Article Type: Task Group Report

Tolerance Limits and Methodologies for IMRT Measurement-Based Verification QA: *Recommendations of AAPM Task Group No. 218*

Moyed Miften* (Chair)

Department of Radiation Oncology, University of Colorado School of Medicine, Aurora, Colorado 80045

Arthur Olch

Department of Radiation Oncology, University of Southern California and Radiation Oncology Program, Childrens Hospital of Los Angeles, Los Angeles, California 90027

Dimitris Mihailidis

Charleston Radiation Therapy Consultants, Charleston, West Virginia 25304

Jean Moran

Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109

Todd Pawlicki

Department of Radiation Oncology, University of California San Diego, La Jolla, California 92093

Andrea Molineu

Radiological Physics Center, UT MD Anderson Cancer Center, Houston, Texas 77030

Harold Li

Department of Radiation Oncology, Washington University, St. Louis, Missouri 63110

Krishni Wijesooriya

Department of Radiation Oncology, University of Virginia, Charlottesville, Virginia 22908

Jie Shi

Sun Nuclear Corporation, Melbourne, Florida 32940

Ping Xia
Department of Radiation Oncology, The Cleveland Clinic, Cleveland, Ohio 44195

Nikos Papanikolaou
Department of Medical Physics, University of Texas Health Sciences Center, San Antonio, Texas 78229

Daniel A. Low
Department of Radiation Oncology, University of California Los Angeles, Los Angeles, California 90095

**Corresponding Author:*
Moyed Miften
Dept of Radiation Oncology
University of Colorado School of Medicine
Aurora, CO 80045
email: moyed.miften@ucdenver.edu

Abstract

Purpose: Patient-specific IMRT QA measurements are important components of processes designed to identify discrepancies between calculated and delivered radiation doses. Discrepancy tolerance limits are neither well defined nor consistently applied across centers. The AAPM TG-218 report provides a comprehensive review aimed at improving the understanding and consistency of these processes as well as recommendations for methodologies and tolerance limits in patient-specific IMRT QA.

Methods: The performance of the dose difference/distance-to-agreement (DTA) and γ dose distribution comparison metrics are investigated. Measurement methods are reviewed and followed by a discussion of the pros and cons of each. Methodologies for absolute dose verification are discussed and new IMRT QA verification tools are presented. Literature on the expected or achievable agreement between measurements and calculations for different types of planning and delivery systems are reviewed and analyzed. Tests of vendor implementations of the γ verification algorithm employing benchmark cases are presented.

Results: Operational shortcomings that can reduce the γ tool accuracy and subsequent effectiveness for IMRT QA are described. Practical considerations including spatial resolution, normalization, dose threshold, and data interpretation are discussed. Published data on IMRT QA and the clinical experience of the group members are used to develop guidelines and recommendations on tolerance and action limits for IMRT QA. Steps to check failed IMRT QA plans are outlined.

Conclusion: Recommendations on delivery methods, data interpretation, dose normalization, the use of γ analysis routines and choice of tolerance limits for IMRT QA are made with focus on detecting differences between calculated and measured doses via the use of robust analysis methods and an in-depth understanding of IMRT verification metrics. The recommendations are intended to improve the IMRT QA process and establish consistent, and comparable IMRT QA criteria among institutions.

Keywords: IMRT QA, tolerance limits, γ, DTA

TABLE OF CONTENTS

I. Introduction and statement of the problem

a. Background and importance

The idea of conformal radiotherapy using intensity modulation was presented in the early 1990's[1-3] with delivery systems first contemplated and shortly thereafter developed[4, 5]. By the mid 1990's, reports of the

first clinical experience with intensity modulated radiation therapy (IMRT) were published[6, 7] with the commissioning and quality assurance (QA) of such systems published simultaneously[8, 9]. IMRT delivery techniques can be divided into two broad categories, fixed and moving gantry. Fixed-gantry IMRT delivery employs step-and-shoot (or segmental), sliding window (dynamic), or compensator-based methods. Moving gantry IMRT delivery includes spiral or sequential tomotherapy[4, 10] and volumetric modulated arc therapy (VMAT). The tomotherapy technique uses specialized binary multi-leaf collimators (MLCs), with only open and closed positions. VMAT requires dynamic MLC delivery along with continuously variable or quantized dose rate and gantry motion[11-13]. IMRT dose distributions are typically much more heterogeneous than those of 3-dimensional (3D) plans, employing complex fields that incorporate different degrees of modulation. Since the inception of IMRT, procedures for the delivery system and patient-specific IMRT plan QA have emerged[10] based on measurement and calculation techniques including independent monitor unit (MU) calculations for IMRT[14]. IMRT QA verification is an important process employed to check the accuracy of IMRT plan dose calculations and to detect clinically relevant errors in the radiation delivery, thereby ensuring the safety of patients and fidelity of treatment.

Several components must be addressed when clinically implementing IMRT such as acceptance and commissioning of the delivery device and treatment planning system (TPS), and development and implementation of a comprehensive IMRT QA program. An early American Association of Physicists in Medicine (AAPM) report on IMRT clinical implementation described delivery systems and pre-treatment QA[15]. In 2009, additional details concerning IMRT commissioning were addressed including tests and sample accuracy results for different IMRT planning and delivery systems[16]. In 2011, strengths and weaknesses of different dosimetric techniques and the acquisition of accurate data for commissioning patient-specific measurements were addressed[17]. A comprehensive White Paper on safety considerations in IMRT was also published, which clearly specified that pre-treatment validations were necessary[18] for patient safety, but the goal of the White Paper was not to explicitly address how that validation should be done. Other possibilities besides measurements have been published including, independent computer calculations, check-sum approaches, and log file analysis[14, 19-24]. Several professional organizations [AAPM, American College of Radiology (ACR), American Society for Radiation Oncology (ASTRO)][15, 16, 18, 25] have strongly recommended patient-specific IMRT QA be employed as part of the clinical IMRT process. A series of New York Times articles highlighted to the general public the hazard to patients when patient-specific IMRT QA was not performed after a change to a patient's treatment plan was made[26, 27].

While the value of patient-specific IMRT QA has been debated among physicists,[19, 28-30] especially whether computational methods can replace physical measurements, measurement-based patient-specific IMRT QA methods are widely used and are the core element of most IMRT QA

programs. In many centers, a QA measurement is routinely performed after a patient's IMRT plan is created and approved by the radiation oncologist. The treatment plan consisting of MLC leaf sequence files (or compensators) as a function of gantry angle and MUs from the patient's plan is computed on a homogeneous phantom to calculate dose in the QA measurement geometry. The physical phantom is irradiated under the same conditions to measure the dose. The calculations and measurements are compared and approved or rejected using the institution's criteria for agreement. If the agreement is deemed acceptable, then one infers that the delivered patient plan will be accurate within clinically acceptable tolerances. This phantom plan does not check the algorithm's management of heterogeneities, segmentation errors, or patient positioning errors. The details of methods used to evaluate the agreement between measured and calculated dose distributions (e.g., how a γ evaluation has been implemented), however, are often poorly understood by the medical physicists. For example, if the tolerance limits have not been thoroughly evaluated, it will be difficult to assess with any degree of confidence that these limits were clinically appropriate. To this end, the AAPM Therapy Physics Committee (TPC) formed Task Group (TG)-218 with the following charges:

1. To review literature and reports containing data on the achieved agreement between measurements and calculations for fixed-gantry IMRT, VMAT, and tomotherapy techniques.
2. To review commonly used measurement methods: composite of all beams using the actual treatment parameters, perpendicular composite, and perpendicular field-by-field. Discuss pros and cons of each method.
3. To review single-point (small-averaged volume), 1D and 2D analysis methodologies for absolute dose verification with ion-chamber and 2D detector arrays, mainly performed with dose differences comparison, distance-to-agreement (DTA) comparison between measured and calculated dose distributions, and a combination of these two metrics (γ method).
4. To investigate the dose difference/DTA and γ verification metrics, their use and vendor-implementation variability, including the choice of various parameters used to perform the IMRT QA analysis.

The objective of this report is to address these charges. The report provides recommendations on tolerance limits and measurement methods. Specifically, various measurement methods are reviewed and discussed. The dose difference/DTA and γ verification metrics are examined in-depth. Data on the expected or achievable agreement between measurements and calculations for different types of planning and delivery systems are reviewed. Results from a test suite developed by TG-218 to evaluate vendors' dose comparison software under well-regulated conditions are presented. Recommendations on the use of γ analysis routines and choice of tolerance limits are made

b. Uncertainties in the IMRT planning and delivery process

Acceptance criteria for initial machine and TPS commissioning are well established[31, 32]. The acceptance criteria for patient-specific IMRT QA, however, are more difficult to establish because of large variations among IMRT planning systems, delivery systems, and measurement tools[33-36]. There are many sources of errors in IMRT planning and delivery. In terms of treatment planning, the error sources can include modeling of the: MLC leaf ends, MLC tongue and groove effects, leaf/collimator transmission, collimators/MLC penumbra, compensator systems (scattering, beam hardening, alignment), output factors for small field sizes, head backscatter, and off-axis profiles. They can also include a selection of the dose calculation grid size and the use and modeling of heterogeneity corrections. Accurate IMRT TPS beam modeling is essential to reduce the uncertainties associated with the TPS planning process and consequently ensure good agreement between calculations and measurements when performing patient-specific verification QA[37, 38].

Spatial and dosimetric delivery system uncertainties also affect IMRT dose distribution delivery accuracy. These uncertainties include: MLC leaf position errors (random and systematic), MLC leaf speed acceleration/deceleration, gantry rotational stability, table motion stability, and beam stability (flatness, symmetry, output, dose rate, segments with low MUs). In addition, differences and limitations in the design of the MLC and accelerators, including the treatment head design, as well as the age of the accelerator/equipment can have an impact on the accuracy of IMRT delivery techniques[37, 38].

Another source of uncertainty among clinics using measurement-based patient-specific IMRT QA programs is the measurement and analysis tools used to interpret the QA results[39-41, 42, 43]. These software tools have several parameters that must be chosen to perform the analysis and the results can vary significantly depending on those choices. One example is the selection of whether to use global or local dose normalization to compare measured and calculated dose distributions.

c. Tolerances and action limits

Quality measures are employed to validate system performance[44, 45], such as IMRT QA. In this report, action limits are defined as the amount the quality measures are allowed to deviate without risking harm to the patient[35] as well as defining limit values for when clinical action is required. An example for IMRT QA is the decision not to treat the patient if the comparison between a point-dose measurement and the planned value exceeds a predefined acceptance criterion (e.g. ±5%). These limits will depend on whether one is using relative or absolute dose differences and/or explicitly excluding low dose regions from the analysis. Action limits should be set based on clinical judgment regarding the acceptability of a specific quality measure deviation.

Tolerance limits are defined as the boundaries within which a process is considered to be operating normally, that is, subject to only random errors. Results outside of the tolerance limits (or

trends moving rapidly towards these boundaries) provide an indication that a system is deviating from normal operation. The measurement results that lie outside the tolerance limits should be investigated to determine if their cause could be identified and fixed. The intent of this approach is to fix issues before they reach clinically unacceptable thresholds or action limits. When using action and tolerance limits, it is assumed that a careful commissioning process was followed. During the commissioning process, systematic errors should be identified and eliminated to the degree possible. This approach can also inform the choice of action limits when ambiguity exists about the clinical impact of exceeding action limits. This report will provide recommendations on one process-based method to choose these limits and accounts for both random errors and any residual systematic errors from the commissioning process.

II. Dose difference, DTA, γ analysis and verification metrics

Dose distributions are almost always represented as arrays of points, each defined by a location and dose value. The spacing between the points is the spatial resolution of the distribution and does not need to be the same in all spatial dimensions or locations. The spatial resolution of the dose distribution plays an important role in its display and evaluation. Coarse dose distributions may require some type of interpolation to display in an easily interpretable form, such as isodose lines or dose color washes. Dose distribution resolution also plays a role in dose distribution comparisons. Some comparison techniques are degraded by coarse resolution, so interpolation is employed.

This discussion of dose comparison techniques will assume that there are two distributions, termed a “reference” and an “evaluated” dose. The reference distribution is typically the one against which the evaluated distribution is being compared; although the specific mathematics and limitations of the comparison techniques may require these roles to be reversed. Some of the comparison techniques are invariant with respect to selection of the reference and evaluated distribution and some are not.

The process of dose comparison is part of a clinical workflow, in which the goal is to determine if the reference and evaluated dose distributions agree to within limits that are clinically relevant. The question of clinical relevance involves more than the dose itself, it also involves the dose gradients as well as dose errors resulting from spatial uncertainties. There is therefore a need to understand both the spatial and dosimetric uncertainties when conducting dose distribution comparisons. The spatial analog to the dose difference is the DTA, which in general refers to the distance between common features of the two dose distributions.

The positional accuracy specification of a steep dose gradient region should at least in part be based on the accuracy of patient positioning. Setting IMRT QA tolerances tighter than the ultimate clinical requirement will lead to unnecessary effort in attempting to reduce respective errors. Finally, in some cases the spatial uncertainty can be related simply to experimental error. Even if the user insists on having zero spatial error in a calculation, or if the calculation is being used for an extremely accurate dose

delivery process, the dose distribution measurements have some spatial uncertainty. Therefore, the DTA criterion can also be partly defined based on the measurement error.

a. Challenges for comparing dose distributions

On the surface, comparing dose distributions would appear to be straightforward. The distributions are no more than arrays of numbers, and a straightforward method for comparing is to calculate their numerical difference. However, in steep dose gradient regions, the dose difference is extremely sensitive to spatial misalignments. This sensitivity leads to large dose differences that easily exceed the dose difference criteria even for clinical irrelevant spatial misalignments.

A common method for comparing dose distributions is to overlay their contours. This technique provides a rapid and qualitative method for comparing the distributions. If the distributions agree exactly, the contours will overlay and if not, they will separate. The separation distance will depend on two factors; the difference in the doses and the local dose gradients. When the gradients are steep, contours move only slightly with changes in dose, so even large dose errors will correspond to small contour separations. Therefore, comparing contours in steep dose gradient regions provides little insight as to the dose differences because it takes very large differences to significantly move the lines. On the other hand, even small dose differences will move isodose lines far in low dose gradient regions. The only places where contour plots easily provide quantitative information are where isodose lines cross or superimpose. If the isodose lines are the same values, the distributions agree exactly at those locations. If two different isodose lines cross, for example the 50% line from one distribution and the 60% line from the other distribution, the dose difference is known at the crossing point. Otherwise, superimposed isodose contours provide little quantitative information.

Figure 1 shows an example of superimposed isodoses from Brulla-Gonzalez[46]. The isodoses represent measured dose distributions from radiochromic film and a 2D dosimeter. The correspondence between the two dose distributions is clear, and shows that there is no extensive disagreement, although disagreements of a few percent are difficult to determine. Figure 2 shows two dose distributions that greatly disagree[47]. The fact that they disagree is instantly clear from the vastly different isodose lines. In this case, additional quantitative dose analysis may be unnecessary.

One of the more challenging aspects of phantom-based dose distribution comparisons is the difference between the phantom and patient doses due to their differing geometries[48]. For measurements intended to evaluate planned dose accuracy, the comparison criteria would ideally be based on clinical organ-by-organ tolerances. For example, the tumor dose tolerance specification might be 3%, while a looser criterion of 10% might be acceptable to some muscle receiving 10 Gy and similarly with spatial tolerances. The spatial accuracy requirement might be 2 mm at the edge of the spinal cord, but 5 mm or more in the muscle. Because the measurements are conducted in phantoms, the planned fluence

distribution does not lead to the clinical dose distribution, even if there are no planning or delivery errors, simply because the dose deposited in the phantom has a different pattern than the dose deposited in the patient. Even if the spatial locations of the organs and tumor are superimposed on the phantom, the dose distributions will typically not conform to them because of the differences in the attenuation and scatter properties between the phantom and patient. Therefore, for evaluations such as patient-specific QA in phantom, we have conventionally relied on more generic acceptance criteria, based on overall goals of dosimetric and spatial accuracy in the domain in which we are able to measure.

b. Dose difference test

The dose difference test is the most straightforward test to understand and interpret. The dose difference at location ($\vec{r}$) is the numerical difference δ between the evaluated dose $D_e(\vec{r})$ and the reference dose $D_r(\vec{r})$ at that location. Mathematically the dose difference can be written as

$$\delta(\vec{r}) = D_e(\vec{r}) - D_r(\vec{r})$$

Note that the doses are sampled at the same positions. This analysis is straightforward when the dose distribution elements occupy the same locations (i.e. same grid resolution), but spatial interpolation is required when they do not. The dose difference test is invariant to within a sign with respect to the selection of the reference and evaluated distributions; all that happens if they are swapped is that the sign of the dose difference changes.

The dose difference test is excellent at providing the user insight as to the concordance between the two distributions in low dose gradient regions. In these regions, the dose changes slowly with location, and the dose difference is indicative of the disagreement between the two distributions independent of spatial uncertainties. Therefore, the spatial error tolerances, or DTA criteria, can be ignored. The opposite is true in regions of steep dose gradient. Even large dose errors will result in effectively moving the dose only a short distance in a steep dose gradient region. Therefore, if there is a non-zero DTA tolerance, it may not be violated even for a large dose difference.

Because small spatial shifts can lead to large dose differences, regions may fail a dose difference test but still be clinically acceptable because the shift is small. Since the clinical reality is that the spatial relationship between the planned and delivered dose almost always matters, the dose difference distribution alone is insufficient to determine if the dose distributions agree to within the clinical tolerances. Figure 3a shows an example of two dose distributions, a film-measured and calculated plane from an IMRT treatment plan, shown in gray-scale[49]. The dose difference is shown in Figure 3b and highlights two regions of discrepancy. The regions to the left and right correspond to opposite signs of the dose discrepancy and have differences of up to 15%, all within the steep dose gradient regions. Dose differences such as these are very large considering the 3% dose difference criterion selected by the investigators.

c. DTA test

Van Dyk et al[32] used the concept of distance-to-agreement (DTA) in 1993 for treatment planning QA. They specified that the distance between two dose distributions rather than the dose difference should be the acceptance criterion in steep dose gradient regions.

Harms et al[50] codified the Van Dyk et al[32] distance criteria into an algorithm. They defined the DTA for a point in the reference distribution as the closest location in the evaluated dose distribution with the same dose as the point in the reference distribution. Unlike the dose difference test, this algorithm required a search of the evaluated dose distribution to identify the closest distance to the point in the reference distribution that had the same dose as that point, equivalent to finding the closest distance of the evaluated distribution isodose line.

The DTA evaluation is ideal for determining the separation between steep dose gradient regions. However, as a comparison between dose distributions, it becomes oversensitive in low dose gradient regions, where even a small dose difference causes the relevant isodose line to move far from the reference point. Because of this, and because most dose distributions are dominated by low dose gradient regions, DTA distributions are difficult to interpret and by themselves not very useful.

