

**TECHNICAL NOTE**

# Technical note: A modified gamma evaluation method for dose distribution comparisons

Liting Yu<sup>1,2</sup> | Tanya Kairn<sup>1,2</sup> | Jamie Trapp<sup>2</sup> | Scott B. Crowe<sup>1,2</sup><sup>1</sup>Royal Brisbane & Women's Hospital, Herston, QLD, Australia<sup>2</sup>Queensland University of Technology, Brisbane, QLD, AustraliaAuthor to whom correspondence should be addressed. Liting Yu  
E-mail: Nancy.yu@health.qld.gov.au**Abstract**

**Purpose:** In this work we have developed a novel method of dose distribution comparison, the inverse gamma (IG) evaluation, by modifying the commonly used gamma evaluation method.

**Methods:** The IG evaluation calculates the gamma criteria (dose difference criterion,  $\Delta D$ , or distance-to-agreement criterion,  $\Delta d$ ) that are needed to achieve a predefined pass rate or gamma agreement index (GAI). In-house code for evaluating IG with a fixed  $\Delta D$  of 3% was developed using Python (v3.5.2) and investigated using treatment plans and measurement data from 25 retrospective patient specific quality assurance tests (53 individual arcs).

**Results:** It was found that when the desired GAI was set to 95%, approximately three quarters of the arcs tested were able to achieve  $\Delta d$  within 1 mm (mean  $\Delta d$ :  $0.7 \pm 0.5$  mm). The mean  $\Delta d$  required in order for all points to pass the gamma evaluation (i.e., GAI = 100%) was  $4.5 \pm 3.1$  mm. The possibility of evaluating IG by fixing the  $\Delta d$  or  $\Delta D/\Delta d$ , instead of fixing the  $\Delta D$  at 3%, was also investigated.

**Conclusion:** The IG method and its indices have the potential to be implemented clinically to quantify the minimum dose and distance criteria based on a specified GAI. This method provides additional information to augment standard gamma evaluation results during patient specific quality assurance testing of individual treatment plans. The IG method also has the potential to be used in retrospective audits to determine an appropriate set of local gamma criteria and action levels based on a cohort of patient specific quality assurance plans.

**KEY WORDS**

dose distribution comparison, gamma evaluation, patient specific quality assurance

## 1 | INTRODUCTION

Contemporary radiation therapy techniques can involve the use of complex radiation fields delivered with moving or static gantries to deliver modulated dose distributions. These techniques, which include intensity modulated radiation therapy (IMRT), volumetric modulated arc therapy (VMAT) and helical tomotherapy (HT), require careful quality assurance tests, to ensure that the planned dose

distributions can be delivered by the treatment system<sup>1–5</sup>. Patient specific quality assurance (PSQA) is often performed prior to the radiation treatment, where the treatment beams are delivered to a phantom and the radiation doses are verified against those predicted in the treatment planning system (TPS). Dose comparisons are used to compare the evaluated dose distribution from measurements against the reference dose distribution from TPS calculations. Currently the gamma index method<sup>6</sup> is the predominant dose

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Journal of Applied Clinical Medical Physics* published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine.

comparison method being used clinically, although several alternative methods have been proposed<sup>7–12</sup>.

Derived from the dose difference ( $\Delta D$ ) test and the distance-to-agreement ( $\Delta d$ ) test<sup>10</sup>, the gamma index method calculates the difference between two dose grids in a combined spatial-dose domain<sup>6</sup>. The result of the gamma test can be summarized by a single percentage value, usually referred to as the “pass rate” or “gamma agreement index” (GAI), which describes the percentage of points in the two dose distributions that agree within specified  $\Delta D$  and  $\Delta d$  (producing a gamma value less than or equal to 1.0). Gamma evaluation results can also be plotted as a two-dimensional gamma distribution with desired spatial resolution, as well as histograms, so that the locations of regions of disagreement can be identified and investigated<sup>6</sup>. The gamma evaluation method has the advantage of producing a quantitative measure based on both dose and spatial criteria, so that large dose differences occurring in high-dose-gradient regions do not disproportionately affect the results of the comparison. The gamma evaluation method has, however, been criticized for being less clinically intuitive than more conventional dose-comparison methods<sup>8</sup>, being sensitive to dose grid resolution<sup>13</sup> and having poor sensitivity and specificity to clinical dosimetric inaccuracies (when evaluated in terms of global dose difference)<sup>14–20</sup>.

Several alternative dose comparison methods have been proposed, to avoid some of the perceived weaknesses of the gamma index method<sup>7–12</sup>. Specifically, several algorithms have been proposed which attempt to account for the differing levels of biological relevance associated with comparison results in different regions of the dose distribution. For example, the normalized agreement test (NAT)<sup>7</sup>, maximum allowed dose difference (MADD) method<sup>8</sup> and the divide and conquer (D&C) gamma method<sup>11</sup> all vary tolerances in dose-differences in high-dose or high-dose-gradient regions that may correspond to clinically important regions. A concept of radiobiological gamma index (Sumida method)<sup>12</sup> has been proposed to integrate radiobiological parameters such as tumor control and normal tissue complication probabilities into gamma index calculation and produces more clinically relevant results. The proposed DVH-based analysis<sup>15,21–24</sup> allows different criteria to be used in each volume depending on clinical significance or required precision, which could be more relevant than simply judging the overall agreement<sup>25</sup>. These and other alternative dose comparison methods share the disadvantage of providing results that are difficult to compare and benchmark against historical data or other sources (other radiation oncology centres<sup>26</sup>, auditing bodies<sup>27</sup> or established quality assurance guidelines<sup>28,29</sup>), given the widespread adoption and acceptance of the gamma evaluation method.

