Title
Technical Note: Assessing the performance of monthly CBCT image quality QA

Journal
Medical Physics, 46(6)

ISSN
0094-2405

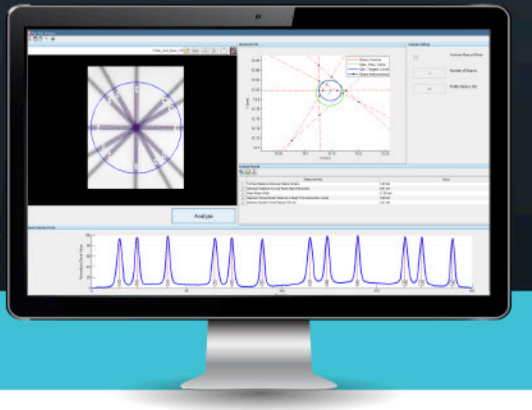
Authors
Manger, Ryan P
Pawlicki, Todd
Hoisak, Jeremy
et al.

Publication Date
2019-06-01

DOI
10.1002/mp.13535

Peer reviewed

COMPREHENSIVE MACHINE QA SOLUTIONS



FROM RADIOLOGICAL IMAGING TECHNOLOGY, INC.

RIT provides extensive machine QA capabilities, including comprehensive software packages that can be used to perform a full suite of measurements in accordance with TG-142, TG-148, and/or TG-135. RIT's automated routines allow you to perform daily, monthly, and annual QA with efficiency and precision, all while giving you confidence knowing your delivery performance is optimized.

TG-142: LINEAR ACCELERATORS

RIT is the single-vendor software solution that performs and trends every test recommended in TG-142.

Perform comprehensive quality assurance of Varian, Elekta, and all linear accelerators with confidence and ease, using your EPID and RIT software. Several packages feature RIT's popular 3D Winston-Lutz Isocenter Optimization routine that will help your SRS/SBRT delivery with its sub-millimeter, fully-automated accuracy. RIT offers automated tests for Star Shot Analysis, Radiation vs. Light Field, MLC accuracy, and many others.

TG-148: HELICAL TOMOTHERAPY®

RIT offers a comprehensive test suite for helical TomoTherapy® and Radixact® machines, in accordance with TG-148.

Designed with the TG-148 report in mind, the RITG148+ and RIT *Complete* software packages analyze the standardized tests for helical TomoTherapy machine QA. These include Static & Rotational Output Consistency, Jaw Centering and Alignment, Overhead Laser Positioning, Interrupted Treatment, and all others recommended for daily, monthly, and annual QA. The software will also analyze image quality using the TomoTherapy Cheese phantom.

TG-135: CYBERKNIFE® ROBOTIC RADIOSURGERY

RIT provides a comprehensive test suite for CyberKnife® and all robotic radiosurgery, in accordance with TG-135.

The RITG135 and RIT *Complete* software packages contain five fully-automated QA tests for CyberKnife® machines: End-to-End Test, AQA Test, Iris Test, Laser Coincidence, and MLC (Garden Fence) Test for the M6 Collimator. Combining a user-friendly interface with automated film detection algorithms, the software eliminates the need for manual manipulation or alignment of images, and drastically reduces the time required to perform these tests.

TomoTherapy®, Radixact®, and CyberKnife® are registered trademarks of Accuray, Inc.

VISIT **RADIMAGE.COM** TO DEMO RIT'S ADVANCED RANGE OF MACHINE QA, MLC QA, PATIENT QA, AND IMAGING QA ROUTINES.

+1(719)590-1077, OPT. 4

SALES@RADIMAGE.COM

Connect with us on social media
@RIT4QA

Technical Note: Assessing the performance of monthly CBCT image quality QA

Ryan P. Manger,[a] Todd Pawlicki, Jeremy Hoisak, and Gwe-Ya Kim
Department of Radiation Medicine and Applied Sciences, University of California San Diego, 3855 Health Sciences Dr., La Jolla, CA 92093, USA

(Received 14 January 2019; revised 11 March 2019; accepted for publication 2 April 2019; published 24 April 2019)

Purpose: To assess the performance of routine cone-beam computed tomography (CBCT) quality assurance (QA) at predicting and diagnosing clinically recognizable linac CBCT image quality issues.

