



# Statistical methods to evaluate the correlation between measured and calculated dose using a quality assurance method in IMRT

Abdulhamid Chaikh<sup>1,2</sup>, Charline Desgranges<sup>1</sup>, Jacques Balosso<sup>1,2</sup>

<sup>1</sup>Department of Radiation Oncology and Medical Physics, Grenoble University Hospital, Grenoble, France

<sup>2</sup>Université Grenoble-Alpes, Grenoble, France

Received June 09, 2015; Revised November 01, 2015; Accepted November 13, 2015; Published Online November 27, 2015

## Original Article

### Abstract

**Purpose:** The objective of this study is to validate a procedure based on a statistical method to assess the agreement and the correlation between measured and calculated dose in the process of quality assurance (QA) for intensity-modulated radiation therapy (IMRT). **Methods:** Fifty-six fields for head and neck cancer treatment from 10 patients were analyzed. For each patient a treatment plan was generated using Eclipse TPS®. To compare the calculated dose with the measured dose a CT-scan of solid water slabs (30 × 30 × 15 cm<sup>3</sup>) was used. Absolute dose was measured by a pinpoint ionization chamber and 2D dose distributions using electronic portal imaging device dosimetry. Six criteria levels were applied for each field case (3%, 3 mm), (4%, 3 mm), (5%, 3 mm), (4%, 4 mm), (5%, 4 mm) and (5%, 5 mm). The normality of the data and the variance homogeneity were tested using Shapiro-Wilk's test and Levene's test, respectively. The Wilcoxon signed-rank paired test was used to calculate *p*-values. The Bland-Altman method was used to calculate the limit of agreement between calculated and measured doses and to draw a scatter plot. The correlation between calculated and measured doses was assessed using Spearman's rank test. **Results:** The statistical tests indicate that the data were not normally distributed, *p* < 0.001, and had a homogenous variance, *p* = 0.85. The upper and lower limits of agreement for absolute dose measurements were 6.44% and -6.40%, respectively. The Wilcoxon test indicated a significant difference between calculated and dose measured with the ionization chamber, *p* = 0.01. Spearman's test indicated a strong correlation between calculated and absolute measured dose, with correlation coefficient  $\rho = 0.99$ . Therefore, there is a lack of correlation between dose difference for absolute dose measurements and gamma passing rates for 2D dose measurements. **Conclusion:** The statistical tests showed that the commonly accepted criteria using gamma evaluation are not able to predict the dose difference for a global treatment plan or per beam. The current QA method provides inadequate protection of the patient. The method described here provides an overall analysis for dosimetric data from calculation and measurement, and can be quickly integrated into QA systems for IMRT.

**Keywords:** IMRT; Quality Assurance; Statistical Method

### Introduction

The objective of radiation therapy is to obtain the highest probability of tumor control or cure with the least amount of morbidity and toxicity to normal tissues, namely the organs at risk (OAR). Currently, numerous different irradiation techniques are available and can be used to irradiate tumors, such as three dimensional radiation therapy (3DRT), intensity-modulated radiation therapy (IMRT), volumetric-modulated arc therapy (VMAT) and tomotherapy, etc. IMRT/VMAT have the benefit of protecting the OAR while giving higher doses to the target volumes. These techniques use a multi leaf collimator (MLC), where the leaves are individually moving during the irradiation producing a kind of crawling gap during the beam shot. This requires more monitor units compared to 3DRT. This highly complex design needs a systematic and specific quality assurance (QA) method to

compare the calculated dose given by the TPS, with the measured dose, that will actually be delivered, for each field.

Currently, there are a wide range of recommendations for QA in IMRT measurements, such as AAPM (American Association of Physicists in Medicine) and ESTRO (European Society for Radiotherapy & Oncology) guidelines, etc.<sup>1,2,3,4</sup> One possibility for QA can be performed by applying the calculated treatment planned for each field of a specific patient to a virtual solid water phantom with simple geometry. Then the calculated doses or the beam fluence are compared with the measured dose or fluence at the linear accelerator. The QA method may consist of two types of comparisons: i) point dose measurements (absolute dose) using a small ionization chamber; ii) comparison with planer

Corresponding author: Abdulhamid Chaikh; Department of Radiation Oncology and Medical Physics, Grenoble University Hospital, Grenoble, France.

Cite this article as: Chaikh A, Desgranges C, Balosso J. Statistical methods to evaluate the correlation between measured and calculated dose using a quality assurance method in IMRT. *Int J Cancer Ther Oncol* 2015; 3(4):3411. DOI: 10.14319/ijcto.34.11

dose represented by the fluence distributions in 2D, using a 2D dosimeter.

