



Review paper

## Treatment plan verification: A review on the comparison of dose distributions



Stefanos Diamantopoulos<sup>a,\*</sup>, Kalliopi Platoni<sup>a</sup>, Georgios Patatoukas<sup>a</sup>, Pantelis Karaiskos<sup>b</sup>, Vassilis Kouloulis<sup>a</sup>, Efstathios Efstathopoulos<sup>a</sup>

<sup>a</sup>2nd Department of Radiology, University General Hospital "Attikon", National and Kapodistrian, University of Athens, Greece

<sup>b</sup>Medical Physics Laboratory, Medical School, National and Kapodistrian University of Athens, Greece

### ABSTRACT

**Purpose:** The aim of this review article is to provide a useful reference for dose comparison techniques within the frame of treatment plan verification. Each technique is presented with a general description given along with advantages and disadvantage and the rationale for its development.

**Methods:** The review was conducted in PubMed from 1993 to 2019 including articles referring to the methodology of dose comparison for treatment plan verification.

**Results:** The search identified thirty-one dose comparison methods that were categorized according to the number of physical parameters that take into account for dose comparison.

**Conclusions:** Among the available methods for the comparison of two dose distributions, the  $\gamma$ -analysis (gamma analysis) has been widely adopted as the gold standard in verification procedures. However, due to various intrinsic limitations of gamma index, the development of a better metric taking into account both statistical and in clinical parameters is required.

### 1. Introduction

Since the introduction of Intensity Modulated Radiation Therapy (IMRT), the need for extensive and patient-specific quality tests prior to treatment has become an indispensable part of daily routine. This requirement is highlighted in the recommendations by the American Association of Physicists in Medicine (AAPM) [1] and European Society of Radiotherapy and Oncology [2]. For the treatment delivery of IMRT and VMAT, various compartments of the linear accelerator are co-operating in coordinated movements, a complex process that can possibly introduce unexpected errors [3]. The dosimetric outcome of such a treatment is practically impossible to be tested with conventional methods or oversimplified calculations.

Therefore, various solutions and dosimetric systems were introduced in clinical routine, incorporating detectors placed in two (2-D) and three (3-D) dimensional layouts, "in silico" machine log files quality control etc [4–6]. These solutions are aiming at adequately capturing the dosimetric outcome of these complex movements in multiple dimensions.

Whatever the sophistication of the dosimetric equipment, in the end, every verification procedure comes down to a dose distribution comparison: Machine-delivered (reference) versus calculated by a

Treatment Planning System-TPS dose distribution (evaluation). This comparison is, in general, unstraightforward and each result should be carefully reviewed. Every analysis-comparison of the distributions tries to answer two fundamental inquiries: how large (quantitative aspect) and clinically significant (qualitative aspect) are these deviations.

The first question is the easiest to be answered, while the second one is undoubtedly the trickiest. A variety of mathematical and statistical methods have been developed to provide practical answers to both questions in a clinical environment. Each one is aiming to carefully compare the reference and the delivered dose distribution in an effective way, which would elevate the level of confidence in treatment execution.

Nevertheless, dose comparison remains not a trivial task to be evaluated, as the means of comparison are primarily statistical, while the results should be clinically relevant as well. Furthermore, statistical tests do not provide information regarding the importance or the source of disagreement. These constitute the principal reasons which led Childress et al. [7] to describe the ideal dose comparison index. According to their analysis the ideal index should have the following seven characteristics: a) biological significance, b) physical meaning, c) fast in terms of computational time, d) independence of image/volume size and represented dose range, e) direct comparability between

\* Corresponding author.

E-mail address: [diamstef@med.uoa.gr](mailto:diamstef@med.uoa.gr) (S. Diamantopoulos).

<https://doi.org/10.1016/j.ejmp.2019.10.029>

Received 17 April 2019; Received in revised form 10 October 2019; Accepted 15 October 2019

Available online 06 November 2019

1120-1797/ © 2019 Associazione Italiana di Fisica Medica. Published by Elsevier Ltd. All rights reserved.

institutions, and f) consistency over time.

The combination of the aforementioned characteristics is formidable to meet in a single index. As this ideal index has not however been developed, it is commonplace practice to evaluate distributions with more than one method or acceptance limits depending on the case.

A dose distribution within a certain volume should be evaluated both in dose and space domain. Treatment plan evaluation methods that consider only the parameter of dose include: dose difference ( $\Delta D$ ), isodose contours superimposition and for further analysis of spotted mismatches, dose profile superimposition, dose difference image and dose difference histograms (based on the dose difference image) that indicate the presence of systematic or random errors [8]. Treatment plan verification using dose volume histograms (DVHs) enables straight comparison of the measured data from the verification process with the calculated ones from the TPS. However, DVHs do not provide spatial information while integral histograms mask errors in small volumes [9].

Information concerning the space domain is given by distance-to-agreement (DTA) [10]. DTA is defined as the distance between a dose point in the reference distribution and the most adjacent point in the evaluated dose distribution that appears to have the same dose value or at least a dose value within a tolerance limit. In low dose gradient areas, dose comparison methods performed better than DTA whereas DTA provides more valid results in high dose gradient regions. The combination of the dose difference method and DTA (composite analysis) utilizing respectively two independent acceptance criteria, was first introduced by Harms et al [11].

The aim of this paper is to provide an overview of evaluation techniques and approaches that have been developed for plan verification. All the methods described are presented with their formalism wherever possible along with their advantages and disadvantages.

## 2. Materials and methods

An extensive review of the academic literature was conducted in PubMed including published papers dated from 1993 up to 2019. The keywords used were (treatment plan verification) and (techniques) and (IMRT) and (Radiation Therapy) and (Radiation Oncology). From the identified articles, only those referring to the methodology of dose comparison were included in the current review. Articles concerning specific plan verification tools or clinical cases were excluded. Finally, the number of citations of these papers was recorded, as a measure of the support of the Medical Physics community towards each method, using Scopus.

## 3. Results

The search identified fifty-three (53) relevant papers and reports that revealed thirty-one (31) dose comparison methods. The current study presents all the methods that analyze verification plans both in dose and space domain and the upgrades of the most popular method, gamma index.

### 3.1. Analysis in dose and space domain

#### 3.1.1. Normalized agreement test (NAT)

Normalized agreement test (NAT) [7] was developed as a supplementary tool to dose comparison indices as it was designed to include clinical correlation to the statistical tests. A NAT matrix consists of pixels that represent the percent deviation from the criteria set for plan verification.

In this approach, the authors distinct and evaluate PTV area differently from the rest of the image based on its biologic significance. In a normalized dataset, areas with dose lower than a fixed percentage of 75% of the maximum computed dose value are considered healthy tissue, while areas with dose values above 75% are PTV.