This study investigates a modified gamma evaluation method, the “inverse gamma” (IG) evaluation, which calculates the gamma evaluation criteria ( $\Delta D$  or  $\Delta d$ ) that would be needed to achieve a predefined GAI. Li et al.<sup>30</sup> proposed a similar approach where the passing percentage is fixed and combination of  $\Delta D$  and  $\Delta d$  was calculated; however, has not been implemented into clinical QA. It is expected that the modified IG method proposed in this study will provide additional information for clinical PSQA, to augment

standard gamma evaluation results by providing users with an indication of the minimum  $\Delta d$  for which a specified GAI can be achieved, when (for example) the  $\Delta D$  is set to 3%.

## 2 | MATERIALS AND METHODS

### 2.A | Standard gamma index method

The gamma index at a point  $r_r$  is defined as<sup>6</sup>:

$$\gamma(r_r) = \min_{\{r_e\}} \{\Gamma(r_e, r_r)\} \quad (1)$$

where

$$\Gamma(r_e, r_r) = \sqrt{\frac{\delta^2(r_e, r_r)}{\Delta D^2} + \frac{r^2(r_e, r_r)}{\Delta d^2}} \quad (2)$$

where  $\delta(r_e, r_r)$  is the dose difference between evaluated and reference doses at point  $r$ ,  $\Delta D$  is the dose difference criterion,  $r(r_e, r_r)$  is the spatial distance between evaluated and reference dose points, and  $\Delta d$  is the distance-to-agreement criterion. The GAI is calculated as the percentage of points for which eq. (2) results in a gamma value less than or equal to 1.0, indicating agreement within the specified  $\Delta D$  and  $\Delta d$ . The gamma index method implemented in this study used the global gamma normalization where the  $\Delta D$  is normalized to the global maximum dose.

### 2.B | Inverse gamma with fixed $\Delta D$ (IG $_{\Delta D}$ )

IG $_{\Delta D}$  calculates the minimum distance-to-agreement criterion ( $\Delta d$ ) that is needed to achieve a specified GAI, given a fixed value of the dose difference criterion ( $\Delta D$ ). The fixed  $\Delta D$  used in this study was selected to be 3%, denoted as IG $_{\Delta D = 3\%}$ . The IG algorithm performs iterative global gamma calculations with  $\Delta d$  increasing from 0 mm in 0.1 mm increments, until the specified GAI is reached and the required minimum  $\Delta d$  is reported. The resulting  $\Delta d$  can be denoted as  $\Delta d_{\text{GAI} = 100\% \text{ or } 95\%, \Delta D = 3\%}$ . Clinically this value would then be used to compare against a tolerance. The time required to perform the IG calculations is dependent on the number of iterations required. Lower  $\Delta d$  values require less time to calculate than higher  $\Delta d$  values. On average it takes a few minutes to run on a desktop PC, which is practical in clinical settings.

As an example, the  $\Delta D$  was fixed at 3% in this work and the values of  $\Delta d$  required to achieve GAI values of 95% and 100% were investigated, for a pre-existing set of VMAT PSQA results. The  $\Delta D$  of 3% was chosen for this work because it is very widely recommended and used. Recent surveys<sup>31,32</sup> suggested that 3%/3 mm are currently the most commonly used gamma evaluation criteria, and the AAPM's task group report on IMRT commissioning (TG-119)<sup>28</sup> used 3%/3 mm and the AAPM's more recent task group report on modulated radiotherapy quality assurance (TG-218)<sup>29</sup> recommended 3%/2 mm. Investigations of cranial and extra-cranial stereotactic radiotherapy have reported the use of 3%/2 mm, 3%/1.5 mm, 3%/1 mm and 3%/0.3 mm criteria<sup>33–40</sup>.

## 2.C | Application

In-house gamma evaluation code was developed using Python v 3.5.2, following the method proposed by Low et al.<sup>6</sup>. The code was validated by establishing agreement with the commercial SNC Patient software package within  $\pm 1.5\%$  [mean 0.3% difference], which falls within the range of variations due to minor differences in algorithm implementation<sup>29</sup> between commercial gamma evaluation software packages reported by TG-218. The in-house code was also validated against the square-field evaluation as per Low & Dempsey<sup>6</sup>.

The code was then modified to include the IG algorithm. A set of pre-existing PSQA results, consisting of 53 arcs from 25 VMAT treatment plans measured using the ArcCheck (Sun Nuclear Corporation, Melbourne, USA) helical diode arrays, were arbitrarily selected and used to validate the performance of the code, by attempting to duplicate the gamma evaluation results produced during conventional PSQA tests using the SNC patient software (version 6.2.2) with Van Dyk global gamma analysis<sup>41</sup> and 2D distance-to-agreement. The lower dose threshold (LDT) was set to 5%, which is consistent with the LDT used in the gamma evaluation.

The VMAT PSQA plans were delivered on Varian IX Clinac linacs (Varian Medical Systems, Palo Alto, USA) and measured using the ArcCheck model 1220, which is a cylindrical water-equivalent phantom with 1386 diode detectors arranged in a spiral pattern. The detector array has its diameter and length of 21 cm. Measurements were taken with a solid PMMA insert placed in the ArcCheck's central cavity. The device was consistently set up so that the centre of the cylinder was at the isocentre. Comparisons were performed between the ArcCheck measured doses and the Eclipse (Varian Medical Systems, Palo Alto, California) calculated doses extracted by the SNC Patient software.