Point dose measurements are done for 2-3 points and the tolerance is 2-3% between measured and calculated dose. To compare the fluence distribution delivered by the LINAC with the calculated dose, a global evaluation can be done using the gamma index ( $\gamma$ ).<sup>5,6</sup> Quality radiation treatment relies on the agreement between calculated and measured dose, which depends on the precision of the dose calculation algorithm, the calibration of the ionization chamber, validation thresholds, either global or local  $\gamma$  evaluation, percentage tolerances for dose difference (DD) and distance to agreement (DTA) value, etc. Advanced 'type b' algorithms such as the Anisotropic Analytical Algorithm and Collapsed Cone Convolution are more accurate and show less dose difference compared with former algorithms.<sup>7</sup> These complex and time consuming 2D protocols are not completely accurate and have drawbacks that should be highlighted and addressed to improve them.

Currently, the residual 5% non-conformity in the  $\gamma$  evaluation accounts for 50 mL per liter of irradiated tissue, in which we control neither the exact anatomical position nor the tolerance consequences for the clinical outcome. Moreover, anatomical information is completely excluded from these QA methodologies.

However, good treatment QA requires high accuracy of both the calculated and measured dose in the 3D anatomical space of the patient. Since it remains impossible to generate a real anatomical assessment of the treatment by the treatment beams without the patient, presently the beam by beam validation procedure depends on the gamma criteria. In this study we explore several gamma criteria including dose difference and distance to agreement in order to ascertain the best relevant criteria for IMRT-QA.

## Methods and Materials

The implementation and validation of the statistical methods consisted of 3 successive steps including the generation of treatment plans, dosimetry measurements using a homogenous phantom and then the statistical analysis.

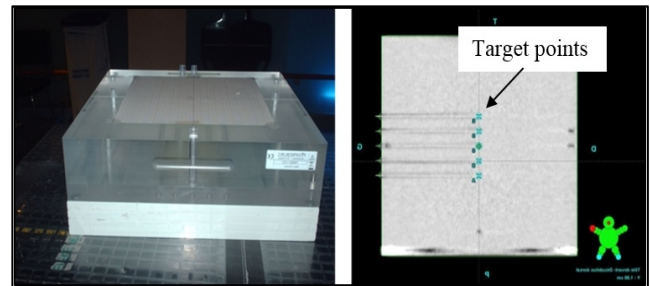
### Treatment plans

This study is based on 56 fields which were used to treat 10 patients with head and neck cancer. The computed tomography (CT) images of each patient were loaded into the Eclipse® Treatment Planning System TPS (Varian, version 8.9). Clinicians delineated the anatomic borders of target structures and OAR. Then treatment plans were generated with 5 to 10 fields for each patient. Dose calculations were performed, using the pencil beam convolution (PBC) algorithm with heterogeneity correction by the modified

Batho method (PBC-MB). The patients were treated by IMRT and 6MV photon beams (Clinac 600, Varian) using the monitor units calculated in these plans.

### Phantom measurements

We consistently made a comparison of the doses calculated with Eclipse® TPS and the measured doses at the accelerator in the same conditions. For this a QA method for IMRT was used consisting of measuring point by point absolute doses, as well as studying the 2D dose distribution (fluence map). To compare the calculated dose with the measured dose a CT-Scan of solid water slabs ( $30 \times 30 \times 15 \text{ cm}^3$ ) was done. **Figure 1** shows the solid water phantom and its CT-Scan which was used to compare the calculated doses with measured ones. All parameters of the treatment plan were applied, except the gantry angle that was always set at 0 degree. The measurements were made on a Clinac600® accelerator.



**FIG. 1:** Solid water phantom (left) and its CT Scan (right panel). This phantom has been used to compare the calculated and measured doses.

### Absolute dose

A Pinpoint chamber,  $0.0125 \text{ cm}^3$  was used to measure the absolute dose for four points in every field. These points were selected in the flat domain of the dose profile calculated with the TPS. The difference between calculated dose ( $D_c$ ) and measured dose ( $D_m$ ) was calculated as follows:

$$DD\% = (D_m - D_c) \times 100 / D_c \quad (1)$$

To validate a specific field the tolerance threshold was 3 points out of 4 having 3% or less difference.

### Gamma passing rates

To compare the fluence map obtained from the TPS with the dose fluence delivered by the accelerator, electronic portal imaging device (Epid) dosimetry was used. The consistency of measurements with calculations was assessed using the  $\gamma$  index. Six criteria levels (DD, DTA) were applied for each field (3%, 3 mm), (4%, 3 mm), (5%, 3 mm), (4%, 4 mm), (5%, 4 mm) and (5%, 5 mm). In all cases, gamma analysis was done with relative and absolute evaluations.

