



Original Research Article

## Automated contouring and statistical process control for plan quality in a breast clinical trial



Hana Baroudi<sup>a,b,\*</sup>, Callistus I. Huy Minh Nguyen<sup>b</sup>, Sean Maroongroge<sup>c</sup>, Benjamin D. Smith<sup>d</sup>, Joshua S. Niedzielski<sup>b</sup>, Simona F. Shaitelman<sup>d</sup>, Adam Melancon<sup>b</sup>, Sanjay Shete<sup>a</sup>, Thomas J. Whitaker<sup>b</sup>, Melissa P. Mitchell<sup>d</sup>, Isidora Yvonne Arzu<sup>d</sup>, Jack Duryea<sup>b</sup>, Soleil Hernandez<sup>a,b</sup>, Daniel El Basha<sup>a,b</sup>, Raymond Mumme<sup>b</sup>, Tucker Netherton<sup>b</sup>, Karen Hoffman<sup>d</sup>, Laurence Court<sup>a,b</sup>

<sup>a</sup> The University of Texas MD Anderson Cancer Center UTHealth Houston Graduate School of Biomedical Sciences, Houston, TX, USA

<sup>b</sup> Department of Radiation Physics, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>c</sup> Department of Radiation Oncology, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>d</sup> Department of Breast Radiation Oncology, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

### ARTICLE INFO

#### Keywords:

Automated segmentation  
Radiotherapy clinical trial  
Breast cancer  
Plan quality assurance

### ABSTRACT

**Background and purpose:** Automatic review of breast plan quality for clinical trials is time-consuming and has some unique challenges due to the lack of target contours for some planning techniques. We propose using an auto-contouring model and statistical process control to independently assess planning consistency in retrospective data from a breast radiotherapy clinical trial.

**Materials and methods:** A deep learning auto-contouring model was created and tested quantitatively and qualitatively on 104 post-lumpectomy patients' computed tomography images (nnUNet; train/test: 80/20). The auto-contouring model was then applied to 127 patients enrolled in a clinical trial. Statistical process control was used to assess the consistency of the mean dose to auto-contours between plans and treatment modalities by setting control limits within three standard deviations of the data's mean. Two physicians reviewed plans outside the limits for possible planning inconsistencies.

**Results:** Mean Dice similarity coefficients comparing manual and auto-contours was above 0.7 for breast clinical target volume, supraclavicular and internal mammary nodes. Two radiation oncologists scored 95% of contours as clinically acceptable. The mean dose in the clinical trial plans was more variable for lymph node auto-contours than for breast, with a narrower distribution for volumetric modulated arc therapy than for 3D conformal treatment, requiring distinct control limits. Five plans (5%) were flagged and reviewed by physicians: one required editing, two had clinically acceptable variations in planning, and two had poor auto-contouring.

**Conclusions:** An automated contouring model in a statistical process control framework was appropriate for assessing planning consistency in a breast radiotherapy clinical trial.

## 1. Introduction

Clinical trials in radiation therapy drive treatment advancements, impacting patient outcomes. Hence, maintaining consistent patient treatment in trials by following a quality assurance (QA) protocol is crucial. In fact, failure to comply to a QA protocol leads to poor quality of radiotherapy treatment in trials, which is associated with increased trial failure rates, reduced survival, and a possible increase in toxicity

[1–4].

The pre-treatment review of radiotherapy plans for clinical acceptability and protocol compliance is especially important in the QA process, as most of the errors that violate protocol compliance are found in the planning process [5]. Taylor et al. conducted a failure modes and effect analysis on clinical trial QA methods and concluded that failures in pre-treatment contouring and plan review were a high risk to the patient [6]. In a gastrointestinal radiotherapy clinical trial, the

\* Corresponding author at: The University of Texas MD Anderson Cancer Center, Department of Radiation Physics, 1400 Pressler Street, FCT8.6014, Houston, TX 77030, USA.