The performance of the standard gamma evaluation calculations by the in-house code was verified by the Low and Dempsey method<sup>6</sup> and established agreement with the output produced by the SNC Patient software. The in-house code was then used to calculate  $IG_{\Delta D = 3\%}$  for all 53 measurements. When calculating  $IG_{\Delta D = 3\%}$ , the target GAI was first defined as 100%, which represents the extreme situation where all points must pass the gamma test. The  $IG_{\Delta D = 3\%}$  analysis was then repeated, with the target GAI set to 95%, which corresponds to two standard deviations and is the action level most commonly used<sup>31</sup>.

## 3 | RESULTS

The mean and standard deviation (SD) of  $IG_{\Delta D = 3\%}$  (when GAI was set to 100% and 95%), compared with their original gamma values, of the 53 VMAT arcs were calculated and displayed in Table 1. The detailed values for each individual arc were displayed in Table A1 in the Appendix.

Figure 1 illustrates an example (arc 1) of the 2D global gamma distribution versus  $\Delta d$  distribution using  $IG_{\Delta D = 3\%}$  with GAI 100%

and 95%. Regions of high and low geometric uncertainties can be easily identified, which is not easily achievable by performing multiple gamma evaluation with varying  $\Delta d$  criteria.

Figure 2 illustrates the results of the IG analysis of the VMAT PSQA results, showing the  $\Delta d$  required to achieve the specified GAI values for each of the 53 arcs, when the  $\Delta D$  is fixed at 3%. The mean  $\Delta d$  required to achieve a GAI of 100% was  $4.5 \pm 3.1$  mm. The number of arcs that achieved a GAI of 100% only when  $\Delta d$  was 10 mm or more provides a graphic indication of the clinical unsuitability of requiring that all points pass the gamma evaluation. The mean  $\Delta d$  required to achieve the more-conventional GAI of 95% was  $0.7 \pm 0.5$  mm, with the majority arcs (75.5%) requiring  $\Delta d$  less than 1 mm.

Examination of the results in Fig. 2 indicates that if initial PSQA testing of these arcs had used gamma evaluation criteria of 3%/2 mm, all arcs would have achieved a GAI greater than 95%. Data in Fig. 2 also indicate that if initial PSQA testing of these arcs had used gamma evaluation criteria of 3%/1 mm, then 75.5% of the arcs would have achieved a GAI greater than 95%.

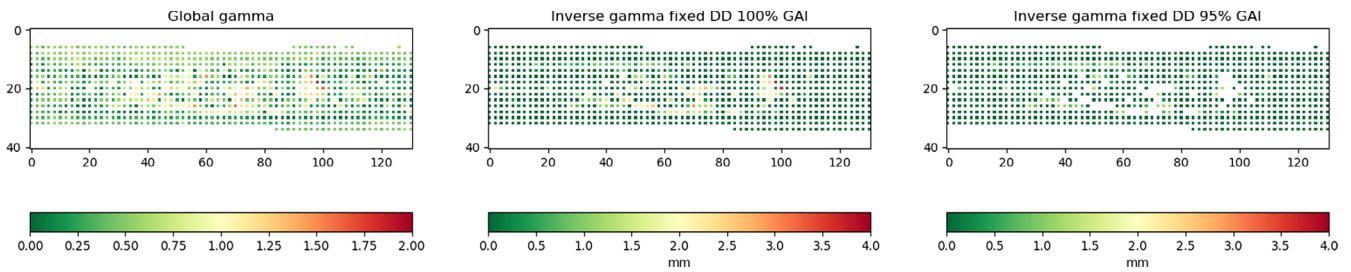
The highest  $\Delta d$  calculated is from arc 2 using GAI of 100%, which indicated that a  $\Delta d$  of nearly 17 mm is needed for all points to pass gamma when the  $\Delta D$  value is set to 3%. However, when GAI of 95% is used,  $\Delta d$  has been significantly dropped to only 1.6 mm. This behavior will be discussed in the next section.

## 4 | DISCUSSION

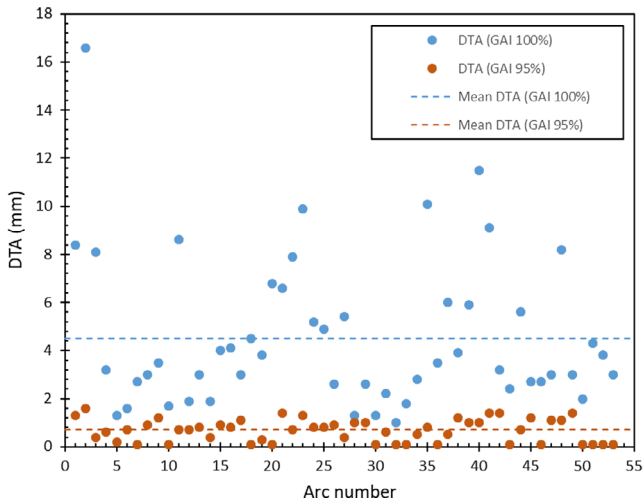
Standard gamma evaluation method cannot provide definitive results (whether in the form of a single numeric GAI value or a 2D gamma map) to answer questions about the magnitude or source of error (whether because of the disagreement in  $\Delta D$  or  $\Delta d$ ). The AAPM's task group report on PSQA (TG-218)<sup>29</sup> has recently recommended that reports of PSQA results include the percentage of gamma values greater than 1.5, as a means to provide an indication of the magnitude of disagreement, but the gamma value of 1.5 is also difficult to interpret – it may be caused by a difference in dose or position or both. DVH based analysis<sup>15,21–24</sup> has potential possibilities to evaluate the clinical importance of any observed differences by seeing the effect on specific structures, which could be more relevant than simply judging the overall agreement<sup>25</sup>. The IG method proposed in this study provides a means to generate two-dimensional maps of  $\Delta D$  or  $\Delta d$  results (in the same way that two-dimensional maps of gamma values are currently provided by commercial gamma evaluation software packages), so that regions of high and low

**TABLE 1** Mean and standard deviation of gamma and  $IG_{\Delta D = 3\%}$  at GAI 100% and 95% calculated from 25 previous VMAT PSQA data (53 arcs).

