

# A management method for the statistical results of patient-specific quality assurance for intensity-modulated radiation therapy

Satoshi Nakamura\*, Hiroyuki Okamoto, Akihisa Wakita, Rei Umezawa, Kana Takahashi, Koji Inaba, Naoya Murakami, Toru Kato, Hiroshi Igaki, Yoshinori Ito, Yoshihisa Abe and Jun Itami

Department of Radiation Oncology, National Cancer Center Hospital, Chuo-ku, Tsukiji 5-1-1, Tokyo, 104-0045, Japan

\*Corresponding author. Department of Radiation Oncology, National Cancer Center Hospital, Tsukiji 5-1-1, Chuo-ku, Tokyo, 104-0045, Japan.

Tel: +81-3-3542-2511; Fax: +81-3-3545-3567; Email: satonaka@ncc.go.jp

Received January 31, 2016; Revised June 27, 2016; Editorial Decision October 4, 2016

## ABSTRACT

There are many reports concerning patient-specific quality assurance (QA) for intensity-modulated radiation therapy (IMRT). However, reports about the statistical results of QA are lacking. Management methods for the results of the QA are needed, even though we have the ESTRO group recommendation that a tolerance limit of 1.96 standard deviation (SD) be established in each institution. The purpose of this study was to establish a management method for determining the tolerance limit and to report the statistical results of patient-specific QA. From April 2006 to March 2015, five linacs in the National Cancer Center, Tokyo, Japan, were used to treat 1185 patients with IMRT. Patient-specific QA was performed using an ion chamber, films, and some detectors. To establish a management method for the results, differences between the measured and calculated doses in the ion chamber were analyzed for each linac, each phantom, and each treatment site. The overall mean dose difference was  $0.5 \pm 1.3\%$ , and the mean dose difference in each linac was  $0.6 \pm 1.2\%$ ,  $0.9 \pm 1.3\%$ ,  $-0.4 \pm 1.4\%$ ,  $-0.1 \pm 1.2\%$  and  $-0.1 \pm 0.9\%$ . The difference between linacs and between treatment sites was significant ( $P < 0.001$  and  $0.01$ , respectively). The proportion of the dose difference within  $\pm 3\%$  was 97.7%, and that was improved from 2006 to 2014. The results of the patient-specific QA should be managed for each linac and each treatment site in order to decide the suitable tolerance limit. Reports of statistical results will be helped if a new tolerance limit and action level will be considered.

**KEYWORDS:** Ppatient-specific QA, IMRT, management method, statistical analysis, tolerance limit

## INTRODUCTION

Patient-specific quality assurance (QA) for intensity-modulated radiation therapy (IMRT) is important, and a lot of QA devices and techniques have been used for patient-specific QA for IMRT [1–7]. The main issue discussed in the scientific literature concerns the accuracy of patient-specific QA—statistical analyses of the results of patient-specific QA have not been reported sufficiently [1, 2] (although reports of the QA result for various facilities have helped other facilities to validate the accuracy of their IMRT treatment).

The tolerance limit for point dose verifications has been recommended by ESTRO and AAPM as the sum of the averaged difference

and 1.96 standard deviations (SDs) [8, 9]. Thus, the tolerance limit can be established from the trend of the patient-specific QA result in each institution. However, few reports have been published about applying statistical methods to patient-specific QA [10–12]. For example, no recommendations about the management method for patient-specific QA exist, though Pulliam *et al.* have reported that their results for patient-specific QA were improved year by year because of an innovation in the treatment planning system (TPS) and improvements in the treatment planning process [2]. The 1.96 SDs limit may have been better applied to each treatment site. A detailed standard management method for the results of patient-specific QA is required.

The purpose of this study was to establish a management method for determining 1.96 SDs from the results of patient-specific QA for IMRT, and to report the statistical results of patient-specific QA.