### Statistical analysis

We used the Shapiro-Wilks test to assess the normality of data and Levene’s test to assess whether the assumption of homogeneity of variance was fulfilled. A Wilcoxon signed-rank test was used to calculate the *p*-value. A bilateral statistical test was realized with an error  $\alpha = 5\%$ , corresponding to a 95% confidence interval. The dose difference was considered significant if  $p < 0.05$ . To measure the strength of the relationship between the calculated and measured dose, Spearman’s rank non-parametric test used to calculate correlation coefficient ( $\rho$  -values); then correlation maps were generated. All statistical tests used in this study were performed with the R programming language.<sup>8,9</sup> The upper limit of agreement (ULA) and lower limit of agreement (LLA) were calculated using the following equations,<sup>10,11,12</sup>:

$$ULA = |\text{average}| + 2 \times SD \quad (2)$$

$$LLA = |\text{average}| - 2 \times SD \quad (3)$$

For the  $\gamma$  analyses, where perfect agreement produced a passing rate of 100%, the confidence limit was defined as:

$$ULA = (100 - \text{average}) + 2 \times SD \quad (4)$$

$$LLA = (100 - \text{average}) - 2 \times SD \quad (5)$$

where, average was the mean percentage of points passing the  $\gamma$  criteria and SD the standard deviation.

### Results

#### Absolute dose

The tolerance condition for 3 points out of 4 having 3% or less of difference was respected for all fields and the average dose difference was  $0.02 \pm 3.2\%$ . The Shapiro-Wilks test showed a significant deviation from normality for DD with  $p < 0.001$ . Levene’s test indicated that the calculated and measured doses did not have significantly different variances, showing homogeneous variance,  $p = 0.85$ . Accordingly, the

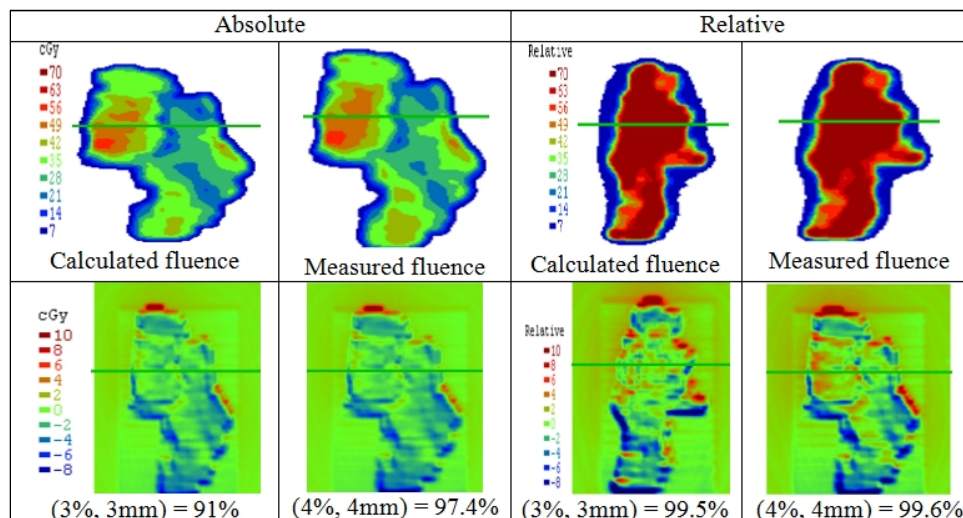
non-parametric Wilcoxon test was used to calculate the *p*-value showing a significant difference between calculated and measured dose of  $p = 0.01$ .

#### Gamma passing rate

**Table 1** summarizes the dosimetry and statistical results for 10 patients treated with 56 beams. It can be seen that ULA and LLA were always below 2% for all tolerance criteria. **Figure 2** shows calculated, measured dose and  $\gamma$  passing rates for absolute and relative evaluations using (3%, 3 mm) and (4%, 4 mm) criteria for a left anterior oblique beam. It can be seen in **Figure 2** that the tolerance level, 95% of pixels having  $\gamma < 1$ , is not met using an absolute evaluation with (3%, 3 mm), but was validated when an absolute evaluation with (4%, 4 mm) was used. Thus, this case shows a minor deviation. **Figure 3** shows the  $\gamma$  passing rates using all the tolerance levels for one patient treated with five beams including one posterior, one anterior and three oblique beams.

**TABLE 1:** The dosimetry and statistical results for 10 patients treated with 56 beams. SD shows the standard deviation. ULA and LLA were the upper and lower level of agreement, respectively. The *p*-values were calculated using the Wilcoxon signed rank test.

