Abstract

In the study, a local approach to setting reference tolerance values for the distance-to-agreement (DTA) component of the gamma index is proposed. The reference tolerance values are calculated in simulations, following a dose delivery model presented in a previous work. An analytical model for determining the quantiles of DTA distribution is also proposed and verified. It is shown that the distributions of DTA values normalized with either quantiles or standard deviation of DTA distributions are universal over analyzed plans and points within a single plan. This enables statistically sound inference about the quality of dose delivery. In particular, based on the normalized distributions the comparison of planned and delivered doses can be formulated within the framework of statistical inference as a problem of multiple statistical testing. For every evaluated point \( P \) of a plan, one may formulate and test a null hypothesis that there is no delivery error against an alternative hypothesis that there is a delivery error in \( P \). It is also shown that the proposed approach is more sensitive than the current standard approach to shift errors in high dose gradient regions.

1. Introduction

The gamma index analysis (Low et al 1998) is a current standard method used in radiotherapy for both machine quality control purposes and treatment plan evaluation (Hussein et al 2017a, 2017b). It was originally proposed as a method for defining dose delivery error in an evaluated point of dose delivery; if the value of gamma index in an evaluated point exceeds some threshold \( \gamma_{Th} \) (usually set to 1), then a dose delivery error is reported in this point. The gamma index analysis is summarized in the form of a gamma index passing rate; the percentage of points with gamma index values less than \( \gamma_{Th} \). However, the gamma index analysis provides more rich feedback as it also returns the spatial locations of dose delivery errors, in contrast to other plan evaluation approaches, for example, dose volume histograms (DVHs) (Visser et al 2014). The formula for gamma index involves some user-selected constants: the dose difference (DD) tolerance and the distance-to-agreement (DTA) tolerance. Most frequently, the DD tolerance is set in the range from 1%–3% of the maximal planned dose while DTA tolerance is set in the range from 1 mm to 3 mm for advanced treatment plans (Blanck et al 2015). If the passing rate exceeds some user-selected passing rate criterion, then the plan delivery is accepted for further steps of radiotherapy procedure. The passing rate criterion is set in conjunction with the DD and DTA tolerances. Clearly, setting more stringent DD or DTA tolerance forces selection of a lower value of the gamma passing rate criterion.
The selection of the constants necessary for conducting gamma index analysis is not an obvious task with any erroneous decisions posing a risk of harm to a treated patient. This is one of the reasons making conclusions on the quality of dose delivery based on gamma index values is not an easy and straightforward task. Moreover, it seems that the selection of DD and DTA tolerances and the passing rate criterion are inconsistent among different studies and radiotherapy centers. For example, recommendations of AAPM Task Group No. 218 (Miften et al 2018) contain a thorough review of these inconsistencies.

As claimed in (Li et al 2011), DD and DTA tolerances and the passing rate criterion are set empirically and there are few studies focused on developing physically or statistically sound methods for selecting these constants. In the study of (Li et al 2011) surface-based distance was proposed to measure the difference between planned and delivered dose distributions. Then, the gradient factor is used to select appropriate DTA and DD tolerances. The importance of gradient information for the selection of appropriate DTA and DD tolerances was further considered in the study of (Low et al 2013). In particular, it was shown in (Li et al 2011) that for steep dose gradients gamma index dose comparison reduces to the DTA analysis. Consequently, to make correct conclusions about the quality of dose delivery, the ratio of DD to DTA tolerances must be adjusted to gradients present in the plan.

In recently published recommendations of AAPM Task Group No. 218 (Miften et al 2018) it is stated that ‘tolerance limits are defined as the boundaries within which a process is considered to be operating normally, that is, subject to only random errors.’ It is further recommended in (Miften et al 2018) that an action limit for an evaluated quantity is dependent on the standard deviation of this quantity, as calculated for a system operating normally, subject to only random errors. A statistical approach to DD tolerance selection following this recommendation was demonstrated in (Tulik et al 2019). It has been shown in (Tulik et al 2019) that if the DD tolerance is set locally proportional to the standard deviation of delivered doses, then for a clinically relevant range of passing rates the true positive rate is higher than for a current standard method. The statistical approach to action-level selection in radiotherapy is not limited to dosimetric procedures. In a study of (Baran et al 2019) a statistical framework for conducting tests of linac geometry was described.

Setting DD tolerances locally based on features of an assessed plan of dose delivery is not uncommon (Stojadinovic et al 2015, Tulik et al 2019). In contrast, setting DTA tolerances locally based on features of an assessed plan of dose delivery, to the best of our knowledge, has not been considered in spite of possibly serious consequences of wrong selection of these tolerances (Low et al 2013). Keeping DTA tolerances constant over an entire field of the assessed plan had until recently some justification. In particular, quality control is based on the comparison of doses measured by specialized phantoms, for example, 2D dosimetric arrays. According to a traditional approach an evaluated plan is recomputed to the geometry of a measurement device and then the measurement results are directly compared with the recomputed plan. Consequently, any dose gradient present for inhomogeneous anatomy of a patient is not directly reflected in gradients of a plan recomputed for typically homogeneous geometry of a phantom. Therefore, controlling the shift of isodoses in steep gradient regions in a plan recomputed for a phantom has only limited consequences for the precision of dose delivery to a real patient.