## MATERIALS AND METHODS

From April 2006 to March 2015 in our institute, five linacs [Clinac 21 EX ('A'), two Clinac iX silhouettes ('B' and 'C'), one Clinac iX KONA ('D') and TrueBeam ('E'), Varian Medical Systems, Palo Alto, CA, USA] were used to treat 1185 patients with IMRT. In each plan, patient-specific QA of the IMRT was performed before the treatment. Our QA program consists of two main verifications: point dose measurements in a phantom (using an ion chamber) and planar dose measurements (using a film and a detector array). The intensity-modulated delivery techniques used were sliding window IMRT and volumetric-modulated arc therapy (VMAT, RapidArc). We began the clinical use of VMAT in January 2010. In total, 67.6% of the IMRT treatment was VMAT. This study focused on the point dose measurements. The ion chamber used in study was calibrated relative to the absorbed dose to water for reference beam quality, using  $^{60}\text{Co}$  gamma-rays according to Japanese dose protocol [13]. The types of chamber used were PTW model 30013 (the 'farmer' chamber, active volume  $0.6\text{ cm}^3$ ) and PTW model 31016 (the 'pin-point' chamber, active volume  $0.016\text{ cm}^3$ ). The location of the point dose measurements were determined in a high-dose region with reference to the dose distribution calculated by the TPSs. The charge collected by the ionization chamber was measured using electrometers: a UNIDOSwebline (PTW, Freiburg, Germany) or a RAMTEC Smart (Toyo Medic Co., Ltd, Tokyo, Japan). The charge was corrected for temperature, pressure, and linac output variations. Additionally, linac output variance was adjusted every week, and the variance was within 1%. Depending on the target size and treatment site, the type of the chamber and the phantom were selected. Four types of phantom were used in our institute: a cylindrical phantom (custom-made), an IMRT Head & Torso Freepoint Phantom (Computerized Imaging Reference Systems, Virginia, USA), a spherical phantom (custom-made) and a cubic phantom (Tough water, Kyoto Kagaku Co. Ltd, Kyoto, Japan). All of the phantoms were made of tissue-equivalent homogeneous material. Eclipse (Varian Medical Systems, Palo Alto, CA, USA) was employed as the TPS. In the evaluation period, the TPS version used changed as follows: version 6.0 was used until December 2009, version 8.6 from January 2010 to December 2011, version 10.0 from January 2012 to December 2013, and version 11 from January 2014 to the present. The measurement dose by the ion chamber was compared with the calculated dose by the TPS. The calculated dose was defined as the mean dose in the contour segmentation of a sensitive volume of the ion chamber. Then, each IMRT plan was copied to each phantom in which the sensitive volume of the chamber was contoured, and the dose distribution was calculated with the TPS. Thus, the type of chamber was determined by the flatness of the dose distribution in the sensitive volume of each chamber. If the standard deviation of the dose distribution in the sensitive volume of the farmer chamber was within 2%, it was chosen in the measurement. Otherwise, the pin-point chamber was chosen. In the brain IMRT plan, the spherical phantom and the

cubic phantom were used. In prostate, gynecology, whole-pelvis and prostate, abdomen, and bone IMRT plans, the IMRT Head & Torso Freepoint Phantom and the cubic phantom were used. In the head-and-neck IMRT plan, the cylindrical phantom and the cubic phantom were used. The measurement point in the QA was defined at the dose-stable area with high or low dose. Therefore, the measurement point was not always at the isocenter of the IMRT plan.

According to the ESTRO group's proposed value for the tolerance limit [8], treatment plans can be passed if the measured-to-calculated ion chamber dose agrees within  $\pm 3\%$ . To establish a new management method, we analyzed those QA results. Bartlett's test was used to examine whether the results of patient QA had a normal distribution in each of the linacs and in each of the phantoms. From the results, we used one-way analysis of variance (ANOVA) as a parametric test and the Kruskal-Wallis test as a non-parametric test. If the QA results showed significant differences between the phantoms, Bartlett's and ANOVA or Kruskal-Wallis tests were used to analyze the data from each treatment site to verify whether the QA results showed significant differences between the treatment sites as well as between the phantoms. The treatment sites were classified into eight sites: brain, prostate, gynecology, whole pelvis and prostate, head and neck, abdomen, vertebra, and 'others'. In order to identify a reason for the QA results showing a significant difference, a comparison of frequency of usage of the phantoms was performed with the Pearson  $\chi^2$  test.

Dong *et al.* and Pulliam *et al.* showed that the QA result did differ significantly between the treatment sites [1, 2]. Thus, relationships between the treatment site and the average speed of the multileaf collimator (MLC) motion in each IMRT plan, or the average gap width of each MLC pair in each IMRT plan, were analyzed. The speed of the MLC motion and the gap width were calculated for each DICOM plan. The speed of the MLC motion and the gap width of the MLC were calculated as follows:

$$(\text{average MLC speed}) = \frac{\sum_{i,m}^{n-1} ((P_{i+1,m} - P_{i,m})^2)^{0.5}}{\sum_{i,m}^{n-1} (((GA_{i+1} - GA_i)^2)^{0.5} \times GS_i)},$$

$$(\text{average MLC gap width}) = \sum_{i,m}^{n,l/2} (OP_{i,m} - P_{i,m}) / (n/2 \times l),$$