$\gamma$ criteria	evaluation	average	SD	ULA	LLA	<i>p</i> -value
3%, 3mm	relative	99.8	0.005	0.21	0.18	1.3E-06
	absolute	98.5	0.01	1.53	1.46	
4%, 4mm	relative	100	0.001	0.003	-0.003	1.8E-06
	absolute	99.2	0.01	0.82	0.77	
5%, 3mm	relative	100	0.0008	0.001	-0.001	1.20E-05
	absolute	99.5	0.008	0.5	0.48	
4%, 4mm	relative	100	0.0006	0.001	-0.001	0.0002
	absolute	99.8	0.005	0.2	0.18	
5%, 4mm	relative	100	0	0	0	0.001
	absolute	99.9	0.003	0.10	0.09	
5%, 5mm	relative	100	0	0	0	0.09
	absolute	100	0.001	0.002	-0.002	



**FIG. 2:** 2D dose distribution for calculated dose and measured dose by Epid for the left anterior oblique beam showing  $\gamma$  passing rates for absolute and relative evaluations using (3%, 3 mm) and (4%, 4 mm) criteria.

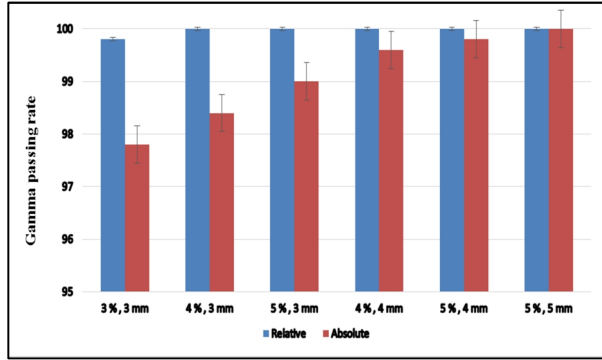


FIG. 3:  $\gamma$  passing rates for one patient using relative and absolute evaluations with all  $\gamma$  criteria.

### Statistical correlation

#### Absolute dose

The data obtained from calculated and measured doses showed a strong correlation with correlation coefficient  $\rho = 0.99$ . Figure 4 shows the Bland–Altman plot presenting the calculated and measured point doses as a function of dose difference. The solid line in Figure 4 represents the average, 0.02%, the two dashed lines represent ULA and LLA, given by equations 2 and 3, at 6.44% and -6.40%. It can be seen that for lower doses, the dose difference can reach 15%.

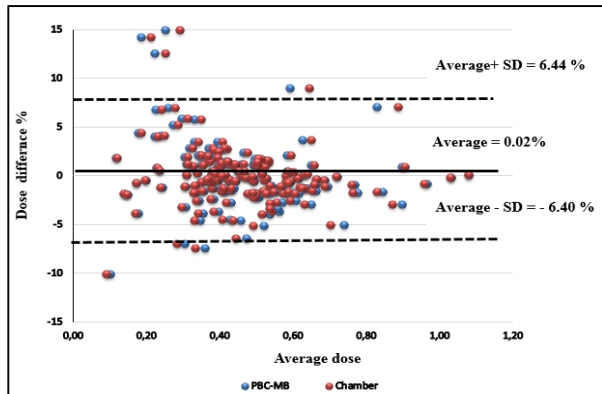


FIG. 4: Bland–Altman plot showing dose difference as a function of average doses for calculated and measured absolute points. The solid line represents the average of 0.02%, the two dashed lines represent ULA and LLA (6.44% and -6.40%) as given by equations 2 and 3.

#### Absolute dose and gamma passing rates

The data demonstrate that there is a lack of correlation between  $\gamma$  passing rates and DD for all criteria with correlation coefficient  $\rho$  ranging from 0 to 0.29. Figure 5 shows an example of correlation plots between DD and  $\gamma$  passing rate, for all beams, using absolute and relative evaluations with (3%, 3 mm).

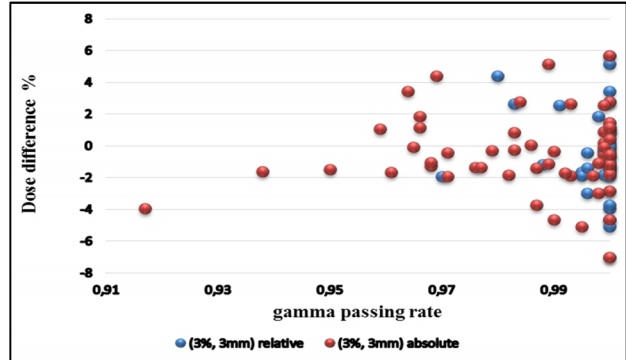


FIG. 5: Correlation plot between  $\gamma$  passing rates and dose difference.

#### Correlation between absolute and relative gamma passing rates

Figure 6 shows the correlation maps for  $\gamma$  passing rates generated for all gamma criteria ranging from 0 to  $\pm 1$ . The  $\rho$ -values are color coded and the results of  $\rho$ -values were between  $[-1; +1]$ . The white, blue and red coloring on the correlation map show respectively very weak correlation, strong positive correlation and strong negative correlation. The correlation coefficient  $\rho$ -values for both criteria (5%, 4 mm) and (5%, 5 mm) with relative evaluations were not calculable. We observed a strong correlation between  $\gamma$  passing rates using (3%, 3 mm) and  $\gamma$  passing rate using (4%, 4 mm) for relative and absolute evaluations, as shown in Figure 6, with dashed boxes. Therefore, using criteria (4%, 4 mm) the tolerance of 95% of pixels having  $\gamma < 1$  was met. This means that  $\gamma$  criteria (4%, 4 mm) can be used in the case of minor deviations to compare calculated and measured fluence in 2D.