Methods: Monthly automated linac CBCT image quality QA data were acquired on eight Varian linacs (Varian Medical Systems, Palo Alto, CA) using the CATPHAN 500 series phantom (The Phantom Laboratory, Inc., Greenwich, NY) and Total QA software (Image Owl, Inc., Greenwich, NY) over 34 months between July 2014 and May 2017. For each linac, the following image quality metrics were acquired: geometric distortion, spatial resolution, Hounsfield Unit (HU) constancy, uniformity, and noise. Quality control (QC) limits were determined by American Association of Physicists in Medicine (AAPM) expert consensus documents Task Group (TG-142 and TG-179) and the manufacturer acceptance testing procedure. Clinically recognizable CBCT issues were extracted from the in-house incident learning system (ILS) and service reports. The sensitivity and specificity of CATPHAN QA at predicting clinically recognizable image quality issues was investigated. Sensitivity was defined as the percentage of clinically recognizable CBCT image quality issues that followed a failing CATPHAN QA. Quality assurance results are categorized as failing if one or more image quality metrics are outside the QC limits. The specificity of CATPHAN QA was defined as one minus the fraction of failing CATPHAN QA results that did not have a clinically recognizable CBCT image quality issue in the subsequent month. Receiver operating characteristic (ROC) curves were generated for each image quality metric by plotting the true positive rate (TPR) against the false-positive rate (FPR).

Results: Over the study period, 18 image quality issues were discovered by clinicians while using CBCT to set up the patient and five were reported prior to x-ray tube repair. The incidents ranged from ring artifacts to uniformity problems. The sensitivity of the TG-142/179 limits was 17% (four of the prior monthly QC tests detected a clinically recognizable image quality issue). The area under the curve (AUC) calculated for each image quality metric ROC curve was: 0.85 for uniformity, 0.66 for spatial resolution, 0.51 for geometric distortion, 0.56 for noise, 0.73 for HU constancy, and 0.59 for contrast resolution.

Conclusion: Automated monthly QA is not a good predictor of CBCT image quality issues. Of the available metrics, uniformity has the best predictive performance, but still has a high FPR and low sensitivity. The poor performance of CATPHAN QA as a predictor of image quality problems is partially due to its reliance on region-of-interest (ROI) based algorithms and a lack of a global algorithm such as correlation. The manner in which image quality issues occur (trending toward failure or random) is still not known and should be studied further. CBCT image quality QA should be adapted based on how CBCT is used clinically. [https://doi.org/10.1002/mp.13535]

Key words: CBCT, control limits, quality assurance, tolerances

1. INTRODUCTION

The American Association of Physicists in Medicine (AAPM) Task Group (TG)-142[1] provides recommendations on the quality control (QC) of megavoltage radiation therapy equipment and builds on recommendations presented in the report of AAPM TG-40[2] by adding tests for newer technology such as multileaf collimators (MLCs) and on-board imaging (OBI) devices. Cone-beam computed tomography (CBCT) is one of the OBI technologies presented in this report along with daily, monthly, and annual metrics and discrete tolerances. The report of TG-179[3] provides further guidance on the QC of CT-based image guided radiation therapy (IGRT) systems.

Metrics presented in these reports are related to image quality — spatial resolution, contrast, Hounsfield Unit (HU) constancy, uniformity, and noise. The purpose of CBCT image quality assurance (QA) is to ensure that the CBCT image quality has not degraded beyond its performance at commissioning. The goal of CBCT image quality QA is to prevent clinically recognizable CBCT image quality issues.

Numerous commercial QA phantoms and software solutions have been developed to offer semi- or fully-automated analysis of CBCT image quality. Yet there are no peer-reviewed publications studying the sensitivity and specificity of these metrics to clinically recognizable image quality issues. CBCT image quality issues may present as ring artifacts, excessive image noise, non-uniformity, or other phenomena that result in degraded clinical utility which are easily detected by human observation.[4] The purpose of this work was to determine whether or not CBCT image quality QA is a good predictor of clinically recognizable image quality issues.

2. MATERIALS AND METHODS

Monthly automated CBCT image quality data were acquired on eight Varian linacs (two TrueBeam, two Trilogy, three 23iX, and a 21 iX) over the course of 34 months using the CATPHAN 500 series phantom (The Phantom Laboratory, Greenwich, NY). CBCT scans were acquired using Pelvis mode — half-fan with a half bowtie filter and a full trajectory. Slice thickness was set to 2.5 mm. The scan diameter or field of view varied depending on the linac from 25 to 30 cm. The CBCT datasets were analyzed using the CATPHAN QA module of Total QA system (Image Owl, Inc., Greenwich, NY) to determine the following CBCT image quality metrics: geometric distortion, spatial resolution, HU constancy, uniformity, and noise. The methods of analysis used in the CATPHAN QA module of Total QA are similar to those presented in the Catphan manual and can be accessed at http://help.imageowl.com.