This perspective has however changed since the introduction of measurement-guided dose reconstruction (Nelms et al 2012, Stasi and Bresciani 2012, Visser et al 2014). Measurement-based dose reconstruction delivers tools to estimate doses delivered to a patient anatomy based on the results of measurements made with conventional 2D phantoms. Then, the dose estimated for the volume of a patient body is compared with the dose planned for a patient using commonly used approaches such as DVH or gamma index. With the availability of such precise dose estimates the arguments listed in the previous paragraph do not hold any more and precise control over the location of gradients must become a vital part of a reliable treatment plan quality control procedure.

Here, we propose a local approach to setting reference tolerance values for the DTA component of gamma index. The motivation for setting reference tolerance for DTA locally is as follows. The dose delivery plan contains regions of both high and low dose gradient. Regions of high gradient may be related to organ or target borders while regions of low gradient may correspond to regions of homogeneous dose, for example, organ/target interiors. Small spatial shifts in dose delivery within organ boundary regions may possibly lead to delivering high doses to organs at risk and small dose to a planned target. Therefore, the quality control should be more sensitive to spatial shift errors within high dose gradient regions. On the other hand, higher spatial shifts in homogeneous dose regions are less dangerous because for relatively homogeneous doses they...
lead to relatively small changes in dose delivered compared to planned dose. For the above reasons, it is reasonable to expect that tolerance values for DTA will vary, depending on a local dose gradient magnitude.

Table 1 contains a list of symbols used in this work.

| Symbol | Description |
|--------|-------------|
| P, Q   | Points      |
| \(P_{ij}\) | Point at position \((i, j)\) |
| \(d(P, Q)\) | Distance between points \(P\) and \(Q\) |
| \(D(P)\) | Planned dose at point \(P\) |
| \(DD(P)\) | Measured dose at point \(P\) |
| \(DD(P, Q)\) | Equal to \(D(P) - D_M(Q)\) |
| \(\Delta_{DD}(P)\) | Dose difference tolerance at point \(P\) |
| \(\Delta_{DTA}(P)\) | Distance-to-agreement tolerance at point \(P\) |
| \(\gamma(P)\) | Gamma index at point \(P\) |
| \(GP\) | Gamma index passing rate |
| \(DTA(P)\) | Explained below equation (2) |
| \(\sigma_{DTA}(P)\) | Explained below equation (4) |
| \(P_{TH}\) | Probability threshold |
| \(f_r\) | Probability density function of dose at point \(P\) |
| \(Pr(P, d)\) | Probability that the dose measured at point \(P\) belongs to an interval defined below equation (7) |
| \(\Theta_{PPH}(P)\) | Distance \(d\) at which \(Pr(P, d)\) at point \(P\) is equal to \(P_{TH}\) |

\[ \gamma(P) = \min_{Q \in N_P} \sqrt{\frac{d(P, Q)^2}{\Delta_{DTA}(P)^2} + \frac{DD(P, Q)^2}{\Delta_{DD}(P)^2}}, \]  

where \(\Delta_{DD}\) and \(\Delta_{DTA}\) are selected tolerances for, respectively, DD and DTA, the point \(P\) belongs to some neighborhood \(N_P\) of \(P\) and \(d(P, Q)\) is the distance between \(P\) and \(Q\). The tolerances can be set either globally, that is, equal for all points in a plan, or locally in which case both \(\Delta_{DD}\) and \(\Delta_{DTA}\) can vary between points.

The values of gamma index are thresholded using a user-selected threshold value \(\gamma_{TH}\). In clinical practice, typically \(\gamma_{TH}\) is equal to 1. If \(\gamma(P)\) is greater than \(\gamma_{TH}\), then, by definition, \(P\) is a point of an erroneous dose delivery, otherwise \(P\) is a point of a correct dose delivery. Eventually, the fraction of points with correctly delivered doses is also calculated (gamma index passing rate GP) with the value of 100%, meaning that all the deviations between the planned and measured doses fall below \(\gamma_{TH}\). The plan is rejected if the value of GP is below a certain threshold, for example, 90% (Li et al 2011) or above 95% (Sothmann et al 2016), although even plans with a high value of GP may not be clinically acceptable (Miften et al 2018).