E-mail address: [hbaroudi@mdanderson.org](mailto:hbaroudi@mdanderson.org) (H. Baroudi).

<https://doi.org/10.1016/j.phro.2023.100486>

Received 4 May 2023; Received in revised form 21 August 2023; Accepted 21 August 2023

Available online 23 August 2023

2405-6316/© 2023 The Authors. Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

implementation of a prospective treatment plan review reduced the unacceptable plan deviation rate from 42% to 2–4% [7].

Yet, many radiotherapy trial protocols lack pre-treatment QA [8]. Even when required, QA compliance can be heterogenous [7,8]. This may be due to the difficulty and high cost associated with experts evaluating every plan and each contour for proper delineation of targets, dose coverage to targets, and appropriate doses to normal tissues. Furthermore, for breast radiotherapy specifically, there can be an additional challenge in plan review. Many physicians treating breast cancer may not delineate target contours, as traditional 3D conformal radiation therapy (3DCRT) treatment techniques, which are still common, do not require target contours, as they are based on the patient's overall geometry. This poses more difficulty in verifying the consistency of the treatment plans. QA requirements also differ based on the complexity of treatment [6]—for example, volumetric modulated arc therapy (VMAT) in comparison to 3DCRT.

In this work, we propose an automated approach to pre-treatment plan verification in breast radiotherapy for different modalities. This approach consists of, first, developing an automated contouring model for breast target structures and, second, using the model to retrospectively evaluate 3DCRT and VMAT plans in a breast radiotherapy clinical trial via a statistical process control (SPC) approach to investigate this method's potential in assessing the consistency of treatment plans. The approach is applicable to clinical trials (as tested here) and routine clinical plan review and is independent of and does not require any clinical contours. In this work, we aim to develop a novel approach to automatic plan quality review for breast radiotherapy treatment plans.

## 2. Materials and methods

### 2.1. Auto-Contouring model

A set of 104 post-lumpectomy breast cancer patients with regional lymph node involvement who received post-operative radiation was selected retrospectively from the University of Texas MD Anderson Cancer Center patient dataset. Forty-nine patients had left-sided treatment, and 55 had right-sided treatment. The dataset included only patients treated in a supine position; 9 patients had one arm above the head, and 95 had two arms above the head. The patients' planning computed tomography (CT) scans were obtained with no contrast, and the selection included an equal mix of breath-hold and free-breathing scans. A total of 101 patients were scanned using a GE LightSpeed 16 Slice CT, and 3 patients were scanned using a Siemens Somatom Definition Edge. These scans had a median slice thickness of 3 mm and a median pixel spacing of 1.2 mm. The use of this data was approved by the institutional review board of MD Anderson (IRB: PA16-0379).

To ensure consistent training data, the MD Anderson breast radiation oncology group set consensus guidance for manually contouring the breast target structures. Nine radiation oncology residents followed those instructions and delineated the following structures: breast clinical target volume, supraclavicular lymph node clinical target volume, and internal mammary lymph node clinical target volume (CTV). Contouring consistency was maintained by peer review where each resident's contours were reviewed by another. Only the mentioned contours were used for this study, as they were consistently treated with no resection in all clinical trial patients.

Two models were trained with nnU-Net deep learning algorithm [9], with one model for the 55 patients with right-sided treatment and a second model for the 49 patients with left-sided treatment. Both models used an 80/20 training/testing ratio with 5-fold cross-validation. The selected algorithm used a 3D full-resolution U-Net architecture with a stochastic gradient descent optimizer and a combined loss function of Dice with cross-entropy. Quantitative metrics, namely the Dice similarity coefficient (DSC) and mean surface distance (MSD), were used to evaluate the test dataset (21 patients) by comparing the automatically generated contours to the manual ones. In addition, the model was

assessed qualitatively by two breast radiation oncology attending physicians, that did not participate in the delineation of the training data, on the clinical acceptability of the automated contours using a 5-point Likert scale (5: use as is, 4: use with minor edits, 3: use with major edits, 2: requires major edits, 1: unusable). We define our criteria for clinical acceptability as a contour receiving a score of 3 or above.