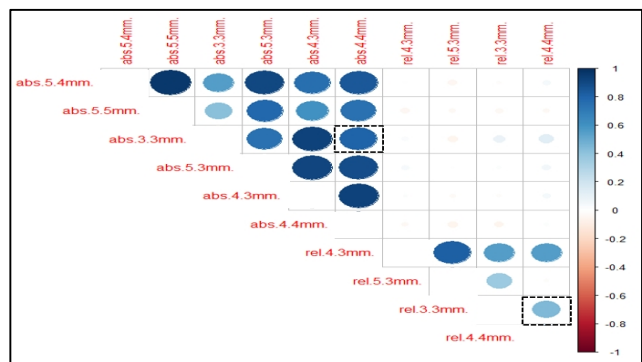


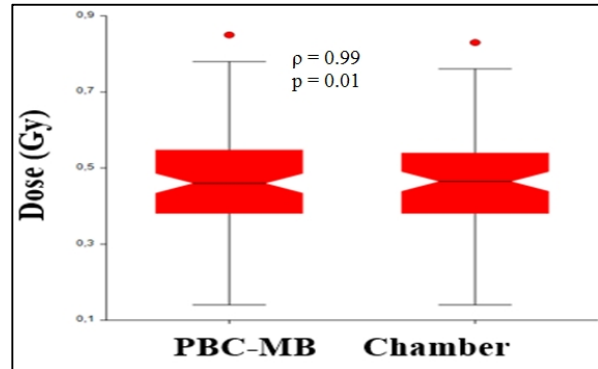
FIG. 6: Correlation map between gamma passing rates using all gamma criteria for all beams, “abs” and “rel” mean absolute and relative evaluations. The correlation values calculated with Spearman’s rank test ranged from 0 to  $\pm 1$ . The white, blue and red coloring in the correlation map show respectively very weak correlation, strong positive correlation and strong negative correlation. The dashed boxes show a strong correlation between  $\gamma$  criteria (3%, 3 mm) and  $\gamma$  (4%, 4 mm) for both relative and absolute evaluations, justifying the use of (4%, 4 mm) for minor deviations.

## Discussion

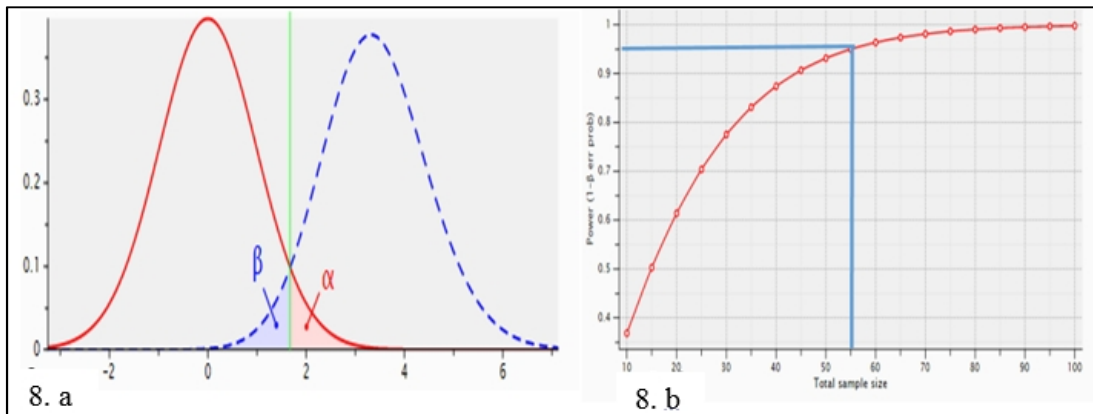
### Influence of dosimetry data on confidence limits

The statistical analysis of 188 points showed that 82.4% were within the threshold of acceptable tolerance of  $\pm 3\%$ . Spearman's rank test showed an excellent correlation between calculated and measured absolute doses with  $\rho = 0.99$ , as shown in **Figure 7** where median values are 0.4 for both calculated and measured doses. However, our data do not have a normal distribution; in this case the multiplying factor "2" in equations 2, 3, 4 and 5 was used to calculate ULA and LLA. Using a factor 1.96 instead of 2 in these equations would more precisely give the values of the 95% interval if these distributions were truly normal. Both confidence limits depend on assumptions such as normal distribution, sample size, standard deviation and average difference. Moreover, the difference between measured and calculated doses depends also on the precision of dose calculation algorithm. Hence, the validation of the treatment plan for each field is based on the assessment of the dose calculation. Currently, numerous different algorithms are available. The algorithms were categorized into two groups 'type a' or 'type b' according to the electron transport calculation. More recently, 'type c' algorithm is available, such as Acuros XB, Varian's Eclipse

