ORION SCHOLAR JOURNALS

International Journal of Scientific Research Updates

Journal homepage: https://orionjournals.com/ijsru/

ISSN: 2783-0160 (Online)

(RESEARCH ARTICLE)

Check for updates

Bland and Altman analysis: Machine performance check output

Aime M. Gloi *, Vladimir Stankovich, Benjamin Rodriguez and Stanley Mayas

Genesiscare, Radiation Oncology Modesto, California, USA.

International Journal of Scientific Research Updates, 2023, 05(02), 069–075

Publication history: Received on 24 April 2023; revised on 06 June 2023; accepted on 09 June 2023

Article DOI: https://doi.org/10.53430/ijsru.2023.5.2.0046

Abstract

Background: The aim of this study was to evaluate the variation in the daily output of a linac Varian Edge TrueBeam over a month, using a Daily QA3 (QA3) and machine performance check (MPC), after Task Group 51protocol and baseline data had been established.

Methods and Materials: The daily output data were collected for all photons (6, 10 MV, and 6, 10 FFF) and electrons (6, 9, 12, and 15 MeV) energies. A Bland-Altman-Plot and intraclass correlation coefficients were used to assess agreement and reliability.

Results: MPC data showed small random deviations of ≤ 2% and agreed with the QA3 measurements. The maximum mean absolute difference between the QA3 and MPC data was observed with 6 MeV (0.5633), with limits of agreement of (−0.5392–1.6658).

Conclusion: These results suggest that MPC data closely track QA3 data and that the two forms are interchangeable.

Keywords: Machine performance check (MPC); Bland-Altman method Daily QA3

1 Introduction

In radiation therapy, the purpose is to precisely deliver a prescribed dose to a tumor, while protecting the surrounding normal tissue outside the target. To ensure accuracy and proper dose delivery, linear accelerator (linac) quality assurance (QA) testing is performed daily, weekly, monthly, annually, and contrasted to the relevant guidelines[1, 2, 3].

Usually, QA tests of linac output are carried out on the Daily QA3 (QA3) before any patient treatment. Recently, a novel system has been added to the TrueBeam Edge 2.0 platform named machine performance check (MPC). It is an application that relies on a fully integrated and automated imaging system comprised of an electronic portal imaging device (EPID), kilovoltage (kV), megavoltage (MV), and an on-board imager (OBI). MPC uses the EPID system to assess beam output constancy. Several studies were undertaken to validate MPC with other output measurement devices. Barnes et al[4] evaluated MPC beam output change with an ionization chamber (IC) whereas others [5, 6] were focused on geometry tests using (IsoCal) phantom.

In this study, the variation in daily output in terms of photon and electron energies was monitored using the QA3 system and MPCs over a month. The interchangeability between QA3 and MPC data was assessed using Bland-Altman approach [7], which is commonly employed to investigate and visualize the agreement between the quantitative outcomes of two methods. While this method is very valuable for two devices studies, there are several limitations to the classical Bland-Altman plot.

* Corresponding author: Aime M. Gloi

MPC tracks a wide range of linac parameters automatically, including the isocenter, output, collimation, gantry, couch, and multi-leaf collimator. A variety of these factors have been shown to act as predictors and warnings for subsequent failures. Hence, the purpose of this work was to evaluate one of these parameters: the variation in the daily output of a linac TrueBeam over a month, using a QA3 and MPC. This report is intended to contrast the MPC output against standard QA3 and offer an insight into MPC output stability and sensitivity to linac drifts.

2 Material and methods

Output measurements from an Edge TrueBeam (Varian Medical Systems, Palo Alto) linac were carried out on clinically available photons (6 and 10 MV, 6 and 10 FFF) and electrons (6, 9, 12, 15 MeV). The linac was calibrated annually using the Task Group 51 protocol [8]. Subsequent baseline data for the QA3 system (Sun Nuclear Corporation; Melbourne, USA) and MPC were acquired. Finally, daily output constancy checks were performed over thirty days using QA3 and MPC.