Because the DTA test involves a search, the DTA value is not invariant to the selection of which distribution is selected as the reference. The reference distribution can have any resolution and dimensionality because the DTA is calculated point by point in the reference distribution, but the evaluated distribution usually has at least the same or greater resolution and dimensionality than the reference distribution.

d. Composite test

Given that the dose difference and DTA tests were complementary in their sensitivity to low and steep dose gradient regions, respectively, it made sense to combine the two so one could determine if a reference point had passed both the dose difference and DTA tests. Harms et al[50] termed this the composite test. A reference point was said to have passed if it passed either the dose difference or DTA tests. Only if it failed both tests was it determined that it had failed the comparison. While the composite test automatically managed both steep and low dose gradient regions, it suffered from being strictly a pass-fail test. If a point failed, the test did not indicate by how much.

e. γ test

The lack of insight as to the magnitude of failure with the composite test led Low et al[43, 51] to generalize the test. They treated the dose distribution comparison from a geometric perspective by evaluating the displacement between the reference and evaluated distributions. This evaluation was conducted

independently for each reference dose point. Similar to the DTA test, the dimensionality of the reference distribution could be a single point, while the evaluated distribution needed to be at least one dimensional.

The question of the displacement between dose distributions was complicated by the fact that there were n+1 degrees of freedom, where n referred to the spatial dimensionality of the comparison (e.g. a film plane contains two spatial and one dose dimension). The dose distribution could be thought of as an n-dimensional sheet within the n+1 dimensional space. The problem with determining a displacement in that space was that one of the axes was dose, while the others were distance. A displacement measurement was meaningless in this multiple-quantity space.

In order to allow the measurement to be defined, the dose and displacement scales were renormalized to be unitless by dividing them by the dose (ΔD) and DTA (Δd) criteria, respectively.

The displacement between two points, $\vec{r}_r$ and $\vec{r}_e$ in the reference and evaluated distributions, respectively, in the renormalized space was termed Γ,

$$\Gamma(\vec{r}_e, \vec{r}_r) = \sqrt{\frac{r^2(\vec{r}_e, \vec{r}_r)}{\Delta d^2} + \frac{\delta^2(\vec{r}_e, \vec{r}_r)}{\Delta D^2}} \quad (1)$$

where $r(\vec{r}_e, \vec{r}_r)$ was the distance between the reference and evaluated points, and $\delta(\vec{r}_e, \vec{r}_r)$ was the dose difference. The minimum displacement was defined as γ

$$\gamma(\vec{r}_r) = \min\{\Gamma(\vec{r}_e, \vec{r}_r)\} \forall \{\vec{r}_e\} \quad (2)$$

Values of γ between 0 and 1 indicated that the comparison passed with respect to the dose and distance criteria. Values greater than 1 indicated failure. Because γ was the displacement between the two distributions, γ was essentially the radius between the reference point and the evaluated distribution, so the pass/fail criteria were essentially the circle, sphere, or hypersphere in 1, 2, and 3 dimensional dose distribution comparisons, respectively. This was similar to the composite test, and in fact comparisons between the two tests showed little in the way of differences between points that passed and failed, although the γ test was shown to be slightly more forgiving than the composite test for clinical dose distributions[43].

While the γ test provided more than a pass/fail comparison, this did not itself allow a straightforward interpretation of the test's meaning. The most effective method of gaining an intuitive understanding of the test's performance was to examine how the test behaved in two extreme conditions, one of near zero dose gradient and one of steep dose gradients.

Figure 4 shows examples of 1D dose distributions with low and steep dose gradients. The γ calculation found the closest approach of the evaluated dose distribution to the reference dose distribution. With low dose gradients, the vector connecting the reference point to the evaluated distribution lay nearly parallel to the dose axis (Figure 4a). The dose difference test could be interpreted as the distance between the two distributions along the dose axis, which is what the γ test defaulted to in

the conditions of a zero dose gradient. Therefore, the γ test defaulted to the dose difference test (within the normalization of the dose difference criterion) in low dose gradient regions, precisely where the dose difference test was most useful.

Figure 4b shows the γ test under conditions of steep dose gradients. In this case, the γ vector lies nearly parallel to the distance axis, or distance axes for 2D and 3D dose distributions. The DTA test could be interpreted as the closest point where the evaluated dose distribution crossed the distance axes (with the origin placed at the reference point). Therefore, as the gradient increased, the γ test became the DTA test as normalized by the DTA criterion.

The main benefit of the γ comparison tool was that it automatically reduced the sensitivity of dose distribution comparisons in steep dose gradient regions. Figure 3c shows an example of γ for the dose distribution comparisons shown in Figure 3b.

f. Other tools

A number of IMRT QA evaluation tools have been developed and reported in the literature[52-54]. The gradient compensation method was developed by Moran et al[52]. They computed the local dose gradient for each point in the dose distribution. A user-selected distance parameter was chosen to allow for geometric uncertainties due, for example to experimental error or calculation grid resolution. The dose gradient at each point was multiplied by this distance parameter to yield a dose value corresponding to the resulting uncertainty in dose due to the spatial uncertainty. Dose differences in excess of this uncertainty would be displayed and analyzed. The gradient compensation method would remove dose differences that might be due to the spatial uncertainty. Presumably, the remaining differences would not be due to spatial errors and the physicist could evaluate the magnitude and clinical relevance of those errors.

The normalized agreement test (NAT) values and NAT index were defined by Childress and Rosen[54]. The NAT index represented the average deviation from the percent dose difference (ΔDm) and DTA criterion (Δd_m) for every pixel calculated, ignoring measurement areas having errors less than a set criterion. They developed an algorithm that started with computing the dose difference and DTA values. If the dose difference for a point was less than the criterion, the NAT value was set to zero. If the DTA value was less than the criterion, the NAT was set to zero. If the calculated dose was less than 75% of the maximum calculated dose, the pixel was assumed to be outside the PTV, and if the measured dose was less than the computed dose, the error was assumed to have no biological significance so the NAT was again set to zero. If, on the other hand, the measured dose was greater than the computed dose, or if the percent dose was greater than 75%, the NAT value was computed as D_{scale} *(δ-1)*, where δ was the lesser of the ratios $ABS(\Delta D/\Delta D_m)$ or $\Delta d/\Delta d_m$, and D_{scale} was the greater of the computed or measured dose at the pixel of interest divided by the maximum computed dose.

Bakai et al[53] developed a dose distribution comparison tool based on gradient-dependent local acceptance thresholds. The method took into account the local dose gradients of a reference dose distribution for evaluation of misalignment and collimation errors. These contributed to the maximum tolerable dose error at each evaluation point to which the local dose differences between comparison and reference data were compared. They identified two weaknesses of the γ test that they were addressing. First, that with an exhaustive search, the γ tool would take a considerable amount of time to calculate, especially for 3D dose distributions. Second, interpolation would be required if the dose distribution spacing was insufficiently fine. They concluded that the search process inherent in the γ evaluation should be avoided and so they defined an alternative process.

First, the dose axis was rescaled to units of distance by multiplying the dose by the ratio of the DTA to dose difference criteria. Without interpolation, for relatively large grid spacings and steep dose gradients, the tool overestimated the value of γ. The difference between the evaluated and reference doses was divided by a quantity called τ' which was related to the local dose gradient, resulting in the dose distribution comparison tool χ. They compared dose distributions calculated using the χ and γ tools and showed both methods gave essentially the same results, however, the calculation of χ was more efficient than the calculation of γ.

g. Practical considerations

Users of the γ tool should understand its performance in some detail. While the mechanics of the calculation are relatively straightforward, there are operational details that can reduce its effectiveness and accuracy.

1. Normalization

Normalization plays a critical role in the interpretation of dose comparison results. The dose difference criterion is a case in point. The dose difference criterion is typically described as the percentage of the maximum dose for one or both of the dose distributions being compared (global normalization), or the percentage of the prescription dose. It also can be described as a local dose percentage (local normalization). Specifically in global normalization, the dose difference between any measured and calculated dose point pair is normalized using the same value for all point pairs, often the maximum planned dose point. On the other hand, in local normalization, the dose difference for all point pairs is normalized to the planned dose at the local point. Selecting the local normalization allows one to have the same relative tolerances in the target and critical structure volumes. However, it will also cause the low dose regions to have unrealistic dose accuracy requirements.

The process of selecting the normalization involves compromises that are based in part on the differences between the patient and phantom geometries. Ideally, one would like to have clinically relevant criteria. The target volumes will receive nearly homogeneous doses in composite delivery, which are also near the maximum dose. Criteria for dose accuracy within the target volume can be straightforwardly defined, say at 3%, because the choice of using global vs. local normalization will make little difference when the target is concerned. However, the phantom and target shapes are different, so the dose level and dose heterogeneity in the phantom measurement will be different than the calculation. One way to manage the difference in target shapes is to reconstruct the dose in the patient using the measured data and thus determine dose delivery and calculation errors in the patient geometry (see section IV.e). This allows the γ passing thresholds to be clinically guided.

The problem is compounded when determining critical structure dose tolerances. The dose difference criterion would ideally be customized for each organ and the dose level found in that organ. For example, the dose tolerance in the spinal cord for a predicted cord dose of 45 Gy could be tighter than the tolerance when the expected cord dose was 20 Gy. This would allow the physicist to select clinically relevant criteria so they would detect only clinically relevant errors. However, changes in dose gradients and absolute dose distribution differences between the measurement and calculation within the organ make the process of selecting organ-specific dose difference criteria more difficult. Also, such a comparison between the measurement and calculation is only possible for composite type delivery. Furthermore, customizing the dose difference criterion for each organ with the commonly used IMRT QA methods is currently not feasible, especially with the uncertainties from the calculated beam penumbra near inhomogeneities.

Like the dose difference criterion, the DTA criteria can also be varied as desired. The concept of the allowance of spatial dose errors comes from the reality that there are always spatial uncertainties in patient setup and in internal tumor and organ positions. Like dose errors, the DTA should be a function of the clinical necessity of placing steep dose gradients. These can be functions of both organ and dose. For example, the DTA criterion for the dose gradient near the spinal cord might be tighter than the DTA criterion for non-critical normal tissues. Steep dose gradients often exist in such tissues for fixed-gantry IMRT due to the entry or exit fluence, yet the absolute spatial accuracy of those gradients might not be clinically important. Discovering where these dose gradient errors are going to be located in the patient with the commonly used IMRT QA methods is difficult.

2. Spatial resolution

Spatial resolution has a great impact on the accuracy of comparing dose distributions. As shown in Low et al,[43] without interpolation, the spatial resolution of the evaluated distribution impacts the accuracy of the γ calculation. Figures 5a and 5b show representations of a 1D dose distribution in a steep dose

gradient region. The distribution in Figure 5a is drawn such that the evaluated dose point closest to the reference point is closer than any interpolated points. Figure 5b differs from 5a in that the closest evaluated dose point lies farther than the interpolated evaluated dose distribution. Without interpolation, the determination of γ in this case would be greater than the closest distance of the interpolated distribution. The error in the γ calculation is a function of the local dose gradient, the spacing between evaluated dose points, and the DTA criterion. Coincidentally, in clinical radiation therapy, the spacing of measured or calculated doses is typically similar to the DTA criterion. For example, many dose distribution calculations are conducted on $3\times3\times3$ mm^3 grids, while the DTA criterion is often 3 mm. When the dose spacing and DTA criteria are similar, the calculated error in γ in steep dose gradients is large if no interpolation is used. Assuming the dose difference is 0 in the γ calculation, the maximum error in γ is roughly the ratio of the spacing to the DTA criterion, so for example γ will be accurate to within 0.33 in a steep dose gradient region if the spacing resolution ratio is 3:1. Therefore, a good rule of thumb is that the resolution of the evaluated dose should be no greater than 1/3 the DTA criterion. Interpolation can be used to satisfy this rule. Not following this rule may degrade dose distributions comparison accuracy.

This problem with spatial resolution had been assumed to be significant only in the case of steep dose gradients. However, as shown in Figure 5c, when the evaluated dose distribution has a low dose gradient region, similar errors can occur if the location of one of the evaluated points does not coincide with the reference point. Interpolation can reduce these errors, but decrease the γ calculation speed performance. Ju et al[55] described an algorithm that reduced the interpolation to a geometric problem. They proposed that each evaluated dose distribution line, surface, or hypersurface (for 1D, 2D, and 3D dose distributions) be subdivided into simplexes, namely the representations of line or surface elements with the smallest possible number of vertices. These constituted line segments, triangles, and tetrahedra. When this subdivision was completed, the closest distance between the reference point and each simplex could easily be computed using only matrix inversion. This sped up the process of calculating γ considerably as well as caused each calculation to be implicitly linearly interpolated, removing the errors caused by coarse evaluated dose distribution spacing.

3. Interpretation

Both the dose difference and DTA criteria can also be used as universal error bars when comparing two measurements. The DTA criterion can be used to allow for measurement positioning error, for example positioning a phantom to the lasers, and the ability to position film within a phantom.

While the ideal method for setting the dose difference and DTA criteria would be organ and dose specific and based on heterogeneous phantoms mimicking patient geometries, this is not currently

practical. One simple way of managing the current standard of practice is to apply dose thresholding for the γ analysis. The doses smaller than a user-selected value are not included in the γ or other analyses. This allows the user to focus on greater, clinically relevant doses.

Two dimensional dose distributions can have thousands of points being evaluated (a typical 20×20 cm^2 film scanned with a 0.5 mm^2 resolution yields 160,000 points). There are a few ways of reviewing the resulting comparison data. The dose difference and γ distributions can be presented as iso-maps or colorwashes. The γ distribution can also be summarized by a γ histogram.

The pass/fail criteria are selected in advance of the γ calculation, but care should be taken when reviewing the γ results. In many cases, some points will not pass the γ test and using the exact number that can pass as the sole determinate of quality is challenging and may not yield clinically useful results. A point that fails may not indicate a severe problem. Most of the time, the γ criteria are fairly strict. One of the advantages of the γ distribution is that it provides an indication not only that the point failed, but by how much. A γ value of 1.01 is indicative of a failure but for a 3% dose difference and 3 mm DTA, a γ value of 1.01 could indicate a dose difference of 3.03% in a low dose gradient region or a DTA of nearly 3.03 mm in a steep dose gradient region. Both of these are examples of failures, but failures that exceed tolerances by 0.03% and 0.03mm in the low and steep dose gradient regions, respectively. A point that fails the γ test by 0.03% or 0.03mm needs to be considered differently than a point that fails by a substantially wider margin. Therefore, the user should look not only at the percentage of points that fail, but also make an analysis of the maximum γ value, the percentage of points that exceed a γ value of, say 1.5, the γ histogram and possibly other statistical values. Examining γ calculations with different dose difference/DTA criteria, e.g. 4%/3mm, 3%/3mm, 3%/2mm, 2%/3mm, and 2%/2mm…etc, can also help the user understand the sources of discrepancies and their impact.

Without dose comparison statistics specific to tumor or organ systems, one needs to remember that any histogram or statistical analysis neglects the spatial information. It should be noted that the γ test could underestimate the clinical consequences of certain dose delivery errors when the entire dose distribution is evaluated together. This was demonstrated in the work of Nelms et al[48] where the specific dose delivery error they evaluated caused the high dose regions to be delivered outside of the γ tolerance while the dose delivery accuracy for lower doses was within tolerance. This correlation of dose error and dose meant that the errors were clustered in the high dose region, which corresponded to the tumor. A comparison of the dose-volume histograms clearly showed a systematic discrepancy between the dose distributions in the target, while the γ statistics (relative numbers of points with $\gamma > 1$) appeared to be clinically acceptable. Another reason for the acceptable γ statistics was that the ratio of tumor to irradiated volumes was very small, so even if the tumor is incorrectly irradiated, the fraction of points that failed the γ test might also be small. This highlights the fact that: 1) γ statistics should be provided in a

structure by structure basis and 2) the γ distribution should be reviewed rather than relying only on distilled statistical evaluations such as γ histograms.

III. Review of measurement methods

Several methods can be used to perform pre-treatment patient-specific IMRT verification QA measurements as shown in Figure 6. The most common methods in clinical practice are: a) true composite (TC), b) perpendicular field-by-field (PFF), and c) perpendicular composite (PC). For each of these methods, the patient's plan is recomputed onto a phantom that exists both physically and within the planning system. The TPS calculates the dose in the phantom (hybrid plan) in the same geometry as for the subsequent measurements[56]. In a survey on planar IMRT QA analysis with 2D diode array devices, Nelms and Simon[57] found that 64% respondents reported using the field-by-field method and 58% performed absolute dose analysis. The survey results showed that 76.3% of the clinics used 3%, 2.9% used 4%, and 15.1% used 5% dose difference limits. In addition, it was noted that 82.7% used 3mm, 5.0% used 4mm and 2.6% used 5mm DTA limits. They reported 34.5% used 0-5%, 36.7% used 5-10% and 28.8% used≥10% standard lower threshold dose limits. In the following sections, each method is described along with the types of data that are obtained. Further, the pros and cons of each method are discussed. For all methods described below, the recommendations of TG-120[17] can be utilized for dosimetric methods.

a. True composite (TC)

The TC method simulates the treatment delivery to the patient. The radiation beams are delivered to a stationary measurement device or phantom on the couch using the actual treatment parameters for the patient, including MUs, gantry, collimator, couch angles, jaws, and MLC leaf positions. The method has been used most often by physicists performing film dosimetry although more recently, diode and chamber array devices have been used. Typically with film dosimetry, an ion chamber (IC) is placed inside the phantom and irradiated along with one or more sheets of ReadyPack EDR2 (Eastman Kodak Company, Rochester, NY) or Gafchromic EBT film (International Specialty Products, Wayne, NJ), providing simultaneous measurements of absolute IC dose and relative planar doses [58-62] (Figure 6a). The measured IC reading can be converted to dose by taking the ratio of a reference field reading to a known dose (e.g., 10x10 cm^2 at reference depth) in phantom.

The film or detector array is usually positioned in a coronal orientation on the couch (Figure 6b) but can be in a sagittal orientation (Figure 6c) (or transverse plane for film) or a rotated plane. Because the recorded doses are from all the beams in the plan at their planned positions, the dose distribution mimics the dose distribution inside the patient, distorted and modified only by the difference between the

patient and phantom external contours and a lack of heterogeneities in the phantom. Within the film or detector array, uniform high dose regions will be present along with similar dose gradients and low dose regions that occur in the patient's plan.

Detector arrays designed for perpendicular irradiation have been used to integrate the dose during the TC irradiation. Additional phantom material surrounding the array has been used to obtain at least 5 cm depth in all directions. Since the 2D arrays are designed for perpendicular measurements, the array detector's radiation response is angularly dependent. This angular dependence is caused by beam attenuation from internal electronics, device encapsulation materials, diode detector packaging materials, and air cavities[63-65]. The diode array may have significant angular dependence within $\pm 10^0$ of the horizontal axis, primarily due to air cavities between detectors[64, 65]. The angular dependence may be smeared out when using beams from many angles, such as with VMAT delivery. However, caution should be taken when using 2D arrays for IMRT QA when more than 20% of the dose comes from the lateral direction[64]. Another limitation for diode or ionization chamber arrays comes into play for non-coplanar beams, which can irradiate the active electronics of the device for certain field sizes and beam angles. Compared to film dosimetry, these ion chamber/diode arrays have much lower spatial resolution. This becomes a disadvantage in measuring doses of very small tumor volumes or with steep dose gradients, as well as commissioning IMRT systems. It is important to note that with arrays, unlike films, an independent IC reading is not essential because the QA analysis can be performed in absolute dose mode.