### 2.2. Consistency evaluation

A total of 127 patients were selected retrospectively from the Shortening Adjuvant PHoton Irradiation (SAPHIRE) clinical trial (NCT02912312), an MD Anderson trial that investigated the occurrence of lymphedema after hypofractionated versus conventionally fractionated radiation therapy for patients with locally advanced invasive breast cancer requiring radiation to the regional nodes. The dataset included only patients who had undergone breast-conserving surgery. A total of 105 patients had adjuvant 3DCRT using tangential fields on the breast target with regional lymph node irradiation, and the remaining 22 patients had VMAT to the same targets. All patients received either a conventional dose regimen of 50 Gy to the breast and 45 Gy to the lymph nodes (mandatory treatment of supraclavicular and internal mammary nodes) in 25 fractions or a hypofractionated regimen of 40.5 Gy to the breast and 37.5 Gy to the lymph nodes in 15 fractions. These exported plans were approved for clinical use, having been through our institution's extensive peer review process that on average involves 5 breast radiation oncology attending physicians. The use of this data was also approved by the institutional review board of MD Anderson (IRB: PA16-0379).

Statistical process control (SPC) was selected as the plan verification method. This technique has long been used for monitoring the quality of a process in the manufacturing industry. In radiotherapy specifically, SPC has been used for QA, e.g., monitoring linac beam quality and machine performance [10–15] and verifying intensity-modulated radiation therapy plans [16–19]. More recently, SPC was used for monitoring the performance of an automated planning tool, the Radiation Planning Assistant [20]. The success of SPC in the different aspects of QA in radiotherapy is a positive indicator for its use in the present work's target coverage consistency evaluation.

SPC uses statistical tools to assess the performance of a process [21]. This is done by modeling a mathematical variable. Control Charts, a common SPC tool, define accuracy as the mean of this variable and precision as the spread. The spread is usually set at three times the standard deviation from the mean covering over 99% of a normally distributed data [21]. These distances, labeled as upper and lower control limits, detect errors for current and future datapoints. If a process is steady, all points should be within those control limits, not following any trends. In our work, we used SPC to assess the consistency of dose coverage. We focused on average dose to automated contours and flagged any point that lied outside of the control limits for possible inconsistencies.

The first step in this methodology was the application of our auto-contouring model to all 127 CT scans from the clinical trial to obtain breast target contours. This is done to measure target doses independent of clinical contours and in case they don't exist. Four patients from the clinical trial were not analyzed: one due to the failure of the auto-contouring model to produce usable contours, one due to a problem in exporting a plan related to data transfer between different treatment planning software, and two due to the use of two different treatment plans, which made obtaining the total dose to auto-contours difficult. Thus, 102 patients given 3DCRT and 21 patients given VMAT were included in the final analysis. Control charts were plotted for the 3DCRT- and VMAT-treated patients separately to assess the consistency of the target coverage between the different methods. Specifically, we used the ratio of the mean dose of clinical target volumes to the prescribed dose as the mathematical variable to be monitored across individual patients. The plots were assessed for process control, i.e., the consistency of the

target coverage and the presence of outliers. Two attending physicians reviewed the plans of the outlier points and assessed whether the plan variation was clinically acceptable. If the outliers were caused by poor automated contours, the limits were recomputed without those data points to better represent the sample's distribution.

### 3. Results

#### 3.1. Auto-Contouring model validation

The mean DSC, between auto-contoured and manual structures, was above 0.7 for all structures, and the average MSD was below 3 mm (Table 1).

As for qualitative assessment, attending physician review scored the majority of automated contours (20 of 21) for breast CTV, axilla levels I, II, and III, supraclavicular nodes, and internal mammary nodes as clinically acceptable (receiving a Likert score of three or more). The one breast auto-contour receiving a score of two (major edits) had a cranial border 2.7 cm superior of the acceptable limit.