NAT value is calculated over the NAT matrix in a two-part equation:

$$i. \text{NATvalue} = 0 \quad (1)$$

when:

- a) a pixel's value within dose ( $\Delta D_m$ ) or space ( $\Delta d_m$ ) acceptance criteria
- b) a pixel's value falls outside acceptance criteria in healthy tissue area but has a measured value lower than the computed one, assuming that this disagreement will actually benefit the patient.

ii. In any other pixel NAT value is given by the following equation:

$$\text{NATvalue} = D_{\text{scale}} \times (\delta - 1) \quad (2)$$

where:

$\delta$  is the lesser value of the  $|\Delta D / \Delta D_m|$  or space  $|\Delta d / \Delta d_m|$  term and:

$$D_{\text{scale}} = \frac{\text{computed or measured dose (whichever is greater at the pixel)}}{\text{maximum computed dose}} \quad (3)$$

Employing the average of the NAT values and  $D_{\text{scale}}$  over the whole matrix the NAT index can be calculated as follows:

$$\text{NATindex} = \frac{\text{Average NATvalue}}{\text{Average of the } D_{\text{scale}} \text{ matrix}} \times 100 \quad (4)$$

This index represents the average deviation from acceptance dose and space criteria ignoring areas of acceptance, and is expressed as a sole value analogously to the dose uniformity concept.

#### 3.1.2. Gradient compensation method

Moran et al. [12] developed this method/tool to evaluate local dosimetric differences based on the dose gradient at each point in the field. Therefore, the local dose gradient is computed for each point in the dose distribution. To account for geometric uncertainties (dose grid dimensions or shift of the distributions), the user specifies a distance parameter (e.g. 1 mm) which in fact represents the geometric tolerance that should be accepted in each dose point. Each dose gradient is multiplied by this parameter to produce a dose value that is relevant to the uncertainty of this point. In this way, dose differences that could be attributed to geometric uncertainties are removed from the analysis, including only discrepancies originating from other sources. The authors state this method could assist dose comparison as it reveals the magnitude and clinical relevance of the differences; nevertheless, it should be used in conjunction with other dosimetric comparison tools.

#### 3.1.3. Fractal analysis

This method was proposed by Wu et al. [13] for 2D verification procedures including films (or portal dosimetry) and is supposed to provide more valid results than visual inspection which is susceptible to errors. This interesting idea incorporates the fractal theory to match calculated and measured data. According to their theory, fractal dimension, which represents a ratio of the changes of a fractal pattern with the scale which is measured, is a statistical index of complexity and a unique fingerprint for each isodose contour. The compared isodoses are considered identical when their fractal dimensions are within 1% which is translated to a 2% dose difference. However, this theoretical concept seems that has not been applied yet in a clinical environment.

#### 3.1.4. Gamma index

Dose distributions as described earlier need to be compared both in dose and space domain. Nevertheless, sensitivities of dose and space comparison seem to complement each other in low and high dose gradient region, respectively. Reviewing dose distributions in a two-step procedure with different acceptance criteria could be a laborious task and more importantly, results could be misleading. The need for a unified measure to compare dose distribution was profound. The problem that prevented this unification was that dose deviation ( $\Delta D$ ) and distance to agreement (DTA) metrics are expressed in different units

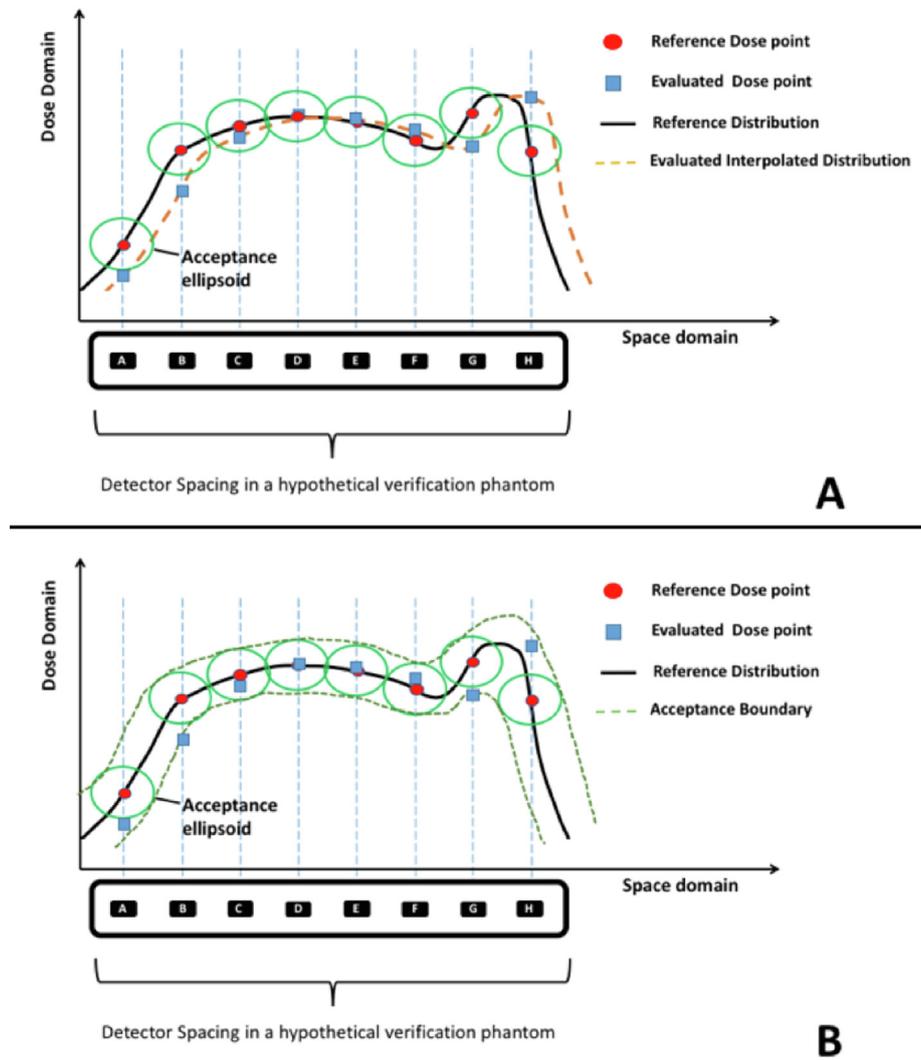


Fig. 1. A. Coarsely spaced detectors (in larger spaces than the radius of the acceptance ellipsoid) in a measurement device can have a severe impact on the final gamma calculation. The measured dose in detectors A, B, G, H could erroneously produce failed gamma values, a problem that can be eliminated with dose point interpolation. Fig. 1B. depicts the elegant alternative to time-consuming interpolation, offered by methods such as  $\chi$ -evaluation and delta envelope. The creation of a constant acceptance barrier which follows the shape of the original distribution could classify correctly all dose points.