The measurement plane for a film or array can be placed either inside the high dose volume or in a plane that samples doses received by a particular critical structure. A common position for film is immediately anterior or posterior to the ion chamber. The mean dose to the ion chamber volume (the chamber volume is contoured in the planning system phantom) as well as the 2D dose distribution in the same plane(s) as the film or array is calculated. The percent relative differences between the measured and calculated chamber doses are then compared to the acceptance criteria. The film dose image is registered to the 2D TPS dose image using pin pricks on the film or other fiducial marks which relate the film to the linac isocenter while the array dose image is aligned to the planning system dose image by deliberate positioning of the isocenter relative to the array origin. After registration, an isodose overlay and/or γ analysis is performed[8, 66]. If an IC is used, it is often placed at the isocenter when it lies in a uniform high dose region. If the isocenter lies in a non-uniform dose region, the IC can be placed away from the isocenter in a more uniform dose region. Note that an ion chamber reading alone without a 2D dose plane measurement is not sufficient for detecting errors other than at that single point[39].

Advantages and Disadvantages

There are three main advantages to the TC method. The first is that the measurement includes inaccuracies of the gantry, collimator, couch angles, and MLC leaf positions with gantry angle (gravity effects) as well as the attenuation of the couch top. The second advantage is that the resulting planar dose distribution is closely related to the dose that will be delivered to the patient, so that the relationship between the high dose region and organs at risk lying in the same plane can be assessed. Third, there is only one dose image to analyze (per plane of interest). The main disadvantage is that portions of many beams will not traverse the film or detector plane. This is particularly true if the film/detector plane is transverse and irradiated by only one pair of MLC leaves. Thus, not every part of every beam is sampled. However, detector devices designed to measure VMAT beams such as ArcCheck (Sun Nuclear Corporation, Melbourne, FL) or Delta4 (ScandiDos, Uppsala, Sweden) generally sample the entire beam area.

b. Perpendicular field-by-field (PFF)

In this method, the gantry is fixed at 0 degree (pointing down) for all beams and the collimator is fixed at the nominal angle (Figure 6d). PFF is used most often with diode or chamber arrays although film and EPIDs have been used as well. Gantry mounting fixtures are available for some arrays so that the actual gantry angle can be used during irradiation to include the gravity effects on the MLC leaves (Figure 6e). The TPS calculates the dose to the same plane as the measurement detector and that dose plane is registered to the measured dose image using pin pricks or other fiducial marks in the case of film or by aligning the array dose image to the planning system dose image by their common center in the case of 2D arrays. A comparison of the planned vs. measured dose is then performed for each field. These analyses can be performed in absolute dose mode so that an independent IC reading is not needed. Isodose and profile overlays are also used to compare the dose distributions. When an IC is used, similar to TC delivery, the chamber is typically placed at the isocenter in a uniform high dose region.

Advantages and Disadvantages

The advantage of the PFF method is that it samples every part of every field as the dose from each of the IMRT fields is delivered and analyzed separately. Field-by-field analysis may reveal some subtle delivery errors. It prevents the dose washout that can occur in a composite measurement geometry when under-dosing in some areas in part of one beam may be compensated by an increased dose in the same region by another beam. PFF may be more stringent because the dose distribution from each beam is so highly modulated, small differences in dose and its location can cause large differences in the analysis result. To a greater extent than for TC, the agreement between the calculations and measurements is more dependent on the normalization values for relative dose analysis or shifts from the initial registration (Table 1). In addition, the significance of a summation of discrete dose errors in each beam image, which

commonly occurs, is not generally known. In this situation, analysis results can be misleading as suggested by a number of studies which have found a poor to moderate correlation between field-by-field 3%/3mm DTA results and the actual measured to calculated 3D dose differences in the patient or phantom[48, 67-70]. While these findings may cast doubt on the value of these measurements, it emphasizes that the method and results should be carefully interpreted. It also emphasizes that clinical interpretation of QA failure results is a challenging process.

c. Perpendicular composite (PC)

The PC method is similar to the PFF method described above in (b), except that the dose is integrated for all the perpendicular fields, resulting in a single dose image for analysis (Figure 6d), and making the method faster than PFF. The same measuring equipment and analysis methods are used.

Advantages and Disadvantages

The advantage of the PC method is that all portions of each beam are incorporated in a single image. EPIDs can be used if each beam's dose image is acquired separately and then added together later. Using the EPID to obtain an integrated image for VMAT is considered PC. The disadvantage is that the method may mask some dose delivery errors, such as those in the scattered regions, and dose errors from any one beam, within the composite, may be obscured by the superposition of the other beam doses[71]. Further, for VMAT delivery the dose rate variation vs. gantry may be obscured using this method and errors caused by using uniform dose rate delivery vs. non-uniform dose rate delivery was not caught using PC. The dose distribution is unrelated to that in the patient and can be difficult to interpret clinically, and may not have large areas of uniform dose despite the compositing effect.

d. Method selection

It is difficult to choose one of the three methods as clinically preferable for IMRT QA. The clinically preferred method would be the one that best identifies significant differences between delivered and calculated doses. The TC method has the advantage of providing a 2D dose slice out of the 3D dose distribution while the PFF method only provides information on each beam with limited ability to interpret the 3D dose consequence of regional errors in a single beam. Of course, if the per-field γ result is perfect for every beam (100% passing) using a specific acceptance criteria, then it is highly likely that the delivered dose will closely match the calculated dose, although Nelms et al showed that even that scenario is not always true[48]. The PC method has the distinct disadvantage of potentially masking errors due to the summation. Regardless of the advantages of any particular 2D based method, none of the methods discussed provides information about the 3D dose deviation in the patient.

When comparing the γ failure rate results for the same plan tested by each of the three methods, little if any correlation has been observed. For example, AAPM TG-119[16] reported confidence limits as baseline values for IMRT commissioning based on the measurement results for a set of test cases from several institutions. The cases were designed to test overall IMRT planning and delivery accuracy. Using the same set of test plans, the TG-119 report showed a difference in the confidence limit of 12.4% for TC vs. 7% for PFF[16]. This means that care must be exercised when attempting to relate results from one clinic using one method to that of another clinic that employs a different method.

IV. Review of methodologies for absolute dose verification

A number of techniques can be used for absolute dose verification. Single-point measurements such as those obtained using individual ICs can be considered dose distributions of zero dimension. Radiographic and radiochromic films, diode arrays, chamber arrays, and EPID provide 2D distributions, and 3D dose distributions are measured within gel and other 3D dosimeters.

Absolute dose can be verified using point dose techniques and 2D measurements. Hybrid versions of 3D measurement techniques have recently become commercially available for clinical use. The AAPM TG-120 report thoroughly describes the tools and measurement techniques used for both IMRT commissioning and patient-specific measurements[17]. The methodologies are briefly reviewed below. The reader is directed to the TG-120 publication for more specific information[17]. Film dosimetry, including calibration, scanning, and other aspects are not discussed here; the reader is instead referred to the report from AAPM TG-69[72].

a. Single-point (small-averaged volume) method

The most basic measurement that can be made for patient-specific plan verification is a point dose measurement. Typically, cylindrical ICs and stylized QA phantoms are used for these measurements. Though commonly referred to as point measurements because a single number is reported, these measurements are actually small-volume measurements over the volume of the chamber. They can be used to verify that a certain MU setting results in the correct absolute dose. Verification of absolute dose can be for either the target or for critical structures, depending on the chamber placement in the QA phantom.

Chamber volumes may vary from 0.007cc to 0.6cc, and leakage current must be adequately taken into consideration[17, 73]. The chamber should be placed in a region of uniform dose and should have adequate spatial resolution. A good rule of thumb is that the maximum and minimum doses across the IC (as reported in the TPS dose statistics for the IC volume) should be within 5% of the mean chamber dose to minimize the effects of volume averaging over a gradient region. The size of the IC is not particularly critical and one can make arguments for using a 0.6cc chamber or using a smaller chamber size[74].

Specifically, the potential problem with small volume chambers is that to the extent there are dose gradients near the chamber, small positioning errors will cause large apparent measured dose differences relative to the treatment plan dose. The larger volume chamber will average these gradients and be less sensitive to positioning errors as long as the chamber volume has been correctly contoured in the TPS.

Clinically significant doses such as the dose in the PTV or in an OAR region should be measured. When possible, the measured dose should be compared to a planned dose to the volume of the chamber, instead of a point dose that corresponds to the middle of the active volume of the chamber as depicted in the QA plan. When a chamber is selected for patient-specific QA, a cross calibration technique can be used to determine the dose response of the chamber chosen[17, 75].

b. 2D methods

2D measurements can be made to obtain a more representative picture of the dose delivery distribution, corresponding to a planar dose. Commercially available dosimeters that fall in the 2D methods category include ionization chamber arrays, diode arrays, EPIDs, and film. They are typically used to measure the 2D dose distributions associated with a specific IMRT field. These detectors can provide relative dose comparisons against the QA plan. They can also be calibrated to report absolute dose measurements but caution should be exercised when using EPIDs and film as they are not ideal absolute dosimeters. Calibration and commissioning should be performed before use, following the vendor recommendations and any published guidelines. Planar dose distributions can be exported from the TPS at the plane of measurement and an absolute or relative dose comparison can be performed against the measurement results from the array. While 2D planar arrays are typically used for single field measurements, not for measuring the TC IMRT plan[17, 76], rotational devices have been used for measuring the TC VMAT plans and some arrays have been inserted in slab phantoms for rotational IMRT QA measurements.

Film can also be used as a 2D measurement device. Both radiographic (e.g. EDR2) and radiochromic (e.g. EBT2) films have been used. Care must be taken to accurately convert the optical density to dose from a film measurement[72, 77-79]. Commissioning of each film batch should include characterizing the dose response, sensitivity, and uniformity of the film. The amount of noise in the film should be considered when determining the dose to be delivered to the film. Radiographic film requires a dark room and a processor. Radiochromic film has the advantage of being less light sensitive than radiographic film, but still must be stored in a light-proof envelope and read at a particular time after exposure to account for the reverse fading effect. For plans delivering more than about 7 Gy per fraction, the MUs may have to be proportionally reduced to avoid EDR2 film saturation[62], while Gafchromic EBT2 films have shown to work for doses of 1 cGy – 10 Gy in the red color channel, and up to 40 Gy in the green color channel. Both types of film measurements can be normalized to a dose reading from an IC to avoid some of the uncertainties in obtaining absolute dose from film light transmission readings.

Although more difficult to use, film can measure TC planar dose distribution with a higher spatial resolution than detector arrays. Film may have the ability to reveal clinically-relevant differences between treatment plans and deliveries that might not be otherwise detected using only per-field measurements. However, detector arrays measure the doses at the individual detector locations more accurately than film due to the uncertainties associated with film processing and densitometry.

c. 3D methods

3D measurement methods, such as 3D gels[80] and PRESSAGE dosimeters[81, 82], are used to provide a TC measurement of the delivered IMRT/VMAT dose in a 3D high-resolution volume. 3D volume gel dosimeters are becoming more refined and have been used by several research groups in an investigational clinical setting[83, 84] but they are not yet widely available for patient-specific QA measurements[85]. Although the published studies show the versatility and the added value of gel polymers as true 3D dosimeters, there remain limitations with this technique, including the stability and manufacturing of the gel, the calibration of the system, the reading of the gel after it has been irradiated and the fact that the gel can only be used once. In addition to gel dosimeters, some newly developed detector arrays that have non-planar geometries are capable of measuring both per-field and composite dose for both IMRT and VMAT QA[86-90], although not in a true 3D, high spatial resolution volume.

d. Comparison of single-point, 2D and 3D methods

The strength of point dose measurements with ICs is that they can be used to verify the accuracy of the MU calculations conducted by the TPS because they measure the absolute dose rather than just the relative dose. These measured doses typically have smaller uncertainty than film or detector array measurements and can also be used when scaling relative 2D measurements. The limitation is that ICs measure the dose at only one point (actually a small volume-averaged region), and while this measured dose is compared against the treatment plan, it does not provide enough information to validate overall plan accuracy.

2D and 3D dose measurement methods give a more comprehensive picture of plan delivery than point dose measurements. Although initially the intent was to use the 2D arrays for per-field dose comparisons, most of the manufacturers have recently incorporated directional response corrections for the detectors. 2D arrays, although an improvement over the single-point method, are limited to a dose comparison in a plane and not a 3D volume. Because 2D or 3D measurements have the ability to show relative dose distributions, it can be tempting to ignore the absolute dose information they measure. When they are used in this manner, considerable differences between dose delivery and dose calculation may go undetected. Therefore, 2D and 3D arrays should always be calibrated and used to measure the absolute dose. The calibration methods and frequency should follow the recommendations from the

manufacturers. The array calibration frequency will also depend on the usage of the array device. A dose calibration measurement compared against a standard dose should be performed before each measurement session to factor the variation of the detector response and accelerator output into the QA measurement.

e. Reconstruction methods

The quantitative comparison between the measured dose in a phantom and the corresponding QA treatment plan has been traditionally based on indices such as the DTA, dose difference and the γ index. New developments in the QA devices and their associated analysis software modules allow the conversion of the measured dose distribution data to 3D absolute dose distributions in the patient anatomy. Thus, a more clinically relevant comparison between the clinical treatment plan, including dose distributions and DVHs for all delineated structures, can be made against the treatment plan reconstructed from the measurements.

There are several commercially available approaches for the 3D reconstruction methods:

i. Forward calculation algorithm

This method treats the measurement data as a fluence map. The software uses the CT data along with a forward calculation algorithm such as pencil beam or collapsed cone[91,92], to reconstruct the 3D dose in the patient. The measurement data can be from an EPID[93], 2D diode or 2D chamber arrays[91, 94, 95]. In essence, these implementations use the patient-specific beams that are then delivered to the detector, typically in the absence of a phantom or patient. The measured data is corrected for the response of the detector and is subsequently used as the input fluence map to a forward dose calculation in the patient anatomy. This operation requires an independent dose calculation platform[91,92] but can also be done using the TPS itself[41], thus removing any ambiguities in the analysis that may result from differences in dose calculation algorithms between the TPS and the independent calculation platform. However, using the TPS dose calculation algorithm for both calculations may mask out beam modeling errors or algorithm limitations for IMRT beams.

ii. Plan dose perturbation

Plan dose perturbation (PDP) uses the difference between the measured and TPS calculated dose in phantom to perturb the 3D TPS calculated patient dose and create a corrected 3D dose distribution in the patient geometry. This method does not require a forward calculation algorithm. It relies on the measurement to create the perturbation matrix for correcting the TPS generated plan[69, 70, 96-98]. These measured data can be obtained using 2D and 3D arrays.

iii. 3D reconstruction in phantom

A rotating phantom with an embedded 2D array was recently introduced commercially, such that the array is always perpendicular to the beam axis for VMAT deliveries. A 3D dose distribution reconstructed from measurements is used for comparison with the patient QA plan[99]. The reconstruction algorithm is based on percent depth dose curves entered to the system as part of the commissioning process. While the dose is reconstructed in 3D, it is done in the phantom and not in the patient CT data.

iv. Per-fraction patient transmission-based QA

This is an EPID based measurement method that collects data during patient treatment to verify delivery. Images are collected while the patient is treated, and are used to reconstruct dose in the patient for each treatment. The measured transmission fluence can be corrected (e.g. for the EPID response and for the scatter from patient) and backprojected[92, 100, 101]. The deconvolved and backprojected primary fluence can then be forward transported through the patient anatomy and the 3D dose can be calculated in the patient and compared against the treatment plan for each fraction.

V. Review of reported IMRT measured vs. calculated agreement

Patient-specific IMRT QA has been performed and analyzed in several modes: i) an absolute dose to a detector, measured in a high dose region with a low dose gradient[40], ii) absolute planar dose, either normalized to a locally calculated point dose, the global calculated maximum dose of the plane, or the maximum dose of the entire treatment plan[17], iii) a DTA measure[32, 50], iv) γ analysis[43, 51].

The reported IMRT QA agreement between measurements and calculations, including absolute dose agreement and γ passing rates for various tolerance limits, are summarized in Table 2. Table 2 shows absolute dose measurements with ICs agreeing with expected values to within 5% and 2D measurements having γ pass rates > 90% using 3%/3mm DTA (global normalization)[102, 103]. Furthermore, recent studies have reported nearly 100% passing rates for 3%/3mm[64, 104-106] and > 95% for 3%/2mm or 2%/2mm[105, 106] in moderately/complex modulated plans.

a. Delivery Methods

1. Fixed-gantry IMRT

With fixed-gantry IMRT delivery, early patient-specific IMRT QA methods were performed using single ICs combined with a few other selected point doses[40, 75] and/or one or two planar films[107]. With the availability of commercial IMRT QA devices, the elimination of radiographic film and film processors, and the need for time efficient methods, the majority of IMRT QA is currently conducted using 2D diode or ionization chamber array devices[76, 108-110]. If an electronic portal imaging device (EPID)[111-114] is used, then no additional dosimetry equipment may be needed.

Based on an early generation IMRT treatment planning system with a pencil beam calculation algorithm, Dong et al[75] reported that a mean percentage difference between the calculated and measured point doses for 1591 points of 751 clinical IMRT cases was 0.37% ± 1.7%, with a range of -4.5% to 9.5%. Irradiating a head and neck (HN) anthropomorphic phantom, 71 institutions out of 250 failed the initial credential test designed by Imaging and Radiation Oncology Core (IROC) Houston using the criteria of 7% for the TLD in the PTV and 4 mm DTA for the film in the high dose gradient area between the PTV and the OAR[33]. A recent update of the IROC HN phantom irradiation study reported the results from 1139 irradiations by 763 institutions from 2001 to 2011[115]. 81.6% of the irradiations passed the criteria, 13.7% irradiations failed only the TLD criteria, 1.8% failed only the film criteria, and 2.9% failed both sets of criteria. Only 69% of the irradiations passed a narrowed TLD criterion of ±5%. The IROC data show that institutions who were sufficiently confident in their IMRT QA planning and delivery processes to participate in the RTOG IMRT trials, varied in their capability to deliver the phantom-planned dose distributions. More importantly, the data show the significance of performing patient-specific IMRT verification QA and emphasizes the importance of properly commissioning IMRT in both the treatment planning and delivery systems.

From a single institution experience, Both et al[102] analyzed 747 fluence maps generated from three commonly used treatment planning systems and measured with a 2D diode array using 6MV photon beams. They found for relative doses that the average passing rate using 3%/3mm with 10% dose threshold criteria for prostate and other cases was 99.3% ± 1.41% and for HN cases was 96.22% ± 2.89%. For absolute point doses, they found that the average percentage dose error for prostate and other cases was 0.419% ± 0.42% and for HN cases was 1.41% ± 1.1%. The differences between the prostate/other cases and HN cases were statistically significant.

2. Volumetric modulated arc radiotherapy

Partly due to its decreased delivery time compared against static gantry IMRT, VMAT is becoming the preferred technique for IMRT delivery. In Teke's[116] study, ten clinically acceptable VMAT treatment plans were calculated in a phantom using a Monte Carlo (MC) dose calculation algorithm and actual delivery log files. Measurements using a Farmer ionization chamber (PTW-Freiburg, Freiburg, Germany) with an active volume of 0.6cc agreed with both MC and TPS calculations to within 2.1%. Analyzing the detailed machine log files, they also confirmed that leaf position errors were less than 1 mm for 94% of the time and there were no leaf errors greater than 2.5 mm. The mean standard deviations (SDs) in MU and gantry angle were 0.052 MU and 0.355°, respectively, for the ten cases they analyzed. This study demonstrated that accurate VMAT delivery and stable machine performance were achievable. As a result, expectation of good agreement between predicted and measured dose for a VMAT delivery is warranted.