where  $n$  is the number of control points,  $l$  is the number of MLCs within the jaw size,  $P$  is the MLC position in each control point and each MLC within the jaw size,  $GA$  is the gantry angle at each control point,  $GS$  is the gantry speed at each control point, and  $OP$  is the MLC position on the opposing side of the MLC to  $P$ . In order to evaluate the relationship between the results of QA and the two parameters, the average MLC speed and average gap width were converted to one dimensional (1D) data using principal component analysis. Then, correlation between the converted 1D data and the results of patient-specific QA was evaluated. Additionally, in order to compare the difference in field dosimetry skills between clinical medical physicists, the results of their QA were compared. From May 2015 to May 2016, 499 plans that were used to treat patients were analyzed to avoid a potential limitation (the planning protocol

was the same in each treatment site during that period). All of the delivery techniques were VMAT, and the same TPS version was used (version 11.0). Bartlett’s test was used to examine whether the average MLC speed and the average gap width for each of the treatment sites and the results of the QA for each of the clinical medical physicists were normally distributed. Using these results, we used one-way ANOVA as a parametric test and Kruskal–Wallis test as a non-parametric test. Since the maximum gantry rotation speed in linac E was faster than in the others, the average speed of the MLC motion in the TrueBeam was omitted from analyses in which the average speed of the MLC motion was involved. A  $P$ -value  $< 0.05$  was considered to be statistically significant.

**RESULTS**

In all cases, the overall mean difference between the measured and the calculated dose was  $0.5 \pm 1.3\%$  (mean  $\pm$  SD). Figure 1 shows the distribution of the point dose differences for all cases. The pass rate for the tolerance limit was 97.7% for the point dose measurement. Figure 2 shows the trend in the pass rate year by year. The pass rate improved from 2006 to 2014. The mean dose differences between the measured and calculated doses in linacs A to E were  $0.6 \pm 1.2\%$ ,  $0.9 \pm 1.3\%$ ,  $-0.4 \pm 1.4\%$ ,  $-0.1 \pm 1.2\%$  and  $-0.1 \pm 0.9\%$  for the whole period. Figure 3 shows the mean dose difference for each linac. The QA result for each linac was not normally distributed ( $P < 0.001$ ). Therefore, we used a Kruskal–Wallis test as a non-parametric test. It was found that the mean dose difference between the measured and calculated doses differed significantly between linacs ( $P < 0.001$ ). On the other hand, the QA results for each phantom

were normally distributed ( $P = 0.14, 0.20, 0.62, 0.40$  and  $0.12$  for linacs A, B, C, D and E, respectively). The mean dose difference between the measured and calculated doses for each phantom differed significantly between linacs A, B, C and D ( $P < 0.01$ ), using an ANOVA test. However, in linac E, the measured and calculated doses did not differ ( $P = 0.22$ ). Therefore, the QA result for each treatment site was analyzed for linac E. The QA result for linac E was not normally distributed ( $P < 0.001$ ), and the mean dose difference differed significantly ( $P = 0.03$ ) between each treatment site. In addition, the QA result in linac B was not normally distributed ( $P < 0.001$ ), and the mean dose difference differed significantly ( $P < 0.001$ ) between each treatment site. Figure 4 compares the mean dose difference for the treatment sites in linac E (Fig. 4a) and linac B (Fig. 4b).

The proportions of the frequency of usage of the cylindrical phantom, the IMRT Head & Torso Freepoint Phantom, the spherical phantom and the cubic phantom were 72.4, 10.4, 17.2 and 0.0% in linac A, 35.7, 58.8, 4.3 and 1.2% in linac B, 39.9, 54.8, 1.9 and 3.4% in linac C, 31.9, 19.1, 48.9 and 0.0% in linac D, and 43.0, 57.0, 0.0 and 0.0% in linac E, respectively. Frequency of usage of the phantoms had a significant difference ( $P < 0.001$ ) in each phantom. The average speeds of the MLC motion for brain, prostate, gynecology, whole pelvis and prostate, head and neck, abdomen, bone, and others were  $6.3 \pm 1.2$ ,  $5.5 \pm 0.7$ ,  $6.4 \pm 1.6$ ,  $7.5 \pm 0.7$ ,  $6.8 \pm 1.2$ ,  $6.9 \pm 1.1$ ,  $5.0 \pm 1.3$  and  $6.2 \pm 1.2$  mm/s, respectively. Figure 5 shows the average speeds of the MLC motion for each treatment site, which were not normally distributed ( $P < 0.001$ ). Therefore, we used a Kruskal–Wallis test as a non-parametric test. It was found that the average speed showed a significant difference between each treatment site ( $P < 0.001$ ). The average gap widths