The interchangeability of the output variation measurement between MPC and QA3 was assessed using the Bland-Altman method. This analysis determines the bias, or the mean difference between the two devices, as a measure of accuracy. The limits of agreement (LoA) were said to be good if 95% of the data lies within 1.96 SD of the mean difference between the two methods and given by:

$$\text{LoA} = (\text{bias}) \pm t_{\alpha,\, n-1} * \sigma \quad \text{………..(1)}$$

where σ is the standard deviation of the differences, n is the sample size, and $t_{\alpha,\, n-1}$ is the t-value corresponding to the degree of freedom ($n-1$) for a type I error (α) of 0.05. In general, the LoAs are defined as the mean difference ± 1.96 σ. The confidence interval (CI) for the bias defines the limits for the bias in the target population, whereas the LoA refers to the spread of the differences in a specific study. The CI of the bias is calculated as follows:

$$\text{CI} = (\text{bias}) \pm t_{\alpha,\, n-1} * \sigma / \sqrt{n} \quad \text{…………….(2)}$$

CI decreases with increasing sample size. In contrast, the LoA does not decrease as the sample size increases. However, this method is subject to multiple sources of variability. The main idea is to ensure that the variability between the measurements is small enough to allow us to use observations from a single device or to use another device in case of malfunction.

The normality of the MPC data for all energies was assessed based on the Anderson-Darling (AD) [9] statistic, which can be used as an alternative to the Kolmogorov-Smirnov test to check for the data distribution behavior based on the following hypotheses:

H_0: Data are sampled from a population that is normally distributed.

H_1: Data are sampled from a population that is not normally distributed. The decision to reject the null hypothesis (H_0) is dependent on the p-value with a specified significance level of 5%. For a p-value greater than 0.05, the test is considered normally distributed.

The equation for AD is as follows:

$$AD^* = AD\left(1 + \frac{0.75}{N} + \frac{2.25}{N^2}\right) \quad \text{…………(3)}$$

where $F(yi)$ is the cumulative distribution function for the specified distribution

A two-way ANOVA without replication and a t-test were used to analyze daily output variation between the QA3 and MPC data. The precision of the QA3 and MPC data was calculated as follows:

$$\text{Precision} = \sqrt{2} * \frac{(\sigma_{QA3} + \sigma_{MPC})}{2} \text{………………(4)}$$

where σ_{QA3} and σ_{MPC} are the standard deviations of the data collected from QA3 and MPC, respectively. The precision is dependent on the σ of the original measurements.

Furthermore, two-way mixed effect intraclass correlation coefficients (ICC) were conducted comparing daily output variation between QA3 and MPC based on the following criteria: (slight, ICC ≤ 0.2; fair, 0.2 < ICC ≤ 0.4; moderate, 0.4 < ICC ≤ 0.6; good, 0.6 < ICC ≤0.8; excellent, ICC > 0.8). A p-value of less than 0.05 was considered significant. For clinical purposes, a 3 % limit difference was set between the two measurements.

3 Results

The output variation of the QA3 and MPC was graphically compared using a Bland-Altman plot. The mean difference is plotted against the average values, and the LoAs were determined. The normality of the mean differences between the QA3 and MPC datasets for all energies were examined using an AD test, resulting in a p-value of ≥ 0.05, as shown in Table 1. Hence, H_0 cannot be rejected, as there is insufficient evidence to conclude that the data does not follow a normal distribution. As a result, the mean difference dataset is uniformly distributed. An example of the normal probability plot is shown in Figure 1 for 6 MeV.

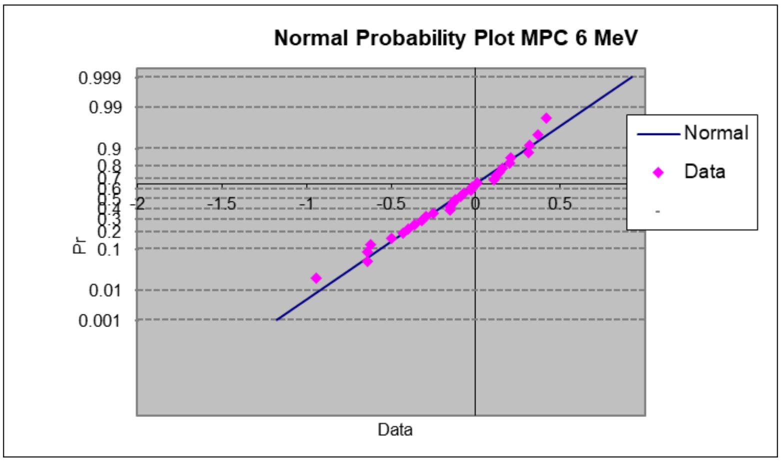


Figure 1 Normal Probability plot 6 MeV

Table 1 Normality test for MPC and QA3 mean difference

Energies	6MV	6FFF	10MV	10FFF	6e	9e	12e	15e
μ	0.049	0.065	0.1110	0.0530	0.5630	0.2170	0.2320	0.2990
σ	0.2880	0.436	0.3270	0.3120	0.5630	0.3740	0.3470	0.4020
AD test statistic	0.4596	0.9217	0.7032	0.3953	0.8485	0.5744	0.5338	0.6387
AD* test statistics	0.4722	0.9470	0.7225	0.4062	0.8718	0.5901	0.5484	0.6563
p-value	0.2439	0.0166	0.0595	0.3506	0.0255	0.1239	0.1582	0.0867