Many investigators have reported on experimental validation of VMAT delivery using various 1D, 2D and 3D dosimeters, including a helical diode array, a biplanar diode array and an ionization chamber array[64, 89, 117]. VMAT delivery is available on commercial gantry-based linear accelerators, with the properly installed software and hardware. The key for accurate delivery is precise synchronization of MLC motion, gantry motion, and dose rate adjustment (when employed). During treatment planning, VMAT plans are discretized into many static beams, varying from 2 to 7 degrees apart. Therefore, delivery accuracy and thus VMAT QA results may depend on discretization resolution and plan complexity. Feygelman et al[118] reported that when calculated using 4º spacing, the overall mean dose errors at a single position measured using an ionization chamber were 0.5%±1.4% and −0.3%±1.4% for the PTV and OAR, respectively, using the TG-119 phantom geometries. The γ passing rate (3%/3mm), measured for absolute dose with a biplanar diode array, was 98.2%±1.6% (range 94.5%–99.9%). Using 6º control point spacing, highly modulated plans exhibited large dosimetric errors [e.g. γ(3%/3mm) passing rates below 90% and ion chamber point dose errors of 6–12%], while simple cases had acceptable results.

Mancuso et al[105] performed a systematic comparison of fixed-gantry IMRT and VMAT patient-specific measurements using the TG-119 phantom geometries. They reported 2%/2mm and 3%/3mm average γ passing rates of > 97% and > 98%, respectively, for both IMRT and VMAT plans with no statistically significant differences between the two modalities. They suggested that it was appropriate to use fixed-beam IMRT action levels for VMAT.

3. Flattening filter free IMRT

Flattening filters were designed to create homogeneous dose distributions for conformal treatment plans. The flattening filter significantly attenuates the dose rate and adds scattered dose. For IMRT, inverse planning methods are used to generate non-flat beams for treatment planning. The increased use of hypofractionated treatments has resulted in the development of flattening filter free beams to increase the dose rate and consequently decrease treatment time, and to reduce the scattered radiation that increases the total body dose. Inverse planning methods are used for planning with these beams. Lang et al[104] presented pre-treatment QA data for 224 cases from four centers, measured with different verification devices to assess the reliability of flattening filter free beam delivery for IMRT and VMAT techniques. They found excellent agreement between dose calculation and dose delivery for these beams, with an average passing rate of 99.3%±1.1% for IMRT and 98.8%±1.1% for VMAT using γ tolerance limits of 3%/3mm. For 52 of the cases, a dose verification at a single position was performed with an IC, either a PinPoint chamber or a Farmer chamber, with a mean dose deviation of only 0.34%. They found that the passing rate was independent of the maximum dose rate used during irradiation of the arc. However,

during irradiation of the arc, the passing rate decreased with increasing ratio of the MU to the dose per fraction, indicating that highly modulated plans had slightly worse QA results.

4. Tomotherapy

With Tomotherapy delivery (Accuray, Sunnyvale, CA), the dose delivery precision relies on synchronization of the dynamic components including gantry, couch and MLC[119, 120]. Broggi et al[121] reported that, over a two-year period and based on monthly checks, gantry-couch synchronization was better than 2 mm, and gantry-MLC synchronization was within 1°. For the same period, dosimetric tests showed an average rotational output variation of −0.0±1% (1 SD). It was noteworthy that several groups reported a dose rate or output variation of 1%-2% during a treatment day or an IMRT QA session[65, 122, 123]. Consequently, the machine output fluctuations were likely inherent in IMRT QA deliveries, which should be considered when analyzing the IMRT QA measurement data.

Phantom based dosimetric measurements are recommended by AAPM TG-148[119] for both commissioning and clinical use. According to TG-148, an ionization chamber measurement is expected to agree with calculations to within 3%; for planar dosimetry using criteria of 3%/3mm DTA, a γ passing rate of at least 90% can be achieved. Care needs to be taken with regard to the angular dependence due to Tomotherapy's rotational dose delivery. Commercial electronic array dosimeters, made of either ionization chamber or diodes, have an anisotropic dose response as a function of beam incident angle[63, 65, 87]. Geurts et al reported their clinic's experience with Tomotherapy QA using a biplanar diode array dosimeter for over 250 clinical cases including breast, HN, colorectal, and prostate cancers[123]. For γ criteria of 3%/3mm, they reported an average of 97.5% (range 90.0%–100%) agreement between calculations and measurements, suggesting both the excellent measurement capability of advanced dosimeters and the delivery accuracy of Tomotherapy units. In addition, a significant positive correlation between IMRT QA and daily output QA was found, whereas there was no correlation between the γ passing rates with treatment variables such as PTV volume, fractional dose, field size, modulation factor, pitch, or off axis distance.

b. Considerations when using the γ test passing rates for evaluation

In clinical practice, many centers use the γ index[51] to avoid the oversensitivity of dose difference and DTA tests to steep and low dose gradient regions, respectively, as we discussed in section II. Some investigators[48, 68], however, have highlighted situations where the γ index is insensitive in detecting dose errors using single field measurements and certain delivery errors such as incorrect leaf positions.

Yan et al[124] introduced random and systematic MLC leaf position errors. With 3%/3mm criteria and a 90% passing rate, planar measurements using a 2D diode array with 7 mm detector spacing were

only able to detect leaf position errors greater than 2 mm. Mu et al[125] analyzed twelve simple direct machine parameter optimization (DMPO) based plans with the total number of segments limited to 50 and five complex plans generated using conventional two-step optimization with ≥ 100 segments. For a 1 mm systematic error, the average changes in D95% were 4% for simple plans versus 8% for complex plans. The average changes in D0.1cc of the spinal cord and brain stem were 4% in simple plans versus 12% in complex plans. They concluded that for induced systematic MLC leaf position errors of 1 mm, delivery accuracy of HN treatment plans could be affected, especially for highly modulated plans.

Kruse investigated the sensitivity of the fraction of points that failed the γ test with 2%/2 mm and 3%/3mm criteria using EPID and ionization chamber array measurements[67]. He used three HN treatment plans and created a second set by adjusting the dose constraints to create highly modulated delivery sequences. The treatment plans were computed on a cylindrical phantom, and EPID and ionization chamber array measurements were acquired and compared against calculation. They found that the highly modulated plans with aggressive constraints had many points that differed from the calculations by more than 4%, with one point differing by 10.6%. Using the ionization chamber array, the fraction of points that passed the 2%/2mm criteria were between 92.4% to 94.9% for the original plans and 86.8% to 98.3% for the highly modulated plans. Similar results were found using the 3%/3mm criteria with the same ionization chamber array. They concluded that the fraction of pixels passing the γ criteria from individually irradiated beams was a poor predictor of dosimetric accuracy for the tested criteria and detector methods.

Nelms et al created four types of beam modeling errors, including wrong MLC transmission factors and wrong beam penumbra[48]. The error-free plans were compared with error-induced plans. Using γ criteria of 3%/3mm, 2%/2mm, and 1%/1mm criteria, they found only weak to moderate correlations between conventional IMRT QA performance metrics and clinically relevant dose-volume histograms differences. Several recent studies demonstrated that common phantom-based IMRT QA techniques are not highly sensitive to some MLC leaf position errors[67, 124, 126-128] or to clinically meaningful errors[67, 48, 68-70, 127, 129].

Kry et al compared IROC Houston's IMRT head and neck phantoms results with those of in-house IMRT QA for 855 irradiations performed between 2003 and 2013[130]. The sensitivity and specificity of IMRT QA to detect unacceptable or acceptable plans were determined relative to the IROC Houston phantom results. Depending on how the IMRT QA results were interpreted, they showed IMRT QA results from institutions were poor in predicting a failing IROC Houston phantom result. The poor agreement between IMRT QA and the IROC Houston phantoms highlighted the inconsistency in the IMRT QA process across institutions. McKenzie et al investigated the performance of several IMRT QA devices in terms of their ability to correctly identify dosimetrically acceptable and unacceptable IMRT patient plans, as determined by IROC-designed multiple ion chamber phantom used

as the gold standard[131]. Using common clinical acceptance thresholds (γ criteria of 2%/2 mm, 3%/3 mm, and 5%/3 mm), they found that most IMRT QA devices performed very poorly in terms of identifying unacceptable plans.

These studies highlighted the importance of adopting tighter tolerances, performing a thorough analysis, having programs for routine QA of the accelerator and MLC, as well as developing new methods to supplement measurement-based patient-specific QA[70, 132]. In addition, these studies highlighted the challenges of using γ test passing rates for evaluating treatment plan acceptability and showed that clinical analysis of IMRT QA failure results is a challenging task. As discussed in section II.g.3, the γ test could underestimate the clinical consequences of certain dose delivery errors when the γ test is summarized in aggregate and when more detailed examination of the γ distribution is not conducted. The γ test passing rate summarization has no spatial sensitivity, similar to dose volume histograms, and the location and clustering of the failed points is not considered along with the passing rate. Also, field-by-field evaluation and dosimetric comparison might obfuscate clinically relevant dose errors and make correlating test results with clinical acceptability difficult. This is especially important because most comparisons are unable to reach 100% passing and so clinical criteria allow a fraction of points to fail the γ test. IMRT QA evaluation of plans that have large regions of low dose cause the fraction of failed points to appear small even when the area of failed points is large compared to the high dose regions, and thus resulting in the γ test passing easily.

c. Passing rates for given tolerances and corresponding action limits

A number of groups have suggested metrics to assess the clinical acceptability of IMRT QA verification plans. Table 3 shows confidence limits (CL), action limits (AL), tolerance limits (TL) and corresponding γ thresholds reported in the literature. Palta et al[35] proposed confidence limits and action levels for a range of dose regions for IMRT plan validation. The confidence limit was calculated as the sum of the absolute value of the mean difference and the SD of the differences multiplied by a factor of 1.96 (|mean deviation|+1.96 SD). The mean difference used in the calculation of confidence limit for all regions was expressed as a percentage of the prescribed dose according to the formula, $100\% \times (D_{calc}-D_{meas})/D_{prescribed}$. The confidence limit formula was based on the statistics of a normal distribution which expects that 95% of the measured points will fall within the confidence limit. The confidence limit values were derived from the results of an IMRT questionnaire from 30 institutions and reflected how the institutions judged the clinical significance of tolerance limits used for IMRT QA. The values were given as follows: i) confidence limit of ±10% or 2mm DTA and action level of ±15% or 3mm DTA for the high dose, high gradient region, ii) confidence limit of ±3% and action level of ±5% for the high dose, low gradient region, and iii) confidence limit of ±4% and action level of ±7% for the low dose, low gradient region.

Palta et al[35] suggested that IMRT treatment plans should not be used clinically if the measured and calculated doses differed by more than the action level values.

Using the confidence limit formalism of Palta et al[35], TG-119 reported confidence limits of ±4.5% for a high dose point in the PTV and ±4.7% for a low dose point in an avoidance structure, both measured using an IC. Confidence limits of ±12.4% and ±7%, respectively, were reported for 2D composite dose measurements made with film and arrays, corresponding to 87.6% and 93% γ passing rates (3%/3mm), respectively. Basran and Woo[133] examined the discrepancies between calculations and 2D diode array measurements for 115 IMRT cases. They reported acceptable tolerance limits of 3% overall, and 5% per-field, for absolute dose differences (independent of disease site). They recommended γ thresholds ≥ 95% for non-HN cases and ≥ 88% for HN cases using 3%/3mm. The ESTRO[38] report on Guidelines for the Verification of IMRT reported the experience of a number of European centers. For IC verification measurements, the report recommended a tolerance limit of 3% and an action limit of 5%.

Low and Dempsey[43] in 2003 proposed the need for fairly broad tolerances. They reported that for typical clinical use at the time, the fraction of points that exceed 3% and 3mm was often extensive, so they used 5% and 2–3mm as γ tolerance values for IMRT clinical evaluations. Childress et al[66] in 2005 analyzed 850 films resulting from IMRT plan verification and reported a "preferred" γ index tolerance criteria of 5% and 3mm.

A number of groups suggested using a combination of the mean γ value, maximum γ value exceeded by a given percentage of measurement points (e.g. 1%), and the fraction of γ values above one (P > 1) to analyze the γ distributions and make judgments on the agreement between measurements and calculation based on clinically driven criteria[66, 134-136]. For example, Stock et al[134] used a γ evaluation (3%/3mm) relative to maximum dose for 9 IMRT plans to decide the acceptability of IMRT verification QA. They considered a plan to meet their pass criteria if the average γ, maximum γ, and P > 1 were < 0.5, < 1.5, and 0-5%, respectively.

De Martin et al[137] analyzed the γ histograms (4%/3mm) for 57 HN IMRT plans using γ mean values, γ_{Δ} (where γ_{Δ} was defined as γ_{mean} + 1.5 SD(γ)), and the percentage of points with γ < 1, γ < 1.5 and γ > 2. They accepted the IMRT verification QA depending on the confidence limit values. They reported γ_{Δ} < 1 and confidence limits of 95.3%, 98.9% and 0.4% for the percentage of points with γ < 1, γ < 1.5 and γ > 2, respectively, for their newly installed linac.

Baily et al[103] compared measured dose planes with calculations for 79 HN and 25 prostate IMRT fields. Passing rates were calculated using dose difference/DTA, γ evaluation, and absolute dose comparison with both local and global normalization. They reported the passing rate spread for the individual prostate and HN fields with the greatest differences observed between global and local

normalization methods. For 2%/2mm and 3%/3mm (10% dose threshold), the prostate γ passing rates were 80.4% and 96.7% for global normalization and 66.3% and 90.8% for local normalization, respectively. On the other hand, the HN passing rates were 77.9% and 93.5% for global normalization and 50.5% and 70.6% for local normalization, respectively.

Carlone et al[138] investigated the use of receiver operating characteristic (ROC) methods in order to set tolerance limits for γ evaluations. They used a group of 17 prostate plans that was delivered as planned and a second group of 17 prostate plans that was modified by inducing random MLC position errors. The errors were normally distributed with $\sigma \sim \pm 0.5$, ± 1.0, ± 2.0, and ± 3.0 mm. A total of 68 modified plans were created and evaluated using five different γ criteria (5%/5mm, 4%/4mm, 3%/3mm, 2%/2mm, 1%/1mm). The dose threshold used during the γ evaluation process was not reported. All plans were delivered on a 2D detector array system with 7 mm detector spacing. Plots of the fraction of fields with a passing rate greater than a user-defined threshold ranging between 0% and 100% were plotted against pass rate percentage. Plots were generated for each combination of the five γ criteria and four σ. A total of 20 ROC curves were then generated by varying the pass rate threshold and computing for each point the fraction of failed modified plans, and the fraction of passed unmodified plans. ROC evaluation was performed by quantifying the fraction of modified plans reported as "fail" and unmodified plans reported as "pass". Optimal tolerance limits were derived by determining which criteria maximized sensitivity and specificity. Specifically, an optimal threshold was identified by the point on the area under the ROC curve closest to the point where sensitivity and specificity equaled 1.

While the γ criteria were able to achieve nearly 100% sensitivity/specificity in the detection of large random MLC errors ($\sigma > 3$ mm), sensitivity and specificity decreased for all γ criteria as the size of error to be detected decreased below 2 mm. The optimal passing threshold values for 2%/2mm were 78.9% (σ=3 mm), 84.6% (σ=2 mm), and 89.2% (σ=1 mm). The optimal passing threshold values for 3%/3mm were 92.9% (σ=3 mm), 96.5% (σ=2 mm), and 98.2% (σ=1 mm). Based on the ROC analysis, Carlone et al concluded that the predictive power of patient-specific QA was limited by the size of error to be detected for the IMRT QA equipment used in their center.

Bresciani et al[106] evaluated the variability of local and global analysis for Tomotherapy plans using 3%/3mm, 2%/2mm, and 1%/1mm, each with both local and global normalization. They reported mean passing rates for local (global) normalization of 93% (98%) for 3%/3mm, 84% (92%) for 2%/2mm, and 66% (61%) for 1%/1mm. They investigated the effect to excluding points below a 5% or 10% dose threshold and found that the choice between these thresholds did not affect the passing rate. They concluded that the variability in passing rates observed in their work showed the need to establish new agreement criteria that could be universal and comparable between institutions.

Pulliam et al performed 2D and 3D γ analyses for 50 IMRT plans by comparing collapse-cone convolution TPS (evaluated) and MC (reference) dose distributions[139]. The analysis was performed using a variety of dose difference (5%, 3%, 2%, and 1%) and DTA (5, 3, 2, and 1 mm) acceptance criteria, low-dose thresholds (5%, 10%, and 15%), and grid sizes (1.0, 1.5, and 3.0 mm). A small difference between 2D and 3D γ passing rates of 0.8% for 3%/3mm and 1.7% for 2%/2mm was reported with no low-dose threshold and a 1 mm grid size. 3D γ analysis produced better agreement than the corresponding 2D analysis. The additional degree of searching increased the percent of pixels passing γ by up to 2.9% in 3D analysis. The greatest difference between 2D and 3D γ results was caused by increasing the dose difference and DTA criteria.

VI. Vendor survey and algorithm testing

In order to better understand the commercial implementation of IMRT QA γ analysis software, TG-218 contacted the vendors and provided them with a set of questions displayed in Table 4 and test cases. The tests examined vendor implementations of the γ verification algorithm employing benchmark cases developed by TG-218.

a. Vendor survey

The vendor questionnaire in Table 4 included questions about the implementation of the γ analysis tool for IMRT QA in their software product. Eight vendors were contacted, and responses were recorded: 3DVH and SNC Patient for MapCHECK and ArcCHECK (Sun Nuclear Corporation, Melbourne, FL), Portal Dosimetry with EPID (Varian Medical Systems, Palo Alto, CA), RIT 113 (Radiological Imaging Technology, Inc, Colorado Springs, CO), IMSure (Standard Imaging Inc, Middleton, WI), Delta4 (ScandiDos, Uppsala, Sweden), VeriSoft with Seven29 2D array (PTW, Freiburg, Germany), COMPASS and OmniPro-IMRT with MatriXX (IBA dosimetry, Schwarzenbruck, Germany). The responses are summarized in Table 5. Two items that stand out in this table for improvement are that not every vendor is providing interpolation between points for dose maps, and not every vendor is providing dose difference/DTA analysis.

b. Testing of vendors algorithms

In addition to the difference in QA tool designs, such as the detector type and electronics, the implementation of the same IMRT QA analysis method, such as composite test and γ index, can vary from different vendors or even different products from the same vendors. The differences could come from the interpolation algorithm and resolution, DTA search points and radius, normalization, etc., as indicated in the survey results in Table 4. These differences can lead to different QA results.

There is no general specification for testing the dose analysis tools provided in the dosimetry software used for such comparisons. Two tests were provided to the vendors for the evaluation of their dose comparison software under well-regulated conditions that span clinically relevant doses and gradients, and to determine if they are producing the correct result. Specifications for reference and evaluated dose distributions were given to test the basic functionality of their dose comparison tools, including the dose difference and γ tools.

There were two types of distributions provided: 1) mathematically defined distributions and 2) distributions based on a clinical treatment plan. In both cases, the distributions were 2D with 0.5 mm resolution. The mathematically defined dose distribution for a circular field contained three distinct regions: firstly, a high and homogeneous dose area in the central region set to 200 cGy with flat gradient; secondly, a linear dose gradient (50% cm^{-1}) next to the central region, and thirdly a homogeneous low dose region set to 8 cGy surrounding the high dose and linear gradient regions. The evaluated dose distributions for the mathematically defined distribution was the reference distributions perturbed by modifying either the radiation dose gradient or the dose levels. Figure 7 shows examples of the reference and evaluated dose distributions. Also tested was a dose distribution acquired from a clinical IMRT treatment plan and the measured 2D film dose shown in Figure 8.