TPS (Varian Medical Systems, Palo Alto, CA).<sup>7,13</sup> In this study PBC-MB density correction have been used. PBC-MB is based on 'type a' which does not take into account the changes in lateral electron transport.<sup>14-18</sup> However, Acuros XB is more accurate and recommended for heterogeneity correction.<sup>19-21</sup>



**FIG. 7:** Boxplot for dose difference obtained from 56 fields indicating the minimum, median as a black line (in the red part) and maximum values, as well as the 25<sup>th</sup> and 75<sup>th</sup> percentiles. The median values were 0.4 for calculated and measured doses. This representation is not adapted to show the differences but the internal distribution of values.



**FIG. 8:** 8.a. distributions of  $\alpha$  and  $\beta$  generated for the Wilcoxon test, with  $\alpha = 0.05$ , power = 0.95, effect size = 0.46 and sample size = 56 and figure 8.b. presents the sample size as a function of power, it can be seen  $n = 56$  represents a power of 95% (GPower software).

In order to properly assess the correlation between calculated and measured doses, the power of the statistical test ( $1 - \beta$ ) was calculated with the sample size used in this study,  $n = 56$ .<sup>22,23</sup> This power had been calculated according to the requirements of the Wilcoxon test with  $\alpha = 0.05$ , effect size = 0.46, leading to a power = 0.95. **Figure 8.a** shows the distribution of  $\alpha$  and  $\beta$  generated by the Wilcoxon test under the above conditions and **Figure 8.b**. shows the sample size as a function of power generated by the Wilcoxon test. It can be seen that for  $n = 56$  the power is of 95%.

For 2D gamma evaluations, we observed that the both confidence limits ULA and LLA for the absolute evaluation were higher than the relative evaluation when varying the  $\gamma$  criteria (as shown in **Table 1**). This is because the absolute

evaluation is more sensitive than the relative evaluation. Furthermore, by increasing the  $\gamma$  criteria from (3%, 3 mm) to (5%, 5 mm) the  $p$ -value increased to reach a non-significant difference,  $p > 0.05$ , between absolute and relative evaluations with (5%, 5 mm). However, there was no correlation between absolute dose and  $\gamma$  passing rates for absolute evaluation.

### Limit of gamma evaluation

Currently the most common  $\gamma$  criteria for IMRT is 3% for DD, 3 mm for DTA and the treatment per beam can be validated if at least 95% of pixels have  $\gamma \leq 1$ . The fundamental limitation of the  $\gamma$  evaluation is that the measurements are performed using a solid water phantom, in contrast to on the patient who presents an heterogeneous medium. The evaluation is

done for a unique rotation angle, equal to zero, for all fields. We question the relevance of this method using a homogenous phantom. It is troubling that using the correlation test, the  $\gamma$  passing rates do not predict the clinical impact resulting from the difference between calculated and measured doses on the patient since the anatomic locations of discrepancies are not displayed.

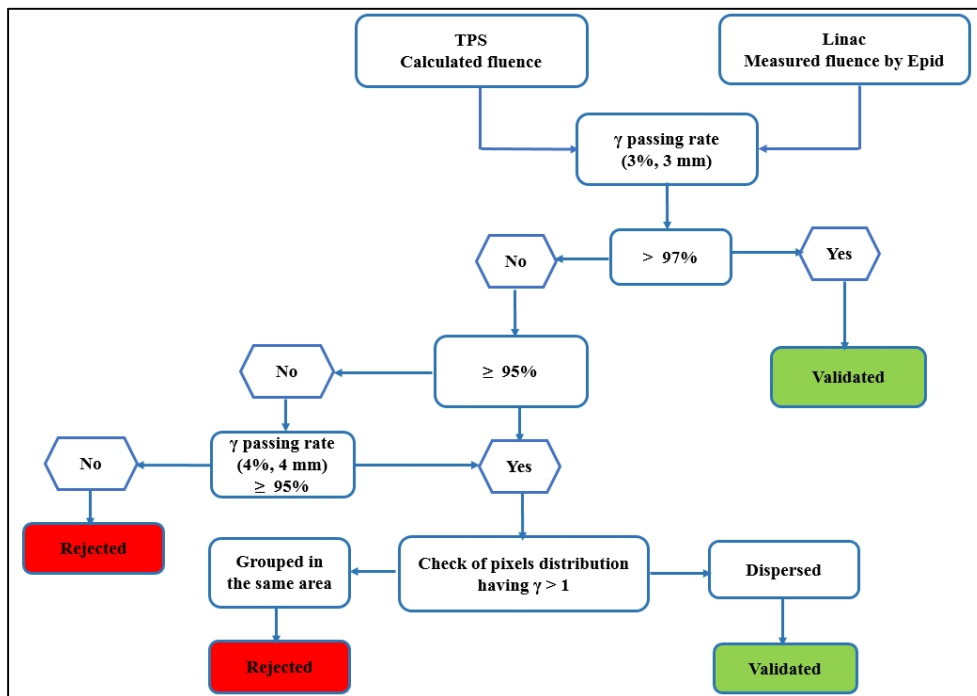
As can be seen from **Figure 2** the large difference ranged from 8 to 10 cGy, presented in blue and red coloring. Therefore, the Bland –Altham plot in **Figure 4** shows that a larger difference in dose of 10 to 15% was associated with low doses ranging from 0.1 to 0.5 Gy.