(Gy and mm respectively).

Low et al. in 1998 [14,15] introduced gamma index and offered an elegant solution to this problem. Dose difference and distance to agreement metrics were merged into a unitless quantity, by dividing each metric by its corresponding acceptance limit. For example, 3% for  $\Delta D$  and 3 mm for DTA.

$$\gamma(\vec{r}_{ref}, \vec{r}_m) = \min \left\{ \sqrt{\frac{|\vec{r}_{ref} - \vec{r}_m|^2}{DTA^2} + \frac{|D(\vec{r}_{ref}) - D(\vec{r}_m)|^2}{\Delta D^2}} \right\} \quad (5)$$

where

$|\vec{r}_{ref} - \vec{r}_m|$  is the distance and  $|D(\vec{r}_{ref}) - D(\vec{r}_m)|$  is the absolute dose difference between the reference and evaluated points.

In this way, an acceptance ellipse is created around each point of the reference dose distribution, which should encompass the measured dose distribution for the point to pass gamma test and is valid both in shallow or steep dose gradient regions (Fig. 1A). Mathematically, this is expressed by gamma being less than or equal to unity. According to this unified criterion, points are categorized by passing and failing. In order for the whole distribution to pass the test, a certain percentage of successfully tested points should be reached as advocated by AAPMs TG-119 Report [1] and later publications. Usually, this percentage is set to 90 or 95%.

Gamma can be calculated using local or global normalization. To perform the first one, local dose differences are being taken into account in gamma formula. To perform global normalization, the local dose difference is divided by a normalization dose value that could be defined anywhere in the dose distribution. Both methods have advantages and limitations. Local gamma highlights differences in high dose gradient areas and low dose regions, while global gamma hides these differences, highlighting instead errors in high dose areas [16].

### 3.1.5. Gamma histograms

The concept of gamma index can be easily extended to provide additional tools for verification evaluation. Therefore, analogously to dose volume histograms (DVHs), gamma histograms (GH) could be produced indicating the percentage of the voxels that are covered by a certain gamma value [17]. There are two types of such histograms a) frequency gamma volumes histograms (fGVHs) and b) cumulative gamma volume histograms (cGVHs). If the dosimetric data is obtained by using a two-dimensional detector or a film, areas of interest can be used instead of volumes, which would produce frequency gamma area histograms (fGAHs) and cumulative gamma area histograms (cGAHs) respectively.

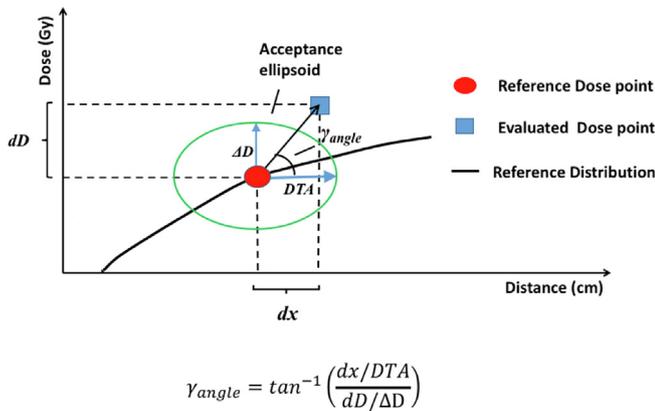


Fig. 2. An example of gamma angle calculation.

### 3.1.6. Gamma angle

Gamma angle method [18,19] uses gamma index's vectorial properties to point out if a given mismatch is attributed to dose discrepancy or spatial distance (Fig. 2). If the discrepancy is only due to dose difference, gamma index vector would be parallel to dose axis, while if it is attributed to spatial difference, gamma vector would be parallel to DTA axis. Considering as the zero angle the first scenario, gamma vector should always be between 0 or  $\pi/2$ . Gamma angle is a helpful tool in dose distribution analysis, with angles closer to zero indicating the dominance of the dosimetric influence to the result and angles closer to  $\pi/2$  spatial mismatches.

$$\gamma_{angle} = \tan^{-1} \left( \frac{\frac{|\vec{r}_{ref} - \vec{r}_m|}{DTA}}{\frac{|D(\vec{r}_{ref}) - D(\vec{r}_m)|}{\Delta D}} \right) \quad (6)$$

For a gamma vector position, the percentage of each contribution is given from the following equations:

$$\Delta D \text{influence}(\%) = \cos(\gamma_{angle}) * 100 \quad (7)$$

$$DTA \text{influence}(\%) = [1 - \cos(\gamma_{angle})] * 100 \quad (8)$$

## 3.2. Gamma index upgrades - in search of a "better" gamma index

The elegant simplicity and universality of this unitless approach established gamma index as the gold standard for treatment verification. Nevertheless, the methodology of gamma index has some pitfalls and drawbacks if it is to be applied in clinical routine, where non-continuous data exists and a clinical decision follows each measurement. Through the years a group of variants of the primary technique has been developed to provide results faster, highlight clinical important mismatches, provide a reasonable strategy for setting acceptance criteria, and disengage results from the dose grid dependence. The following section summarizes these efforts.

### 3.2.1. Dose grid dependence limiting methods

**3.2.1.1. "Box" method.** Introduced by Jiang et al. [20], the box method combines advantages from dose difference test, composite analysis, and gamma analysis. The acceptance area forms an error "box" in dose and space domains where the side lengths corresponds to dose and space acceptance limits.

The main idea is based on the conversion of the spatial constitute of gamma index to a dose equivalent measure. To achieve that, the "equivalent dose tolerance" concept was invented that corresponds to gamma's spatial tolerance. Under this scope, a new unified tolerance limit is determined, under the term of "maximum allowed dose difference" (MADD). Using MADD, the spatial deviation can be expressed in a single-valued dose limit. On the other hand, dose acceptance criterion

can vary in magnitude depending on the point under examination. This "adaptive" index is produced when dose difference is being scaled by the ratio of MADD to dose criterion (normalized dose difference (NDD) concept). Therefore, MADD value (which is insensitive to dose grid resolution) is stricter in low than in high dose gradient areas, providing a more reasonable dose comparison method.

**3.2.1.2. Chi-evaluation acceptance Interval-Acceptance test tube.** Bakai et al. [21] tried to limit the dependence of grid resolution by modifying the original gamma test acceptance region. In their approach, a new evaluation factor chi ( $\chi$ ) was developed by using a dose gradient dependent tube that engulfs continuously the reference distribution profiles. The radius of the tube is determined by dose and spatial tolerance limits while the curvature of the tube depends on dose local gradient. Evaluation points with values smaller than  $|\chi|$  pass the test similarly to gamma analysis. In this way, the detection of false negative results can be restricted. In addition, the "creation" of this tube around the reference points renders meaningless the need for continuous interpolation. Therefore, calculations with this method can become 120 times faster than the original gamma index method, provided that both reference and evaluated distributions have the same grid resolution. Under any other circumstance, an initial point interpolation has to be performed in order to equalize spatial density. The latter is the most significant drawback of this technique, compared to gamma which can be applied in any case. Furthermore, using only local dose gradient makes this method inaccurate in the case of non-zero second derivatives. A similar approach was applied by Bak et al. [22,23] who proposed the modified dose difference (MDdiff) method. They defined a dimensionless factor  $\beta$  which depends on the dose gradient and the  $\Delta D$  to DTA criteria ratio. Furthermore, they proposed as a critical value to accept a point MDdiff =  $(1/2)\Delta D$ .