53 VMAT Arcs	Gamma pass rate (%) 2%, 2 mm, 5% LDT	$\Delta d_{GAI = 100\%, \Delta D = 3\%}$ (mm)	$\Delta d_{GAI = 95\%, \Delta D = 3\%}$ (mm)
Mean $\pm$ SD	97.9 $\pm$ 1.5	4.5 $\pm$ 3.1	0.7 $\pm$ 0.5



**FIG. 1.** Global gamma distribution map (left).  $IG_{\Delta D} = 3\%$  (GAI = 100%)  $\Delta d$  distribution map (middle).  $IG_{\Delta D} = 3\%$  (GAI = 95%)  $\Delta d$  distribution map (right).



**FIG. 2.** DTA values ( $\Delta d$ ) required to achieve GAI values of 95% (darker/orange plot symbols) and 100% (lighter/blue plot symbols), with a set  $\Delta D$  of 3%.

dosimetric or geometric uncertainty can be easily identified. Now with the ability of some commercial software reconstructing 2D dose map onto 3D CT dataset, potentially the shifts identified here might allow correlating the identified geometrical shifts with the borders of organs at risk (OAR) and planning target volumes (PTVs). This method may help address the challenge of implementing DVH or anatomical-structure-based analysis methods into the clinical workflow<sup>25</sup>. For example, the IG method could potentially allow the  $\Delta d$  results identified at the periphery of target structures to be evaluated for compliance with stereotactic criteria, while the  $\Delta d$  throughout the rest of the dose distribution are evaluated more leniently.

TG-218 also states that the  $\Delta d$  should be a function of the clinical necessity of placing steep dose gradients, yet discovering the locations of the dose gradient errors in the patient is difficult with the commonly used IMRT QA methods<sup>29</sup>. The IG method proposed in this study will tackle this difficulty by searching for a single  $\Delta d$  for the structures that provides a pass without running multiple gamma evaluations for each structure. This method also allows a simple check of the value against a tolerance table.

This IG method has been proposed as a means to provide additional information, to augment standard gamma evaluation results

during PSQA testing of individual treatment plans. For example, if the local PSQA action threshold is a GAI of 95%, and a given treatment plan achieves a GAI of 94%, using criteria of 3%/2 mm, instead of iteratively changing the  $\Delta d$  criterion by manual trial and error, the proposed  $IG_{\Delta D} = 3\%$  method can be used as an additional test of verification more efficiently. This method can be used to identify the minimum  $\Delta d$  for which a GAI of 95% is achievable for that plan. The two-dimensional  $\Delta d$  distribution map can also identify the region of high and low  $\Delta d$  values. In cases where it is concerned that there are points having deviations larger than the selected gamma criteria, however, are not able to be identified due to using GAI of 95%, the GAI of 100% can be used to eradicate this common weakness of the conventional gamma evaluation method. It is worth mentioning that the weakness of using GAI of 100% is that it is very sensitive to noise. In cases where there are problems with even a single pixel (noise, dead pixel, etc), the IG method could potentially keep doing iterations until reaching the maximum specified gamma criterion, which is both time consuming and giving less meaningful results (for example arc 2). Therefore, care needs to be taken when using GAI of 100%. The GAI values of near 100%, for example 99%, 99.5%, are highly recommended in order to avoid the disturbance of random noise in the dataset.

Similarly, the IG method may be used to retrospectively audit cohorts of PSQA results (producing results similar to Fig. 2), so that the suitability of existing or proposed gamma evaluation criteria and action thresholds can be evaluated. The data produced by applying IG analysis to aggregated PSQA results may be evaluated using statistical process control (SPC) methods, such as those have been used to identify uncontrolled behaviors in linac and tomotherapy quality assurance results<sup>42–45</sup>.

The 3%  $\Delta D$  criterion was selected in this study because it is widely used, in combination with a range of  $\Delta d$  values<sup>28,29,31,33,35–40</sup>, but the use of 3% as a  $\Delta D$  criterion is not universal<sup>26,34</sup>. Similarly, GAI values of 100% and 95% were selected as examples of extreme and commonly-used PSQA action levels, respectively. The IG method can, however, be used with any other  $\Delta D$  or GAI values, to fit the needs of individual centres.

The IG method also offers more variability and flexibility than simply fixing the  $\Delta D$  criterion. The algorithm is amenable to using a fixed  $\Delta d$  to identify the  $\Delta D$  that would produce a specified GAI. Alternatively, the ratio of  $\Delta D$  to  $\Delta d$  can be fixed, enabling the IG

algorithm to identify the pair of criteria (related to each other via a set ratio) that produce a specified GAI. (Examples of the use of these two additional forms of the IG index are provided in the appendix.)

For centres that are committed to using the standard gamma evaluation method for PSQA, the IG indices developed in this study could be used to investigate or justify the choice of gamma criteria for ongoing PSQA use, based on combination of the local treatment technique and measuring device.

## 5 | CONCLUSIONS

A novel dose comparison method called the inverse gamma (IG) method has been developed. The  $IG_{\Delta D = 3\%}$  index has been tested on 25 retrospective VMAT PSQA plans (53 arcs). This index was proven useful to quantify the minimum  $\Delta d$  based on given  $\Delta D$  of 3% in order to pass a given GAI. This method has the potential to be implemented clinically to perform additional analysis of failed plans and to provide those who prescribe, plan and test modulated radiotherapy treatments with more detailed dosimetric information about the reliability with which planned doses can be delivered. The IG method also has the potential to be used in retrospective internal and inter-departmental audits, to evaluate the suitability of the local gamma evaluation criteria.