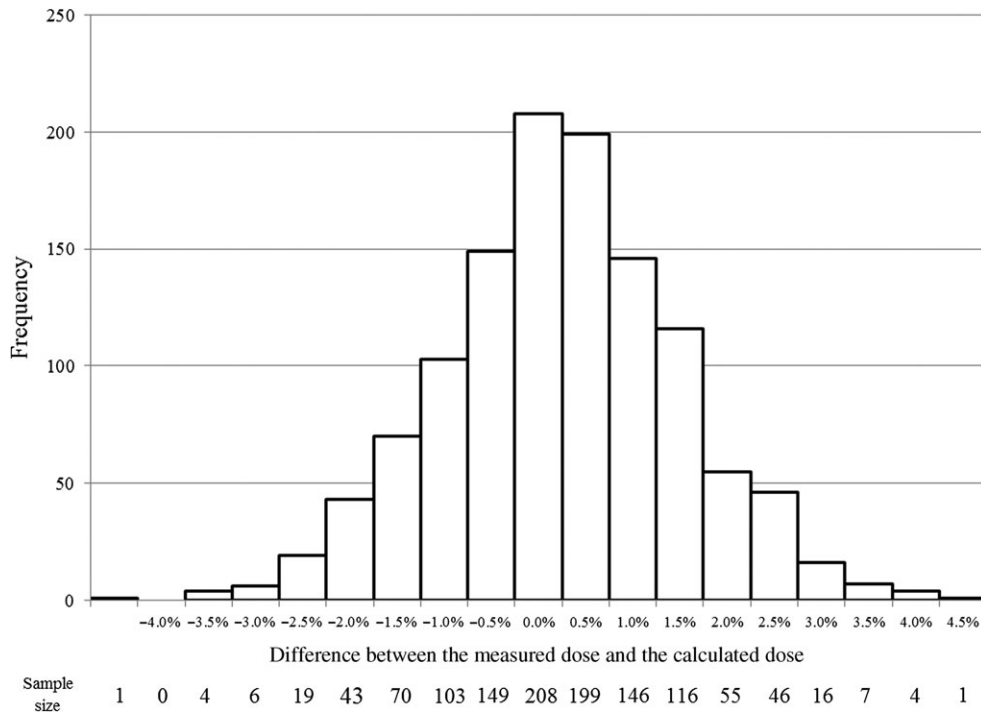
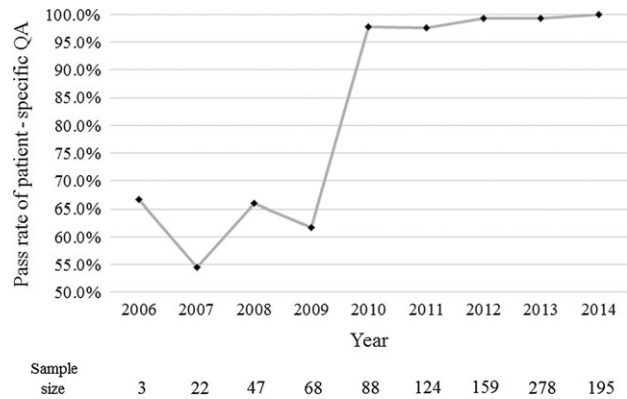


Fig. 1. Distribution of the percentage dose difference between the dose measured with an ion chamber and the dose calculated with a treatment planning system.

for brain, prostate, gynecology, whole pelvis and prostate, head and neck, abdomen, bone, and others were  $41.2 \pm 11.8$ ,  $19.3 \pm 4.6$ ,  $46.1 \pm 19.9$ ,  $39.6 \pm 9.6$ ,  $36.7 \pm 10.3$ ,  $38.8 \pm 16.4$ ,  $27.6 \pm 12.9$  and  $32.2 \pm 15.5$  mm, respectively. Figure 6 shows the average gap width for each treatment site. The average gap widths for the various treatment sites were not normally distributed ( $P < 0.001$ ). Therefore, we used a Kruskal–Wallis test as a non-parametric test. It was found that the average gap width showed a significant difference between each treatment site ( $P < 0.001$ ). The converted 1D data from the average MLC speed and the average gap width were compared with the results of the patient-specific QA in the head and neck because the number of QA results for the head and neck was greatest. The coefficient of correlation between the results of the QA and the converted data was  $-0.19$ . Thus, the correlation was not strong.