Table 2 summarizes the mean differences (μ) between the two pairs of measurements (QA3 and MPC) as well as their mean standard deviation (σ), the LoAs, and the CI from the Bland-Altman plots (Figure 2). These results show that over thirty days, the variation in the output of all energy beams as measured by QA3 and tracked by MPC was within 2%, with the highest variation at 6 MeV (0.5633) with LoAs of −0.5392 to 1.6658. The smallest mean difference is observed with 6 MV with LoAs of −0.516 to 0.6142, suggesting that the interchangeability was robust. The agreement between the QA3 and MPC measurements was assessed statistically using correlation, the intraclass coefficient, ANOVA, and a t-test (Table 3). A pairwise comparison and heat map visualization of the bias between QA3 and MPC is presented in Figures 3a and 3b for all energies. The legend is color-coded to illustrate the differences in correlation between and within energies. Further, a dendrogram associated with a heat map depicting hierarchical clustering is obtained using QA3 and MPC bias measurements. The Euclidean distances between clusters were determined to find the contribution of the individual energy to the average similarity between groups and within groups. A colored-coded Z-score (a numerical measurement that describes a value associated with the mean of a group of values) is shown in the legend.

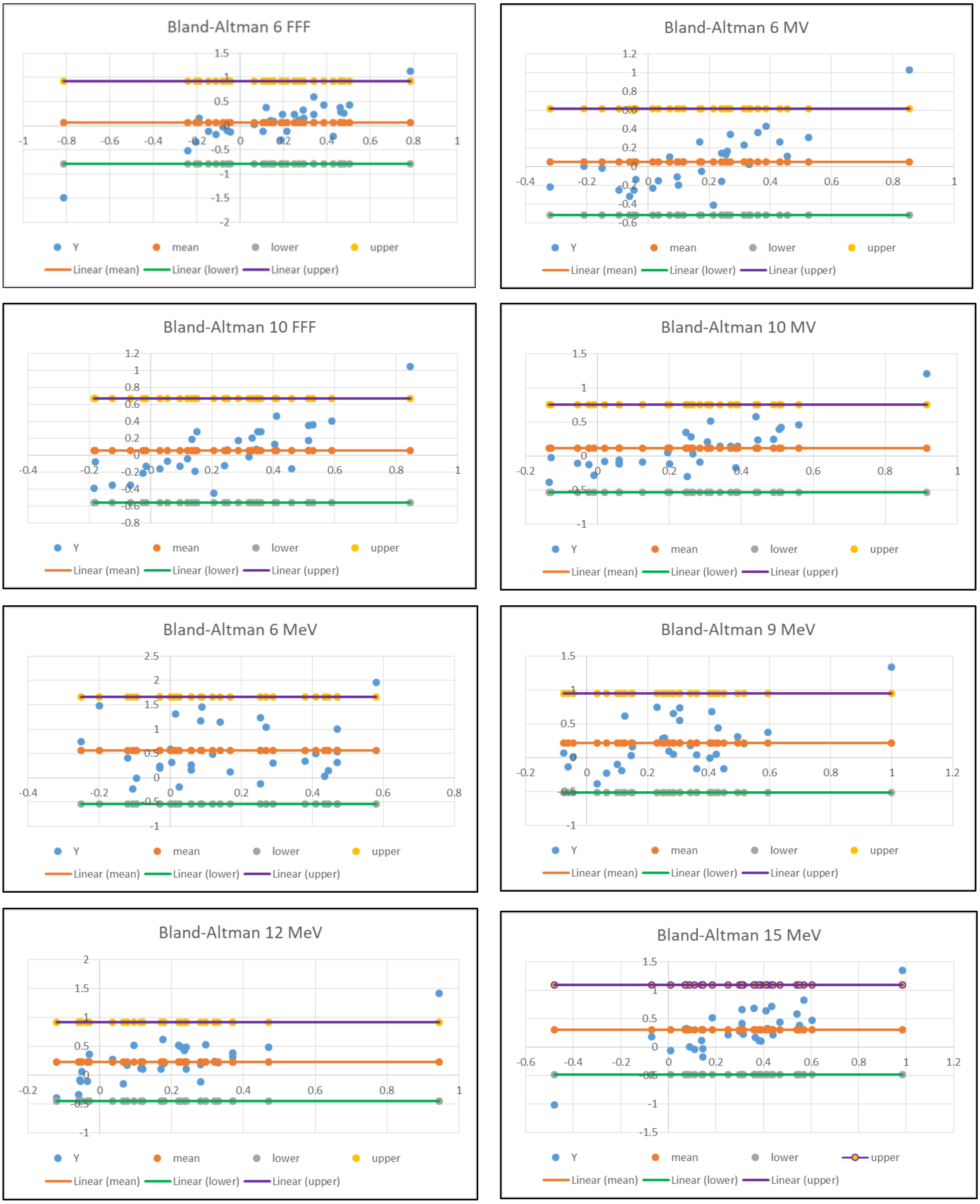


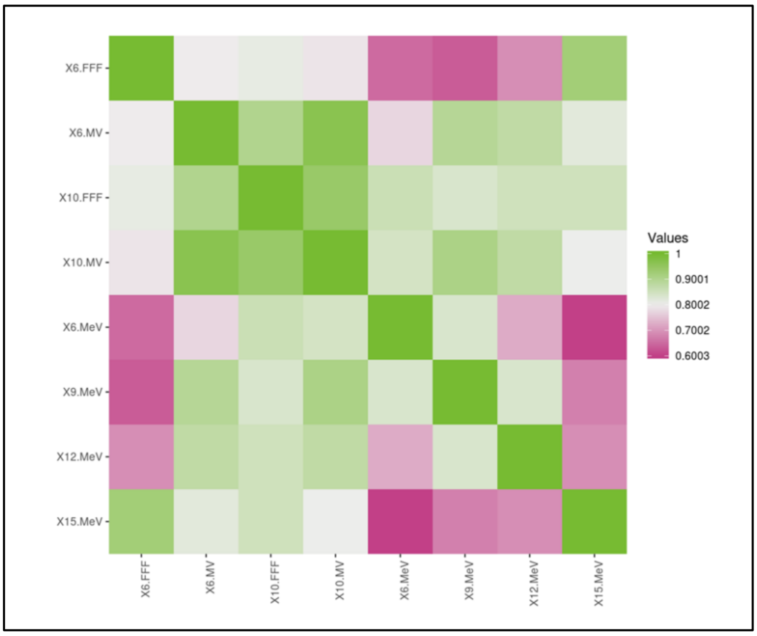
Figure 2 (a-h). Bland-Altman Plots showing the paired differences against the average for MPC and QA3

Table 2 Results of Bland–Altman bias and limits of agreement (LoA) for all dependent variables

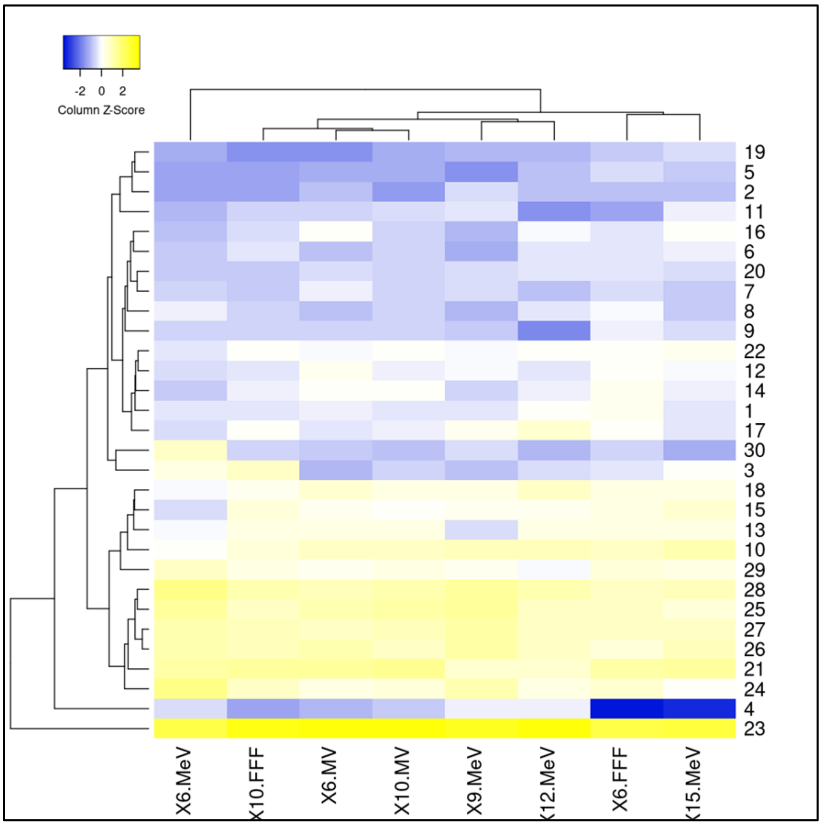
Energies	μ	σ	LoA	CI
6 FFF	0.0653	0.4356	-0.7886-0.9192	-0.0905-0.22143
6 MV	0.0490	0.2884	-0.5162-0.6142	-0.05420-0.1522
10 FFF	0.0533	0.3122	-0.5587-0.6653	-0.05841-0.16507
10 MV	0.1106	0.3266	-0.5296-0.7509	-0.00627-0.22754
6 MeV	0.5633	0.5625	-0.5392-1.6658	0.36203-0.7641
9 MeV	0.2170	0.34741	-0.5162-0.9502	0.08312-0.35087
12 MeV	0.2316	0.3471	-0.4487-0.9119	0.107379-0.35582
15 MeV	0.2986	0.4019	-0.4892-1.0865	0.15481-0.442515