**3.2.1.3. Delta envelope.** A similar method to chi-evaluation was developed by Blanpain and Mercier [24] termed "delta-envelope". Being consistent with the initial gamma index idea, authors tried to expand chi-evaluation to a more accurate technique even where the dose gradient changes (Fig. 1.B) in no interpolated data. This method requires the pre-calculation of a tolerance envelope using the minimal and maximal doses allowed per reference dose point. All reference points are taken into account, forming a continuous envelope that delimits the tolerance area which is independent of local gradient. Because of the continuity of the envelope, interpolation is unnecessary. The shape of the envelope could become narrower on its external border in high dose gradient regions, where the density of reference points becomes sparser. Even in this case, interpolation to correct this drawback is of lesser importance. Furthermore, the authors suggest three similar but different in interpretation indices,  $\delta_a$ ,  $\delta_b$ ,  $\delta_c$ , which are analogous to  $\gamma$ -index, in order to provide useful information about the large deviations outside the envelope. The absence of interpolation offers reduced computational time even for the case of  $\delta_c$  index where multiple envelopes have to be computed.

### 3.2.2. Calculation time reduction methods

**3.2.2.1. Fast algorithm for gamma evaluation.** A real gamma evaluation can take from minutes to hours to be calculated, especially for complex treatment plans. To speed up gamma calculation process, Wendling et al. [3] designed an algorithm for fast gamma analysis. The idea is to reduce calculation time by limiting search area for points that could pass the gamma criteria in a sphere with a chosen radius. However, the actual gamma formula does not change. The maximum search distance (radius of the sphere) should be much larger than the DTA criterion. A presorted table of distances is calculated only once from the center towards the periphery of the sphere with increasing steps in the space domain. For the interpolation of values of the evaluated dose distribution, the surrounding points of a certain point in dose matrix are weighted using interpolation factors and then summed. For even

further time saving, the calculation can be stopped for points that are improbable to result in a minimal gamma (points in a distance that itself alone raises gamma above unity).

**3.2.2.2. Geometric interpretation of the gamma index.** In 2008 Ju et al. [25] proposed “geometric interpretation of the gamma dose distribution comparison technique” as an interpolation-free calculation variation of the original gamma analysis. This technique provided an accurate and fast way of computing gamma which overcame the aforementioned limitations. The authors pointed out the interpolation problem of gamma which is caused by the high dependence on dose grid size or spatial resolution. For accurate results, one of the distributions should have dose points spaced in significantly smaller distances than DTA criterion, which is not always the case especially in high dose gradient regions. If not so, the closest point to the reference dose distribution may not be the actual closest one as it would have been in case of a continuous dose surface, leading to false negative results. To avoid this pitfall, the authors used simplexes – line segments, triangles, and tetrahedral for one, two and three-dimensional dose distributions respectively- to subdivide the evaluated distribution. In this way, the closest distance between the reference distribution and the simplexes of the evaluated distribution can be computed using matrix algebra without the highly time-consuming process of interpolation. This method proved more efficient than the interpolation as it produces equally accurate results even with poorer dose grid resolutions in reduced time. Finally, to further reduce calculation time, authors suggest limiting the search for minimum distance around any given point to an area according to DTA criterion and the maximum gamma value that needs to be recorded:

Area of search = Maximum gamma value to be recorded  $\times$  DTA criterion

Using the conventional interpolation method and the original gamma index, similar accuracy levels could be achieved for at least 16 times finer grid resolution.

**3.2.2.3. GPU method.** Gu et al. [26] employed graphical processing unit (GPU) instead of traditional central processing unit (CPU) calculations to speed up gamma results. GPUs are composed of hundreds of cores that can handle thousands of threads simultaneously in contrast to CPUs that consist of few cores. Therefore, the main idea was to perform minimum distance search for each voxel, which is a highly demanding task, using GPU. Gamma was calculated by using the geometric technique proposed by Ju et al. [25] and the pre-sorting technique mentioned above by Wendling et al. [3]. By this combination, the authors managed a calculation time reduction in the order of 45–70 times compared to CPU calculations. Similar results were found by Persoon et al. [27] with the use of GPU reducing time  $57 \pm 15$  times over conventional CPU calculations [16].

**3.2.2.4. Filter cascade method.** In University Hospital of Leuven, Belgium, an algorithm for rapid and accurate comparison of IMRT dose distributions has been developed, based on gamma index methodology. The so-called filter cascade method [28], evaluates distributions following a three-step sequence. On the first level, the algorithm searches for measured points that seem to fulfill the smaller-than-unity gamma equation, without further investigating which point produces the lowest gamma value, as indicated from gamma index theory, saving computational time. In the second step, only the points that failed the initial check are considered to avoid false negatives especially in steep gradient areas. There, while the evaluated distribution fulfills acceptance criteria, the sampling points could fall outside the gamma ellipsoid. Based on the previous assumption, that only a binary pass/fail result matters instead of the actual gamma value, authors proposed the following to avoid time consuming calculations: if dose difference ( $\Delta D$ ) for at least two data points (points A & B) is of different sign, then the  $\Delta D$  value becomes zero

somewhere (point C) in between. Therefore, gamma at point C is lesser than unity. These points are classified as accepted and the rejected ones are propelled to the subsequent step. The third and final filter examines if the remaining rejected points laying on the outer boundary of the acceptance ellipsoid have a different sign of a point that lies within. In that case, linear interpolation is performed between these two points and if any part of the line intersects the ellipsoid, the point is accepted. This methodology can be extended to even more levels, although the authors state there is no significant possibility for each rejected data point to have been misclassified after this triple filter cascade method.

**3.2.2.5. Fast Euclidean distance transformation.** This method was designed in another attempt to reduce vast calculation times required in the classic gamma distribution. In this approach [29], a pre-calculation of a gamma indices table is performed via Euclidean distance calculations, based on the reference distribution. The evaluated distribution is compared to the reference by searching this table's values. Hence, there is a hundredfold speeding up of calculations in two dimensions and around  $10^4$ - $10^5$  times for three-dimensional dose matrices. Furthermore, the creation of such a gamma table enables the easy calculation of the derivatives of gamma in dose and space domain. This could result in the determination of the dominant source of errors in the evaluated distributions and the guidance to corrective actions.