The mean dose difference for the three clinical medical physicists was  $0.2 \pm 1.2$ ,  $0.0 \pm 1.1$  and  $0.2 \pm 1.1\%$ , respectively, and the

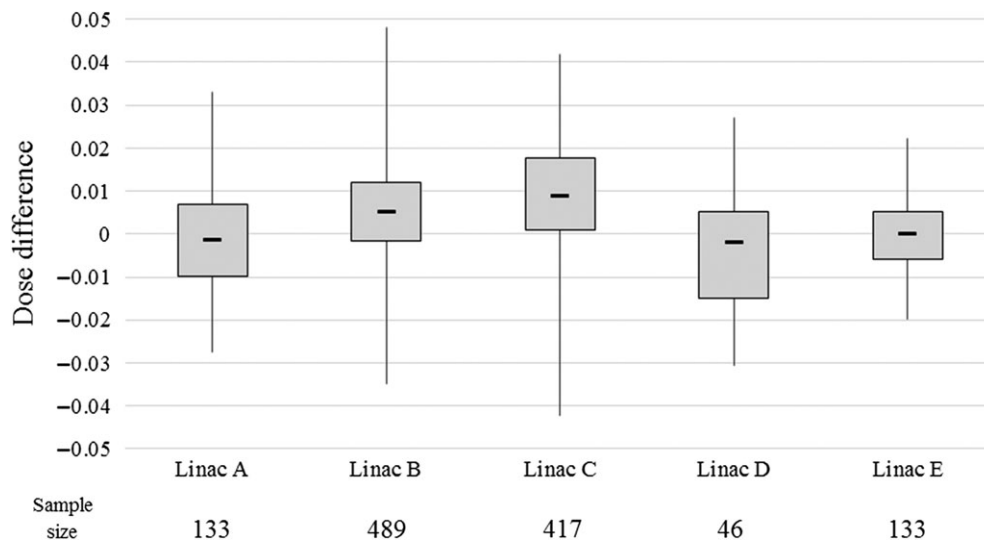


**Fig. 2. Pass rate for the patient-specific QA for all linacs. The criteria was the tolerance limit proposed by the ESTRO group, i.e. treatment plans can be passed if the measured and calculated ion chamber doses agree within  $\pm 3\%$ .**

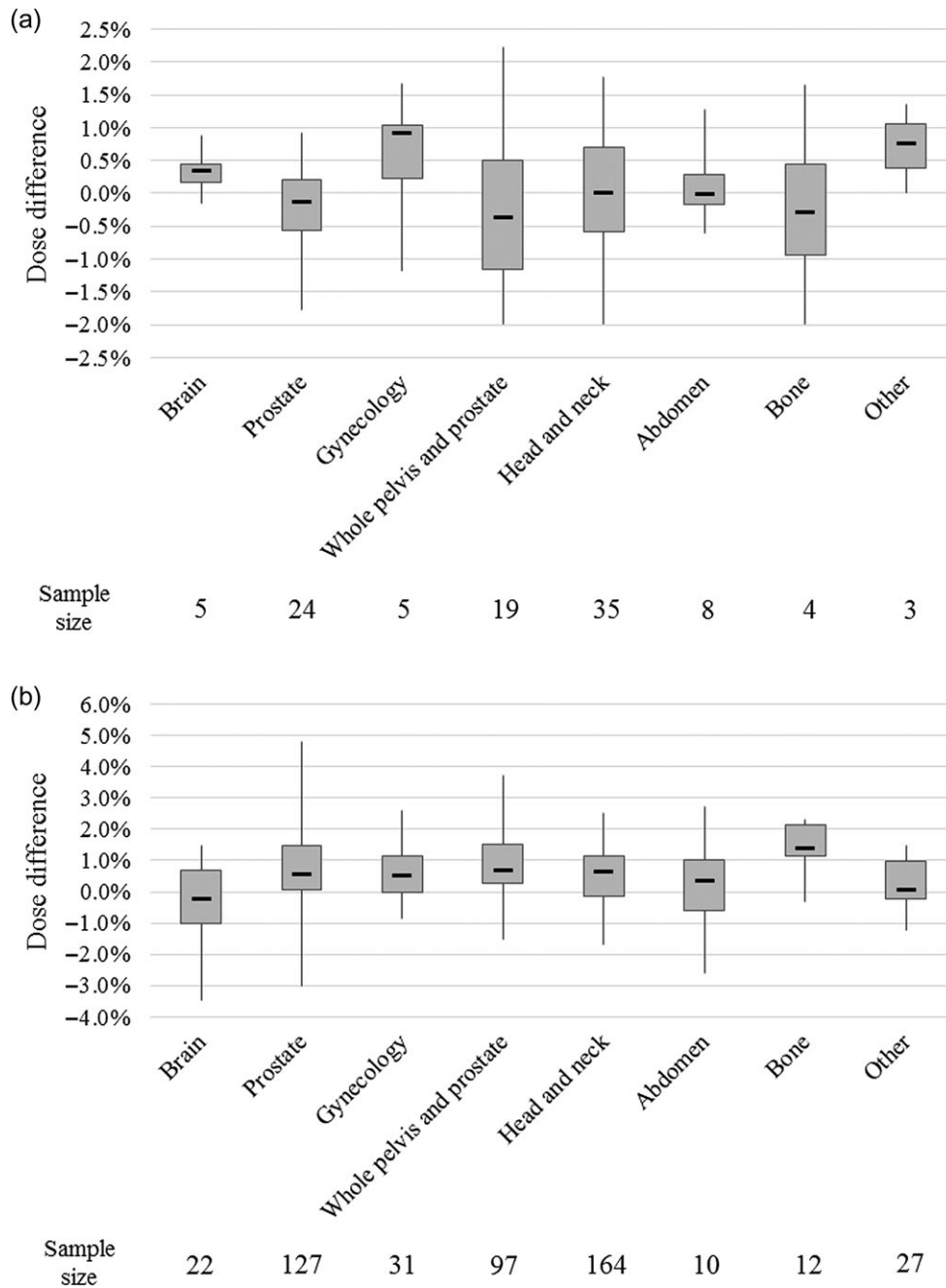
QA results for each clinical medical physicist were normally distributed ( $P = 0.57$ ). The mean dose difference between the measured and calculated doses obtained by each clinical medical physicist did not differ significantly from that of the other clinical medical physicists ( $P = 0.16$ ), as determined from an ANOVA test.