## CONFLICT OF INTEREST

The authors have no relevant conflict of interest to disclose.

## REFERENCES

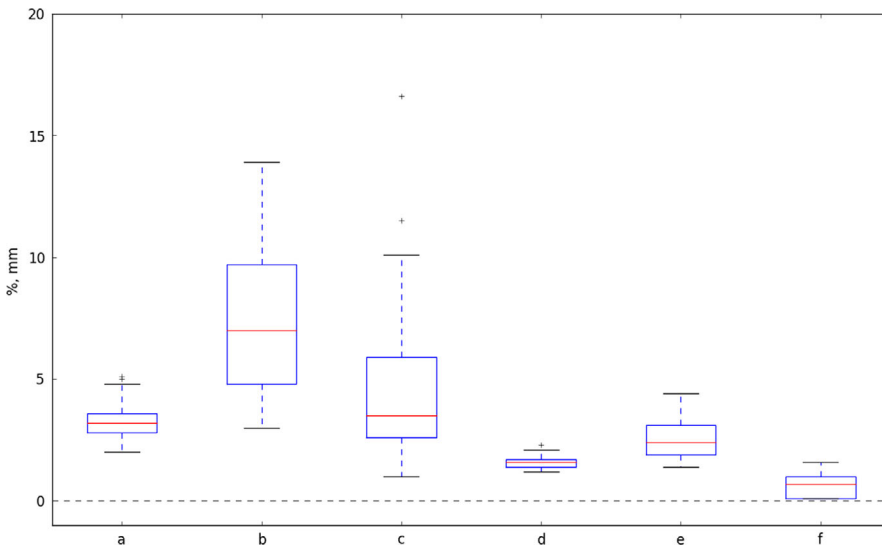
- LoSasso T, Chui CS, Ling CC. Comprehensive quality assurance for the delivery of intensity modulated radiotherapy with a multileaf collimator used in the dynamic mode. *Med Phys*. 2001;28:2209–2219.
- Mell LK, Mehrotra AK, Mundt AJ. Intensity-modulated radiation therapy use in the US, 2004. *Cancer*. 2005;104:1296–1303.
- Cedric XY, Tang G. Intensity-modulated arc therapy: principles, technologies and clinical implementation. *Phys Med Biol*. 2011;56:R31.
- Bridge P, Carmichael MA, Brady C, Dry A. A snapshot of radiation therapy techniques and technology in Queensland: an aid to mapping undergraduate curriculum. *J Med Radiat Sci*. 2013;60:25–34.
- Bridge P, Dempsey S, Giles E, et al. Practice patterns of radiation therapy technology in Australia: results of a national audit. *J Med Radiat Sci*. 2015;62:253–260.
- Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med Phys*. 2003;30:2455–2464.
- Childress NL, Rosen II. The design and testing of novel clinical parameters for dose comparison. *Int J Radiat Oncol Biol Physics*. 2003;56:1464–1479.
- Jiang SB, Sharp GC, Neicu T, Berbeco RI, Flampouri S, Bortfeld T. On dose distribution comparison. *Phys Med Biol*. 2006;51:759.
- Thomas SJ, Cowley IR. A comparison of four indices for combining distance and dose differences. *Int J Radiat Oncol Biol Physics*. 2012;82:e717–e723.
- Harms WB, Low DA, Wong JW, Purdy JA. A software tool for the quantitative evaluation of 3D dose calculation algorithms. *Med Phys*. 1998;25:1830–1836.
- Stojadinovic S, Ouyang L, Gu X, Pompoš A, Bao Q, Solberg TD. Breaking bad IMRT QA practice. *J Appl Clin Med Phys*. 2015;16:154–165.
- Sumida I, Yamaguchi H, Kizaki H, et al. Novel radiobiological gamma index for evaluation of 3-dimensional predicted dose distribution. *Int J Radiat Oncol Biol Physics*. 2015;92:779–786.
- Depuydt T, Van Esch A, Huyskens DP. A quantitative evaluation of IMRT dose distributions: refinement and clinical assessment of the gamma evaluation. *Radiother Oncol*. 2002;62:309–319.
- Nelms BE, Zhen H, Tomé WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys*. 2011;38:1037–1044.
- Zhen H, Nelms BE, Tomé WA. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Med Phys*. 2011;38:5477–5489.
- Kruse JJ. On the insensitivity of single field planar dosimetry to IMRT inaccuracies. *Med Phys*. 2010;37:2516–2524.
- Carrasco P, Jornet N, Latorre A, Eudaldo T, Ruiz A, Ribas M. 3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification. *Med Phys*. 2012;39:5040–5049.
- Heilemann G, Poppe B, Laub W. On the sensitivity of common gamma-index evaluation methods to MLC misalignments in Rapidarc quality assurance. *Med Phys*. 2013;40:031702.
- Nelms BE, Chan MF, Jarry G, et al. Evaluating IMRT and VMAT dose accuracy: practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. *Med Phys*. 2013;40:111722.
- Kry SF, Molineu A, Kerns JR, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. *Int J Radiat Oncol Biol Phys*. 2014;90:1195–1201.
- Hauri P, Verlaan S, Graydon S, Ahnen L, Klöck S, Lang S. Clinical evaluation of an anatomy-based patient specific quality assurance system. *J Appl Clin Med Phys*. 2014;15:181–190.
- Kadoya N, Saito M, Ogasawara M, et al. Evaluation of patient DVH-based QA metrics for prostate VMAT: correlation between accuracy of estimated 3D patient dose and magnitude of MLC misalignment. *J Appl Clin Med Phys*. 2015;16:179–189.
- Opp D, Nelms BE, Zhang G, Stevens C, Feygelman V. Validation of measurement-guided 3D VMAT dose reconstruction on a heterogeneous anthropomorphic phantom. *J Appl Clin Med Phys*. 2013;14:70–84.
- Feygelman V, Stambaugh C, Opp D, Zhang G, Moros EG, Nelms BE. Cross-validation of two commercial methods for volumetric high-resolution dose reconstruction on a phantom for non-coplanar VMAT beams. *Radiother Oncol*. 2014;110:558–561.
- Stambaugh C, Ezzell G. A clinically relevant IMRT QA workflow: Design and validation. *Med Phys*. 2018;45:1391–1399.
- Crowe SB, Sutherland B, Wilks R, et al. Relationships between gamma criteria and action levels: Results of a multicenter audit of gamma agreement index results. *Med Phys*. 2016;43:1501–1506.
- Hussein M, Clementel E, Eaton DJ, et al. A virtual dosimetry audit—Towards transferability of gamma index analysis between clinical trial QA groups. *Radiother Oncol*. 2017;125:398–404.
- Ezzell GA, Burmeister JW, Dogan N, et al. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. *Med Phys*. 2009;36:5359–5373.
- Miften M, Olch A, Mihailidis D, et al. Tolerance Limits and Methodologies for IMRT Measurement-Based Verification QA: Recommendations of AAPM Task Group No 218. *Med Phys*; 2018;45:e53–e83.
- Li H, Dong L, Zhang L, Yang JN, Gillin MT, Zhu XR. Toward a better understanding of the gamma index: investigation of parameters with a surface-based distance method. *Med Phys*. 2011;38:6730–6741.
- Barber J, Vial P, White P, et al. A survey of modulated radiotherapy use in Australia & New Zealand in 2015. *Australas Phys Eng Sci Med*. 2017;40:811–822.