#### 3.2. Statistical process control charts

Fig. 1 shows the SPC charts of breast CTV, supraclavicular node CTV, and internal mammary nodes CTV with all patients and control limits for 3DCRT and VMAT methods. For both methods, the control limits were wider for the lymph nodes than for the breast, indicating an increased variability in dose coverage of the lymph nodes' automated contours between patients. Notably, the control limits were different and narrower for VMAT than for 3DCRT (average of the relative standard deviation (standard deviation over the mean of the data), 4.2% for 3DCRT and 1.4% for VMAT).

In total, three outliers were flagged from the charts. For one plan (auto-contoured supraclavicular node mean dose, 89%), the reviewing physicians indicated that they would prefer to have improved coverage of the supraclavicular CTV (e.g. by changing the gantry angle) (Fig. 2). The other outliers were poor auto-contours (auto-contoured breast CTV mean dose, 64%, and 87%). The control limits for the breast CTV were then recomputed excluding the falsely flagged points. Two additional outliers were flagged (Fig. 3), both representing clinically acceptable planning variations (auto-contoured breast mean dose, 92% and 105%).

### 4. Discussion

In this work, we investigated the use of SPC and an auto-contouring model to assess the consistency of breast target dose coverage in a clinical trial, comparing different treatment approaches. Overall, a low flagging rate was expected in this preliminary study, as the plans had already been approved for clinical use. We were, however, able to identify one outlier using this approach (a supraclavicular node 3DCRT field for which modification was suggested). We also showed an increase in variability in the supraclavicular and internal mammary lymph nodes auto-contours' dose. This could be attributed to contouring variability for these small structures in addition to the use of a single treatment field

instead of opposing ones that are less sensitive to contouring changes. We also demonstrated that different treatment approaches required distinct tolerance levels, with traditional plans (3DCRT tangents) showing more variability than VMAT. Using the auto-contouring tool and SPC provided insights into dose coverage variability in a clinical trial dataset. It also successfully flagged a plan for review, highlighting its potential for QA in treatment planning.

Usually, pre-treatment QA involves labor intensive dose checks for individual patients plans. This work automates this process, turning it into an independent, population-based reviewed, thus saving time and costs. Implementing our automated QA tool in clinical trials' protocols could standardize pre-treatment plan QA, ensuring consistent patient care and reliable trial results. This approach also benefits non trial patients undergoing the same treatment regimen (e.g., 3DCRT or VMAT). This process is also translatable to different disease sites using different auto-segmentation models.

Peer review is often used for plan checks, but it is not very effective in identifying detectable errors (55% detection rate for physician chart rounds [22] and 32% [23] and 38% [24] for physics plan checks). Automated plan checks, however, have been shown to improve this detection rate [24,25]. Our method for automating part of the plan check could enhance the detection of plan inconsistencies. The latest work in automating plan checks includes automated contour QA, as in the work of Rhee et al. [26], who developed an independent automated contouring system to verify auto-contours. Similarly, automated plan QA was demonstrated by Gronberg et al. [27], who created a dose prediction model to identify suboptimal plans. In the current work, we investigated a unique approach by using an automated contouring system independent from clinical contours to verify treatment plans automatically.