### 3.2.3. Biological-Anatomy related methods

**3.2.3.1. Anatomy corresponding method.** This method was proposed to link gamma information to patient anatomy [30]. To overcome the absence of regional information in simple gamma analysis, the authors suggested a method that cross-examines gamma results on patient's digital reconstructed radiographs (DRRs). The analysis is performed field by field by applying each gamma map on its corresponding treatment field DRR. Fluences and images are acquired by 2D-detectors like films or portals. This method is restricted to static-field IMRT plans and can be optimally performed in cases that the position of target volumes and organs at risk are directly related to bony structures depicted in DRRs, such as head and neck treatments.

**3.2.3.2. Gamma plus (+) index.** An extension of the original gamma index was developed by Stathakis et al. [31] toward the incorporation of radiobiological concepts in the plan verification process. The new index named gamma plus ( $\gamma +$ ) index and is expressed mathematically as follows:

$$\gamma(\vec{r}_{ref}, \vec{r}_m) = \min \left\{ \sqrt{\frac{|\Delta r_{ij}|^2}{DTA^2} + \frac{|\Delta F(D_{i,j})|^2}{\Delta D^2}} \right\} \quad (9)$$

where  $|\Delta r_{ij}|$  is the spatial shift of the voxel (i,j) and  $|\Delta F(D_{i,j})|$  is the absolute dose difference between the reference and evaluated points of a function  $F(D)$  of physical dose which includes radiobiological information. This information could be the Equivalent Uniform Dose (gEUD), the equivalent to 2 Gy per fraction dose (EQD<sub>2</sub>) or biologically effective uniform dose ( $D$ ).

Prior to any gamma calculation, the physical dose map has to be converted to a corresponding radiobiological matrix, and therefore  $\alpha/\beta$  values should be assigned to all voxels. Authors used 10 Gy as  $\alpha/\beta$  in PTV region, 3 Gy at OARs area, and EQD<sub>2</sub> for dose conversion. This index is more clinically relevant than the original gamma analysis as radiobiological effects of the dose are taken into account.

**3.2.3.3. Radiobiological gamma index (RGI).** Sumida et al [32–34] introduced the term of Radiobiological Gamma Index (RGI) in an effort to transfuse clinical meaning to the physical gamma index (PGI). In their analysis, voxels of the dose distribution matrix that failed gamma analysis are further investigated by calculating their TCP and NTCP values. The RGI can be calculated by the following equations:

$$RGI = PGI \text{ (if } PGI < 1) \quad (10)$$

$$RGI = PGI \times n \text{ (if } PGI > 1) \quad (11)$$

$n$  is a factor with different values for target area and healthy tissue area.

Therefore, for the  $i$ -th voxel:

$$n = \begin{cases} 1 + |TCP_i - ToleratedTCP| \text{ for the target volumes} \\ 1 + |NTCP_i - ToleratedNTCP| \text{ for the healthy tissues} \end{cases} \quad (12)$$

The radiobiological gamma passing percentage is defined as the number of voxels with an RGI lower or equal to unity divided by the total number of voxels in each organ multiplied by 100. This approach enables medical physicists to spot clinically significant divergences that could remain hidden under the classic gamma analysis as pointed out by Nelms et al [35] and Zhen et al [9].

### 3.2.4. Reasonable criteria of acceptance setting methods

**3.2.4.1. Surface-based distance method.** As dose and space difference limits in gamma analysis are rather empirically determined, Li et al [36] proposed a technique in order to obtain proper criteria depending on the complexity of the reference distribution. Considering the reference and evaluation dose distribution as surfaces, they defined the dose gradient factor according to the equation:

$$a = \frac{1}{\text{mean}(|\nabla D(x, y, z)|)} \quad (13)$$

Utilizing this factor and the original gamma equation they reached to the following equation:

$$\Delta D_m = \frac{\Delta d_m}{a} \quad (14)$$

which means that a) for different treatment plans, with different dose gradients, acceptance criteria should also vary according to mean dose gradient and b) dose and space criteria are not independent of each other, but are related through the dose gradient factor  $a$ .

Dose gradient factor should be determined prior to dose and space criteria definition and consequently gamma analysis. In this way, gamma analysis becomes more relevant to actual IMRT and QA system capabilities than simply setting a pair of predetermined criteria.

Finally, the authors note that the detection limit of gamma index (for a certain pair of dose and space acceptance criteria) varies linearly (linear coefficients always  $< 1$ ) relatively to the global shift of the distributions.

**3.2.4.2. Squared gamma method.** Squared Gamma ( $\gamma^2$ ) method [37] is a simple method that could be used as a supplement to the gamma index method without further modifications or exhaustive calculations. Authors have shown that the squared gamma index distribution can have similar properties to the statistical chi-squared distribution with one degree of freedom when gamma acceptance criteria are replaced by the standard deviations of dose and space uncertainties. Therefore, squared gamma method can be used to signify the statistical significance of the measured deviations from gamma analysis and as a consequence to determine in a more statistical manner the plan acceptance tolerance rate.

**3.2.4.3. Gamma index modification with uncertainty features.** This probabilistic modification of gamma index analysis was introduced as a more reliable technique with respect to the primary method as it takes into account the uncertainties related to the hardware and software involved in the process [38]. It enables tolerance levels to be adjusted close to the achievable accuracy and uses simple and physically meaningful parameters to characterize experimental devices, computations, and their uncertainties.

The measured distribution is considered suitable for treatment when every point satisfies the probability test. In case of a passing rate

smaller than 100%, the failures cannot be attributed just to measurement uncertainties, pointing out a problem with the dose delivery procedure.

**3.2.4.4. Unbinned Multivariate test.** This method [39] aims at establishing a global criterion to determine whether a mismatch detected by the gamma method is attributed to the uncertainty associated with the measuring system or it is an actual error. To perform such an analysis a statistical test should be used like  $\chi^2$  test, which is widely used in physics. If this test is to be applied to dose verification, data from reference and evaluation distributions should be represented in histograms. Nevertheless, the number of the bins of the histograms may alter the result of the  $\chi^2$  test. Therefore, the authors proposed an alternative statistical test,  $T^2$ , similar to  $\chi^2$  test, but without the unwanted bin dependency. In particular, this test is described by the following formula:

$$T^2 = \frac{1}{N_d} \sum_{i=0}^{N_d} \left( \frac{\gamma_{i,exp} - \bar{\gamma}_{i,mean}}{\sigma_{i,\gamma}} \right)^2 \quad (15)$$

where,

$N_d$  is the number of points or detectors of the system

$\gamma_{i,exp}$  is the expected gamma value for the  $i_{th}$ -detector

$\bar{\gamma}_{i,mean}$  is the average measured gamma values for the  $i_{th}$ -detector and  $\sigma_{i,\gamma}$  is the standard deviation of  $\gamma_i$

The parameters,  $\bar{\gamma}_{i,mean}$  and  $\sigma_{i,\gamma}$  must be calculated beforehand. The positioning error and the intrinsic uncertainties of the detectors are simulated (considering Gaussian positional uncertainties and independent detectors), compared against the reference dose matrix and producing a simulated gamma matrix. This process is repeated numerous times in order to calculate the required parameters. In this way, the whole dosimetric system is characterized. The statistical significance is set to the cutoff of  $p = 0.01$ . As the authors state, although the test is more powerful than common gamma criteria, its main drawback lays in determining highly localized discrepancies, such as unexpected peaks.