**DISCUSSION**

The pass rate of the patient-specific QA improved almost every year. This result corresponded with the result of Pulliam *et al.* [2]. The pass rate dramatically improved in 2010. This was associated with an upgraded version of the TPS and revision of the parameters of the multileaf collimator. This is the first study to reveal that the results of patient-specific QA depend on the treatment machine. There may be a number of reasons for this. First, it might be associated with the commissioning method for each linac. Szpala *et al.* reported that a single value for the partial transmission through rounded leaf ends did not exist [14]. In our institution, virtual leaf parameters, which are adjusted to decrease the difference between the measured dose and the calculated dose, have been applied to each linac in the TPS. Therefore, the commissioning accuracy affects the tolerance limit because the sum of the average deviation contribute to it. According to the ESTRO group, the confidence limit is defined as the sum of the average deviation and 1.96 SDs. Kielar *et al.* reported that the ideal dosimetric leaf gap parameter depended on the target size, and that it was determined by the actual treatment plan verification to ensure dosimetric accuracy [15]. Thus, the ideal parameter depended on the conditions of the actual treatment plan, such as field size and target size. Therefore, the QA result depends on the commissioning method. Second, the machine performance (e.g. maximum gantry speed) differed between the Clinac series and TrueBeam. The Clinac series cannot apply a Jaw-Tracking VMAT technique, whereas the TrueBeam can. Third, the results of the patient-specific QA for each linac might be subliminally separated by the treatment sites because the treatment site was considered when selecting the linac and the beam energy. The



**Fig. 3. Box plot of the mean difference between the measured and calculated doses in each medical linac.**



**Fig. 4. (a) Box plot of the mean difference between the measured and calculated doses at each treatment site in linac E, (b) Box plot for the mean difference between the measured and calculated doses at each treatment site in linac B.**

beam energy for each linac is different. The beam energy is 6 and 10 MV X-rays in linac A, 6 and 15 MV X-rays in linacs B and C, 4 and 10 MV X-rays in linac D, and 4, 6, 10 and 15 MV X-rays (and 6 and 10 MV X-rays without flattening filter) in linac E. The proportions of the patient-specific QA for each treatment site were not the same for each linac because the frequency of usage of the phantom varied significantly in the present study. The difference in the proportion might be associated with the significant difference in

the results for patient-specific QA for each linac. Therefore, it is necessary that the results of patient-specific QA are managed separately for each linac because both the distribution and mean of the difference between the measured and calculated doses of the patient-specific QA showed a significant difference between linacs. The results of the patient-specific QA for each phantom differed significantly in four of the linacs. The other linac did not show a significant difference between each phantom—which might be