The expression for gamma index is fairly complicated as it depends on planned and delivered doses at reference and evaluated points, distance between reference and evaluated points, tolerances \(\Delta_{DD}\) and \(\Delta_{DTA}\) and requires searching for a minimum over some set of reference points. This is one of the reasons making conclusions on the quality of dose delivery based on gamma index values is not a straightforward task. Here, we focus on how the distance component and tolerance of DTA influence the percentage of points GP passing the quality control procedure. To separate the effect of the dose component from the effect of the distance component onto quality control results, we consider a limit of a very small \(\Delta_{DD}\). When continuous interpolations of measured realizations are considered one has in this limit:

\[ \lim_{\Delta_{DD} \to 0} \gamma(P) = \frac{DTA(P)}{\Delta_{DTA}(P)}, \]  

where \(DTA(P)\) is the distance between the pixel \(P\) and the closest position \(Q\) in the interpolated measurement, so that \(DD(P, Q) = 0\) (which may lie between some pixels). This limit holds because the agreement is exact at point \(Q\). Note that the ratio of DTA to \(\Delta_{DTA}\) is also used in a compositional approach to quality control (Harms et al 1998).
2.1. Probability-based approaches to DTA action levels

It would be plausible to state formally the comparison of planned and delivered doses within the framework of statistical inference as a problem of multiple statistical testing. In particular, for every evaluated point \( P \) we formulate a null hypothesis that there is no delivery error against an alternative hypothesis that there is a delivery error in \( P \). Formal statistical testing of the presence of a delivery error in some evaluated point must be based on an appropriate universal function—a test statistic that does not depend on particular features of a plan. In particular, it would be plausible to define a random variable \( \Gamma \) depending on the dose planned in \( P \) and doses measured in some neighborhood of \( P \), so that \( \Gamma \) is independent of \( P \) and thus also independent of a particular plan being delivered to a patient.

For example, if \( X \) is a normal random variable with expected value equal to \( m_X \) and standard deviation equal to \( s_X \), then a random variable \( X_U \) defined as:

\[
X_U = \frac{X - m_X}{s_X},
\]

is distributed in accordance with a standard Gaussian distribution, that is, the distribution of \( X_U \) is universal, independent of a particular distribution of \( X \). The existence of such universal distributions derived from measured data is a prerequisite of a model-based statistical inference.

Inspired by the analogy from statistical inference, we propose that the distance component in the definition of gamma index be normalized in the following way:

\[
\Gamma = \lim_{\Delta \to 0} \frac{\gamma (P)}{\Delta D_{TA} (P)} = \frac{D_{TA} (P)}{\sigma_{DTA} (P)},
\]

where \( \sigma_{DTA} (P) \) is the standard deviation of the distribution of DTAs calculated in \( P \); for example, over multiple dose deliveries of the same plan. In the following section, we demonstrate that \( \Gamma \) has indeed the property of being a universal statistic.

The standard deviation \( \sigma_{DTA} (P) \) for DTA can be in principle determined from simulations. For example, based on the noise model described in (Tulik et al 2019) and summarized in the next section, simulated dose delivery data can be generated and then the DTA can be calculated for every point in a plan using, for example, the algorithm described in (Harms et al 1998). Data sufficient for determining \( \sigma_{DTA} (P) \) in every evaluated point can be collected by repeating this procedure multiple times. It is also possible to estimate from the simulations the probability density function (PDF) of DTA and in particular quantiles corresponding to user-selected probability thresholds \( P_{TH} \). Given the PDF of DTA, \( \sigma_{DTA} (P) \) can be estimated in a standard way.

The simulation-based procedure to estimate the PDF of DTA while being straightforward is computationally expensive. Thus, below we estimate the PDF of DTA using analytical considerations.

Let the dose planned for \( P \) be equal to \( D(P) \) and the dose measured at \( P \) be equal to \( D_M (P) \). It is assumed that the measured dose \( D_M (P) \) at pixel \( P \) is a random number drawn from a Gaussian distribution described by a PDF \( f_d \) with an expected value equal to \( D(P) \) and some standard deviation \( \sigma (D(P)) \). In our previous study (Tulik et al 2019), we have shown that \( \sigma (D(P)) \) grows linearly with dose \( D(P) \) planned in \( P \) with proportionality coefficient equal to approximately 0.1. It is also assumed that doses delivered at different locations are independent. While it is true that this assumption generally does not hold, its influence on the conclusions of our analysis will be tested.

To calculate the quantiles of PDF of DTA, given a dose \( D(P) \) planned for \( P \), we calculate the distance from \( P \) to the isodose \( D(P) \) in the grid of measured doses. To find the probability that isodose \( D(P) \) is found at distance \( d \) from \( P \), we define the following function \( Z(P, d) \):

\[
Z(P, d) = \{ Z(P_i) : d (P_i, P) \leq d \},
\]

where \( Z(P_i) \) is a random number drawn from \( f_d \) for all \( N(P, d) \) pixels \( P_i \), so that their distance \( d(P_i, P) \) to \( P \) is smaller than or equal to \( d \). Given \( Z(P, d) \), we define the following two random variables:

\[
Z_{\min}(P, d) = \min Z(P, d),
\]

\[
Z_{\max}(P, d) = \max Z(P, d).
\]