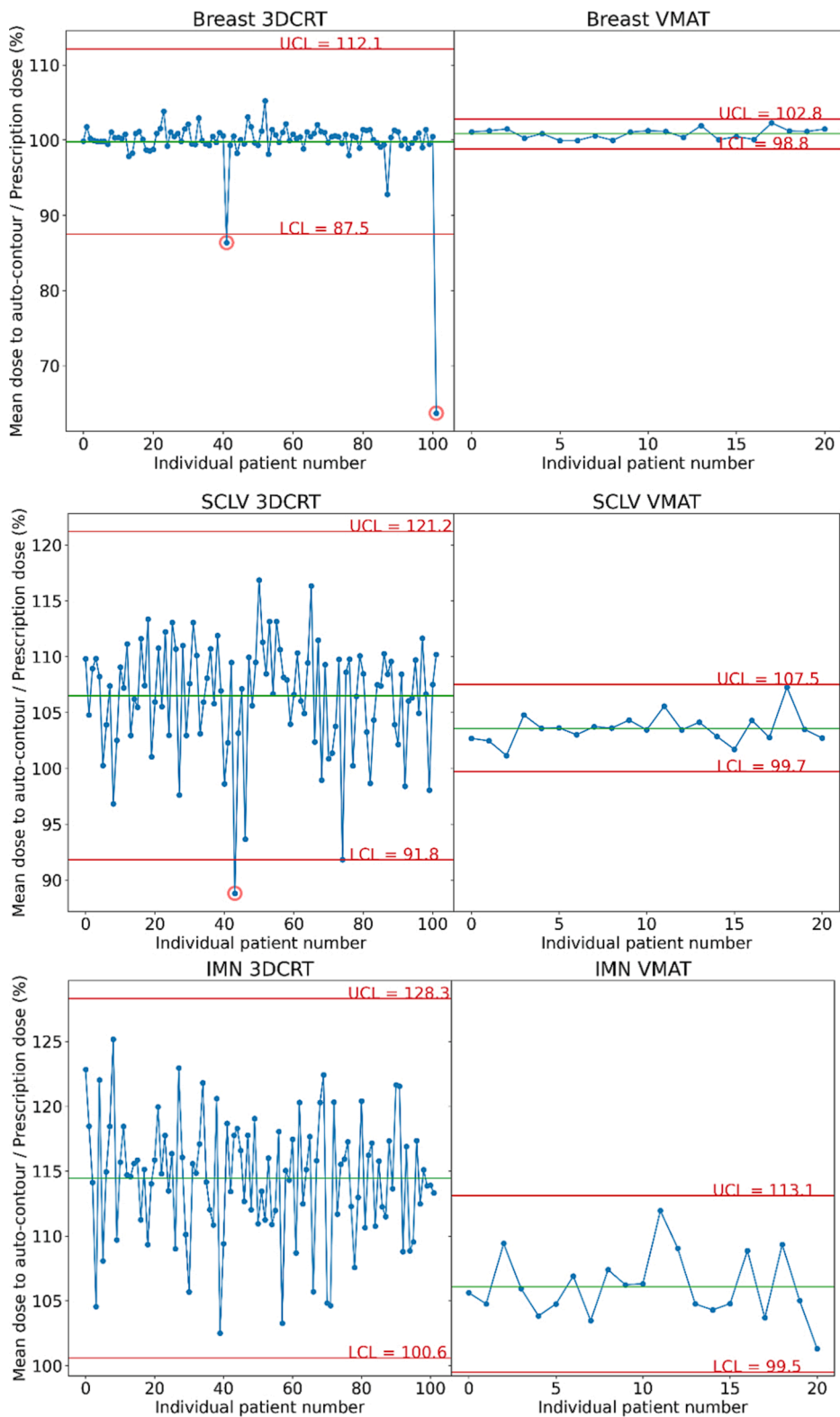
In a similar work, Jung et al. automatically segmented cardiac sub-structures using atlas-based methods to assess dose to structures compared to manual contours for toxicity related QA. [28]. While atlas-based segmentation was explored for breast structures, deep-learning methods showed superior performance [29]. Vaassen et al. also showed that CTV automated contouring of the breast required minimal edits for clinical planning use (<5% DSC difference between automated and edited contour) which validates our physician review results [30]. In addition, SPC presents a more streamlined, patient based, analysis method in comparison to simple dose difference evaluation. This is also true when looking at the work done by Simoes et al. where they compared the difference in PTV coverage for breast using a relative difference of the V95% metric between the dose on the deep learning generated and manual contours [31].