It is interesting to note that the  $\gamma$  evaluation is not able to predict the change in the dose volume histogram either for the target volume or OAR.<sup>24-28</sup> Chaikh *et al.*, 2014, showed for a specific patient, that 95% of pixels had  $\gamma < 1$  with a  $\gamma$  average value less than unity<sup>24</sup>. In this case, we might conclude that there is no difference between the reference plan and the tested plan, but the real changes of dose distribution were not taken into consideration to protect the OAR. It was observed that a  $\gamma$  evaluation with (3%, 3 mm) under-estimated the dose in a small fraction of the target volume and overestimated the dose for OAR.<sup>24</sup> Zhen *et al.* 2011, showed a significant correlation between 3D dose volumes in a patient geometry for QA comparison.<sup>25</sup> In this study, the dose difference

reached up to 15%, while  $\gamma$  evaluation showed 95% of pixels as having  $\gamma < 1$ . It may be that if the calculated dose in TPS shows a plateau, it does not give the same form and may be in the slope of the profile. The statistical analysis confirmed and advocated that the limit of the  $\gamma$  evaluation should predict the change that can be observed on clinical treatment planning. The current QA method using a water phantom is not sufficient to fully protect the patient. Therefore,  $\gamma$  evaluations with (3%, 3 mm) or (4%, 4 mm) for minor deviations, are inadequate, per se, to predict the risk to the OAR.

## Recommendations

In our department, according to our study, the tolerance for (4%, 4 mm) was validated only for minor deviations. It is also interesting to quantify the degree of dose difference according to the level of delivered dose. Much care should be taken during interpretation, since sometimes a small overestimation in dose can dramatically change the safety or OAR. An example of this can be seen in **Figure 2** where the validation of the beam is obtained with a  $\gamma$  passing rate  $> 95\%$  using criteria (4%, 4 mm), however a small area with (a mild) dose difference of 10 cGy represents (as much as) 10% overestimation in dose in a relative evaluation. It is interesting in this case to identify whether the 10% are localized in the PTV or in the OAR.



**FIG. 9:** Flow chart of the dosimetry decisional method including successive steps to validate the IMRT plan per beam.

**Figure 9** summarizes the recommendations and shows that this method can be rapidly used to validate a treatment plan beam by beam. The zones with pixels having  $\gamma > 1$ , should be

identified to verify if the overestimated dose is dispersed or grouped in one same area and with a high or low level of dose.

As a whole, we recommend the following steps to validate a treatment plan beam by beam for IMRT:

- **Case 1:** The treatment is validated if the  $\gamma$  passing rate is at least 97% of pixels having  $\gamma < 1$  with (3%, 3 mm) criteria.
- **Case 2:** The treatment could be validated with caution if the  $\gamma$  passing rate is at least 95% of pixels having  $\gamma < 1$  with (3%, 3 mm) criteria. However, the pixel distribution should be checked to determine whether the 5% of pixels having  $\gamma > 1$  is dispersed or grouped in the same area. If these pixels are dispersed, then the treatment can be validated.
- **Case 3:** Minor deviation if the  $\gamma$  passing rate is at least 95% of pixels having  $\gamma < 1$  with (4%, 4 mm). The distribution of pixels should be checked as recommended in case 2.
- **Case 4:** major deviation if the  $\gamma$  passing rate  $< 95\%$  of pixels having  $\gamma < 1$  with (4%, 4 mm). The treatment should be rejected and a discussion between the medical physicists and oncologists should take place.

## Conclusion

In this study, we used statistical methods to evaluate the correlation between absolute dose and gamma passing rates. According to the strong correlation we propose gamma criteria of (4%, 4 mm), if the (3%, 3 mm) criterion with 95% passing rate is not verified. The statistical methods confirmed the limitations of a global evaluation based on the  $\gamma$  index to predict the change of clinical treatment planning for target volumes or OAR. This method can be easily implemented and allows the clinician and medical physicist to validate the treatment planning based on a visual display of data on tolerance limits and acceptability criteria.