Given \( Z_{\min}(P, d), Z_{\max}(P, d) \) and \( D(P) \), one may ask what is the probability \( Pr(P, d) \) that \( D(P) \in [Z_{\min}(P, d), Z_{\max}(P, d)] \). The probability \( Pr(P, d) \) is a function of point \( P \) and the probed distance \( d \) (figure 1), it is monotonic, equal to zero at \( d = 0 \) and approaches one as \( d \) increases to infinity.
In the plot of \( Pr(P, d) \) versus \( d \) (figure 1), \( P_{TH} \) corresponds to some critical value of \( d \), which we choose as a quantile \( \Theta_{PTH}(P) \) for the tested pixel \( P \). For high gradients of the planned doses, the dose values around \( P \) are more scattered and thus \( Pr(P, d) \) grows faster with distance \( d \) than in the case of small gradient in which case the plot of \( Pr(P, d) \) versus \( d \) is more flat (figure 1). Consequently, setting exactly the same \( P_{TH} \), we get smaller \( \Theta_{PTH}(P) \) for high gradient regions than for small gradient regions. Varying \( P_{TH} \), the other quantiles of the PDF of DTA can be computed from which the PDF of DTA can be estimated.

Given the set \( Z(P, d) \) of \( N(P, d) \) random numbers (equation (6)) corresponding to doses delivered in pixels \( P \), so that \( d(P, P) \leq d \) and assuming that doses delivered at different locations are independent, \( Pr(P,d) \) can be calculated from the following expression:

\[
Pr(P, d) = \sum_{i=1}^{N(P,d)} \sum_{j=1, j \neq i}^{N(P,d)} \int_{-\infty}^{\infty} dx f_{D}(x) \int_{D(P)} dx f_{P}(x) \prod_{k=1, k \neq i, j}^{N(P,d)} \int_{-\infty}^{\infty} dx f_{P}(x), \tag{8}
\]

where \( f_{D} \) is the probability density function of the dose \( D_{M}(P) \) delivered at the location corresponding to pixel \( P \) and \( x \) for \( i \in \{1, 2, \ldots, N(P, d)\} \) corresponds to possible doses at point \( P \). Assuming that each \( f_{D} \) is a Gaussian distribution, \( Pr(P,d) \) can be calculated directly from equation (8).

Instead of calculating \( Pr(P,d) \) directly from equation (8) one can transform it to a form that is easier to compute. It is always true that either \( D(P) \) is smaller than or equal to \( Z_{MIN}(P, d) \), \( D(P) \) is greater than or equal to \( Z_{MAX}(P, d) \) or \( D(P) \) is between \( Z_{MIN}(P, d) \) and \( Z_{MAX}(P, d) \). Hence, the sum of probabilities of these three events is equal to 1. Instead of calculating the probability of the third event, which is equal to \( Pr(P,d) \), the probabilities of the other two events can be calculated and subtracted from 1.

The probability \( Pr_{MIN}(P, d) \) that \( D(P) \) is smaller than \( Z_{MIN}(P, d) \) is equivalently the probability that \( D(P) \) is smaller than all elements of \( Z(P, d) \). Since we assume that distributions of elements of \( Z(P, d) \) are independent, \( Pr_{MIN}(P, d) \) can be expressed as a product of probabilities:

\[
Pr_{MIN}(P, d) = \prod_{i=1}^{N(P,d)} \int_{-\infty}^{D(P)} f_{P}(x) dx. \tag{9}
\]
The probability \( \Pr_{\text{MAX}} (P, d) \) that \( D(P) \) is greater than \( Z_{\text{MAX}} (P, d) \) can be calculated in a similar way:

\[
\Pr_{\text{MAX}} (P, d) = \prod_{i=1}^{N(P,d)} \int_{D(P)} f_P (x_i) \, dx_i. \tag{10}
\]

Finally, \( Pr(P, d) \) can be calculated as

\[1 - \Pr_{\text{MAX}} (P, d) - \Pr_{\text{MIN}} (P, d).\]

### 2.2. Noise model

To simulate the spatial distributions of delivered doses, we use a noise model proposed and verified in our previous study (Tulik et al. 2019). For convenience, some details of this noise model are summarized below. In particular, the plan \( D_{PL} \) is a rectangular matrix of planned doses \( D(P_{ij}) \) for pixels \( P_{ij} \) at coordinates \((i, j)\) in the grid of size \( I \) by \( J \) pixels:

\[
D_{PL} = \{ D \{ P_{ij} \} : i = 1, 2, \ldots , I, j = 1, 2, \ldots , J \}. \tag{11}
\]

To simulate dose delivery with no systematic error, we add noise to \( D_{PL} \); in each \((i, j)\) we generate a random number sampled from a Gaussian distribution with zero average and standard deviation proportional to \( D(P_{ij}) \). The resulting matrix of random errors \( E \) is convolved with a Gaussian smoothing kernel to reproduce autocorrelation of the noise, which can be seen in the real data (Low et al. 2013). The resulting matrix \( SE \) is added to \( D_{PL} \). In agreement with experimental data, we assumed that the standard deviation of the Gaussian noise in \((i, j)\) is equal to 1.25 times \( D(P_{ij}) \) and the size of the smoothing kernel is equal to 3.6 mm. After smoothing the standard deviation of the noise, it is equal to 0.1 times the dose value. A noise model without autocorrelation was also considered.