The tools described in this work may also have wider application in understanding the results of clinical trials. The ACOSOG Z0011 (Alliance) Trial surprisingly showed that axillary dissection showed no benefit for local control or survival as compared to sentinel node biopsy alone in patients with node positive disease. All of these patients received radiation to the whole breast, leading to speculation that the axilla may have inadvertently received radiation, which would have treated the undissected axilla. Jaggi et al. retrospectively assessed the charts from this trial, using proximity of the superior border to the

**Table 1**

Evaluation of auto-contoured structures. Quantitative: mean surface distance and Dice similarity coefficient, and qualitative: Likert scores for the clinical acceptability of breast clinical target volume, supraclavicular nodes, and internal mammary nodes.

Structure	Quantitative		Qualitative				
	Mean Dice Similarity Coefficient	Mean surface distance (mm)	5: Use as is	4: Minor edits that are not necessary	3: Minor edits that are necessary	2: Major edits	1: Unusable
Breast clinical target volume	0.9 ± 0.0	3.5 ± 2.1	10%	10%	75%	5%	0
Supraclavicular nodes	0.7 ± 0.1	2.4 ± 1.0	5%	0	95%	0	0
Internal mammary nodes	0.7 ± 0.1	1.8 ± 1.2	100%	0	0	0	0

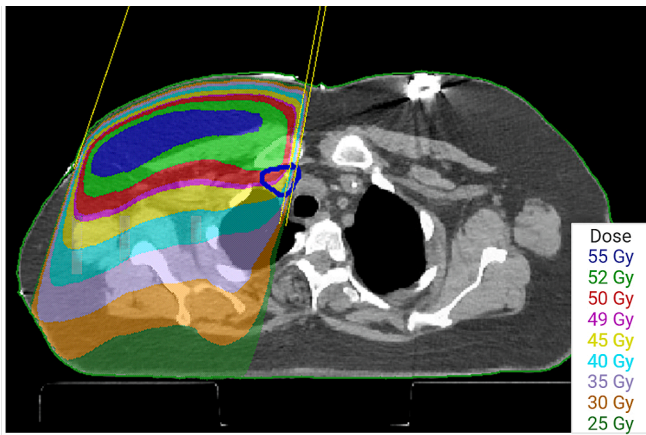


**Fig. 1.** Statistical process control charts of both modalities in each structure. Shown are 3D conformal radiation therapy (3DCRT) and volumetric modulated arc therapy (VMAT) doses to breast critical target volume, internal mammary nodes (IMN), and supraclavicular nodes (SCLV) auto-contours. The green line shows the mean of the points, and the red lines show the upper and lower control limits (UCL and LCL). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

humeral head as a surrogate to assess dose to the axilla [32]. While hypothesis generating, definite conclusions could not be drawn from the analysis. The current proposed method has the potential to robustly answer questions about delivered target and OAR dose to help answer questions about clinical outcome on large randomized trials. Retrospective assessment of DICOM data submitted for routine QA could be subjected to evaluation using AI tools to help answer many clinically

relevant questions regarding optimal dosing on multi-institutional trials.

Our automated contouring model was trained using 50 patients. This is adequate since nn-UNet does not require a large dataset to give acceptable results when the training set is consistent with patients undergoing standardized simulation and imaging protocol such as ours (DSC plateaus starting 25 patients for training) [33,34]. Chung et al. developed a similar automated contouring model for breast targets using



**Fig. 2.** Axial view of the outlier patient with auto-contoured supraclavicular node mean dose of 89%. Automated supraclavicular contour is shown in blue. This patient’s plan was reviewed retrospectively by two attending physicians who concluded that they would have preferred editing the medial supraclavicular field (making it closer to the esophagus). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a 3D U-net algorithm and used DSC as one of their evaluation metrics [35]. Their results were comparable to ours with DSC differences on the hundredth level. Although not addressed in this work, the use of an automated contouring model was important to reduce interobserver variability in delineating breast structures. The extent of this variability has been shown by Li et al. who conducted a study to determine the inter and intra-observer variability in delineating breast targets and normal structures. They determined that the percent overlap between internal mammary contours drawn by 9 physicians from 8 different institutions on a patient scan was as low as 10%. In addition, the supraclavicular nodes contours had a standard deviation of 60% in volumetric variation. The breast contours also showed variation in border delineation between physicians [36]. Knowing that there are variations in manual contours that translate to variations in the dose distribution to these structures, we aimed to select a plan quality metric that is not overly sensitive to these variations on automated contours. In our preliminary work, we investigated the use of different dosimetric parameters such as