The method by Ju et al[55] was used as the benchmark (gold standard) for the commercial calculation evaluation. The percentage of points passing dose difference and DTA tolerances of all combinations of 2%, 3%, 2 mm, and 3 mm, were utilized. For the mathematically defined plans, two low dose thresholds of 4% and 5.5% were also utilized. The dose threshold to exclude low dose areas in the clinical plans was set to zero.

Five vendors responded with results. Table 6 shows the relative γ passing rate for these vendors for the mathematically defined plans. Table 7 shows the relative passing rates from clinical plan shown in Figure 8. For the mathematical and clinical tests, some vendors used the reference image as evaluated and the evaluated image as reference due to their software design. As these data show, the passing rates for the mathematically defined tests were calculated within 0.1% for two vendors, and most other vendors were within 6%. Vendor E had consistently greater passing rates than the other vendors or the gold standard for the mathematical tests. The clinical plan showed more variation than the mathematically defined tests. Comparing the film dose image to the TPS dose, agreement with the gold standard was within 6% across all vendors.

These results indicate that the vendors are not using a standardized approach to implementing the dose comparison tests. Given that the mathematically defined tests showed excellent albeit not perfect agreement, the discrepancies in the clinical case are likely due to the methods used to align the doses or handling of image resolutions. This highlights the fact that the user should understand how their vendor

has implemented the algorithm and should run benchmark test cases against their algorithm to evaluate its accuracy.

VII. Process-based tolerance and action limits

Although not explicitly mentioned in Section I.c., there is a human contribution to every IMRT QA measurement that is a source of variation. Another source of variation is the complexity of each IMRT case, for example, intensity modulation differences between head and neck and prostate IMRT cases. A process view of IMRT QA includes all sources of variation mentioned in Section I.c. as well as the human and case-specific issues. Accounting for all aspects of variation in IMRT QA can be achieved by setting process-based tolerance and action limits.

Action limits should set a minimum level of process performance such that IMRT QA measurements outside the action limits could result in a negative clinical impact for the patient. Tolerance limits refer to the range within which the IMRT QA process is considered to be unchanging. An out-of-control process serves as a warning that the process might be changing. If an IMRT QA measurement is outside the tolerance limits but within action limits, it is left up to the medical physicist to determine whether or not action should be taken.

Action limits come in two categories: 1) those that are universally defined and guided by outcomes data and expert consensus, and 2) those that are locally defined and guided by local experience. For any QA measurement, it is desirable to use universal action limits as these should be directly correlated with treatment outcome. This implies that there is some clinical evidence or a least consensus agreement among experts perhaps guided by summary statistics of retrospective data to inform the choice of those action limits. An example of universally defined action limits are those for treatment machine output because there is a direct correspondence between treatment outcome and output. Exceeding action limits that are locally defined do not necessarily result in harm to a patient when exceeded, but in the interest of good patient care, it is deemed best to keep process performance within those limits. Patient-specific IMRT QA is an example of action limits being set in this fashion. Locally defined action limits may vary from institution to institution or case-type to case-type since those limits are based on local equipment, processes, and case types as well as the experience of the local physicist.

Using methods from statistical process control[140-142], IMRT QA measurements can be used to determine action limits when universal action limits are not appropriate. Action limits determined in this fashion can be procedure-, equipment-, and site-specific for each individual institution and are calculated using the following equation[45],

$$\Delta A = \beta\sqrt{\sigma^2 + (\bar{x} - T)^2} \qquad (3)$$

where ΔA is the difference between the upper and lower action limits, typically written as $\pm A/2$. T is the process target value and σ^2 and $\bar{x}$ are the process variance and process mean, respectively. The constant β is a combination of two factors. One factor originates from the process capability metric, C_{pm}, as a cutoff for an acceptably performing process[143] and is combined with another factor that balances type I errors (rejecting the null hypothesis when it is true) and type II errors (not rejecting the null hypothesis when it is false) when using an IMRT QA measurement to make a decision about process performance. In using IMRT QA measurements to make decisions about process performance, the null hypothesis is that the process is unchanging. Current information suggests that β = 6.0 is an appropriate value to use[45] although this may be refined upon further research. Using equation (3) will likely result in action limits than are wider than what is currently accepted but should allow medical physicists to focus on problems with patient-specific IMRT QA that are likely to have identifiable causes. If the target, T, is known as in the example of patient-specific IMRT QA point dose difference (i.e., 0%) or γ passing rate (i.e., 100%), then the known target value should be used. If the target value is unknown or not defined, then the process average can be used as a best estimate of the target. This latter approach will have the effect of tightening the action limits compared to the former approach.

In this procedure, the process average, $\bar{x}$, and variance, σ^2, are calculated from the IMRT QA measurements over a time period when the process does not display out-of-control behavior. If the process is out-of-control, then one must identify and remove the reason for the out-of-control process behavior and continue monitoring the process until it displays a degree of control for about an additional 20 IMRT QA measurements. Then, control chart limits from an I-chart of individual IMRT QA measurements are used as the tolerance limits. The I-chart is a statistical tool that helps identify any IMRT QA measurement that display abnormal (out-of-control) process behavior. The I-chart has upper and lower limit lines (called control limits) and a center line that are calculated using the IMRT QA measurements[140-142]. Out-of-control process behavior is indicated when any one IMRT QA measurements is outside the upper or lower control limits on the I-chart. The IMRT QA measurements should be somewhat equally distributed above and below the center line. The center line, upper control limit, and lower control limits for an I-chart are calculated using the following equations:

$$\text{center line} = \frac{1}{n}\sum_1^n x \quad (4)$$

$$\text{upper control limit} = \text{center line} + 2.660 \cdot \overline{mR} \quad (5)$$

$$\text{lower control limit} = \text{center line} - 2.660 \cdot \overline{mR} \quad (6)$$

where x is an individual IMRT QA measurement, n is the total number of measurements, and $\overline{mR} = \frac{1}{n-1}\sum_{i=2}^n |x_i - x_{i-1}|$ is the moving range.

In this procedure, the control limits are used at the tolerance limits. Establishing process control is a key element of this procedure because a controlled process is an indication that the process is stable

and suitable for the purpose of IMRT QA. Using the proposed procedure requires a different view of QA such that measurements provide a description of the entire process (people + equipment + procedures) and not just the hardware and software equipment by itself. It is important to note tolerance limits will depend on plan complexity due to a greater case-to-case measurement variability. Therefore, it can be good practice to calculate tolerance limits separately for cases with high plan complexity and those with lower plan complexity, for example, for head and neck IMRT QA compared to prostate IMRT QA.

Two examples of process based tolerance and action limits are provided here to illustrate the procedure described in this section. The processes are VMAT and fixed-gantry IMRT QA and the γ passing rates with 3%/2mm and a 10% dose threshold are used as the QA measure using an ArcCHECK and MapCHECK device, respectively. Considering the VMAT QA example with 20 γ passing rates with an average of 96.66%, standard deviation of 1.739%, and moving range of 1.905%, then the action limit range following equation (3) is 22.6% which translates to an action limit of 100% – 22.6% / 2 = 88.7% (note that γ is bounded at 100% for the upper limit). As long as the process is not out-of-control, the control chart limit will be used as the tolerance limits calculated using equation (6) which is 96.66% – 2.660 · 1.905% = 91.6% for this example. For the fixed-gantry IMRT QA example, the 20 γ passing rates with an average of 95.92%, standard deviation of 3.388%, and moving range of 3.642%, then the action limit is 84.1% and the tolerance limit is 86.2%. In the case of γ passing rates, the upper tolerance (control) limit and action limit are bounded and equal to 100%.

The last step in the procedure is to compare the tolerance limits to the action limits. For example if the γ tolerance limits are lower than the action limits, then either the process needs to be fixed or the action limits lowered (i.e., use a larger value of β in equation (3)). Fixing the process may require new or modified equipment or training of personnel performing the IMRT QA measurements and analysis. Using this standardized procedure for setting action and tolerance limits will allow medical physicists to compare IMRT QA processes across institutions. The full procedure is summarized in Figure 9.

VIII. Recommendations

a. IMRT QA, tolerance limits, and action limits

Published data on IMRT QA results and the clinical experience of the group members were used to develop guidelines and provide recommendations on universal tolerance and action limits for IMRT verification QA using the γ method. This included in-depth literature review of IMRT QA results, analysis of widely used IMRT QA delivery and evaluation methods, and operational details that can improve the effectiveness and accuracy of the γ method. End-to-End QA verification tests for the IMRT TPS and IMRT delivery equipment, along with patient-specific verification QA are required to evaluate

the accuracy of radiation delivery to patients. Tolerances, action limits, and pass/fail criteria should be defined to evaluate the acceptability of IMRT QA verification plans.

We recommend the following terminology as it pertains to IMRT QA delivery methods (see Section III).

- Perpendicular field-by-field (PFF): the radiation beam is perpendicular to the plane of the measurement device. The device can be placed on the couch or attached to the gantry head. The dose from each of the IMRT beams is delivered and analyzed.
- Perpendicular composite (PC): the radiation beam is always perpendicular to the measurement device detector plane. The device can be placed on the couch or attached to the gantry head. The doses from all IMRT radiation beams are delivered and subsequently summed.
- True Composite (TC): all of the radiation beams are delivered to a stationary measurement device in a phantom placed on the couch using the actual treatment beam geometry for the patient, including MUs, gantry, collimator, couch angles, jaws, and MLC leaf positions. This method most closely simulates the treatment delivery to the patient.

We make the following recommendations for IMRT QA verification of dose distributions (fixed-gantry IMRT and rotational IMRT):

- IMRT QA measurements should be performed using the TC delivery method provided that the QA device has negligible angular dependence or the angular dependence is accurately accounted for in the vendor software.
- IMRT QA measurements should be performed using the PFF delivery method if the QA device is not suitable for TC measurements, or for TC verification error analysis.
- IMRT QA measurements should not be performed using the PC delivery method which is prone to masking delivery errors.
- Analysis of IMRT QA measurements and the corresponding treatment plan should be performed in absolute dose mode, not relative dose (the user should not normalize the dose to a point or region, i.e. relative dose mode).
- A dose calibration measurement compared against a standard dose should be performed before each measurement session to factor the variation of the detector response and accelerator output into the IMRT QA measurement.
- Global normalization should be used. Global normalization is deemed more clinically relevant than local normalization. The global normalization point should be selected whenever possible in a low gradient region with a value that is ≥ 90% of the maximum dose

in the plane of measurement. This will provide a more realistic measure of the comparison between the two dose distributions.

- Local normalization is more stringent than global normalization for routine IMRT QA. It can be used during the IMRT commissioning process and for troubleshooting IMRT QA.
- The dose threshold should be set to exclude low dose areas that have no or little clinical relevance but can significantly bias the analysis. An example is setting the threshold to 10% in a case where the critical structure dose tolerance exceeds 10% of the prescription dose. This allows the γ passing rate analysis to ignore the large area or volume of dose points that lie in very low dose regions which, if included, would tend to increase the passing rate when global normalization is used.

Tolerance and action limits (terms were defined in section I.c) are the foundation for a robust IMRT QA verification process. We make the following recommendations regarding tolerance limits and action limits for evaluating the IMRT QA analysis, including measurements. The limits are the same for PFF and TC delivery methods, and assume the tolerance and action limits are coincident with the goals of the treatment plan. If they are not, for example stereotactic radiosurgery (SRS) and stereotactic body radiotherapy (SBRT) cases, tighter tolerances should be considered. The following recommendations are for γ analysis using global normalization in absolute dose:

- Universal tolerance limits: the γ passing rate should be ≥ 95%, with 3%/2mm and a 10% dose threshold.
- Universal action limits: the γ passing rate should be ≥ 90%, with 3%/2mm and a 10% dose threshold.
    - If the plan fails this action limit, evaluate the γ failure distribution and determine if the failed points lie in regions where the dose differences are clinically irrelevant, in which case the plan may be clinically acceptable. If the γ failure points are distributed throughout the target or critical structures and are at dose levels that are clinically relevant, the plan should not be used and the medical physicist should follow the steps outlined in section (b) below. It may be necessary to review results with a different detector or different measurement geometry. For example, if the failure is seen with the TC delivery, a PFF analysis can be valuable to further explore the discrepancies between calculations and measurements.
- Equipment- and site-specific limits can be set following the method described in Section VII.

- If action limits are determined that are significantly lower than the universal action limits recommended above, then action should be taken as outlined in section (b) below to improve the IMRT QA process. From a process perspective, strict adherence to standardized procedures and equipment as well as additional training may also be necessary.

- Tighter criteria should be used, such as 2%/1mm or 1%/1mm to detect subtle regional errors and to discern if the errors are systematic for a specific treatment site or delivery machine.
- For IMRT QA performed with an IC and film, tolerance and action limits for the ion chamber measurement should be within ≤ 2% and ≤ 3%, respectively, and the film γ passing rate limits should be assessed as specified above. An IMRT treatment plan should not be used if the chamber measurement error or the γ passing rate exceeds the universal action limits.
- For any case with γ passing rate less than 100%,
  - the γ distribution should be carefully reviewed rather than relying *only* on distilled statistical evaluations,
  - review of γ results should not be limited to only the percentage of points that fail, but should include other relevant γ values (maximum, mean, minimum, median), as well as a histogram analysis.
  - an analysis of the maximum γ value and the percentage of points that exceed a γ value of 1.5 should be performed. For a 3%/2 mm, a γ value of 1.5 could indicate a dose difference of 4.5% in a low dose gradient region or a DTA of ~3.0 mm in a steep dose gradient region. Both of these are examples of failures, but failures that exceed tolerances by 1.5% and 1 mm in the low and steep dose gradient regions, respectively. Such information should be used to deduce clinical relevance whenever possible (e.g. cluster of failing points near or at the boundary of a tumor and critical structure).
- The IMRT treatment process should be monitored and thoroughly investigated if the γ passing rate is systematically lower than the tolerance limits or higher than the action limits. This includes reviewing dose differences directly without γ criteria or using local dose normalization and tighter dose difference and DTA criteria.
- γ statistics should be reviewed on a structure by structure basis if the user software allows for it. Vendors should include this feature in their future software development.
- Track γ passing rates across patients, especially for the same tumor sites, to look for systematic errors in the system.

- Vendors should implement a γ tracking feature across patients and for the same tumor sites in their future software development.
- Vendors should implement the simplex method for interpolation-free γ calculation and make the γ tool more practical and accurate[55].
- Whenever referring to a γ passing rate, always specify the dose difference (global or local) and DTA criteria and the dose threshold. Without these parameters, the passing rate is meaningless.
- Software tools that can provide a measure of the agreement between measured and calculated DVHs of patient structures are preferred over analysis in phantoms. DVH analysis can be used to evaluate the clinical relevance of the QA results, especially when the γ passing rate fails the tolerance limits or is inconsistent.

The accuracy of IMRT delivery can be affected by differences and limitations in the design of the MLC and accelerators among the different manufacturers, including the treatment head design, as well as the age of the accelerator/equipment. In addition, IMRT dosimetry QA equipment design, tumor sites (e.g. HN vs. prostate), complexity of the IMRT plans, uncertainties, inaccuracies and tolerances in the planning, delivery, and measurement may affect the IMRT QA verification results. For centers with IMRT QA results that are unable to meet the tolerance and action limit values recommended in this report, the center should perform a comprehensive analysis to determine the sources for these differences and remedy them. For example, the use of statistical process control methods can be useful in identifying the outlier cases failing to pass the tolerance limits for in-depth analysis[44, 45]. Also, it can be helpful to perform the TG-119 recommended tests and then compare to the published results or conduct independent tests using the IROC-Houston IMRT phantoms. The recommendations and guidelines provided in this report can be applied for any modulated treatment fields regardless of the system used to generate them. Finally, future research efforts should be focused on further improving the correlation between IMRT QA evaluation metrics and underlying planning or delivery errors.

b. Course of actions to check and evaluate failed or marginal IMRT QA results

IMRT QA verification results may not pass the tolerance limits and/or action limits provided in this report. When encountering a failure, the medical physicist should investigate the potential reasons for the IMRT QA failure following these recommendations.

The following, and/or the items considered necessary, should be checked with respect to each system in the order given:

1. *Setup and Beam*

- Phantom setup
- Correct plan version received by, and/or approved in, the record and verify (R&V) system.
- Correct QA plan generated, correct dose per fraction, correct delivery technique and data transferred from TPS to IMRT QA verification software
- Beam flatness, symmetry, and output on the day of the measurement
- Beam stability when delivering many segments with low MUs[144]
- Accuracy, stability, and calibration of the measurement device
- Detector size and inter-detector spacing with respect to the size of the IMRT fields, especially for SRS and SBRT cases
- Dose value of the global normalization point

2. *IMRT QA software*

- Correct reporting and handling of the plan and measured data
- Recheck values used for dose and DTA tolerance, dose threshold, and registration of the measured and calculated dose distributions

3. *MLC*

- Review results of periodic general patient-specific IMRT QA
- Leaf tolerances (speed, position, acceleration/deceleration)
- Tongue-and-groove effects which may require a measurement with a high resolution detector
- Beam profile data for both collimator- and MLC-defined fields
- Dynamic leaf-gap for rounded-leaf ends
- Intra- and inter-leaf transmission
- Jaw tracking positions (to minimize leaf transmission)

4. *Treatment Planning System (TPS)*

- The amount of modulation and the complexity of intensity patterns in the plan
- The total number of small segments in the plan, including small elongated fields
- The total number of monitor units which affects the total transmission dose and is related to plan complexity
- TPS modeling accuracy for small-fields, including output factors, profiles, and penumbra

- Characterization of the leaf-parameters in the TPS, including MLC transmission, gap and rounded leaf ends
- Minimum MU numbers
- The minimum segment (or beamlet) size in the TPS
- The dose calculation grid size in the TPS for non-Monte Carlo (MC) algorithms and the variance setting for MC algorithms
- The IMRT QA device CT numbers to electron density conversion
- Use of multiple carriage beams
- Gantry-angle spacing for VMAT delivery
- The above treatment planning items should be thoroughly checked as part of the IMRT TPS commissioning process. The commissioning should also include verification of IMRT plans for a full range of clinical cases, dose calculation algorithm and optimization parameters.

If the IMRT verification plan fails and there is more complex modulation than normal in your clinical practice, the planner should consider re-planning the IMRT case and attempt to achieve the planning objectives with less complex intensity patterns. In most systems, the planner can use tools to smooth the patterns during delivery without compromising plan quality.

IX. Conclusions

IMRT is becoming the standard of care for many disease sites. Recommendations were proposed in this report to ensure the appropriate implementation of patient-specific IMRT QA and the appropriate use of QA tools and methodologies. The report provides in-depth analysis on QA tools and practical recommendations on the use of the γ metric for IMRT verification. Vendor differences in γ analysis were presented. It demonstrated the need for the vendors to improve the implementation of the γ tool and remove this additional source of uncertainty from the IMRT QA process. Action limits using γ passing criteria were recommended to define the degree to which the quality measures are allowed to vary without risking harm to the patient and when clinical action is required. Tolerance limits using γ passing criteria were recommended to define the normal operating boundary of the IMRT QA verification process. The recommendations presented in the report provide suggested standards that can be implemented at the clinical level to evaluate the acceptability of patient-specific IMRT QA plans. They are intended to aid in the establishment of universal and comparable criteria among institutions.