32. Abolaban F, Zaman S, Cashmore J, Nisbet A, Clark C. Changes in patterns of intensity-modulated radiotherapy verification and quality assurance in the UK. *Clin Oncol (R Coll Radiol)*. 2016;28:e28–e34.
33. Middlebrook ND, Sutherland B, Kairn T. Optimization of the dosimetric leaf gap for use in planning VMAT treatments of spine SABR cases. *J Appl Clin Med Phys*. 2017;18:133–139.
34. Ji Kim, Park S-Y, Kim HJ, Kim JH, Ye S-J, Park JM. The sensitivity of gamma-index method to the positioning errors of high-definition MLC in patient-specific VMAT QA for SBRT. *Radiation Oncology*. 2014;9:167.
35. Fogliata A, Nicolini G, Clivio A, Vanetti E, Cozzi L. Accuracy of Acuros XB and AAA dose calculation for small fields with reference to RapidArc® stereotactic treatments. *Med Phys*. 2011;38:6228–6237.
36. Zolfaghari S, Crowe SB, Papworth D, West M, Kairn T. Testing the limits of stereotactic radiosurgery for multiple brain metastases. *Australas Phys Eng Sci Med*. 2016;39(4):1097–1098.
37. Ong CL, Dahele M, Slotman BJ, Verbakel WF. Dosimetric impact of the interplay effect during stereotactic lung radiation therapy delivery using flattening filter-free beams and volumetric modulated arc therapy. *Int J Radiat Oncol Biol Phys*. 2013;86:743–748.
38. Hardcastle N, Oborn BM, Haworth A. On the use of a convolution-superposition algorithm for plan checking in lung stereotactic body radiation therapy. *J Appl Clin Med Phys*. 2016;17:1–12.
39. Wen N, Lu S, Kim J, et al. Precise film dosimetry for stereotactic radiosurgery and stereotactic body radiotherapy quality assurance using Gafchromic™ EBT3 films. *Radiat Oncol*. 2016;11:132.
40. Gardner SJ, Lu S, Liu C, Wen N, Chetty IJ. Tuning of AcurosXB source size setting for small intracranial targets. *J Appl Clin Med Phys*. 2017;18:170–181.
41. Van Dyk J, Barnett R, Cygler J, Shragge P. Commissioning and quality assurance of treatment planning computers. *Int J Radiat Oncol Biol Phys*. 1993;26:261–273.
42. Pawlicki T, Whitaker M, Boyer AL. Statistical process control for radiotherapy quality assurance. *Med Phys*. 2005;32:2777–2786.
43. Sanghanthum T, Suriyapee S, Srisatit S, Pawlicki T. Statistical process control analysis for patient-specific IMRT and VMAT QA. *J Radiat Res*. 2012;54:546–552.
44. Binny D, Lancaster CM, Trapp JV, Crowe SB. Statistical process control and verifying positional accuracy of a cobra motion couch using step-wedge quality assurance tool. *J Appl Clin Med Phys*. 2017;18:70–79.
45. Binny D, Mezzenga E, Lancaster CM, Trapp JV, Kairn T, Crowe SB. Investigating output and energy variations and their relationship to delivery QA results using Statistical Process Control for helical tomotherapy. *Phys Med*. 2017;38:105–110.