Incident learning systems have been used in other studies for retrospective quality improvement.[4,5–8] The ILS employed in this study, Radiation Oncology Quality Reporting System — Machine Log (ROQRS-ML), is an in-house developed system used to communicate machine issues to clinical staff and the equipment vendor in real-time and to record past and ongoing machine issues and resolutions.[9] ROQRS-ML was queried for CBCT-related issues occurring between July 2014 and May 2017. These issues were primarily clinical in nature (e.g., a therapist reporting a ring artifact on a patient's CBCT). ROQRS-ML was used consistently for five the eight linacs during this 3-yr period. To address the three linacs where ROQRS-ML was used intermittently, the linac service reports were also queried for CBCT-related work orders (e.g., tube repair or parts replacement).

Table I. TG-142 recommended monthly tests for cone-beam computed tomography (kV and MV)

Metric	Non-SRS/SBRT	SRS/SBRT
Geometric distortion	≤2 mm	≤1 mm
Spatial resolution	Baseline	Baseline
Contrast	Baseline	Baseline
HU constancy	Baseline	Baseline
Uniformity and noise	Baseline	Baseline

HU, hounsfield unit; SRS/SBRT, stereotactic radiosurgery/stereotactic body radiotherapy.

The values of the CBCT image quality metrics from the prior monthly QA were used to assess the performance of CATPHAN QA. The only published recommendations for QC limits on CATPHAN QA are from AAPM TG-142 and TG-179 (Tables I and II). No recommendations are presented for low contrast resolution (contrast), HU constancy, uniformity, or noise because these values are manufacturer-dependent. In this study, the manufacturer-supplied values from acceptance testing were used as the QC limits for contrast resolution at 1% contrast (15 mm diameter), HU constancy (≤±40 HU), and uniformity (≤±40 HU). No QC limits were available for noise.

The QC limits were compared against the ROQRS-ML and service report data to determine their sensitivity and specificity. In this context, sensitivity is defined as the percentage of clinically recognizable CBCT image quality issues following a failing monthly CATPHAN QA. Results are categorized as failing if one or more image quality metrics are outside the QC limits. The specificity of CATPHAN QA was defined as one minus the fraction of failing CATPHAN QA results that did not have a clinically recognizable CBCT image quality issue in the subsequent month.

Since the performance of CBCT image quality QA cannot be fully examined with just one set of QC limits, receiver operating characteristic (ROC) curves were generated for each image quality metric by varying the QC limit and calculating the sensitivity and specificity over that range. True positive rate was defined as the sensitivity, and false positive rate (FPR) was defined as the specificity. The area under the curve (AUC) was calculated for each image quality metric to determine their relative performance.

Table II. TG-179 recommended computed tomography (CT)-based IGRT system image quality metrics.

Frequency	Metric	Tolerance
Monthly or upon upgrade	Scale, distance, and orientation accuracy[a]	Baseline
	High contrast spatial resolution[a]	≤2 mm (or ≤5 lp/cm) Baseline
	Low contrast detectability[a]	Baseline
	Uniformity and noise[a]	Baseline
Annual, if used for dose calculation	CT number accuracy and stability[a]	Baseline

[a]These tests can be performed on a semiannual basis after stability has been demonstrated, 6–12 months after commissioning.