Acknowledgements

We would like to thank Bernard Jones and Tai Dou for their contributions to this report. Also, we would like to thank the vendors for responding to the survey and for performing the tests.

REFERENCES

1. T. Bortfeld, J. Bürkelbach, R. Boesecke and W. Schlegel, "Methods of image reconstruction from projections applied to conformation radiotherapy," Physics in Medicine and Biology **35**, 1423 (1990).
2. S. Webb, "Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator," Physics in Medicine and Biology **36**, 1201 (1991).
3. A. Brahme, "Optimization of stationary and moving beam radiation therapy techniques," Radiotherapy and Oncology **12**, 129-140 (1988).
4. T. R. Mackie, T. Holmes, S. Swerdloff, P. Reckwerdt, J. O. Deasy, J. Yang, B. Paliwal and T. Kinsella, "Tomotherapy: a new concept for the delivery of dynamic conformal radiotherapy," Medical physics **20**, 1709 (1993).
5. J. Stein, T. Bortfeld, B. Dörschel and W. Schlegel, "Dynamic x-ray compensation for conformal radiotherapy by means of multi-leaf collimation," Radiotherapy and Oncology **32**, 163-173 (1994).
6. M. Carol, W. Grant III, D. Pavord, P. Eddy, H. Targovnik, B. Butler, S. Woo, J. Figura, V. Onufrey and R. Grossman, "Initial clinical experience with the Peacock intensity modulation of a 3-D conformal radiation therapy system," Stereotactic and Functional Neurosurgery **66**, 30-34 (1996).
7. C. C. Ling, C. Burman, C. S. Chui, G. J. Kutcher, S. A. Leibel, T. LoSasso, R. Mohan, T. Bortfeld, L. Reinstein and S. Spirou, "Conformal radiation treatment of prostate cancer using inversely-planned intensity-modulated photon beams produced with dynamic multileaf collimation," International journal of radiation oncology, biology, physics **35**, 721 (1996).
8. X. Wang, S. Spirou, T. LoSasso, J. Stein, C. S. Chui and R. Mohan, "Dosimetric verification of intensity-modulated fields," Medical physics **23**, 317 (1996).
9. C. S. Chui, S. Spirou and T. LoSasso, "Testing of dynamic multileaf collimation," Medical physics **23**, 635 (1996).
10. D. A. Low, K. Chao, S. Mutic, R. L. Gerber, C. A. Perez and J. A. Purdy, "Quality assurance of serial tomotherapy for head and neck patient treatments," International Journal of Radiation Oncology* Biology* Physics **42**, 681-692 (1998).

11. C. C. Ling, P. Zhang, Y. Archambault, J. Bocanek, G. Tang and T. Losasso, "Commissioning and quality assurance of RapidArc radiotherapy delivery system," Int J Radiat Oncol Biol Phys **72**, 575-581 (2008).
12. D. G. Kaurin, L. E. Sweeney, E. I. Marshall and S. Mahendra, "VMAT testing for an Elekta accelerator," J Appl Clin Med Phys **13**, 3725 (2012).
13. K. Otto, "Volumetric modulated arc therapy: IMRT in a single gantry arc," Medical Physics **35**, 310-317 (2008).
14. L. Xing, Y. Chen, G. Luxton, J. Li and A. Boyer, "Monitor unit calculation for an intensity modulated photon field by a simple scatter-summation algorithm," Physics in Medicine and Biology **45**, N1 (2000).
15. G. A. Ezzell, J. M. Galvin, D. Low, J. R. Palta, I. Rosen, M. B. Sharpe, P. Xia, Y. Xiao, L. Xing and C. X. Yu, "Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee," Medical Physics **30**, 2089-2115 (2003).
16. G. A. Ezzell, J. W. Burmeister, N. Dogan, T. J. LoSasso, J. G. Mechalakos, D. Mihailidis, A. Molineu, J. R. Palta, C. R. Ramsey, B. J. Salter, J. Shi, P. Xia, N. J. Yue and Y. Xiao, "IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119," Med Phys **36**, 5359-5373 (2009).
17. D. A. Low, J. M. Moran, J. F. Dempsey, L. Dong and M. Oldham, "Dosimetry tools and techniques for IMRT," Medical Physics **38**, 1313-1338 (2011).
18. J. M. Moran, M. Dempsey, A. Eisbruch, B. A. Fraass, J. M. Galvin, G. S. Ibbott and L. B. Marks, "Safety considerations for IMRT: Executive summary," Medical Physics **38**, 5067 (2011).
19. T. Pawlicki, S. Yoo, L. E. Court, S. K. McMillan, R. K. Rice, J. D. Russell, J. M. Pacyniak, M. K. Woo, P. S. Basran, J. Shoales and A. L. Boyer, "Moving from IMRT QA measurements toward independent computer calculations using control charts," Radiother Oncol **89**, 330-337 (2008).
20. J. Fan, J. Li, L. Chen, S. Stathakis, W. Luo, F. D. Plessis, W. Xiong, J. Yang and C.-M. Ma, "A practical Monte Carlo MU verification tool for IMRT quality assurance," Physics in medicine and biology **51**, 2503-2514 (2006).

21. A. Leal, F. Sánchez-Doblado, R. Arráns, J. Roselló, E. C. Pavón and J. I. Lagares, "Routine IMRT verification by means of an automated Monte Carlo simulation system," International Journal of Radiation Oncology*Biology*Physics **56**, 58-68 (2003).
22. A. Agnew, C. E. Agnew, M. W. D. Grattan, A. R. Hounsell and C. K. McGarry, "Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries," Physics in medicine and biology **59**, N49-N63 (2014).
23. D. Rangaraj, M. Zhu, D. Yang, G. Palaniswaamy, S. Yaddanapudi, O. H. Wooten, S. Brame and S. Mutic, "Catching errors with patient-specific pretreatment machine log file analysis," Practical Radiation Oncology **3**, 80-90 (2013).
24. A. M. Stell, J. G. Li, O. A. Zeidan and J. F. Dempsey, "An extensive log-file analysis of step-and-shoot intensity modulated radiation therapy segment delivery errors," Medical Physics **31**, 1593-1602 (2004).
25. A. C. Hartford, J. M. Galvin, D. C. Beyer, T. J. Eichler, G. S. Ibbott, B. Kavanagh, C. J. Schultz and S. A. Rosenthal, "American College of Radiology (ACR) and American Society for Radiation Oncology (ASTRO) Practice Guideline for Intensity-modulated Radiation Therapy (IMRT)," American Journal of Clinical Oncology **35**, 612-617 610.1097/COC.1090b1013e31826e30515 (2012).
26. W. Bogdanich, "Radiation Offers New Cures, and Ways to Do Harm," in *The New York Times*, (New York, 2010), pp. A1.
27. W. Bogdanich, "As Technology Surges, Radiation Safeguards Lag," in *The New York Times*, (New York, 2010), pp. A1.
28. J. C. Smith, S. Dieterich and C. G. Orton, "It is STILL necessary to validate each individual IMRT treatment plan with dosimetric measurements before delivery," Medical Physics **38**, 553-555 (2011).
29. R. A. C. Siochi, A. Molineu and C. G. Orton, "Patient-specific QA for IMRT should be performed using software rather than hardware methods," Medical Physics **40**, 0706011-0706013 (2013).
30. J. J. Kruse and C. S. Mayo, "Comment on “Catching errors with patient-specific pretreatment machine log file analysis”," Practical Radiation Oncology **3**, 91-92 (2012).
31. B. Fraass, K. Doppke, M. Hunt, G. Kutcher, G. Starkschall, R. Stern and J. Van Dyke, "American Association of Physicists in Medicine Radiation Therapy Committee Task

Group 53: quality assurance for clinical radiotherapy treatment planning," Med Phys **25**, 1773-1829 (1998).

32. J. Van Dyk, R. B. Barnett, J. E. Cygler and P. C. Shragge, "Commissioning and quality assurance of treatment planning computers," Int J Radiat Oncol Biol Phys **26**, 261-273 (1993).
33. G. S. Ibbott, D. S. Followill, H. A. Molineu, J. R. Lowenstein, P. E. Alvarez and J. E. Roll, "Challenges in credentialing institutions and participants in advanced technology multi-institutional clinical trials," Int J Radiat Oncol Biol Phys **71**, S71-75 (2008).
34. J. R. Palta, J. A. Deye, G. S. Ibbott, J. A. Purdy and M. M. Urie, "Credentialing of institutions for IMRT in clinical trials," International Journal of Radiation Oncology*Biology*Physics **59**, 1257-1259 (2004).
35. J. R. Palta, S. Kim, J. Li and C. Liu, "Tolerance limits and action levels for planning and delivery of IMRT," in *Intensity-Modulated Radiation Therapy: The State of Art*, edited by J. R. Palta and T. R. Mackie (Medical Physics Publishing, Madison, WI, USA, 2003), pp. 593-612.
36. I. J. Das, C. W. Cheng, K. L. Chopra, R. K. Mitra, S. P. Srivastava and E. Glatstein, "Intensity-modulated radiation therapy dose prescription, recording, and delivery: patterns of variability among institutions and treatment planning systems," Journal of the National Cancer Institute **100**, 300-307 (2008).
37. T. LoSasso, C.-S. Chui and C. C. Ling, "Comprehensive quality assurance for the delivery of intensity modulated radiotherapy with a multileaf collimator used in the dynamic mode," Medical Physics **28**, 2209-2219 (2001).
38. M. Alber, B. Mijnheer, D. Georg, E. S. f. T. Radiology and Oncology, *Guidelines for the Verification of IMRT*. (ESTRO, 2008).
39. N. L. Childress, C. Bloch, R. A. White, M. Salehpour and I. I. Rosen, "Detection of IMRT delivery errors using a quantitative 2D dosimetric verification system," Medical Physics **32**, 153-162 (2005).
40. C. F. Chuang, L. J. Verhey and P. Xia, "Investigation of the use of MOSFET for clinical IMRT dosimetric verification," Medical Physics **29**, 1109-1115 (2002).
41. J. Godart, E. W. Korevaar, R. Visser, D. J. Wauben and A. A. Van't Veld, "Reconstruction of high-resolution 3D dose from matrix measurements: error detection

capability of the COMPASS correction kernel method," Phys Med Biol **56**, 5029-5043 (2011).

42. Z. Han, S. K. Ng, M. S. Bhagwat, Y. Lyatskaya and P. Zygmanski, "Evaluation of MatriXX for IMRT and VMAT dose verifications in peripheral dose regions," Med Phys **37**, 3704-3714 (2010).
43. D. A. Low and J. F. Dempsey, "Evaluation of the gamma dose distribution comparison method," Medical Physics **30**, 2455-2464 (2003).
44. T. Pawlicki, S. Yoo, L. E. Court, S. K. McMillan, R. K. Rice, J. D. Russell, J. M. Pacyniak, M. K. Woo, P. S. Basran, A. L. Boyer and C. Bonilla, "Process control analysis of IMRT QA: implications for clinical trials," Physics in medicine and biology **53**, 5193-5205 (2008).
45. T. Sanghangthum, S. Suriyapee, G.-Y. Kim and T. Pawlicki, "A method of setting limits for the purpose of quality assurance," Physics in medicine and biology **58**, 7025-7037 (2013).
46. L. Brualla-González, F. Gómez, A. Vicedo, D. M. González-Castaño, A. Gago-Arias, A. Pazos, M. Zapata, J. V. Roselló and J. Pardo-Montero, "A two-dimensional liquid-filled ionization chamber array prototype for small-field verification: characterization and first clinical tests," Physics in medicine and biology **57**, 5221-5234 (2012).
47. J. Duan, S. Shen, J. B. Fiveash, I. A. Brezovich, R. A. Popple and P. N. Pareek, "Dosimetric effect of respiration-gated beam on IMRT delivery," Medical Physics **30**, 2241-2252 (2003).
48. B. E. Nelms, H. Zhen and W. A. Tomé, "Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors," Medical physics **38**, 1037 (2011).
49. L. Bogner, J. Scherer, M. Treutwein, M. Hartmann, F. Gum and A. Amediek, "Verification of IMRT: Techniques and Problems," Strahlentherapie und Onkologie **180**, 340-350 (2004).
50. W. B. Harms, Sr., D. A. Low, J. W. Wong and J. A. Purdy, "A software tool for the quantitative evaluation of 3D dose calculation algorithms," Med Phys **25**, 1830-1836 (1998).
51. D. A. Low, W. B. Harms, S. Mutic and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," Med Phys **25**, 656-661 (1998).

52. J. M. Moran, J. Radawski and B. A. Fraass, "A dose-gradient analysis tool for IMRT QA," Journal of Applied Clinical Medical Physics **6**, 62-73 (2005).
53. A. Bakai, M. Alber and F. Nusslin, "A revision of the gamma-evaluation concept for the comparison of dose distributions," Physics in Medicine and Biology **48**, 3543-3553 (2003).
54. N. L. Childress and I. I. Rosen, "The design and testing of novel clinical parameters for dose comparison," International Journal of Radiation Oncology*Biology*Physics **56**, 1464-1479 (2003).
55. T. Ju, T. Simpson, J. O. Deasy and D. A. Low, "Geometric interpretation of the gamma dose distribution comparison technique: Interpolation-free calculation," Medical Physics **35**, 879-887 (2008).
56. J. R. Palta, C. Liu and J. G. Li, "Quality assurance of intensity-modulated radiation therapy," International Journal of Radiation Oncology* Biology* Physics **71**, S108-S112 (2008).
57. B. E. Nelms and J. A. Simon, "A survey on planar IMRT QA analysis," J Appl Clin Med Phys **8**, 2448 (2007).
58. N. L. Childress, M. Salehpour, L. Dong, C. Bloch, R. A. White and I. I. Rosen, "Dosimetric accuracy of Kodak EDR2 film for IMRT verifications," Medical Physics **32**, 539-548 (2005).
59. A. J. Olch, "Dosimetric performance of an enhanced dose range radiographic film for intensity-modulated radiation therapy quality assurance," Medical physics **29**, 2159 (2002).
60. N. Dogan, L. B. Leybovich and A. Sethi, "Comparative evaluation of Kodak EDR2 and XV2 films for verification of intensity modulated radiation therapy," Physics in Medicine and Biology **47**, 4121 (2002).
61. N. L. Childress, L. Dong and I. I. Rosen, "Rapid radiographic film calibration for IMRT verification using automated MLC fields," Medical physics **29**, 2384 (2002).
62. X. Zhu, P. Jursinic, D. Grimm, F. Lopez, J. Rownd and M. Gillin, "Evaluation of Kodak EDR2 film for dose verification of intensity modulated radiation therapy delivered by a static multileaf collimator," Medical physics **29**, 1687 (2002).

63. P. A. Jursinic, R. Sharma and J. Reuter, "MapCHECK used for rotational IMRT measurements: step-and-shoot, TomoTherapy, RapidArc," Med Phys **37**, 2837-2846 (2010).
64. L. Masi, F. Casamassima, R. Doro and P. Francescon, "Quality assurance of volumetric modulated arc therapy: evaluation and comparison of different dosimetric systems," Med Phys **38**, 612-621 (2011).
65. A. Van Esch, C. Clermont, M. Devillers, M. Iori and D. P. Huyskens, "On-line quality assurance of rotational radiotherapy treatment delivery by means of a 2D ion chamber array and the Octavius phantom," Med Phys **34**, 3825-3837 (2007).
66. N. L. Childress, R. A. White, C. Bloch, M. Salehpour, L. Dong and I. I. Rosen, "Retrospective analysis of 2D patient-specific IMRT verifications," Medical Physics **32**, 838-850 (2005).
67. J. J. Kruse, "On the insensitivity of single field planar dosimetry to IMRT inaccuracies," Medical Physics **37**, 2516-2524 (2010).
68. M. Stasi, S. Bresciani, A. Miranti, A. Maggio, V. Sapino and P. Gabriele, "Pretreatment patient-specific IMRT quality assurance: A correlation study between gamma index and patient clinical dose volume histogram," Medical Physics **39**, 7626-7634 (2012).
69. P. Carrasco, N. Jornet, A. Latorre, T. Eudaldo, A. Ruiz and M. Ribas, "3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification," Medical Physics **39**, 5040-5049 (2012).
70. H. Zhen, B. E. Nelms and W. A. Tome, "Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA," Med Phys **38**, 5477-5489 (2011).
71. M. Podesta, Sebastiaan M J J G. Nijsten, Lucas C G G. Persoon, Stefan G. Scheib, C. Baltes and F. Verhaegen, "Time dependent pre-treatment EPID dosimetry for standard and FFF VMAT," Physics in medicine and biology **59**, 4749-4768 (2014).
72. S. Pai, I. J. Das, J. F. Dempsey, K. L. Lam, T. J. LoSasso, A. J. Olch, J. R. Palta, L. E. Reinstein, D. Ritt and E. E. Wilcox, "TG-69: radiographic film for megavoltage beam dosimetry," Medical physics **34**, 2228 (2007).
73. L. B. Leybovich, A. Sethi and N. Dogan, "Comparison of ionization chambers of various volumes for IMRT absolute dose verification," Medical Physics **30**, 119-123 (2003).

74. P. Fenoglietto, B. Laliberté, N. Aillères, O. Riou, J. B. Dubois and D. Azria, "Eight years of IMRT quality assurance with ionization chambers and film dosimetry: experience of the montpellier comprehensive cancer center," Radiation Oncology **6**, 85 (2011).
75. L. Dong, J. Antolak, M. Salehpour, K. Forster, L. O'Neill, R. Kendall and I. Rosen, "Patient-specific point dose measurement for IMRT monitor unit verification," Int J Radiat Oncol Biol Phys **56**, 867-877 (2003).
76. E. Spezi, A. L. Angelini, F. Romani and A. Ferri, "Characterization of a 2D ion chamber array for the verification of radiotherapy treatments," Physics in medicine and biology **50**, 3361-3373 (2005).
77. A. N.-R. Chair, C. R. Blackwell, B. M. Coursey, K. P. Gall, J. M. Galvin, W. L. McLaughlin, A. S. Meigooni, R. Nath, J. E. Rodgers and C. G. Soares, "Radiochromic film dosimetry: Recommendations of AAPM Radiation Therapy Committee Task Group 55," Medical physics **25**, 2093-2115 (1998).
78. T. Bogucki, W. Murphy, C. Baker, S. Piazza and A. Haus, "Processor quality control in laser imaging systems," Medical physics **24**, 581 (1997).
79. O. A. Zeidan, S. A. L. Stephenson, S. L. Meeks, T. H. Wagner, T. R. Willoughby, P. A. Kupelian and K. M. Langen, "Characterization and use of EBT radiochromic film for IMRT dose verification," Medical Physics **33**, 4064-4072 (2006).
80. F. Gum, J. Scherer, L. Bogner, M. Solleder, B. Rhein and M. Bock, "Preliminary study on the use of an inhomogeneous anthropomorphic Fricke gel phantom and 3D magnetic resonance dosimetry for verification of IMRT treatment plans," Physics in medicine and biology **47**, N67-N77 (2002).
81. T. Gorjiara, R. Hill, Z. Kuncic, J. Adamovics, S. Bosi, J.-H. Kim and C. Baldock, "Investigation of radiological properties and water equivalency of PRESAGE® dosimeters," Medical Physics **38**, 2265-2274 (2011).
82. H. Sakhalkar, D. Sterling, J. Adamovics, G. Ibbott and M. Oldham, "Investigation of the feasibility of relative 3D dosimetry in the Radiologic Physics Center Head and Neck IMRT phantom using Presage/optical-CT," Medical Physics **36**, 3371-3377 (2009).
83. C.-S. Wuu and Y. Xu, "Three-dimensional dose verification for intensity modulated radiation therapy using optical CT based polymer gel dosimetry," Medical Physics **33**, 1412-1419 (2006).