### 2.3. Simulations

In the first computational experiment, dose delivery simulation was based on the Gaussian noise model without autocorrelation. In these simulations, standard deviation of delivered doses was equal to 0.1 times the dose value, in agreement with published data (Tulik et al. 2019). One hundred dose deliveries were simulated.

Next, we considered simulation of dose delivery based on the Gaussian noise model with autocorrelation. In these simulations, standard deviation of delivered doses was equal to 1.25 times the dose value and the size of the Gaussian kernel was equal to 3.6 mm, in agreement with published data (Tulik et al. 2019) that together leads to standard deviation of delivered doses effectively equal to 0.1 times the planned dose value, as in the case of uncorrelated noise. Again, one hundred dose deliveries were simulated.

In the final experiment, we considered simulation of dose delivery based on the Gaussian noise model with autocorrelation and a systematic error—a shift by a random vector with components drawn independently from a Gaussian distribution with some non-zero mean and standard deviation equal to 2 pixels. For each considered value of the systematic error 100 random shifts were simulated for each of the 22 plans.

### 3. Results

In figure 2, we show a sample plan together with pictures demonstrating variations of \( \sigma_{\text{DTA}} \) and \( \Theta_{\text{PTH}} \) over the plan. In the figure, we show \( \Theta_{\text{PTH}} \) for \( P_{\text{TH}} = 0.8 \), denoted as \( \Theta_{0.8} \), as calculated directly from the simulations as well as using the analytical approach. The average and standard deviation of \( \sigma_{\text{DTA}} \) for the sample plan were equal to 3.1 and 2.8 pixels, respectively. The average and standard deviation of \( \Theta_{0.8} \) calculated in simulations were equal to 5.4 and 5.9, respectively. The average and standard deviation of \( \Theta_{0.8} \) calculated analytically were equal to 5.5 and 4.5, respectively. In the figure, we also show the regions of high and low gradient determined using k-means clustering algorithm (Kaufman and Rousseeuw 1990). Note the close similarity of figures 2(b)–(d).

In figure 3, we show how quantiles \( \Theta_{\text{PTH}} \) of DTA distribution increase with increasing threshold probability \( P_{\text{TH}} \). Note that there is a very good correspondence between quantiles calculated from simulations and those based on the analytical model. The quantiles were calculated for an exemplary plan shown in figure 2. Similarly good agreement between analytical predictions and numerical results was observed for other analyzed plans. In figure 4, we show the values of the Pearson’s coefficient of correlation between quantiles \( \Theta_{\text{PTH}} \) of DTA distribution calculated from simulations and from the analytical model. In the figure, we also show the values of the Pearson’s coefficient of correlation between \( \sigma_{\text{DTA}} \) and quantiles \( \Theta_{\text{PTH}} \) of DTA distribution calculated from the analytical model as well as the values of the Pearson’s coefficient of correlation between \( \sigma_{\text{DTA}} \) and quantiles \( \Theta_{\text{PTH}} \) of DTA distribution calculated from simulations.
The coefficients of correlation were calculated separately for each of 22 plans and then averaged over the set of plans. Thus, the error bars represent standard deviation of the coefficient of correlation over the set of different plans. Note that the coefficient of correlation between quantiles of DTA distribution calculated analytically and numerically (in simulations) is very high in the entire range of $P_{TH}$. Note also that quantiles of DTA distribution corresponding to $P_{TH}$ greater than or equal to 0.8 correlate very well with $\sigma_{DTA}$, that is, these quantiles could be surrogate measures of the scatter of DTA values and thus they can be used as a factor normalizing DTA in equation (2).
It has been argued in the previous section that normalizing DTA with a position-dependent value of $\sigma_{DTA}$ could be advantageous as it can possibly lead to a distribution of normalized DTA values, which is universal over different plans and different points within the same plan. In figure 5, we show PDFs of DTA normalized with different factors: a constant value of 6 pixels (the normalization of a distance component with a fixed value is current clinical practice), $\sigma_{DTA}$, $\Theta_{PTH}$ of DTA distribution calculated from analytical model and from simulations for $P_{TH} = 0.8$. In the figure, we show PDFs of DTA of all 22 analyzed plans as well as point-wise PDFs of DTA, calculated for randomly selected 20 points of the plan shown in figure 1. In the latter case, the point-wise PDFs were generated based on 10 000 simulated dose deliveries.