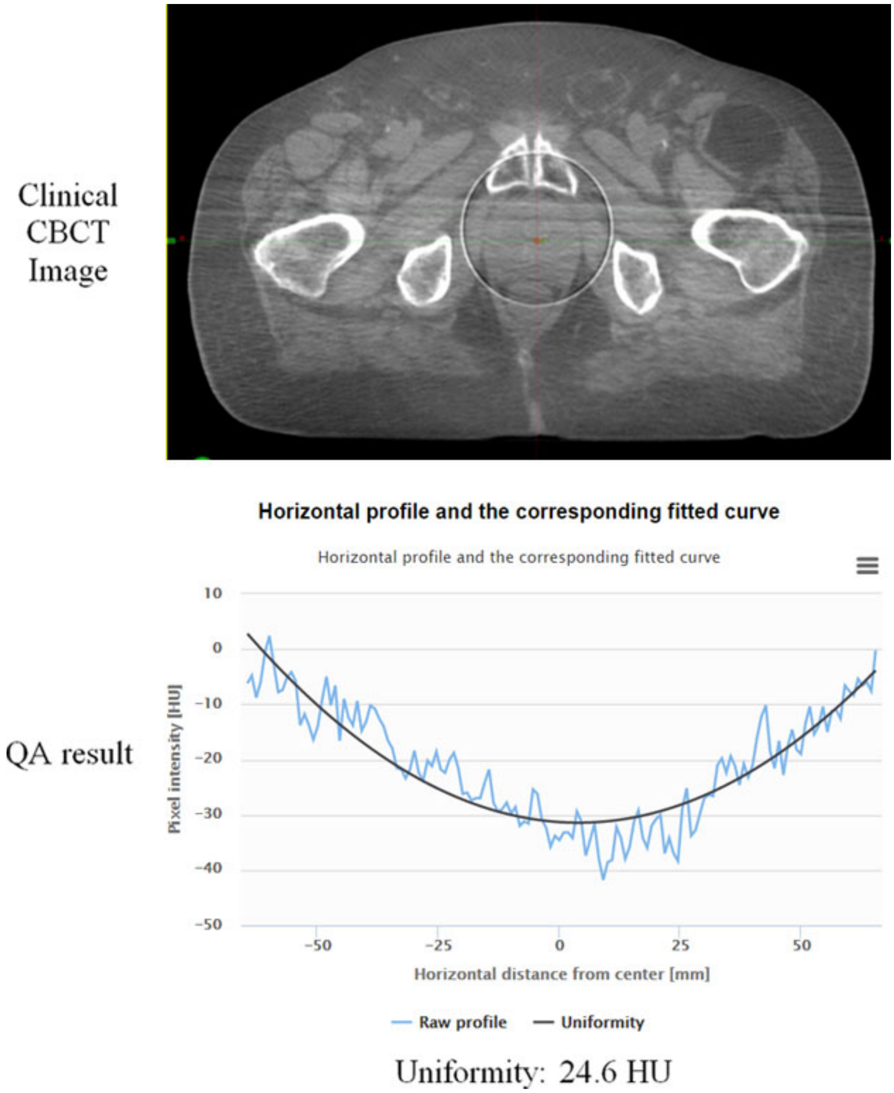


FIG. 1. Example of an artifact discovered clinically that passed the prior quality control. Uniformity results are provided for reference. [Color figure can be viewed at wileyonlinelibrary.com]

3. RESULTS

Between July 2014 and May 2017, there were 23 CBCT image quality issues reported in ROQRS-ML. Of these, 18 issues were discovered by therapists or physicians while using CBCT to set up the patient. The incidents ranged from ring artifacts to uniformity problems. The ROQRS-ML query also captured reports of mechanical breakdowns that could impact image quality upon return to service, particularly if major components of the imaging system such as the x-ray tube were replaced. The sensitivity of the TG-142/179 tolerances was 17% (4 of the 23 prior monthly QCs were out of tolerance).

An example of a scan that was within the TG-142/179 tolerances but was irregular based on visual inspection is presented in Fig. 1. In this case, the ring artifact was initially discovered on a patient scan.

Receiver operating characteristic curves for each image quality metric are presented in Fig. 2, and the AUCs are provided in Table III. Two of the metrics with a low AUC, geometric distortion and contrast resolution, required inverse results to yield an AUC > 0.5. This suggests their results are similar to a random guess. The best predictor of clinically recognizable CBCT image quality issues was uniformity, with an AUC of 0.85. A uniformity QC limit of 7.5 HU resulted in a sensitivity of 0.91 and specificity of 0.71, and a limit of 10.0 HU resulted in a sensitivity of 0.73 and specificity of 0.92. The ROC curves are not presented with error bars as the uncertainty of CATPHAN QA results has not, to our knowledge, been quantified for each of the image quality metrics evaluated.