84. M. Oldham, G. Gluckman and L. Kim, "3D verification of a prostate IMRT treatment by polymer gel-dosimetry and optical-CT scanning," Journal of Physics: Conference Series **3**, 293-296 (2004).
85. M. McJury, M. Oldham, V. P. Cosgrove, P. S. Murphy, S. Doran, M. O. Leach and S. Webb, "Radiation dosimetry using polymer gels: methods and applications," British Journal of Radiology **73**, 919-929 (2000).
86. D. Letourneau, J. Publicover, J. Kozelka, D. J. Moseley and D. A. Jaffray, "Novel dosimetric phantom for quality assurance of volumetric modulated arc therapy," Medical Physics **36**, 1813-1821 (2009).
87. V. Feygelman, K. Forster, D. Opp and G. Nilsson, "Evaluation of a biplanar diode array dosimeter for quality assurance of step-and-shoot IMRT," J Appl Clin Med Phys **10**, 3080 (2009).
88. G. Yan, B. Lu, J. Kozelka, C. Liu and J. G. Li, "Calibration of a novel four-dimensional diode array," Medical Physics **37**, 108-115 (2010).
89. V. Feygelman, G. Zhang, C. Stevens and B. E. Nelms, "Evaluation of a new VMAT QA device, or the "X" and "O" array geometries," J Appl Clin Med Phys **12**, 3346 (2011).
90. J. Kozelka, J. Robinson, B. Nelms, G. Zhang, D. Savitskij and V. Feygelman, "Optimizing the accuracy of a helical diode array dosimeter: A comprehensive calibration methodology coupled with a novel virtual inclinometer," Medical Physics **38**, 5021-5032 (2011).
91. R. Boggula, F. Lorenz, L. Mueller, M. Birkner, H. Wertz, F. Stieler, V. Steil, F. Lohr and F. Wenz, "Experimental validation of a commercial 3D dose verification system for intensity-modulated arc therapies," Physics in medicine and biology **55**, 5619-5633 (2010).
92. W. D. Renner, K. J. Norton and T. W. Holmes, "A method for deconvolution of integrated electronic portal images to obtain fluence for dose reconstruction," 2005 **6**, 22-39 (2005).
93. C. Wu, K. E. Hosier, K. E. Beck, M. B. Radevic, J. Lehmann, H. H. Zhang, A. Kroner, S. C. Dutton, S. A. Rosenthal, J. K. Bareng, M. D. Logsdon and D. R. Asche, "On using 3D γ-analysis for IMRT and VMAT pretreatment plan QA," Medical Physics **39**, 3051-3059 (2012).

94. A. Van Esch, T. Depuydt and D. P. Huyskens, "The use of an aSi-based EPID for routine absolute dosimetric pre-treatment verification of dynamic IMRT fields," Radiotherapy and oncology **71**, 223-234 (2004).
95. Y. Nakaguchi, F. Araki, M. Maruyama and S. Saiga, "Dose verification of IMRT by use of a COMPASS transmission detector," Radiol Phys Technol **5**, 63-70 (2012).
96. A. J. Olch, "Evaluation of the accuracy of 3DVH software estimates of dose to virtual ion chamber and film in composite IMRT QA," Medical Physics **39**, 81-86 (2012).
97. B. E. Nelms, D. Opp, J. Robinson, T. K. Wolf, G. Zhang, E. Moros and V. Feygelman, "VMAT QA: Measurement-guided 4D dose reconstruction on a patient," Medical Physics **39**, 4228-4238 (2012).
98. D. Opp, B. E. Nelms, G. Zhang, C. Stevens and V. Feygelman, "Validation of measurement-guided 3D VMAT dose reconstruction on a heterogeneous anthropomorphic phantom," Journal of Applied Clinical Medical Physics **14**, 70-84 (2013).
99. S. Stathakis, P. Myers, C. Esquivel, P. Mavroidis and N. Papanikolaou, "Characterization of a novel 2D array dosimeter for patient-specific quality assurance with volumetric arc therapy," Medical Physics **40**, 0717311-0717315 (2013).
100. M. Wendling, R. J. W. Louwe, L. N. McDermott, J.-J. Sonke, M. van Herk and B. J. Mijnheer, "Accurate two-dimensional IMRT verification using a back-projection EPID dosimetry method," Medical Physics **33**, 259-273 (2006).
101. M. Wendling, L. N. McDermott, A. Mans, J.-J. Sonke, M. v. Herk and B. J. Mijnheer, "A simple backprojection algorithm for 3D in vivo EPID dosimetry of IMRT treatments," Medical Physics **36**, 3310-3321 (2009).
102. S. Both, J. M. Alecu, A. R. Stan, M. Alecu, A. Ciura, J. M. Hansen and R. Alecu, "A study to establish reasonable action limits for patient specific IMRT QA," Journal of Applied Clinical Medical Physics **8**, 1-8 (2007).
103. D. W. Bailey, B. E. Nelms, K. Attwood, L. Kumaraswamy and M. B. Podgorsak, "Statistical variability and confidence intervals for planar dose QA pass rates," Medical physics **38**, 6053 (2011).
104. S. Lang, G. Reggiori, J. Puxeu Vaquee, C. Calle, J. Hrbacek, S. Klock, M. Scorsetti, L. Cozzi and P. Mancosu, "Pretreatment quality assurance of flattening filter free beams on

224 patients for intensity modulated plans: a multicentric study," Med Phys **39**, 1351-1356 (2012).

105. G. M. Mancuso, J. D. Fontenot, J. P. Gibbons and B. C. Parker, "Comparison of action levels for patient-specific quality assurance of intensity modulated radiation therapy and volumetric modulated arc therapy treatments," Medical Physics **39**, 4378-4385 (2012).
106. S. Bresciani, A. D. Dia, A. Maggio, C. Cutaia, A. Miranti, E. Infusino and M. Stasi, "Tomotherapy treatment plan quality assurance: The impact of applied criteria on passing rate in gamma index method," Medical Physics **40**, 1217111-1217116 (2013).
107. A. J. Olch and M. L. Whitaker, "Validation of a treatment plan-based calibration method for 2D detectors used for treatment delivery quality assurance," Med Phys **37**, 4485-4494 (2010).
108. P. A. Jursinic and B. E. Nelms, "A 2-D diode array and analysis software for verification of intensity modulated radiation therapy delivery," Medical Physics **30**, 870-879 (2003).
109. D. Létourneau, M. Gulam, D. Yan, M. Oldham and J. W. Wong, "Evaluation of a 2D diode array for IMRT quality assurance," Radiotherapy and oncology **70**, 199-206 (2004).
110. S. Amerio, A. Boriano, F. Bourhaleb, R. Cirio, M. Donetti, A. Fidanzio, E. Garelli, S. Giordanengo, E. Madon, F. Marchetto, U. Nastasi, C. Peroni, A. Piermattei, C. J. Sanz Freire, A. Sardo and E. Trevisiol, "Dosimetric characterization of a large area pixel-segmented ionization chamber," Medical Physics **31**, 414-420 (2004).
111. P. B. Greer and C. C. Popescu, "Dosimetric properties of an amorphous silicon electronic portal imaging device for verification of dynamic intensity modulated radiation therapy," Medical Physics **30**, 1618-1627 (2003).
112. B. Warkentin, S. Steciw, S. Rathee and B. G. Fallone, "Dosimetric IMRT verification with a flat-panel EPID," Medical Physics **30**, 3143-3155 (2003).
113. L. N. McDermott, M. Wendling, B. van Asselen, J. Stroom, J.-J. Sonke, M. van Herk and B. J. Mijnheer, "Clinical experience with EPID dosimetry for prostate IMRT pre-treatment dose verification," Medical Physics **33**, 3921-3930 (2006).
114. R. M. Howell, I. P. N. Smith and C. S. Jarrio, "Clinical Implementation of Portal Dosimetry—Establishing Action Levels," Journal of Applied Clinical Medical Physics **9**, 16-25 (2008).

115. A. Molineu, N. Hernandez, T. Nguyen, G. Ibbott and D. Followill, "Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom," Medical Physics **40**, 5330-5337 (2013).
116. T. Teke, A. M. Bergman, W. Kwa, B. Gill, C. Duzenli and I. A. Popescu, "Monte Carlo based, patient-specific RapidArc QA using Linac log files," Med Phys **37**, 116-123 (2010).
117. V. Chandraraj, S. Stathakis, R. Manickam, C. Esquivel, S. S. Supe and N. Papanikolaou, "Comparison of four commercial devices for RapidArc and sliding window IMRT QA," J Appl Clin Med Phys **12**, 3367 (2011).
118. V. Feygelman, G. Zhang and C. Stevens, "Initial dosimetric evaluation of SmartArc - a novel VMAT treatment planning module implemented in a multi-vendor delivery chain," J Appl Clin Med Phys **11**, 3169 (2010).
119. K. M. Langen, N. Papanikolaou, J. Balog, R. Crilly, D. Followill, S. M. Goddu, W. Grant, 3rd, G. Olivera, C. R. Ramsey, C. Shi and A. T. Group, "QA for helical tomotherapy: report of the AAPM Task Group 148," Med Phys **37**, 4817-4853 (2010).
120. C. J. Bailat, T. Buchillier, M. Pachoud, R. Moeckli and F. O. Bochud, "An absolute dose determination of helical tomotherapy accelerator, TomoTherapy High-Art II," Med Phys **36**, 3891-3896 (2009).
121. S. Broggi, G. M. Cattaneo, S. Molinelli, E. Maggiulli, A. Del Vecchio, B. Longobardi, L. Perna, F. Fazio and R. Calandrino, "Results of a two-year quality control program for a helical tomotherapy unit," Radiother Oncol **86**, 231-241 (2008).
122. S. M. Goddu, S. Mutic, O. L. Pechenaya, S. R. Chaudhari, J. Garcia-Ramirez, D. Rangaraj, E. E. Klein, D. Yang, J. Grigsby and D. A. Low, "Enhanced efficiency in helical tomotherapy quality assurance using a custom-designed water-equivalent phantom," Phys Med Biol **54**, 5663-5674 (2009).
123. M. Geurts, J. Gonzalez and P. Serrano-Ojeda, "Longitudinal study using a diode phantom for helical tomotherapy IMRT QA," Med Phys **36**, 4977-4983 (2009).
124. G. Yan, C. Liu, T. A. Simon, L. C. Peng, C. Fox and J. G. Li, "On the sensitivity of patient-specific IMRT QA to MLC positioning errors," J Appl Clin Med Phys **10**, 2915 (2009).
125. G. Mu, E. Ludlum and P. Xia, "Impact of MLC leaf position errors on simple and complex IMRT plans for head and neck cancer," Phys Med Biol **53**, 77-88 (2008).

126. J. D. Gordon, S. P. Krafft, S. Jang, L. Smith-Raymond, M. Y. Stevie and R. J. Hamilton, "Confidence limit variation for a single IMRT system following the TG119 protocol," Medical physics **38**, 1641-1648 (2011).
127. L. Coleman and C. Skourou, "Sensitivity of volumetric modulated arc therapy patient specific QA results to multileaf collimator errors and correlation to dose volume histogram based metrics," Medical Physics **40**, 1117151- 1117157 (2013).
128. A. Fredh, J. B. Scherman, L. S. Fog and P. Munck af Rosenschöld, "Patient QA systems for rotational radiation therapy: A comparative experimental study with intentional errors," Medical Physics **40**, 0317161-0317169 (2013).
129. M. F. Chan, J. Li, K. Schupak and C. Burman, "Using a Novel Dose QA Tool to Quantify the Impact of Systematic Errors Otherwise Undetected by Conventional QA Methods: Clinical Head and Neck Case Studies," Technology in cancer research & treatment **13**, 57-67 (2014).
130. S. F. Kry, A. Molineu, J. R. Kerns, A. M. Faught, J. Y. Huang, K. B. Pulliam, J. Tonigan, P. Alvarez, F. Stingo and D. S. Followill, "Institutional Patient-specific IMRT QA Does Not Predict Unacceptable Plan Delivery," International Journal of Radiation Oncology*Biology*Physics **90**, 1195-1201 (2014).
131. E. M. McKenzie, P. A. Balter, F. C. Stingo, J. Jones, D. S. Followill and S. F. Kry, "Toward optimizing patient-specific IMRT QA techniques in the accurate detection of dosimetrically acceptable and unacceptable patient plans," Medical Physics **41**, 121702 (2014).
132. B. E. Nelms, M. F. Chan, G. Jarry, M. Lemire, J. Lowden, C. Hampton and V. Feygelman, "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 **40**, 1117221-11172215 (2013).
133. P. S. Basran and M. K. Woo, "An analysis of tolerance levels in IMRT quality assurance procedures," Medical Physics **35**, 2300-2307 (2008).
134. M. Stock, B. Kroupa and D. Georg, "Interpretation and evaluation of the γ index and the γ index angle for the verification of IMRT hybrid plans," Physics in Medicine and Biology **50**, 399 (2005).

135. G. J. Budgell, B. A. Perrin, J. Mott, J. Fairfoul and R. I. Mackay, "Quantitative analysis of patient-specific dosimetric IMRT verification," Physics in Medicine and Biology **50**, 103 (2005).
136. M. van Zijtveld, M. L. P. Dirkx, H. C. J. de Boer and B. J. M. Heijmen, "Dosimetric pre-treatment verification of IMRT using an EPID; clinical experience," Radiotherapy and Oncology **81**, 168-175 (2006).
137. E. De Martin, C. Fiorino, S. Broggi, B. Longobardi, A. Pierelli, L. Perna, G. M. Cattaneo and R. Calandrino, "Agreement criteria between expected and measured field fluences in IMRT of head and neck cancer: The importance and use of the< i> γ</i> histograms statistical analysis," Radiotherapy and Oncology **85**, 399-406 (2007).
138. M. Carlone, C. Cruje, A. Rangel, R. McCabe, M. Nielsen and M. MacPherson, "ROC analysis in patient specific quality assurance," Medical Physics **40**, 0421031- 0421037 (2013).
139. K. B. Pulliam, J. Y. Huang, R. M. Howell, D. Followill, R. Bosca, J. O'Daniel and S. F. Kry, "Comparison of 2D and 3D gamma analyses," Medical Physics **41**, 021710 (2014).
140. T. Pawlicki, M. Whitaker and A. L. Boyer, "Statistical process control for radiotherapy quality assurance," Medical Physics **32**, 2777-2786 (2005).
141. S. L. Breen, D. J. Moseley, B. Zhang and M. B. Sharpe, "Statistical process control for IMRT dosimetric verification," Medical Physics **35**, 4417-4425 (2008).
142. K. Gérard, J.-P. Grandhaye, V. Marchesi, H. Kafrouni, F. Husson and P. Aletti, "A comprehensive analysis of the IMRT dose delivery process using statistical process control (SPC)," Medical Physics **36**, 1275-1285 (2009).
143. D. C. Montgomery, *Introduction to Statistical Process Control*, 6th ed. (Hoboken, NJ:Wiley, 2009).
144. M. B. Sharpe, B. M. Miller, D. Yan and J. W. Wong, "Monitor unit settings for intensity modulated beams delivered using a step-and-shoot approach," Medical Physics **27**, 2719-2725 (2000).

Figure Captions

Figure 1. Isodose overlay of two measurements, the solid line from radiochromic film and the dashed line from LIC_{2D}. From Brulla-Gonzalez[46]. The 20%, 50%, 70%, 80%, 90%, and 95% dose levels are shown.

Figure 2. Superimposed isodose distribution for two different dose distributions. The fact that the distributions disagree is clear from the intersecting isodose lines, but a quantitative evaluation of the discrepancy by eye is impossible using this type of display. Image is from Duan et al[47].

Figure 3. a) Film and calculated dose distributions from Bogner et al[49]. b) Dose difference distribution (percent of prescription dose) showing that large dose differences can occur in steep dose gradient regions, even for dose distributions that are otherwise similar. c) γ distribution based 3% dose difference and 3 mm DTA.

Figure 4. Examples of one-dimensional γ dose comparison analyses in low (a) and steep (b) dose gradients. The closest approach that the evaluated distribution makes to the reference point is γ. For low dose gradients, the γ test is essentially the dose difference test. For steep dose gradients, the γ test is essentially the DTA test.

Figure 5. Example of the γ calculation error when the evaluated dose distribution spatial resolution is relatively coarse with respect to the DTA criterion. a) The calculation is correct. b) The calculated value is greater than what would be calculated if interpolation was used. c) Evaluated dose distributions with low dose gradients can have the same error if the evaluated pixel locations differ from the reference pixel.

Figure 6. (a) True composite (TC) delivery on a phantom with an IC placed at a specific depth and a radiographic film at a coronal orientation. (b) TC delivery on a stationary 2D array device placed in the coronal direction on the treatment table. (c) TC delivery on a stationary 2D array device placed in the sagittal direction on the treatment table. (d) Perpendicular field-by-field (PFF) or perpendicular composite (PC) delivery on a stationary 2D array device placed in the coronal direction on the treatment table. (e) PFF or PC delivery on 2D array device mounted on the treatment head.

Figure 7. Dose distributions for the mathematical case (circular-shape field) sent to vendors to test γ calculations. (a) Reference dose distribution (resolution 0.5 mm). (b) Evaluated dose distribution (resolution 0.5 mm). (c) Overlayed dose profiles showing the differences between the two distributions.

Figure 8. Clinical dose distributions from a clinical IMRT plan sent to vendors to test γ calculations. (a) TPS calculated dose distribution (resolution 0.5 mm). (b) Film measurement (resolution 0.5 mm).

Figure 9. Flow chart outlines the procedure for setting tolerance and action limits for IMRT QA.

Table 1. For field-by-field γ analysis based on relative dose, the passing rate is highly dependent on the location of the point of normalization. This table from commercial software system shows passing rates based on either points that maximize the passing rate, the central axis point, or the maximum dose point.

Normalization point X,Y coordinates (mm)	**Pass**	**Fail**	**%Pass**
-25,25	263	3	98.9
-20,30	259	7	97.4
-45,5	248	18	93.2
-30,30	251	19	93.0
-25,35	244	20	92.4
0,0 (CAX)	221	79	73.7
-25,35 (Max)	244	20	92.4

Table 2. IMRT QA measurement results reported in the literature. Results include absolute point dose agreement and γ passing rates for various tolerance limits.