According to the definition of gamma index, a dose is delivered erroneously in a point if the gamma index value in that point exceeds one. We show in figure 6 how the passing rate GP—the fraction of pixels with DTA less than $\Delta_{DTA}$—depends on the selection of a factor normalizing DTA in equation (2). Note that because we consider only random noise with magnitude characteristic to real therapeutic devices, all cases when normalized DTA exceeds one can be classified as false positives. In figure 6(a), we plot the value of GP as a function of a global $\Delta_{DTA}$ changing in the range from 3 to 12 pixels. In figure 6(b), we show the case of $\Delta_{DTA}$ equal to $k(P_{TH})\sigma_{DTA}$ where the factor $k(P_{TH})$ is the quantile corresponding to $P_{TH}$, calculated from

![Figure 3. Quantiles $\Theta_{PTH}$ (in millimeters) of the probability distribution of DTA calculated from the analytical model (a) and from simulations (b).](image)
standard normal distribution. In figures 6(c) and (d), we plot the value of GP as a function of $P_{TH}$ with the corresponding value of $\Delta_{DTA}$ equal to $\Theta_{PTH}$. Clearly, GP must grow with growing $\Delta_{DTA}$. In the case of
Figure 5. PDFs of normalized DTA values calculated for 22 plans (a)–(d) and 20 randomly selected points of a sample plan shown in figure 1. Following normalizing factors \( \Delta_{DTA} \) were used: 6 pixels (a) and (e), \( \sigma_{DTA} \) (b), \( \Theta_{0.8} \) of DTA distribution calculated from the analytical model (c) and (f) and from simulations (d). Orange dotted curves are PDFs of normalized DTA values calculated for a sample plan. Points of each other color and shape correspond to a specific plan.

normalization based on \( \sigma_{DTA} \) and quantiles of the normal distribution, the gamma passing rate is lower than that obtained for normalization using the analytical model or simulations in the entire tested range of \( P_{TH} \).

In figure 6, we show GP for the total plan area (circles) and separately for high (triangles) and low gradient (squares) regions. For global normalization with a fixed factor there is a large discrepancy between GP values determined for the low and high gradient region. This discrepancy changes non-linearly as \( \Delta_{DTA} \) grows. The passing rate for high gradient regions remains high and false positives are reported mainly in low gradient regions. For normalization based on \( \Delta_{DTA} \) and \( \Theta_{PTH} \) the passing rate grows linearly with \( P_{TH} \). For normalization based on \( \Theta_{PTH} \) calculated numerically GP equals \( P_{TH} \) and is the same in high and low gradient regions. For normalization based on \( \Theta_{PTH} \) calculated analytically GP overestimates \( P_{TH} \) by a few percent and is slightly higher in high and low gradient regions. A similar pattern is observed for normalization based on \( k\left(P_{TH}\right)/\sigma_{DTA} \).

The main objective of the next experiment was to check if the analytical model assuming uncorrelated noise can be used in the case of spatial autocorrelation. In figure 7, we show the PDFs of DTA calculated for spatially autocorrelated data, but normalized with \( \Theta_{PTH} \) calculated analytically for uncorrelated data. In figure 7, we also show passing rates calculated for spatially correlated data based on normalization derived from an analytical model with uncorrelated noise. For comparison, we repeat in figure 7(b) the plot from figure 5(c) and in figure 7(d) the plot from figure 6(c). Clearly, using the analytical model of spatially
uncorrelated noise for spatially correlated data makes PDFs of normalized DTA values overlap nearly exactly, but there is almost no change in the dependence of GP on $P_{TH}$. The same pattern is observed if quantiles $\Theta_{PTH}$ or $\sigma_{DTA}$ calculated numerically for uncorrelated data are used as normalization for spatially correlated data.

The results of the last experiment are illustrated in figure 8. We show in it how GP degrades while increasing the mean shift of the systematic error from 2 to 10 pixels. Error bars represent standard deviation calculated based on all 22 considered plans and 100 random shifts per plan. In figure 8, we compare the cases of $\Delta_{DTA}$ set globally to 6 pixels and $\Delta_{DTA}$ set locally to $\Theta_{0.9}$ of DTA distribution calculated from the analytical model with uncorrelated noise. We show GP for the total plan area (circles) and separately for high (triangles) and low gradient (squares) regions. Passing rates in high gradient regions are much higher than average for the global method in contrast to a local normalization of DTA based on $\Theta_{0.9}$.

4. Discussion

Let us discuss how the proposed approach can be used for quality control in radiotherapy. Assume that the DTA calculated for a pixel $P$ is equal to some $d_a(P)$. The quality control for dose delivery involves in particular a decision whether $d_a(P)$ is small enough and dose delivery in $P$ can be accepted or in contrast, is too large and the quality of dose delivery in $P$ is not sufficient.

The procedure of evaluation of $d_a(P)$ can proceed as follows. Set a threshold probability $P_{TH}$, which is the largest tolerated probability of an error of the second kind (that is an error of claiming that $d_a(P)$ fits the assumed statistical model while there are some important, possibly non-random unaccounted factors influencing $d_a(P)$). $P_{TH}$ should be set to a small value as any false negative decision (that is, falsely claiming that a device operates correctly) can potentially lead to accepting a malfunctioning device that can pose a hazard to a treated patient. On the plot of $Pr(P_d)$ versus $d$ (figure 1), $P_{TH}$ corresponds to some critical value of $d$, which we may choose as the reference tolerance $\Delta_{DTA}(P)$ for the tested pixel $P$. Thus, if $d_a(P)$ is smaller than $\Delta_{DTA}(P)$, then we know that the probability of an error of the second kind for $P$ is smaller than the accepted threshold $P_{TH}$ (figure 1).