Table 3 The numerical values of limits of agreement calculated based on tests that include (I) a t-test, (ii) ANOVA two without replication, (iii) ICC test, (iv) Pearson correlation, and (v) precision

Energies	Anova P-value	Pearson correlation	t-test P-values two-tail	ICCC	Precision
6 FFF	0.0345	0.5478	0.4181	0.3320	0.216375
6 MV	0.0033	0.6444	0.3597	0.6244	0.2510
10 FFF	0.0088	0.6119	0.3572	0.4235	0.3342
10 MV	0.0294	0.5352	0.0738	0.3325	0.36017
6 MeV	0.8766	-0.2159	6.6E-06	-0.0488	0.220146
9 MeV	0.1865	0.2012	0.0035	0.1362	0.401872
12 MeV	0.1488	0.3003	0.0010	0.1486	0.256727
15 MeV	0.0835	0.4259	0.0003	0.17528	0.415779



A



B

Figure 3 (a, b). Visualization of pairwise and heat map bias from Bland-Altman analysis between MPC and QA3

4 Discussion

One of the goals of this study was to compare and investigate the interchangeability between QA3 and MPC output variation. Bland-Altman plots were generated for each MPC derived energy output. The results indicated that more than 95% of the points were within the LoAs. The mean difference for all energies is less than 1%, as shown in Table 1, and has a clinically acceptable range. The weakness of the Bland-Altman method is that the consistency of measurement is generally determined visually, without any statistical tests. In this study, a paired t-test was performed to verify that the mean difference was statistically small enough. Strong positive correlations between the MPC and QA3 outputs were exhibited for all photons, and weak correlations for all electron energies (Table 3). A two-way ANOVA without replication revealed that the p-value for all photons was > 0.05, meaning that the variances were not significantly different between the QA3 and MPC outputs. In contrast, the p-values derived from the electron energies were < 0.05, suggesting statistical significance for the variance between MPC and QA3. As pointed out by Mostafa et al.[10], weak correlations for the electron energies may be due to noise and instability in the output, concealing trends that could be alleviated by increasing the data pool.

In summary, an ANOVA was used to extend the paired t-test method in terms of detecting bias in the measurement and showing data consistency within and per energy. An assessment of energy reliability using ICC coefficients gave results that were moderate too good for photon energy, and slight for electron energy. Other factors related to the environment, such as the relative humidity and temperature, can affect the output of the linac[11]. QA3 electronics is affected when used frequently, and as a result would influence the output. Our QA3 vs. MPC results compared well with those of Armoogum et al.[12] in which the MPC data consistently indicated a lower output (−1.08%) than the QA3 device. Similar results were found by Zao et al.[13] for different energies, with values of 0.26 ± 0.37, 0.12 ± 0.44, 0.33 ± 0.42, and 0.14 ± 0.43 for 6 MV, 10 MV, 15 MV, and 10 FFF, respectively. Our results were far better than those reported by Juneja et al.[14], who reported maximum and mean absolute differences between QA3 and MPC of 3.26% and 0.85 ± 0.61%, respectively. The linac output, as measured daily with QA3, was perfectly captured by MPC. Further investigation of the sensitivity of MPC in terms of detecting significant deviations in the output is warranted.