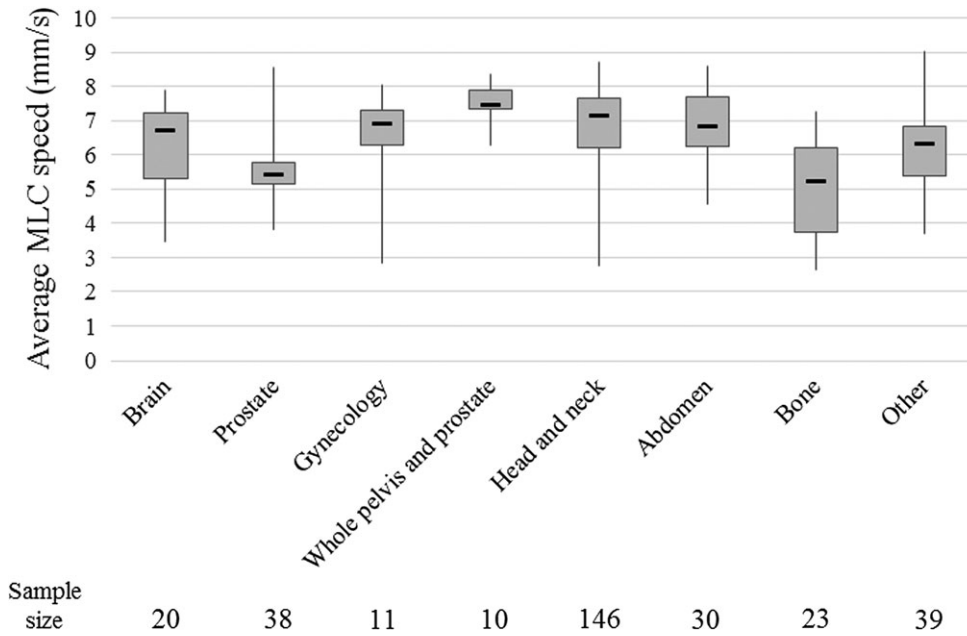


Fig. 5. Box plot of the plan’s average speed of the MLC motion within jaw size at each treatment site.

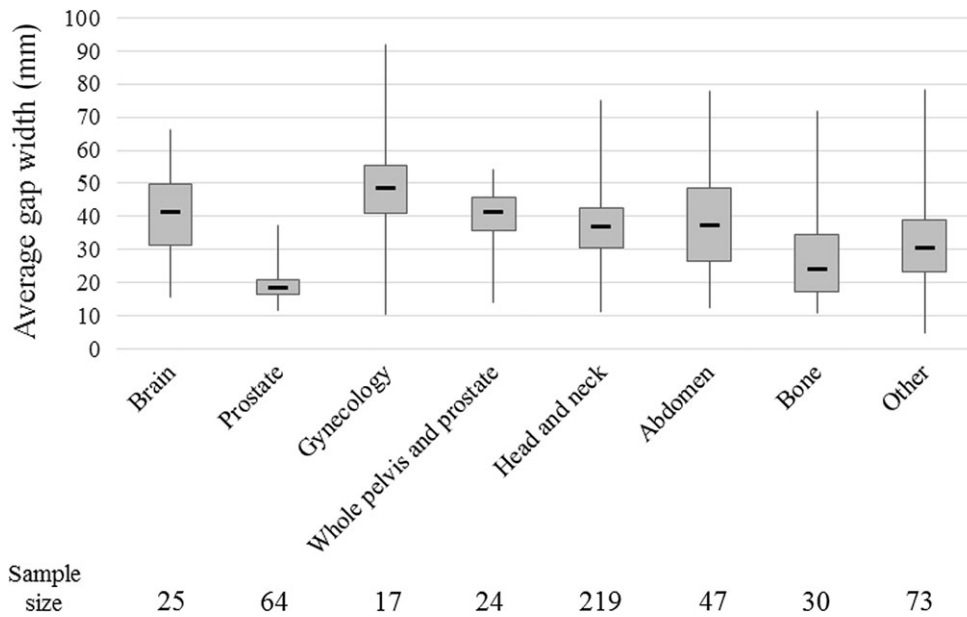


Fig. 6. Box plot of the plan’s average gap width of the MLC within jaw size at each treatment site.

associated with frequency of usage of the phantoms. Actually, in that linac, the results were significantly different for the various the treatment sites. It was suggested that the results of the QA should be managed for each treatment site because the type of phantom used in the QA was associated with the treatment site. These results corresponded with those of previous reports [1, 2]. Carlson *et al.* showed that the parameters of the MLC were used to express the plan complexity [16]. Additionally, Miura *et al.* showed that faster

MLC motion speed induced increasing dosimetric errors [17]. LoSasso *et al.* showed that the gap width of the MLC was also related to the dose error [18]. In this article, the average speed of the MLC and the average gap of the MLC differed significantly between treatment sites. Those parameters might be associated with the significant differences in the results obtained for QA between each treatment site. However, there was little correlation between the results of the QA and the converted 1D data from the average

speed of the MLC motion and the average gap width. This suggests that the results of patient-specific QA should be considered parameters for measurement as well as the plan parameters, though the plan parameters might need to be revised.

According to the results of patient-specific QA, the skills of the medical physicists in our institute were almost the same in terms of QA process. As a limitation of the results, it was conceivable that the materials and methods used in the QA process varied. For example, depending on the plan information, we used several chambers, electrometers, and phantoms. Thus, a small amount of uncertainty due to difference in dosimetric tools was included in the total uncertainty of the results.

