Appl Clin Med Phys* 2020; **21**: 110–6. doi: <https://doi.org/10.1002/acm2.12805>
- Fidanzio A, Porcelli A, Azario L, Greco F, Cilla S, Grusio M, et al. Quasi real time in vivo dosimetry for VMAT. *Med Phys* 2014; **41**: 062103. doi: <https://doi.org/10.1118/1.4875685>
- Woodruff HC, Fuangrod T, Van Uytven E, McCurdy BMC, van Beek T, Bhatia S, et al. First experience with real-time EPID-based delivery verification during IMRT and VMAT sessions. *Int J Radiat Oncol Biol Phys* 2015; **93**: 516–22. doi: <https://doi.org/10.1016/j.ijrobp.2015.07.2271>
- Spreeuw H, Rozendaal R, Olaciregui-Ruiz I, González P, Mans A, Mijnheer B, et al. Online 3D EPID-based dose verification: proof of concept. *Med Phys* 2016; **43**: 3969–74. doi: <https://doi.org/10.1118/1.4952729>
- Bedford JL, Hanson IM. A method to verify sections of arc during intrafraction portal dosimetry for prostate VMAT. *Phys Med Biol* 2019; **64**: 205009. doi: <https://doi.org/10.1088/1361-6560/ab47c8>
- Lukka HR, Pugh SL, Bruner DW, Bahary J-P, Lawton CAF, Efstathiou JA, et al. Patient reported outcomes in NRG Oncology RTOG 0938, evaluating two ultrahypofractionated regimens for prostate cancer. *Int J Radiat Oncol Biol Phys* 2018; **102**: 287–95. doi: <https://doi.org/10.1016/j.ijrobp.2018.06.008>
- Bezjak A, Paulus R, Gaspar LE, Timmerman RD, Straube WL, Ryan WF, et al. Safety and efficacy of a five-fraction stereotactic body radiotherapy schedule for centrally located non-small-cell lung cancer: NRG Oncology/ RTOG 0813 trial. *J Clin Oncol* 2019; **37**: 1316–25. doi: <https://doi.org/10.1200/JCO.2018.06.0622>
- Murray Brunt A, Haviland JS, Wheatley DA, Sydenham MA, Alhasso A, Bloomfield DJ, et al. Hypofractionated breast radiotherapy for 1 week versus 3 weeks (FAST-Forward): 5-year efficacy and late normal tissue effects results from a multicentre, non-inferiority, randomised, phase 3 trial. *Lancet* 2020; **395**: 1613–26. doi: [https://doi.org/10.1016/S0140-6736\(20\)30932-6](https://doi.org/10.1016/S0140-6736(20)30932-6)
- Persoon LCGG, Podesta M, Nijsten SMJJG, Troost EGC, Verhaegen F. Time-resolved versus integrated transit planar dosimetry for volumetric modulated arc therapy: patient-specific dose differences during treatment, a proof of principle. *Technol Cancer Res Treat* 2016; **15**: NP79–87. doi: <https://doi.org/10.1177/1533034615617668>
- Schyns LEJR, Persoon LCGG, Podesta M, van Elmpt WJC, Verhaegen F. Time-resolved versus time-integrated portal dosimetry: the role of an object's position with respect to the isocenter in volumetric modulated arc therapy. *Phys Med Biol* 2016; **61**: 3969–84. doi: <https://doi.org/10.1088/0031-9155/61/10/3969>
- Olaciregui-Ruiz I, Rozendaal R, Mijnheer B, Mans A. Site-specific alert criteria to detect patient-related errors with 3D EPID transit dosimetry. *Med Phys* 2019; **46**: 45–55. doi: <https://doi.org/10.1002/mp.13265>
- South CP, Khoo VS, Naismith O, Norman A, Dearnaley DP. A comparison of treatment planning techniques used in two randomised UK external beam radiotherapy trials for localised prostate cancer. *Clin Oncol* 2008; **20**: 15–21. doi: <https://doi.org/10.1016/j.clon.2007.10.012>
- Dearnaley D, Syndikus I, Sumo G, Bidmead M, Bloomfield D, Clark C, et al. Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: preliminary safety results from the CHHiP randomised controlled trial. *Lancet Oncol* 2012; **13**: 43–54. (supplementary appendix). doi: [https://doi.org/10.1016/S1470-2045\(11\)70293-5](https://doi.org/10.1016/S1470-2045(11)70293-5)