4. DISCUSSION

The relationship between CBCT image quality QA and clinically recognizable image quality issues has not been well studied. Hence, the primary goal of this work was to evaluate the performance of CATPHAN QA at detecting

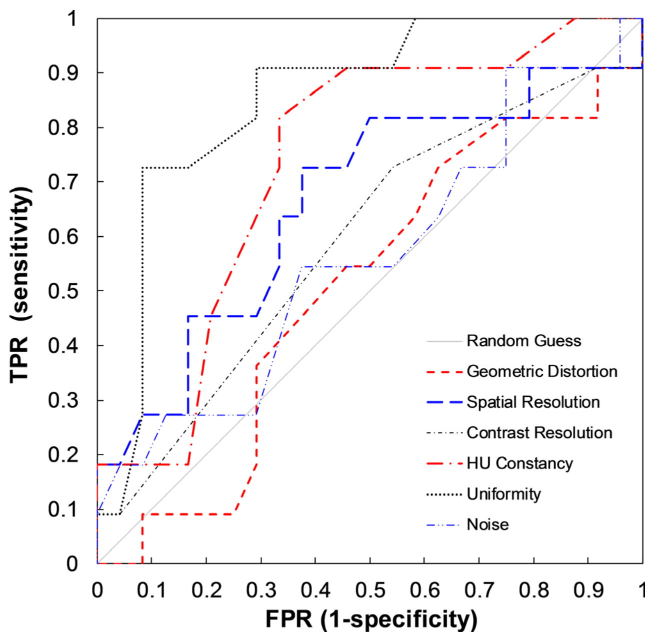


FIG. 2. Receiver operating characteristic curves for the cone-beam computed tomography image quality QA metrics versus a random guess [Color figure can be viewed at wileyonlinelibrary.com]

TABLE III. The area under the curve (AUC) for each cone-beam computed tomgraphy image quality assurance metric

Image quality metric	AUC
Uniformity	0.85
HU constancy	0.73
Spatial resolution	0.66
Contrast resolution	0.59
Noise	0.56
Geometric distortion	0.51

HU, hounsfield unit.

clinically recognizable image quality issues. CATPHAN QA paired with the TG-142/179 and manufacturer QC limits had poor sensitivity and specificity for these clinically recognizable issues. When varying the QC limits to generate ROC curves, all of the image quality metrics besides uniformity were shown to be poor predictors of clinically recognizable image quality issues. There are notable trade-offs when choosing uniformity as a predictor. When adjusting its QC limit to yield a sensitivity of 0.9, the FPR is nearly one-in-three, which is high for a routine QA test. On the contrary, if the QA limit is adjusted to provide a lower FPR (0.08), the sensitivity decreases to 0.71. More research is needed to determine the nature in which CBCT image quality issues occur — randomly versus trending toward failure.

The total QA (TQA) uniformity calculation is based on a scan of a CATPHAN module with uniform density. Five ROIs (four peripheral and one central) are chosen, and the mean and standard deviation of the CT number are calculated (Fig. 3). For instance, in the case presented in Fig. 4, the CATPHAN scan produced a uniformity number of 4.8 HU, which is reasonable based on manufacturer defined limits. Since the QA image was not visually assessed by the physicist, the QA was marked as "passed." The next day, the image quality issue was discovered clinically and reported via ROQRS-ML; a physicist scanned the CATPHAN while troubleshooting the issue, and qualitative inspection of the phantom scan clearly displayed artifacts. Yet, when the phantom scan was analyzed quantitatively, the results of the analysis were within the TG-142/179 QC limits. This is because the algorithm was blinded to the artifact, which occurred outside of its ROIs. Perhaps a global image quality metric such as correlation would improve the sensitivity of CATPHAN.

The potential severity of a CBCT image quality issue varies depending on the clinical use of CBCT. In most clinics, CBCT is only used for patient setup. Daily QA of CBCT-MV coincidence should be adequate in this setting. A more detailed assessment of image quality such as CATPHAN QA would only need to be performed after upgrades or major service in a clinic like this. A clinic that utilizes CBCT for adaptive treatment planning would benefit from monitoring other metrics in addition to what is listed above such as HU constancy and uniformity.

The results of this study beg the question as to whether monthly automated image quality QA has value in a modern radiation therapy clinic where daily image guidance using CBCT is used. In practice, CBCT image quality QC can be performed on every case that uses a CBCT. It is likely that better image quality QC can be achieved by having a robust reporting system for the therapists when they observe an image quality issue and that requires the physicist to follow-up and address any image quality issues that entered into the reporting system. Note that this study was performed at a single institution (multiple clinics), so the results should be viewed in the light of this limitation.