Consider an example shown in figure 9. In this figure, we show PDFs of DTA values normalized with $\Theta_{0.9}$ of DTA distributions calculated from analytical models for all 22 collected plans. The data shown in
Figure 7. PDFs of DTA calculated for spatially autocorrelated (a) and uncorrelated (b) simulated data, normalized with quantiles \( \Theta_{0.8} \) of DTA distribution calculated from the analytical model with uncorrelated noise. Points of each color and shape correspond to a specific plan. Passing rates calculated for spatially autocorrelated (c) and uncorrelated (d) simulated data as a function of a probability \( P_{TH} \) defining DTA tolerance based on the analytical model with uncorrelated noise. Circles: total GP, triangles: GP for high dose gradient regions, squares: GP for low dose gradient regions. Error bars represent standard deviation calculated based on the selected noise model.

Figure 9(a) were generated by adding noise following the model (Tulik et al. 2019) and also (figure 9(b)) by global shifting of the resulting spatial distribution by a random vector with expected value equal to 6 pixels and standard deviation equal to 2 pixels. Both the distributions without and with shifts are fairly universal and can be approximated with an exponential distribution \( P(DTA/\Theta_{0.9}) \):

\[
P(DTA/\Theta_{0.9}) = \lambda \exp \left( -\lambda \frac{DTA}{\Theta_{0.9}} \right).
\]

The rate parameter \( \lambda \) is approximately equal to 0.55 for the data shown in figure 9(a) and to 0.23 for the data shown in figure 9(b). Now, assume that we would like to detect the cases with average systematic shift of 6 pixels with probability of at least 0.9, that is, we accept at most 0.1 probability of an error of a second kind. From the distribution shown in figure 9(b) it then follows that the threshold for \( DTA/\Theta_{0.9} \) is equal to 0.37, that is, if at any pixel the ratio \( DTA/\Theta_{0.9} \) is larger than 0.37, then the pixel does not pass quality control procedure. Unfortunately, with that high power of error detection the probability of an error of the first kind (false positives) would be as high as 0.82. These results demonstrate that the detection of errors based on DTA or more generally on gamma index may not be very effective. This finding is in agreement with the results of our previous study (Tulik et al. 2019), where it was shown that under some circumstances gamma index may behave as a random classifier of dose delivery errors.

In this article, we have shown that with proper normalization of DTA, universal distributions of normalized DTA values can be obtained. This universality holds over sets of plans and sets of points within a plan. This in turn enables statistically sound inference about either the presence or absence of dose delivery errors. In particular, if the noise model is known and some DTA value was found then it is possible to estimate both the power of the testing for the presence of a delivery error and the probability of an error of the first kind.

That kind of inference is not possible for a current standard approach as the distributions of raw DTA values (or DTA values scaled by some constant) are not universal. An alternative in the case of raw DTA would be to generate the distributions in simulations. However, such an approach can be quite expensive.
Figure 8. Passing rate GP as a function of mean shift of a systematic error (in millimeters) for $\Delta_{\text{DTA}}$ set globally to 6 pixels (a) and set locally to $\Theta_{0.9}$ of DTA distribution calculated from the analytical model with uncorrelated noise (b). Circles: total GP, triangles: GP for high dose gradient regions, squares: GP for low dose gradient regions. Error bars represent standard deviation calculated based on the selected noise model.

In contrast, the normalizing factors related to quantiles or standard deviation of DTA distributions can be derived from an analytical model. This model requires specifying for every point $P$ in the plan a minimal and maximal dose in some neighborhood of size $d$ of point $P$, where $d$ is varied in some range. Finding these minimal and maximal values is not expensive—the sequence of minimal and maximal values for growing $d$ can be found by iteratively applying erosion and dilation filters to the original plan. With each iteration minimal and maximal value within a neighborhood of increasing size of every point $P$ is found. We have shown that the predictions of the analytical model quite precisely reproduce the results of simulations.

The advantage of the proposed approach is also its interpretability. In particular, if $\Theta_{P_{TH}}$ is selected as the factor normalizing DTA, then the passing rate for data without systematic error is simply equal to $P_{TH}$. Importantly, the discrepancies between passing rates estimated for high and low dose gradient regions are
low compared to current standard approach (normalizing DTA with some constant). We have also shown that quantiles calculated for uncorrelated noise can be used for data with spatially correlated noise—the PDFs of DTA values normalized using this method remain universal.