- Bedford JL. Sinogram analysis of aperture optimization by iterative least-squares in volumetric modulated arc therapy. *Phys Med Biol* 2013; **58**: 1235–50. doi: <https://doi.org/10.1088/0031-9155/58/5/1235>
- Bedford JL, Hanson IM, Hansen VN. Comparison of forward- and back-projection in vivo EPID dosimetry for VMAT treatment of the prostate. *Phys Med Biol* 2018; **63**: 025008. doi: <https://doi.org/10.1088/1361-6560/aa9c60>
- Poludniowski G, Thomas MDR, Evans PM, Webb S. CT reconstruction from portal images acquired during volumetric-modulated arc therapy. *Phys Med Biol* 2010; **55**: 5635–51. doi: <https://doi.org/10.1088/0031-9155/55/19/002>

28. Hanson IM, Hansen VN, Olaciregui-Ruiz I, van Herk M. Clinical implementation and rapid commissioning of an EPID based *in-vivo* dosimetry system. *Phys Med Biol* 2014; **59**: N171–9. doi: <https://doi.org/10.1088/0031-9155/59/19/N171>
29. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *proceedings of the 14th international joint conference on artificial intelligence; 1995 August; Montreal, Quebec, Canada*. San Francisco, CA: Morgan Kaufmann; 1995.
30. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2nd edition. New York, NY: Springer; 2017.
31. Passarge M, Fix MK, Manser P, Stampanoni MFM, Siebers JV. A Swiss cheese error detection method for real-time EPID-based quality assurance and error prevention. *Med Phys* 2017; **44**: 1212–23. doi: <https://doi.org/10.1002/mp.12142>
32. Fuangrod T, Greer PB, Woodruff HC, Simpson J, Bhatia S, Zwan B, et al. Investigation of a real-time EPID-based patient dose monitoring safety system using site-specific control limits. *Radiat Oncol* 2016; **11**: 106. doi: <https://doi.org/10.1186/s13014-016-0682-y>
33. Mijnheer B, Jomehzadeh A, González P, Olaciregui-Ruiz I, Rozendaal R, Shokrani P, et al. Error detection during VMAT delivery using EPID-based 3D transit dosimetry. *Phys Med* 2018; **54**: 137–45. doi: <https://doi.org/10.1016/j.ejmp.2018.10.005>
34. McCowan PM, McCurdy BMC. Frame average optimization of cine-mode EPID images used for routine clinical *in vivo* patient dose verification of VMAT deliveries. *Med Phys* 2016; **43**: 254–61. doi: <https://doi.org/10.1118/1.4938413>
35. Cools RAM, Dirkx MLP, Heijmen BJM. A novel method for sub-arc VMAT dose delivery verification based on portal dosimetry with an EPID. *Med Phys* 2017; **44**: 5556–62. doi: <https://doi.org/10.1002/mp.12518>
36. Gulliford SL, Webb S, Rowbottom CG, Corne DW, Dearnaley DP. Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate. *Radiother Oncol* 2004; **71**: 3–12. doi: <https://doi.org/10.1016/j.radonc.2003.03.001>
37. Mahdavi SR, Tavakol A, Sanei M, Molana SH, Arbabi F, Rostami A, et al. Use of artificial neural network for pretreatment verification of intensity modulation radiation therapy fields. *Br J Radiol* 2019; **92**: 20190355. doi: <https://doi.org/10.1259/bjr.20190355>
38. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. *Med Phys* 2019; **46**: e1–36. doi: <https://doi.org/10.1002/mp.13264>
39. Bojecho C, Ford EC. Quantifying the performance of *in vivo* portal dosimetry in detecting four types of treatment parameter variations. *Med Phys* 2015; **42**: 6912–8. doi: <https://doi.org/10.1118/1.4935093>