## Conflict of interest

The authors declare that they have no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## Acknowledgement

We thank Dr Alison Foote (Clinical Research Center, Grenoble University Hospital) for language editing.

## References

1. Mijheer B, Georg D. Guidelines for the verification of IMRT. Brussels, Belgium: ESTRO; 2008.
2. Ezzell GA, Burmeister JW, Dogan N, *et al.* IMRT commissioning: multiple institution planning and

- dosimetry comparisons, a report from AAPM Task Group 119. *Med Phys.* 2009;36:5359-73.
3. Thwaites D. Accuracy required and achievable in radiotherapy dosimetry: have modern technology and techniques changed our views? *Journal of Physics: Conference Series.* 2013;444:012006.
4. Low DA, Maron JM, Dempsey JF, Dong L, Oldham M. Dosimetry tools and techniques for IMRT. *Med Phys.* 2011; 38: 1314-38.
5. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys.* 1998;25:656-61.
6. Spezi E, Lewis DG.  $\gamma$  histograms for radiotherapy plan evaluation. *Radiothera Oncol.* 2006;79:224-30.
7. Chaikh A, Balosso J. Should the dose prescription be readjusted when using tissues density corrections algorithms for radiation oncology. *J Case Rep Onc Ther.* 2014;1:01018.
8. Chaikh A, Giraud JY, Perrin E, *et al.* The choice of statistical methods for comparisons of dosimetric data in radiotherapy. *Radiat Oncol.* 2014; 9:205.
9. The R project for statistical computing. Available from <http://www.r-project.org>
10. Bland J Martin, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-10.
11. Hamilton C, Stamey J. Using Bland-Altman to assess agreement between two medical devices -don't forget the confidence intervals. *J Clin Monit Comput.* 2007;21:331-33.
12. Mantha S, Roizen MF, Fleischer LA, *et al.* Comparing methods of clinical measurement: reporting standards for Bland and Altman analysis. *Anesth Analg.* 2000;90:593-602.
13. Knoos T, Wieslander E, Cozzi L, *et al.* Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. *Phys Med Biol.* 2006;51: 5785-807.
14. Papanikolaou N, Battista J, Mackie T, Kappas C, Boyer A. Tissue inhomogeneity corrections for megavoltage photon beams. AAPM Report No 85; Task Group No. 65, 2004.
15. Ahnesjö A, Aspradakis M M. Dose calculations for external photon beams in radiotherapy. *Phys Med Biol.* 1999; 44 :99-155.
16. Batho HF. Lung corrections in cobalt 60 beam therapy. *J Can AssocRadiol.* 1964;15:79-83.
17. El-Khatib E, Battista J J. Improved lung dose calculation using tissue-maximum ratios in the Batho correction. *Med Phys.* 1984;11:279-86.
18. Thomas S J. A modified power law formula for inhomogeneity corrections in beams of high energy x rays. *Med Phys.* 1991;18:719-23.
19. Rana S. Clinical dosimetric impact of Acuros XB and analytical anisotropic algorithm (AAA) on real

- lung cancer treatment plans: review. [Int J Cancer Ther Oncol. 2014;2:02019.](#)
20. Ojala J. The accuracy of the Acuros XB algorithm in external beam radiotherapy – a comprehensive review. [Int J Cancer Ther Oncol. 2014;2:020417.](#)
  21. Rana S, Pokharel S. Dose-to-medium vs. dose-to-water: Dosimetric evaluation of dose reporting modes in Acuros XB for prostate, lung and breast cancer. [Int J Cancer Ther Oncol. 2014;2:020421.](#)
  22. Faul F, Erdfelder E, Lang AG, *et al.* G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. [Behavior Research Methods. 2007;39:175-91.](#)
  23. Faul F, Erdfelder E, Buchner A, *et al.* Statistical power analyses using G\*Power 3.1. Tests for correlation and regression analyses. [Behavior Research Methods. 2009;41:1149-60.](#)
  24. Chaikh A, Giraud JY, Marguet M, *et al.* A decision tool to adjust the prescribed dose after change in the dose calculation algorithm. [Int J Cancer Ther Oncol. 2014;2:020414.](#)
  25. Zhen H. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. [Med Phys. 2011;38: 5477-89.](#)
  26. Nelms BE, Zhen H, Tomé WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. [Med Phys. 2011;38:1037-44.](#)
  27. Daniel W, Benjamin EN, Kristopher A, *et al.* Statistical variability and confidence intervals for planar dose QA pass rates. [Med Phys. 2011;38:6053-64.](#)
  28. Stasi M, Bresciani S, Miranti A, *et al.* Pretreatment patient-specific IMRT quality assurance: A correlation study between gamma index and patient clinical dose volume histogram. [Med Phys. 2012;39:7627-34.](#)