5 Conclusion

Finally, based on a Bland-Altman analysis, QA3 and MPC are interchangeable based on the mean difference, and ongoing research. As MPC becomes more widely accepted and mainstream, QA processes will become more stringent. We joined another group [15] to propose to re-baseline MPC if the difference between MPC and ionization chamber was >1.5% at the monthly comparison measurement, or if consistently >1% quarterly. We suggest that monthly ionization chamber output measurements should be maintained and a monthly QA3 should be added for verification. A threshold of 2% should be the limiting factor, which would ensure that the daily output is within the Task Group-142 protocol [2] tolerance of 3%.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest.

References

[1] Kutcher GJ, Coia L, Gillin M, Hanson WF, Leibel S, Morton RJ, et al. Comprehensive QA for Radiation Oncology Report of AAPM Radiation-Therapy Committee Task-Group-40. Med Phys. 1994; 21 (4):581–618. https://doi.org/10.1118/1.597316

[2] Klein E. TG-142: Quality Assurance for Medical Accelerators. Med Phys. 2011; 38(6). https://doi.org/ 10.1118/1.3613300

[3] Bissonnette J, Balter P, Dong Lei, et al. Quality assurance for image-guided radiation therapy utilizing CT-based technologies: a report of the AAPM TG-179. Med Phys. 2012; 39(4):1946–1963. DOI:10.1118/1.3690466

[4] Barnes, M. P., & Greer, P. B. (2017). Evaluation of the TrueBeam machine performance check (MPC) beam constancy checks for flattened and flattening filter-free (FFF) photon beams. Journal of applied clinical medical physics, 18(1), 139-150.

[5] McDermott GM, Buckle AH. Monitoring Linear Accelerator Output Constancy Using the Ptw Linacheck. Med Dosim. 2011; 36(1):71–4. https://doi.org/10.1016/j.meddos.2009.12.001 WOS:000287617100012. PMID: 20346645

[6] Speight RJ, Esmail A, Weston SJ. Quality assurance of electron and photon beam energy using the BQ-Check phantom. J Appl Clin Med Phys. 2011; 12(2):239–44. https://doi.org/10.1120/jacmp.v12i2. 3366 WOS:000290659500022. PMID: 21587183

[7] Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986, 327, 307–310. DOI: 10.1016/s0140-6736(86)90837-8

[8] PR. Almond, P.J. Biggs, B.M. Coursey, W.F. Hanson, M.S. Huq, R. Nath, and D.W.O. Rogers, AAPM Task Group 51: Protocol for clinical reference dosimetry of high energy photon and electron beams," Med. Phys. 26, 1847-1870 (1999). DOI: 10.1118/1.598691

[9] Pettitt A. Testing the normality of several independent samples using the Anderson-Darling statistic. J Appl Stat. 1977; 26(2):156-161. Doi: 10.2307/2347023

[10] Mostafa D, South C, Earley J. Assessing MPC for Output Checks. Medical Physics Department, St Luke's Cancer Centre, Royal Surrey County Hospital, Guildford, UK

[11] Bartolac S, Heaton R, Norrlinger B, Letourneau D: Seasonal variations in measurements of linear accelerator output. J Appl Clin Med Phys. 2019 Mar; 20(3): 81 88. doi: 10.1002/acm2.12548

[12] Armoogum K, Cornish G. EP-1749: Long-term stability of the Varian Truebeam Machine Performance Check (MPC). Radiotherapy and Oncology. 2018; 127(18): S937–S938. Doi: 10.1016/s0167-8140(18)32058-9

[13] Zhao B, Huang Y, Snyder K, Liu C, Chetty I, and Wen N. Feasibility of using machine performance check for daily photon output constancy, and isocentricity QA. Med Phys 2017; 44(6):2937-2938.

[14] Juneja, B; GAO, S; Balter, P; Nitsch, P. SU-F-T-480: Evaluation of the Role of Varian Machine Performance Check (MPC) in Our Daily QA Routine. DOI: 10.1118/1.4956665

[15] Li Y, Netherton T, Nitsch PL, Gao S, Klopp AH, Balter PA, et al. independent validation of machine performance check for the Halcyon and TrueBeam linacs for daily quality assurance. Journal of Applied Clinical Medical Physics. 2018; 19(5):375–382. doi:10.1002/acm2.12391