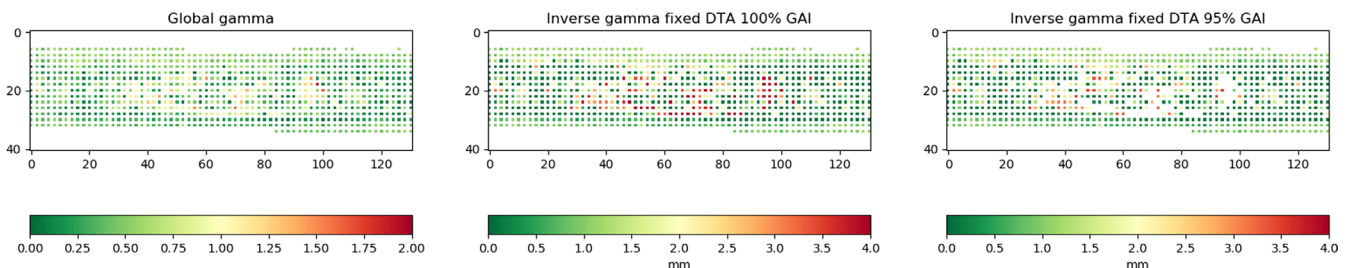
## APPENDIX

The body of this manuscript describes the IG method, using the example of keeping the  $\Delta D$  fixed and identifying the minimum  $\Delta d$

required to achieve a specified GAI for each PSQA comparison (abbreviated to  $IG_{\Delta D}$ ). Two additional forms of the IG algorithm have



**FIG. A1.** Distribution of  $IG_{\Delta D/\Delta d = 1.0}$ ,  $IG_{\Delta d = 1 \text{ mm}}$  and  $IG_{\Delta D = 3\%}$  at GAI 100% and 95% of 53 individual arcs, including mean and standard deviation. From left to right are a)  $IG_{\Delta D/\Delta d = 1.0}$  at 100%, b)  $IG_{\Delta d = 1 \text{ mm}}$  at 100%, c)  $IG_{\Delta D = 3\%}$  at 100%, d)  $IG_{\Delta D/\Delta d = 1.0}$  at 95%, e)  $IG_{\Delta d = 1 \text{ mm}}$  at 95% and f)  $IG_{\Delta D = 3\%}$  at 95% respectively.



**FIG. A2.** Global gamma distribution map (left).  $IG_{\Delta d = 1 \text{ mm}}$  (GAI = 100%)  $\Delta d$  distribution map (middle).  $IG_{\Delta d = 1 \text{ mm}}$  (GAI = 95%)  $\Delta d$  distribution map (right).

**TABLE A1** Results of  $IG_{\Delta D/\Delta d} = 1.0$ ,  $IG_{\Delta d} = 1\text{mm}$  and  $IG_{\Delta D} = 3\%$  calculated for 53 VMAT arcs using 100% GAI and 95% GAI respectively.

Patient	arc	Plan	100% pass rate			95% pass rate		
			%, mm	1mm DTA	3%, mm	%, mm	1mm DTA	3%, mm
1	1	Abdomen	4.3	11.0%	8.4	1.9	3.5%	1.3
	2		3.9	10.5%	16.6	2	3.8%	1.6
2	3	Lung	4	4.4%	8.1	1.5	2.3%	0.4
	4		3.1	7.9%	3.2	1.7	2.6%	0.6
3	5	Brain	2.5	3.4%	1.3	1.4	1.7%	0.2
	6		2.1	4.4%	1.6	1.4	2.0%	0.7
	7		3	3.5%	2.7	1.4	1.7%	0.1
4	8	Boost RT BM	3	7.0%	3	1.4	2.4%	0.9
	9		3.3	13.9%	3.5	1.7	4.1%	1.2
5	10	RTLUNG	2.4	5.7%	1.7	1.5	1.9%	0.1
	11		3.2	5.9%	8.6	1.8	2.5%	0.7
6	12	LT BM BOOST	2.7	4.8%	1.9	1.3	1.8%	0.7
	13		3	8.8%	3	1.4	2.5%	0.8
7	14	Oesophagus	2.6	4.4%	1.9	1.6	2.1%	0.4
	15		3.2	7.7%	4	1.7	2.8%	0.9
8	16	BM Boost	3.3	7.4%	4.1	1.4	1.6%	0.8
	17		3	12.4%	3	1.6	3.2%	1.1
9	18	GOJ	3.5	6.1%	4.5	1.5	1.9%	0.1
	19		3.5	7.2%	3.8	1.7	2.6%	0.3
10	20	Pelvis	3.6	4.3%	6.8	1.8	2.3%	0.1
	21		4	6.0%	6.6	2.3	3.3%	1.4
11	22	RTHN	3.5	6.8%	7.9	1.6	2.4%	0.7
	23		4.7	9.8%	9.9	2	3.5%	1.3
12	24	Pelvis	3.4	8.0%	5.2	1.9	2.8%	0.8
	25		4.8	9.4%	4.9	2	2.9%	0.8
13	26	Pelvis boost	2.9	6.2%	2.6	1.6	2.6%	0.9
	27		4.2	9.3%	5.4	2.1	3.6%	0.4
14	28	Oesophagus	2	4.8%	1.3	1.3	1.7%	1
	29		2.7	11.1%	2.6	1.7	3.0%	1
15	30	LT CHEST	2.3	4.1%	1.3	1.3	1.5%	0.1
	31		2.6	6.2%	2.2	1.6	2.1%	0.6
16	32	Brain	2.3	3.0%	1	1.4	1.8%	0.1
	33		2.4	7.8%	1.8	1.6	2.2%	0.1
17	34	RTHN	3	5.7%	2.8	1.7	2.3%	0.5
	35		4.1	12.8%	10.1	1.6	2.4%	0.8
	36		3.2	4.3%	3.5	1.3	1.7%	0.1
18	37	Rforehead	5.1	7.9%	6	1.7	2.5%	0.5
	38		3.6	9.7%	3.9	2	3.2%	1.2
	39		5	6.6%	5.9	1.7	2.9%	1
19	40	RT BM boost	4.5	12.7%	11.5	1.6	2.4%	1