Author year	**Delivery technique**	**Dosimeter**	**Number of irradiation**	**Reported results**
Dong 2003[75]	Fixed-gantry and serial tomotherapy	IC	751 cases and 1591 measurements	0.37% ± 1.7% (-4.5% to 9.5%)
Both 2007[102]	Fixed-gantry	2D Diode array	747 fields	3%/3 mm relative: 96.22%±2.89% for HN, 99.30%±1.41% for prostate and other sites; absolute point dose error: 1.41%± 1.10% for HN,

				0.419%±0.420% for prostate and other sites
Ibbott 2008[33]	Not specified	Film, TLDs	250 (multi-institution)	179 (72%) pass (7%/4mm absolute/global)
Molineu 2013[107]	Not specified	Film, TLD	1139 irradiations, 763 institutions	929 (81.6%) pass (7%/4mm absolute/global)
Basran 2008[108]	Fixed-gantry	2D diode array	115 plans	3%/3mm absolute/global: 95.5%±3.5% for HN, 98.8%±2.0% for GU, 97.3%±1.6% for lung
Ezzell 2009[16]	Fixed-gantry and Tomotherapy	Film, IC, 2D diode array	10 institutions, 5 from-easy-to-difficult cases per institution	high dose point: -0.2%±2.2%; low dose point: 0.3%±2.2% (composite); per-field: 97.9%±2.5% (3%/3mm absolute/global); composite film: 96.3%±4.4% (3%/ 3mm absolute/global)
Geurts 2009[109]	Tomotherapy	3D diode array	264 plans	3%/3mm: 97.5%, range 90.0%-100%; absolute/relative or global/local not indicated
Langen 2010[110]	Tomotherapy	IC, planar dosimeter	TG-148 member institutions	IC: 3%; planar: >90% (3%/3mm absolute/global); range or SD not given
Masi 2011[64]	VMAT	IC, film, 2D diode array, 2D IC array	50 plans	IC: 1.1%±1.0%; electronic planar: >97.4% (3%/3mm or 3%/2mm absolute/both global and local), range 92.0%-100%; EDR2: 95.1%, range 83.0%-100%; EBT2: 91.1%, range 80.0-98.5%
Baily 2011[103]	Fixed-gantry	2D diode array, EPID	25 prostate fields, 79 HN fields	2%/2mm absolute/global: 80.4% (prostate), 77.9% (HN); 2%/2mm absolute/local: 66.3% (prostate), 50.5% (HN); 3%/3mm absolute/global: 96.7% (prostate), 93.5% (HN); 3%/3mm absolute/local: 90.8% (prostate), 70.6% (HN)
Lang 2012[104]	Fixed-gantry or VMAT with FFF	IC, Film, 3D diode array, 2D IC array	224 plans (52 plans with IC)	99.3%±1.1% (3%/3mm absolute/global); point dose: 0.34% (±2% for 88% of cases)
Mancuso 2012[105]	Fixed-gantry and VMAT	IC, Film or 2D diode array	TG-119 test cases	IC: −0.82% ± 0.48% (IMRT) and −1.89% ± 0.50% (VMAT); Film: 97.6%±0.6% for IMRT, 97.5%±0.8% for VMAT (2%/2mm composite, absolute /global); Diode: 98.7%±0.3% for IMRT and 98.6%±0.4% for VMAT (3%/3 mm absolute /global)
Bresciani 2013[106]	Tomotherapy	3D diode array	73 plans	absolute global: 98%±2% (3%/3mm), 92%±7% (2%/2mm), 61%±11% (1%/1mm); Absolute local (2 cGy local threshold): 93%±6%(3%/3mm), 84%±9%(2%/2mm), 66%±12%(1%/1mm)

Table 3. IMRT verification QA confidence limits (CL), action limits (AL), tolerance limits (TL) and corresponding γ thresholds reported in the literature.

Author year	Delivery technique	Dosimeter	Number of irradiation	Reported/Recommended Tolerance Levels
Palta 2003[35]	Fixed-gantry	Not Specified	Results from an IMRT questionnaire of 30 institutions	CL and AL: ±10%/2mm and ±15%/3mm (high dose, steep gradient); CL and AL: ±3% and ±5% (high dose, low gradient); CL and AL: ±4% and ±7% (low dose, low gradient)
Low 2003[43]	Fixed-gantry	N/A	simulated fields mimicking clinical fields	γ index tolerance criteria: 5%/2-3 mm
Childress 2005[66]	Fixed-gantry	Film	858 fields	γ index tolerance criteria: 5%/3 mm
Stock 2005[134]	Fixed-gantry	Film, IC	10 plans	γ index (3%/3mm): $\gamma_{mean} < 0.5$, $\gamma_{max} < 1.5$, and fraction of $\gamma_{>1}$ 0-5%
De Martin 2007[135]	Fixed-gantry	Film, IC	57 HN plans	γ index (4%/3mm): γ_{Δ} (γ_{mean} + 1.5 SD(γ)) < 1; γ threshold (4%/3mm): $\gamma_{<1} > 95.3\%$, $\gamma_{<1.5} > 98.9\%$, $\gamma_{>2} < 0.4\%$
ESTRO 2008[38]	Fixed-gantry	IC	Not specified	TL: 3% AL: 5%
Basran 2008[108]	Fixed-gantry	2D diode array	115 plans	TL: 3% overall, 3% per-field (independent of disease site); γ threshold (3%/3mm): ≥ 95% (non-HN cases); γ threshold (3%/3mm): ≥ 88% (HN cases)
Ezzell 2009[16]	Fixed-gantry and Tomotherapy	Film, IC, 2D diode array	10 institutions, 5 from-easy-to-difficult cases per institution	CL: ±4.5% (high dose point in PTV); CL: ±4.7% (low dose point in OAR); CL: ±12.4% (film composite), 87.6% passing (3%/3mm); CL: ±7% (per field), 93.0% passing (3%/3mm)
Carlone 2013[136]	Fixed-gantry	2D diode array	85 prostate plans (68 modified with random MLC errors)	γ threshold (2%/2mm): 78.9% (σ~±3 mm), 84.6% (σ~±2 mm), 89.2% (σ~±1 mm); γ threshold (3%/3mm): 92.9% (σ ~±3 mm), 96.5% (σ~±2 mm), and 98.2% (σ~±1 mm).

Table 4. Vendor survey questionnaire on the implementation of IMRT QA γ analysis software.

1. Do you perform interpolation between points in the dose image__, if so, to what resolution_____?

2. Do you re-sample one or both images for the γ analysis____? If so, on what basis and to what resolution_____?

3. Which image is considered the reference image for the γ analysis, plan or measured_____? Is this user selectable?_____

4. Can you use an acquired and plan dose image that are each in standard DICOM RT format?_____

5. What search radius do you use, is it user selectable?_____

6. Do you offer both relative and absolute dose modes?_____

7. Is your dose tolerance part of the γ analysis referred to the local dose or maximum dose or other____? Is that user selectable_____?

8. Do you specify the dose threshold value above which the analysis will take place? If so, what is the dose threshold?___ Is this value user selectable_____?

9. Do you offer plan-to-acquired-dose image auto-registration___? Manual registration___? Assume center of each image is point in common_____?

10. For relative mode, how do you normalize the acquired and plan dose images-
    a. at maximum?,
    b. to an area,
    c. to a user selectable point?
    d. Other ?_____

11. Do you perform % dose difference-DTA(Van Dyk analysis)? _____
12. If so, how do you normalize the acquired and plan dose images-
    a. at maximum of the reference image?,
    b. to an area,
    c. to a user selectable point?
    d. Other ? _____

Table 5. Vendor* responses to the questionnaire on IMRT QA γ analysis software listed in Table 4.**

QA question #	SNC Patient	3DVH	Portal Dosimetry	RIT113	IMSure	Delta 4	VeriSoft	Compass	OP IMRT
1. Perform interpolation between points:	yes	yes	no	no	yes	yes	no	yes	yes
2. Resample one or both images for γ analysis:	no	no	yes	yes	yes	yes	yes	yes	yes
3.Reference image is plan or measurement:	measured	plan	plan	both	plan	measured	measured	plan	both
user selectable	no	no	no	yes	no	no	no	no	yes
4.Can user acquire and plan dose image in DICOMRT:	yes	yes	yes	yes	no	yes	yes	yes	yes
5.DTA search radius (mm):	8	5		10	30	2.5xDTA	3		
user selectable	no	no	no	yes	no	no	yes	no	yes
6.Offer both relative and absolute dose modes:	yes	no	yes	yes	no	yes	yes	yes	yes
7.Dose tolerance part of γ user selectable:	yes	yes	yes	yes	yes	yes	yes	yes	yes
local dose	yes	yes	yes	yes	yes	yes	yes	yes	yes
max dose	yes	yes	yes	yes	yes	yes	yes	yes	yes
8.Dose threshold above γ analysis occurs:	0-100%	0-100%	0-100%	0-100%		0-100%	0-30%	10cGy	0-100%
user selectable	yes	yes	yes	yes	yes	yes	yes	yes	yes
9.Registration between plan and measurement:	yes	yes	yes	yes	no	yes	yes	yes	yes
auto registration	yes	yes	yes	yes		yes	yes	yes	no
manual registration	yes	no	yes	yes		yes	yes	no	yes
assume center of each image as common point	yes	yes	no	no		no		yes	no
10.Relative mode, normalize plan/measurement to:	yes	no	yes	yes	yes	yes	yes	yes	yes
at maximum	yes	no	yes	yes	yes	yes	yes	yes	yes
to an area	no	no	no	yes	yes	yes	yes	no	yes
user selectable point	yes	no		yes	yes	yes	yes	yes	yes

others	yes								
11.Perform % dose difference/DTA(Van Dyk analysis):	yes	yes	no	yes	no	no	no	yes	yes
Normalize to maximum of reference image	yes	yes		yes					yes
Normalize to an area	no	no		yes					yes
Normalize to a user selectable point	yes	no		yes				yes	yes
Others	yes								

*Vendors: 3DVH and SNC Patient for MapCHECK and ArcCHECK (Sun Nuclear Corporation, Melbourne, FL), Portal Dosimetry with EPID (Varian Medical Systems, Palo Alto, CA), RIT 113 (Radiological Imaging Technology, Inc, Colorado Springs, CO), IMSure (Standard Imaging Inc, Middleton, WI), Delta4 (ScandiDos, Uppsala, Sweden), VeriSoft with Seven29 2D array (PTW, Freiburg, Germany), COMPASS and OmniPro-IMRT with MatriXX (IBA dosimetry, Schwarzenbruck, Germany)

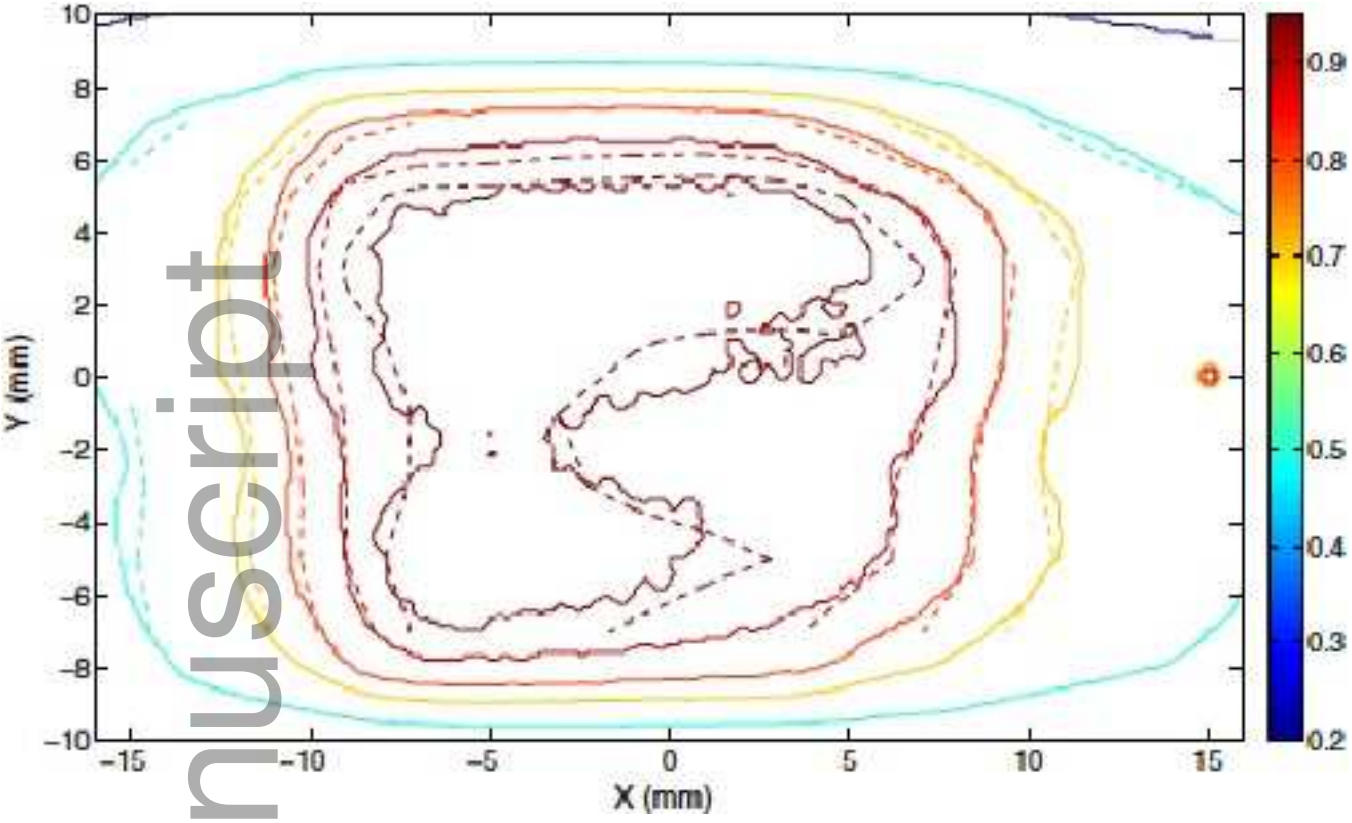
** Survey was conducted in 2017

Table 6. γ passing rates from mathematical test (Figure 7)

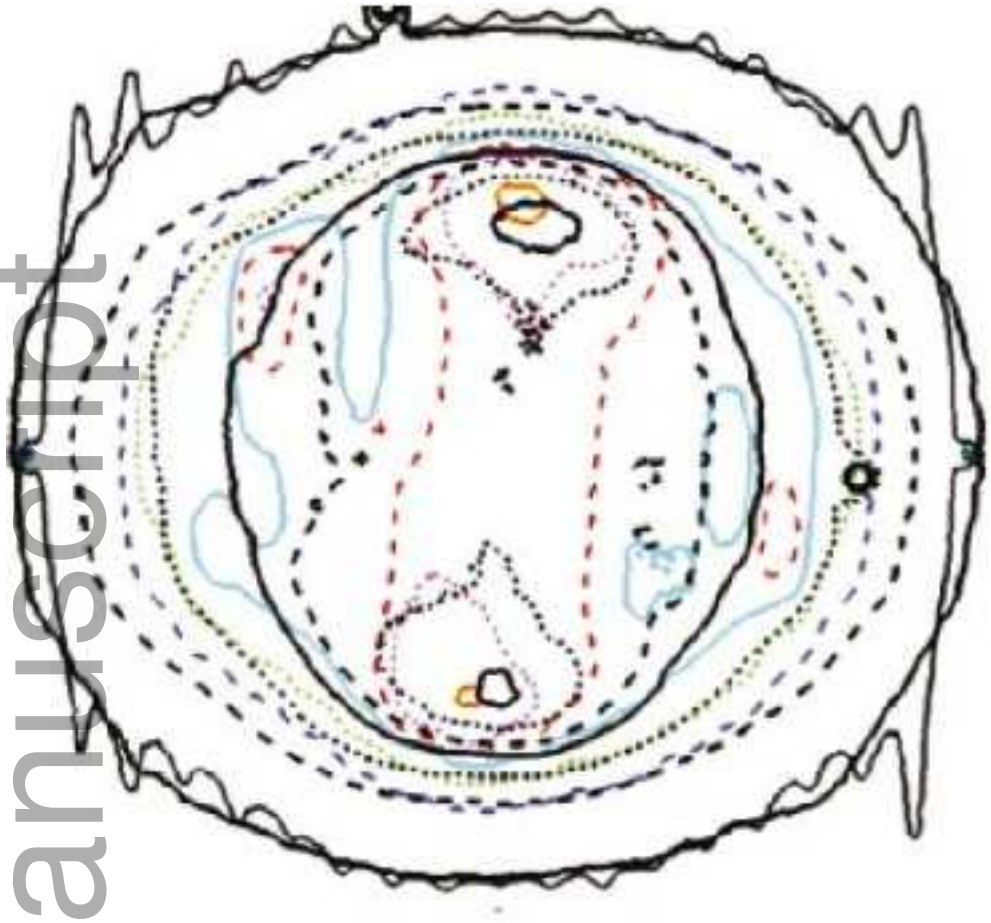
Dose Tolerance (%)	DTA Tolerance (mm)	Dose Threshold (%)	Gold standard (%)	γ passing rate Vendor A (%)	Diff (%)	γ passing rate Vendor B (%)	Diff (%)	γ passing rate Vendor C (%)	Diff (%)	γ passing rate Vendor D (%)	Diff (%)	γ passing rate Vendor E (%)	Diff (%)
3	3	4	95.1	95.1	0.0%	95.3	0.2%	95.1	0.0%	95.1	0.0%	96.4	1.4%
3	2	4	90.2	90.2	0.0%	89.8	-0.4%	90.2	0.0%	90.2	0.0%	91.4	1.4%
2	3	4	93.4	93.4	0.0%	93.7	0.3%	93.4	0.0%	93.4	0.0%	94.7	1.3%
2	2	4	88.3	88.3	0.0%	88.2	-0.1%	88.3	0.0%	88.3	0.0%	89.6	1.5%
3	3	5.5	81.9	81.9	0.0%	81.6	-0.3%	80.4	-1.8%	81.9	0.0%	85.8	4.8%
3	2	5.5	63.1	63.1	0.0%	59.8	-5.2%	61.0	-3.3%	63.1	0.1%	65.8	4.4%
2	3	5.5	75.3	75.3	0.0%	75.1	-0.2%	73.9	-1.8%	75.2	-0.1%	79.0	5.0%
2	2	5.5	55.4	55.4	0.0%	53.6	-3.2%	53.4	-3.5%	55.4	0.1%	58.5	5.7%

Table 7. γ passing rates from clinical test (Figure 8)

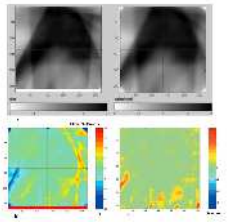
Dose Tolerance (%)	DTA Tolerance (mm)	Dose Threshold (%)	Gold standard (%)	γ passing rate Vendor A (%)	Diff (%)	γ passing rate Vendor B (%)	Diff (%)	γ passing rate Vendor C (%)	Diff (%)	γ passing rate Vendor D (%)	Diff (%)	γ passing rate Vendor E (%)	Diff (%)
3	3	0	98.7	98.5	-0.3%	96.7	-2.1%	98.5	-0.3%	98.5	-0.2%	97.0	-1.8%
3	2	0	97.1	96.5	-0.6%	94.5	-2.7%	96.5	-0.6%	96.5	-0.6%	95.4	-1.7%
2	3	0	96.6	96.1	-0.5%	91.6	-5.2%	96.1	-0.5%	96.1	-0.5%	92.3	-4.5%
2	2	0	93.0	92.1	-1.0%	87.2	-6.2%	92.1	-1.0%	92.1	-1.0%	88.7	-4.6%



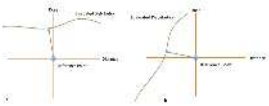
mp_12810_f1.tif



mp_12810_f2.tif



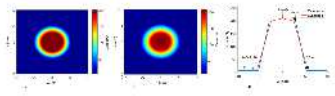
mp_12810_f3.tif



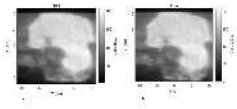
mp_12810_f4.tif

mp_12810_f5.tif

mp_12810_f6.tif



mp_12810_f7.tif



mp_12810_f8.tif

mp_12810_f9.tif