Therefore, the management method needs to consider the accurate measurement. These results for the management method of patient-specific QA are important for both deciding the tolerance limit and also for implementation of statistical approaches. As a side note, in order to consider the accurate measurement, the management method requires that medical physicists maintain high-level skills. Additionally, statistical results for patient-specific QA in many institutes are important for more adequately defining the tolerance limit and action level.

### CONCLUSION

The results of patient-specific QA should be managed separately for each medical linac and each treatment site. Our approach enabled us to define the recommended tolerance limit. In the management protocol, the tolerance limit was established as 1.96 SDs, as determined by the result of the QA for each medical linac and each treatment site. However, the tolerance limit should be  $\pm 3\%$  because of the recommendation from the ESTRO group when there is insufficient data from the patient QA to calculate 1.96 SDs. Additionally, reports on the statistical results of patient-specific QA will be helpful if institution define a new tolerance limit and action levels.

### FUNDING

This work was partially supported by a JSPS Grant-in-Aid for Young Scientists (B) Grant Number 26860410, by the Practical Research for Innovative Cancer Control and by the Medical Research and Development Programs Focused on Technology Transfer: Development of Advanced Measurement and Analysis Systems (SENTAN) from the Japan Agency for Medical Research and Development, AMED (16ck0106039h0003 and 16hm0102046s0504), and by the National Cancer Center Research and Development Fund (26-A-28).

### CONFLICT OF INTEREST

The authors state that there are no conflicts of interest.

### REFERENCES

- Dong L, Antolak J, Salehpour M, et al. Patient-specific point dose measurement for IMRT monitor unit verification. *Int J Radiat Oncol Biol Phys* 2003;56:867–77.
- Pulliam KB, Followill D, Court L, et al. A six-year review of more than 13,000 patient-specific IMRT QA results from 13 different irradiation sites. *J Appl Clin Med Phys* 2014;15:196–206.
- Carrasco P, Jornet N, Latorre A, et al. 3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification. *Med Phys* 2012;39:5040–9.
- Pecharrómán-Gallego R, Mans A, Sonke JJ, et al. Simplifying EPID dosimetry for IMRT treatment verification. *Med Phys* 2011;38:983–92.
- Stathakis S, Myers P, Esquivel C, et al. Characterization of a novel 2D array dosimeter for patient-specific quality assurance with volumetric arc therapy. *Med Phys* 2013;40:071731.
- Li G, Bai S, Chen N, et al. Evaluation of the sensitivity of two 3D diode array dosimetry systems to setup error for quality assurance (QA) of volumetric-modulated arc therapy. *J Appl Clin Med Phys* 2013;14:13–24.
- Hussein M, Rowshanfarzad P, Ebert MA, et al. A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems. *Radiother Oncol* 2013;109:370–6.
- Alber M, Broggi S, Wagter CD, et al. Guidelines for the verification of IMRT. *ESTRO Booklet No. 9*. Brussels, Belgium: ESTRO, 2008, pp. 89–106.
- 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.
- Palaniswaamy G, Brame RS, Yaddanapudi S, et al. A statistical approach to IMRT patient-specific QA. *Med Phys* 2012;39:7560–70.
- Gérard K, Grandhaye JP, Marchesi V, et al. A comprehensive analysis of the IMRT dose delivery process using statistical process control (SPC). *Med Phys* 2009;36:1275–85.
- Breen SL, Moseley DJ, Zhang B, et al. Statistical process control for IMRT dosimetric verification. *Med Phys* 2008;35:4417–25.
- Morishita Y, Kato M, Tanaka N, et al. A standard for absorbed dose rate to water in a  $^{60}\text{Co}$  field using a graphite calorimeter at the National Metrology Institute of Japan. *Radiat Prot Dosimetry* 2013;154:331–9.
- Szpala S, Cao F, Kohli K. On using the dosimetric leaf gap to model the rounded leaf ends in VMAT/RapidArc plans. *J Appl Clin Med Phys* 2014;15:67–84.
- Kielar KN, Mok E, Hsu A, et al. Verification of dosimetric accuracy on the TrueBeam STx: rounded leaf effect of the high definition MLC. *Med Phys* 2012;39:6360–71.
- Carlson JN, Park JM, Park SY, et al. A machine learning approach to the accurate prediction of multi-leaf collimator positional errors. *Phys Med Biol* 2016;61:2514–31.
- Miura H, Tanaooka M, Inoue H, et al. DICOM-RT plan complexity verification for volumetric modulated arc therapy. *J Med Phys Clin Eng Radiat Oncol* 2014;3:117–24.
- LoSasso T, Chui CS, Ling CC. Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy. *Med Phys* 1998;25:1919–27.