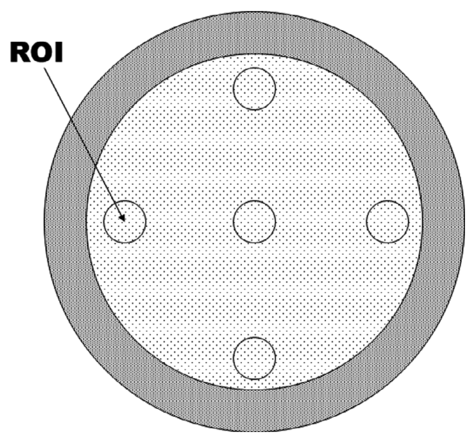
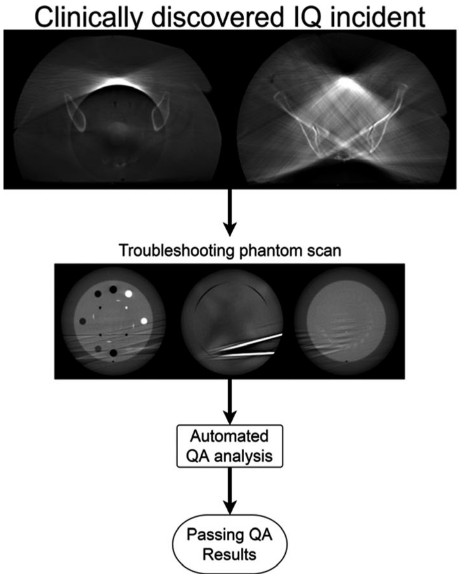


FIG. 3. CATPHAN uniformity module and region-of-interest (ROI)s from the CATPHAN 500 manual



Metric	Result	QC Limit
Geometric distortion (mm)	0.23	≤ 1
Spatial resolution (lp/cm)	6.40	≥ 5
Contrast (mm)	4	≤ 15
Uniformity (HU)	4.8	≤ 40
Noise (HU)	11.6	none given

FIG. 4. Example of a case with clinically unacceptable image quality that yielded passing quality control (QC) results even though artifacts are visible on the QC image set.

5. CONCLUSION

Automated monthly linac CBCT CATPHAN QA that uses ROI-based algorithms are a poor predictor of human observable image quality issues when using the TG-142/179 recommended metrics and tolerances. Of the QA metrics afforded by analysis of the CATPHAN, uniformity is the best predictor of clinically recognizable CBCT image quality issues when using appropriate QC limits. Automated monthly image quality QA can be replaced by a system of reporting and follow-up on any human observable image quality issues.

CONFLICTS OF INTEREST

This investigation was reported at the AAPM 59th Annual Meeting, 2017, Denver, CO. TP is a founding partner of Image Owl, Inc.

[a]Author to whom correspondence should be addressed. Electronic mail: rmanger@ucsd.edu

REFERENCES

1. Klein EE, Hanley J, Bayouth J, et al. Task Group 142 Report: quality assurance of medical accelerators. *Med Phys*. 2009;36:4197–4212.
2. Kutcher GJ, Coia L, Gillin M, et al. Comprehensive QA for radiation oncology: Report of AAPM Radiation Therapy Committee Task Group 40. *Med Phys*. 1994;21:581–618.
3. Bissonnette J-P, Balter PA, Dong L, et al. Quality assurance for image-guided radiation therapy utilizing CT-based technologies: a Report of the AAPM TG-179. *Med Phys*. 2012;39:1946–1963.
4. Yin F-F, Wong J, Balter J., et al. The role of in-room kV x-ray imaging for patient setup and target localization. *Report of AAPM Task Group*; 2009;104.
5. Terezakis SA, Harris KM, Ford E, et al. An evaluation of departmental radiation oncology incident reports: anticipating a national reporting system. *Int J Radiation Oncol Biol Phys*. 2012;85:919–923.
6. Chang DW, Cheetham L, Marvelde L, et al. Risk factors for radiotherapy incidents and impact of an online electronic reporting system. *Radiother Oncol*. 2014;112:199–204.
7. Gensheimer M, Zeng J, Carlson J, et al. Influence of planning time and treatment complexity on radiation therapy errors. *Pract Radiat Oncol*. 2016;6:187–193.
8. Kusano AS, Nyflot MJ, Zeng J, et al. Measurable improvement in patient safety culture: a departmental experience with incident learning. *Pract Radiat Oncol*. 2014;5:229–237.
9. Hoisak J, Pawlicki P, Kim G, Fletcher R, Moore K. Improving linear accelerator service response with a real-time electronic issue reporting system. *J Appl Clin Med Phys*. 2014;15:257–264.