**3.2.4.5. Inverse gamma with fixed  $\Delta D$  ( $IG_{\Delta D}$ ).** Inverse gamma method ( $IG_{\Delta D}$ ) [40] calculates gamma index value by keeping fixed the dose difference criterion ( $\Delta D$ ) and gradually increasing the distance-to-agreement criterion ( $\Delta d$ ) up to the value needed to achieve a specified acceptance rate. This approach is a fast estimation of the magnitude of spatial error which can be compared against a tolerance. In addition, a fixed tolerance ratio of  $\Delta D$  to  $\Delta d$  could be defined, producing different combinations which lead to a specified gamma passing rate.

## 4. Discussion

The crux of this paper is to provide a comprehensive review of dose comparison methods implemented for treatment plan verification. The paper focuses on the methods that the reference dose distribution is compared to the evaluated one. The presentation of dose verification tools, including 2D detectors, 3D phantoms, fluence detectors, machine log files [4–6], etc. is out of the scope of the current study. Low et al [15,41,42] discussed other approaches of dose comparison techniques. A thorough description of gamma index upgrades was also presented by Hussein et al [16]. TG-218 report by AAPM [43] reviewed IMRT dose distribution verification and patient-specific QA, including practical implementation of gamma index and a description of other worth mentioning dose comparison methods. To the best of our knowledge, the current paper is the first to include and summarize all the existing dose comparison techniques for IMRT and VMAT plan verification.

The presented categories of the dose comparison methods concern

**Table 1**  
An overview of all gamma index variants and the limitations of the original gamma technique that are trying to overcome.

Method (Citations)	Biological interpretation	Physical intuition	Dose grid size independence	Calculation time reduction	Reasonable criteria of acceptance	Limitations
Gamma Histograms (67)		+				No spatial information. Supplementary to Gamma Index
Gamma Angle (1 6 2)		+				Supplementary to Gamma Index
Box Method (81)		+	+			No direct spatial information, DTA is translated to a dose equivalent component
$\chi$ -Evaluation (1 4 2)			+			Need for interpolation if distributions are not of the same grid size.
Delta Envelope (22)			+			Possibly inaccurate for slope changing areas
Surface-based distance method (34)					+	High dose-gradient areas may affect the passing rate. Need for interpolation
Gamma <sup>2</sup> (3)					+	Constant dose gradient factor does not represent the sharpness of local dose gradient. Biased to field size
Geometric Interpretation of gamma (68)			+	+	+	Supplementary to Gamma Index Method
Fast Algorithm (1 1 3)				+		Gamma index limitations in terms of selection of proper acceptance criteria and biological significance of errors
GPU Method (45)				+		Need for interpolation.
Filter Cascade Method (3 8 3)				+		See Geometric Interpretation of gamma
Fast Euclidean Distance Transformation (28)	+	+	+	+		No continuous (minimum) gamma value calculation. Pass/fail result discretization of dose values leads to rounding errors (but within the intrinsic noise level)
Anatomy Corresponding Method (8)	+					Applied to 2D distributions only
Gamma + (3)	+					Visual inspection. Better used in parallel with dose-response curves
RGI Index (17)	+					Need for weighting factors per organ to increase sensitivity of the index
Gamma with Uncertainty Features (4)					+	The combination of large dose or space uncertainties with large tolerance limits could lead to accept an inadequate case
Unbinned Multivariate Test (0)					+	Cannot determine highly localized discrepancies such as unexpected peaks
Inverse Gamma (0)					+	Supplementary to Gamma Index Method

methods that analyze verification results equally in dose and space domain. This kind of techniques is dominated by gamma index and its major variants (Gamma Histograms and Gamma Angle) which focus on appointing physical intuition to the results of dose comparison. The rest gamma index upgrades are structured according to the primary limitation of the original gamma index method they are trying to overcome. Therefore, these upgrades are grouped in a) variants that are trying to limit dose grid dependence, b) variants that are setting more reasonable criteria of acceptance, c) methods that reduce calculation time of gamma index and d) methods that correlate mismatches to anatomy or radiobiology of the patient. Nonetheless, the aforementioned categorization is not absolute. Some of the methods may fall under more than one category. For example, Box method limits as well dose grid dependence, similar to the geometric interpretation of gamma index. Furthermore, Fast Euclidean transformation apart from time reduction offers physical a biological interpretation of the results.

Category a) has a total of 313 citations, category b) has 41, c) has 637 and d) has 56. These numbers reveal the major interest of the community was shifted towards time reducing methods followed by dose grid independent methods. Surprisingly, categories b) and d) which represent methods that assist in correlating verification results to clinical practice fall way behind. This trend can be explained as follows. Verification procedure is a quite laborious task, which should be included in a clinical setting. Therefore, most of the researchers focused their efforts on constituting patient specific quality assurance efficient under clinical limitations which means valid (dose grid independent methods) and fast (calculation time reduction methods).

Gamma index analysis offered a successful combination of two metrics for better dose comparison and has been widely adopted in daily routine. The majority of the identified papers shows that gamma index is calculated with 3%/3mm acceptance criteria. AAPM's report TG-218 on tolerance levels and methodologies for IMRT verification QA recommends 3%/2mm [43]. Published studies on stereotactic treatment QA used further reduced values (3%/2 mm, 3%/1.5 mm, 3%/1 mm and 3%/0.3 mm) as Hussein et al. [16] have reported.

Although gamma index in its primal form is specified in space and can be compared to the sources off errors, the passing rate approach used in clinic (usually 95%) includes some intrinsic limitations. This statistical approach of plan evaluation with passing rates fails to be clinically intuitive and easily interpretable, while the rather arbitrary adoption of acceptance criteria does not reflect the QA accuracy that could be achieved [44]. Furthermore, gamma passing rates do not provide spatial information on where the mismatch occurred. It has been reported that there is a weak correlation between gamma index passing rates and PTV coverage in terms of DVH, while there is no correlation with the DVH changes for organs at risk [45]. The lack of anatomical/biological information could have been partially replenished by a faulty dose level indication. However, gamma fails to provide this kind of information as well [9]. In addition, the sign of the mismatch is not being pointed out, e.g. if the measured point has higher or lower dose value than expected.