(Continues)

**TABLE A1** (Continued)

Patient	arc	Plan	100% pass rate			95% pass rate		
			%, mm	1mm DTA	3%, mm	%, mm	1mm DTA	3%, mm
	41		4.4	13.3%	9.1	1.8	4.4%	1.4
20	42	LT BM BOOST	3.1	10.7%	3.2	1.8	3.8%	1.4
	43		2.7	6.6%	2.4	1.5	1.8%	0.1
21	44	LT BM BOOST	3.5	4.8%	5.6	1.5	2.5%	0.7
	45		2.8	10.3%	2.7	1.7	4.1%	1.2
22	46	RT BRAIN	2.9	6.9%	2.7	1.5	1.8%	0.1
	47		3	8.4%	3	1.6	3.1%	1.1
23	48	LT BM BOOST	3.6	11.0%	8.2	1.7	3.4%	1.1
	49		3	12.0%	3	1.9	4.1%	1.4
24	50	Gynae	2.7	3.9%	2	1.2	1.4%	0.1
	51		3.7	7.2%	4.3	1.5	2.1%	0.1
25	52	ABDO + Pelvis	3.4	5.1%	3.8	1.3	1.6%	0.1
	53		3	3.9%	3	1.4	1.6%	0.1
Mean			3.3	7.5%	4.5	1.6	2.6%	0.7
SD			0.7	2.9%	3.1	0.2	0.8%	0.5

been investigated: inverse gamma with fixed  $\Delta d$  ( $IG_{\Delta d}$ ) and inverse gamma with fixed dose-distance-ratio ( $IG_{\Delta D/\Delta d}$ ).

As an obvious analogue to  $IG_{\Delta D}$ ,  $IG_{\Delta d}$  keeps the  $\Delta d$  fixed and calculates the minimum  $\Delta D$  required to achieve a specified GAI for each PSQA comparison. The algorithm performs iterative global gamma calculations of  $\Delta D$  from 0% with 0.1% increment until the minimum  $\Delta D$  were found to achieve a GAI that is no less than the predefined threshold GAI value. To provide an indication of the results achievable using  $IG_{\Delta d}$ , the 53 VMAT PSQA results used in this work were re-evaluated using a fixed  $\Delta d$  of 1 mm, denoted as  $IG_{\Delta d = 1\text{mm}}$  (see Fig. A1). A comparable example of the resulting  $\Delta D$  distribution when applying the  $IG_{\Delta d = 1\text{mm}}$  method is illustrated in Fig. A2. As we can see that Fig. 1 is a  $\Delta d$  distribution map, in which the points with higher values are indication of regions where larger distance-to-agreement disagreement occurs. Likewise, Fig. A2 is a  $\Delta D$  distribution map where the points with higher values indicate regions of larger dose difference disagreement. In comparison, neither  $\Delta d$  nor  $\Delta D$  information is apparent on the gamma distribution map.

The  $IG_{\Delta D/\Delta d}$  concept is based on the gamma index definition (global normalization). The difference is that instead of calculating GAI based on fixed  $\Delta D$  and  $\Delta d$  in the gamma method,  $IG_{\Delta D/\Delta d}$  calculates the minimum  $\Delta D$  and  $\Delta d$  based on fixed GAI and dose-difference ratio ( $\Delta D/\Delta d$ ). The dose-difference ratio used in this study was unity (e.g., 3%, 3 mm, 2%/2 mm, 1%/1 mm, etc), denoted as  $IG_{\Delta D/\Delta d} = 1.0$ . While keeping a constant ratio of 1, the algorithm performs iterative global gamma calculations of GAI using  $\Delta D$  and  $\Delta d$  ranging

from 1%/ 1 mm with 0.1%/mm increment until the resulting GAI equals to or above the predefined threshold GAI value (see Fig. A1).

Figure A1 shows how the IG results can vary, depending on the selection of the GAI and on which specific IG method (fixed  $\Delta D$ , fixed  $\Delta d$  or fixed  $\Delta D/\Delta d$ ) is used for the comparison. Clearly, the  $\Delta D$  and  $\Delta d$  values required to achieve a "passing" result are higher when the GAI action threshold is 100%, compared to when the GAI action threshold is 95%. Similarly, the  $\Delta D$  values required to exceed the GAI action threshold are higher when the  $\Delta d$  is set to 1 mm, than when the  $\Delta d$  is allowed to vary at a constant ratio with the  $\Delta D$ .

Figure A1 also provides an indication of how the indices developed in this study can be used to evaluate locally used or proposed gamma criteria and GAI, as part of a statistical process control (SPC)

process. In Fig. A1, the mean  $\Delta D$  and  $\Delta d$  criteria that result from an  $IG_{\Delta D/\Delta d = 1.0}$  evaluation of the VMAT PSQA results, with a GAI of 95%, is  $1.6 \pm 0.2\%/1.6 \pm 0.2$  mm, which is just within 2%/2 mm. This suggests that if 2%/2 mm was used in a standard gamma calculation, the resulting mean GAI would be slightly higher than 95% (in fact, the results of this study show that it is  $97.9 \pm 1.5\%$ ). This result suggests that using the gamma criteria of 2%/2 mm and GAI of 95% is suitable for the cohort of plans. This finding is consistent with the results estimated using the SPC method<sup>29</sup> for the same cohort of patients. Similarly, Fig. A1 suggests that if a center decided to use a GAI of 100%, then the suitable gamma criteria they should be using is 4%/4 mm, assuming a unity dose-to-distance ratio is preferred.