It is instructive to compare how the current standard approach (normalization of DTA with a constant factor) and the proposed approach (normalizing DTA with either some quantile or standard deviation of DTA distribution) estimate passing rate in the case of errors present in the dose delivery data. We have selected the normalizing factors for the data shown in figure 8 in such a way that an average passing rate over the total plan is almost the same for the current standard and the proposed approach. However, the current standard approach estimates the passing rate in high dose gradient regions overly optimistically—the passing rate for these regions is substantially higher than average. In contrast, the passing rate for high dose gradient regions calculated based on the proposed approach is lower than average. In fact, in the case of errors consisting of shifts of dose delivery patterns, any quality assurance procedure should be especially sensitive to shift errors in high dose gradient regions for the reasons listed in the introduction. The proposed approach has this property while the current standard approach has not, including the composite test and gamma test (Miften et al 2018). While the gamma test is less sensitive to shift errors in low dose gradient regions, this effect, to our knowledge, has not been described in a statistically rigorous way.
Furthermore, we plan to integrate our approach with 3D time-dependent dose reconstruction methods (Nelms et al 2012) to allow more detailed quality assurance, including separate analysis of OAR and PTV regions. The dose gradient factor (Li et al 2011) could also be used as an alternative quantification of the dose gradient.

Finally, we considered a quality testing based only on a ratio of DTA to some normalizing factor, which is a limiting case of gamma index. We may expect that the conclusions following from our research also hold for non-zero DD tolerance. Separating the effects of the DD term from the effect of the distance term for a generic case of non-zero DD and distance tolerances would be however hardly possible.

Acknowledgments

This study was supported by the POIR.04.04.00-00-15ES/18 Project of the Foundation for Polish Science. The POIR.04.04.00-00-15ES/18 project is carried out within the “TEAM-NET” programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

ORCID iDs

Mateusz Baran  https://orcid.org/0000-0001-9667-5579
Krzysztof Rzecki  https://orcid.org/0000-0002-6834-2344
Tomasz Sośnicki  https://orcid.org/0000-0001-7059-7971
Zbigniew Tabor  https://orcid.org/0000-0002-9688-9718

References

Baran M, Rzecki K, Kabat D, Tulik M, Wydra A, Derda Z, Sochaczewska A and Tabor Z 2019 A simulation-based method for evaluating geometric tests of a linac c-arm in quality control in radiotherapy J. App. Clin. Med. Phys. 20 133–42
Blank O, Masi LDamme M C et al 2015 Film-based delivery quality assurance for robotic radiosurgery: commissioning and validation Phys. Med. 31 476–83
Ezzell G A et al 2009 IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119 Med. Phys. 36 5559–73
Harms W B, Low D A, Wong J W and Purdy J A 1998 A software tool for the quantitative evaluation of 3D dose calculation algorithms Med. Phys. 25 1830–6
Hussein M, Clark C H and Nisbet A 2017a Challenges in calculation of the gamma index in radiotherapy—towards good practice Phys. Med. 36 1–11
Hussein M, Clementel E, Eaton D J, Greer P B, Haworth Aishkira S K et al 2017b A virtual dosimetry audit—towards transferability of gamma index analysis between clinical trial QA groups Radiat. Oncol. 125 398–404
Kaufman I. and Rousseuw P J 1990 Finding Groups in Data: An Introduction to Cluster Analysis (New York: Wiley)
Li H, Dong L, Zhang L, Yang J N, Glinin M T and Zhu X R 2011 Toward a better understanding of the gamma index: investigation of parameters with a surface-based distance method Med. Phys. 38 6730–41
Low D, Morele D, Chow P, Dou T H and Ju T 2013 Does the $\gamma$ dose distribution comparison technique difficult to the distance to agreement test in clinical dose distributions? Med. Phys. 40 071722
Low D A, Harms W B, Mutic D and Purdy J A 1998 A technique for the quantitative evaluation of dose distributions Med. Phys. 25 656–61
Miften M et al 2018 Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM Task Group No. 218 Med. Phys. 45 e53–e83
Nelms B E, Opp D, Robinson J, Wolf T K, Zhang G, Moros E and Feygelman V 2012 VMAT QA: measurement-guided 4D dose reconstruction on a patient Med. Phys. 39 4228–38
Soltmann T, Blank O, Poels K, Werner R and Gauer T 2016 Real time tracking in liver SBRT: comparison of CyberKnife and Vero by planning structure-based $\gamma$-evaluation and dose-area histograms Phys. Med. Biol. 61 1677–91
Stasi M, Bresciani S, Miranti A, Maggio A, Sapino V and Gabriele P 2012 Pretreatment patient-specific IMRT quality assurance: a correlation study between gamma index and patient clinical dose volume histogram Med. Phys. 39 7626–34
Stojadinovic S, Ouyang L, Gu X, Pomposa A, Bao Q and Solberg T D 2015 Breaking bad IMRT QA practice J. Appl. Clin. Med. Phys. 16 154–65
Tulik M, Kabat D, Baran M, Kycia R and Tabor Z 2019 Use of statistical approaches to improve the quality control of the dose delivery in radiotherapy Phys. Med. Biol. 64 145018
Visser R, Wauben D Jd Grooth M et al 2014 Evaluation of DVH-based treatment plan verification in addition to gamma passing rates for head and neck IMRT Radiat. Oncol. 112 389–95