Gamma index passing rates are equally susceptible to technical limitations. Many of the follow-up techniques were developed to avoid the interpolation (necessary to produce valid results) of the data. If an exhaustive search of dose points is performed over the whole distribution, this method could be extremely time-consuming. Computational time is equal to the third power of grid size [46]. Otherwise, if the search region is restricted, gamma could be overestimated. Two things have greatly reduced this issue. The first is that Ju et al [25] developed an extremely fast algorithm for calculating gamma index with automated interpolation, and second, that computers are significantly faster than they used to be.

Dose grid dependence is a limitation that every method experiences. Gamma index shows a great dependence to dose grid size relative to the distance criterion (Fig. 1A), resulting in faulty results under certain circumstances [35,47]. Ju's et al. [25] approach limits this problem as

well. Noisy data could lead to undetectable differences by affecting the closest distance where the minimum dose difference could be found [15,41,46]. In addition, the presence of noise biases gamma to lower values in the evaluated distribution. More reliable results could be produced by calculating a full 3D gamma distribution instead of 2D or 2.5D [19,48]. But even in datasets free of noise, gamma could accept points with larger mismatches than the nominal values of the criteria under some combinations of DD, DTA and acceptance passing rate limits [36,49,50].

The limitations of gamma passing rate approach have been revealed even under the scope of the receiver operating characteristic (ROC) analysis [51,52]. It has been shown that the predictive power of patient treatment verification is limited by the size of error to be detected. Even though mismatches that are relatively large ( $> 3$  mm) can be successfully detected, lower sized errors are harder to be spotted.

These constitute the main reasons that since gamma index valuable introduction, many research groups are trying to develop variants, which would eliminate the aforementioned drawbacks and flaws while exploiting its indisputable advantages. These extensions or variations of the original gamma method are described in detail and are listed in Table 1, alongside with the limitation that they try to overcome.

Nowadays, as many of the technical problems have been solved, the community should shift its interest towards methods that could "translate" verification outcome to patient related information. A DVH analysis of the verification result is a step towards this direction; however, the lack of spatial information is a drawback. Additionally, a DVH approach can not be utilized during a TPS commissioning or other routine QAs. In those occasions, methods that take into account experimental uncertainty should be considered to clarify possible mismatches. Whenever gamma analysis is used for patient verification, results should be analyzed with acceptance criteria that are linked to the radiobiological impact of the mismatches. A DVH evaluation of the measured distribution is encouraged to be used if available alongside with a refined gamma index calculation (such as the geometric interpretation of the gamma method [25]). Other radiobiological related methods such as gamma + or RGI are also suggested to complement analysis, especially for cases where the plan "passes" but with a rate close to the limit of acceptance. Furthermore, the analysis is suggested to be performed at least in two different regions of interest (marked as PTV and OARs) by using 90% of the prescription dose as a boundary. Doses from that value and up (high dose region) could be associated with PTV coverage. Below that value mismatches would mostly affect OARs. A further subdivision of the OAR region to high gradient (90%-50%), mean dose (50%-30%) and low dose (30%-20%) zones could provide a better overview of the location of the errors, as already proposed by Stojadinovic et al. [53].

## 5. Conclusions

It is evident that among the available methods for the comparison of two dose distributions, the  $\gamma$ -analysis (gamma analysis) prevails. It has been widely adopted as it offers a global evaluation tool, for every point of the distributions, independently of local dose gradients. It is established as the gold standard in verification procedures and clinical decisions are made upon its results. However, high passing rates in plan verification do not necessarily guarantee accurate dose delivery as per literature. Since gamma index introduction, many efforts have been conducted focusing on refining it. Nevertheless, due to various intrinsic restrictions the generation of an ideal index is still a challenge. The medical physics community should not rest on gamma index familiarity but instead should strive for a better metric taking into account both statistical and in clinical parameters.

## References

- [1] Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al.

- IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. *Med Phys* 2009;36:5359–73.
- [2] Alber M, Georg BMD. E. S. f. T. Guidelines for the Verification of IMRT. *Radiol Oncol* 2008.
  - [3] Wendling M, Zijp LJ, McDermott LN, Smit EJ, Sonke JJ, Mijnheer BJ, et al. A fast algorithm for gamma evaluation in 3D. *Med Phys* 2007;34:1647–54.
  - [4] Siochi RAC, Molineu A, Orton CG. Patient-specific QA for IMRT should be performed using software rather than hardware methods. *Med Phys* 2013;40:070601-n/a.
  - [5] Vazquez-Quino LA, Huerta-Hernandez CI, Rangaraj D. Clinical experience with machine log file software for volumetric-modulated arc therapy techniques. *Proc (Bayl Univ Med Cent)* 2017;30:276–9.
  - [6] Kerns JR, Childress N, Kry SF. A multi-institution evaluation of MLC log files and performance in IMRT delivery. *Radiat Oncol* 2014;9:176.
  - [7] Childress NL, Rosen II. The design and testing of novel clinical parameters for dose comparison. *Int J Radiat Oncol Biol Phys* 2003;56:1464–79.
  - [8] Yeo LJ, Kim J-h. A procedural guide to film dosimetry : with emphasis on IMRT. Madison, Wis.: Med Phys Pub 2004.
  - [9] Zhen H, Nelms BE, Tome WA. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Med Phys* 2011;38:5477–89.
  - [10] Van Dyk J, Barnett RB, Cygler JE, Shragge PC. Commissioning and quality assurance of treatment planning computers. *Int J Radiat Oncol Biol Phys*. 1993;26:261–73.
  - [11] Harms Sr. WB, Low DA, Wong JW, Purdy JA. A software tool for the quantitative evaluation of 3D dose calculation algorithms. *Med Phys* 1998;25:1830–6.
  - [12] Moran JM, Radawski J, Fraass BA. A dose gradient analysis tool for IMRT QA. *J Appl Clin Med Phys* 2005;6:62–73.
  - [13] Wu J-M, Kuo C-M, Chen C-J. Dose verification in intensity modulation radiation therapy: a fractal dimension characteristics study. *Biomed Res Int* 2013;2013. 349437.
  - [14] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys* 1998;25:656–61.
  - [15] Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med Phys* 2003;30:2455–64.
  - [16] Hussein M, Clark CH, Nisbet A. Challenges in calculation of the gamma index in radiotherapy - Towards good practice. *Phys Med* 2017;36:1–11.
  - [17] Spezi E, Lewis DG. Gamma histograms for radiotherapy plan evaluation. *Radiother Oncol* 2006;79:224–30.
  - [18] Stock M, Kroupa B, Georg D. Interpretation and evaluation of the gamma index and the gamma index angle for the verification of IMRT hybrid plans. *Phys Med Biol* 2005;50:399–411.
  - [19] Al Sa'd M, Graham J, Liney GP, Moore CJ. Quantitative comparison of 3D and 2.5D gamma analysis: introducing gamma angle histograms. *Phys Med Biol* 2013;58:2597–608.
  - [20] Jiang SB, Sharp GC, Neicu T, Berbeco RI, Flampouri S, Bortfeld T. On dose distribution comparison. *Phys Med Biol* 2006;51:759–76.
  - [21] Bakai A, Alber M, Nusslin F. A revision of the gamma-evaluation concept for the comparison of dose distributions. *Phys Med Biol* 2003;48:3543–53.
  - [22] Bak J, Choi J, Kim J-S, Park S. Modified dose difference method for comparing dose distributions. *J Appl Clin Med Phys Am Coll Med Phys* 2012;13:3616.
  - [23] Ansbacher W. Modified dose difference method for comparing dose distributions. *J Appl Clin Med Phys* 2012;13:3970-.
  - [24] Blanpain B, Mercier D. The delta envelope: a technique for dose distribution comparison. *Med Phys* 2009;36:797–808.
  - [25] Ju T, Simpson T, Deasy JO, Low DA. Geometric interpretation of the gamma dose distribution comparison technique: interpolation-free calculation. *Med Phys* 2008;35:879–87.
  - [26] Gu X, Jia X, Jiang SB. GPU-based fast gamma index calculation. *Phys Med Biol* 2011;56:1431–41.
  - [27] Persoon LC, Podesta M, van Elmpt WJ, Nijsten SM, Verhaegen F. A fast three-dimensional gamma evaluation using a GPU utilizing texture memory for on-the-fly interpolations. *Med Phys* 2011;38:4032–5.
  - [28] 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;62:309–19.
  - [29] Chen M, Lu W, Chen Q, Ruchala K, Olivera G. Efficient gamma index calculation using fast Euclidean distance transform. *Phys Med Biol* 2009;54:2037–47.
  - [30] Winięcki J, Zurawski Z, Drzewiecka B, Słosarek K. Anatomy-corresponding method of IMRT verification. *Rep Pract Oncol Radiother* 2010;16:1–9.
  - [31] Stathakis S, Mavroidis P, Shi C, Xu J, Kaweloa KI, Narayanasamy G, et al. gamma + index: a new evaluation parameter for quantitative quality assurance. *Comput Methods Progr Biomed* 2014;114:60–9.
  - [32] Sumida I, Yamaguchi H, Kizaki H, Aboshi K, Tsujii M, Yoshikawa N, et al. Novel radiobiological gamma index for evaluation of 3-dimensional predicted dose distribution. *Int J Radiat Oncol Biol Phys* 2015;92:779–86.
  - [33] Sumida I, Yamaguchi H, Das LJ, Kizaki H, Aboshi K, Tsujii M, et al. Evaluation of the radiobiological gamma index with motion interplay in tangential IMRT breast treatment. *J Radiat Res* 2016;57:691–701.
  - [34] Hamatani N, Sumida I, Takahashi Y, Oda M, Seo Y, Isohashi F, et al. Three-dimensional dose prediction and validation with the radiobiological gamma index based on a relative seriality model for head-and-neck IMRT. *J Radiat Res* 2017;58:701–9.
  - [35] Nelms BE, Chan MF, Jarry G, Lemire M, Lowden J, Hampton C, 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-n/a.
  - [36] 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;38:6730–41.
  - [37] Ceberg C. A note on the interpretation of the gamma evaluation index. *J Phys Conf Ser IOP Publish* 2013. 012082.
  - [38] Cutanda Henríquez F, Vargas Castrillón S. A probability approach to the study on uncertainty effects on gamma index evaluations in radiation therapy. *Computat Mathemat Meth Med* 2011. 2011.
  - [39] Ramos Garcia LI, Perez Azorin JF, Aguilar-Redondo PB, Moran-Velasco V. Improving the gamma analysis comparison using an unbinned multivariate test. *Phys Med Biol* 2017;62:N417–27.
  - [40] Yu L, Kairn T, Trapp J, Crowe SB. Technical note: A modified gamma evaluation method for dose distribution comparisons. *J Appl Clin Med Phys* 2019;20:193–200.
  - [41] Daniel ALP. Gamma dose distribution evaluation tool. *J Phys Conf Ser* 2010;250. 012071.
  - [42] Low Daniel AP. Gamma dose distribution evaluation tool. *J Phys Conf Ser* 2010;250. 012071.
  - [43] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM Task Group No. 218. *Med Phys*. 2018;45:e53–83.
  - [44] Basran PS, Woo MK. An analysis of tolerance levels in IMRT quality assurance procedures. *Med Phys* 2008;35:2300–7.
  - [45] Stasi M, Bresciani S, Miranti A, Maggio A, Sapino V, Gabriele P. Pretreatment patient-specific IMRT quality assurance: a correlation study between gamma index and patient clinical dose volume histogram. *Med Phys* 2012;39:7626–34.
  - [46] Schreiner LJ, Holmes O, Salomons G. Analysis and evaluation of planned and delivered dose distributions: practical concerns with  $\gamma$ - and  $\chi$ -Evaluations. *J Phys Conf Ser* 2013;444. 012016.
  - [47] Sunjic S, Ceberg C, Bokulic T. Statistical analysis of the gamma evaluation acceptance criteria: a simulation study of 2D dose distributions under error free conditions. *Phys Med Eur J Med Phys* 2018;52:42–7.
  - [48] Onizuka R, Araki F, Ohno T. Monte Carlo dose verification of VMAT treatment plans using Elekta Agility 160-leaf MLC. *Phys Med Eur J Med Phys*. 2018;51:22–31.
  - [49] Alharthi T, Arumugam S, Vial P, Holloway L, Thwaites D. EPID sensitivity to delivery errors for pre-treatment verification of lung SBRT VMAT plans. *Phys Med* 2019;59:37–46.
  - [50] Alharthi T, Pogson EM, Arumugam S, Holloway L, Thwaites D. Pre-treatment verification of lung SBRT VMAT plans with delivery errors: toward a better understanding of the gamma index analysis. *Phys Med Eur J Med Phys* 2018;49:119–28.
  - [51] Carlone M, Cruje C, Rangel A, McCabe R, Nielsen M, Macpherson M. ROC analysis in patient specific quality assurance. *Med Phys* 2013;40. 042103.
  - [52] Woon W, Ravindran PB, Ekayanake PSV, Lim YY, Khalid J. A study on the effect of detector resolution on gamma index passing rate for VMAT and IMRT QA. *J Appl Clin Med Phys* 2018;19:230–48.
  - [53] Stojadinovic S, Ouyang L, Gu X, Pompoš A, Bao Q, Solberg TD. Breaking bad IMRT QA practice. *J Appl Clin Med Phys* 2015;16:154–65.