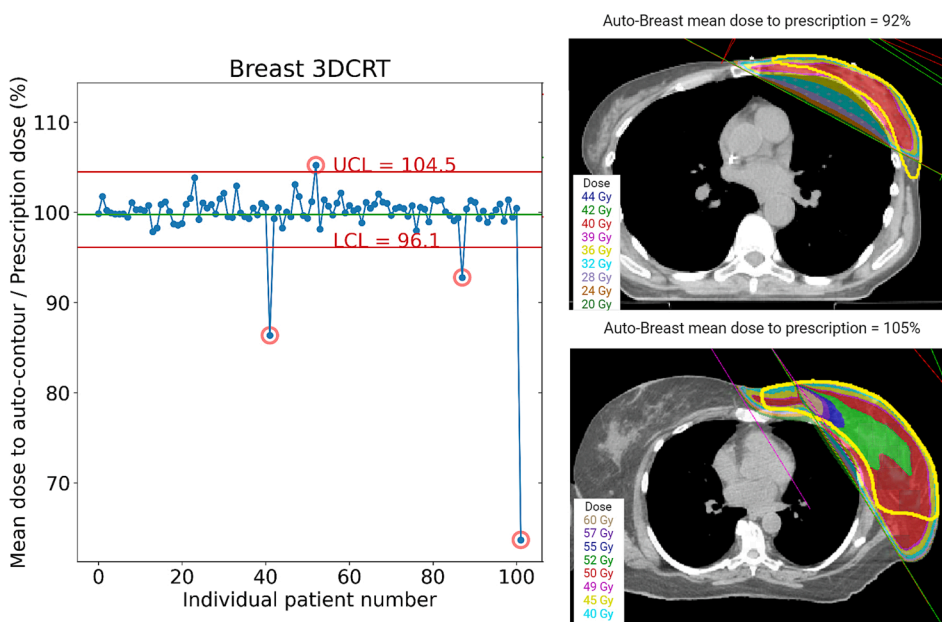
D95%, D90% and V95%, however, the impact of dose variation in the automated contours seems to be exacerbated in the SPC analysis for those metrics such that it hindered the sensitivity of the control charts to detect any errors. For this reason, we focused on mean dose for this current study. Additional parameters could be examined further when more data is available. In addition, the incorporation of three standard deviations as a fundamental approach in SPC studies is widely recognized and accepted across industries. Alternative limits should be explored when a larger dataset from diverse institutions, coupled with an accurate understanding of the true failure rate of the plans is available.

The next step for this work is to expand SPC testing to a larger patient dataset, including patients from other institutions. This will help us assess SPC’s ability to detect unacceptable inconsistencies in cases with more variations in clinical practice. Our evaluation is not sensitive to systematic errors for our internal data, as the dataset was obtained from patients treated with the same institutional protocols, planned on the same treatment planning software and underwent the same quality assurance. Nevertheless, SPC could be effective in identifying systematic differences when analyzing data from various institutions. This dual capacity emphasizes the versatility and credibility of SPC, making it an indispensable tool for ensuring quality control.

In conclusion, we have shown that using an auto-contouring model in an SPC framework can successfully assess planning variability and has potential for identifying inconsistencies in a breast radiotherapy clinical trial.

**CRedit authorship contribution statement**

**Hana Baroudi:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Callistus I. Huy Minh Nguyen:** Methodology, Software, Investigation, Writing – review & editing. **Sean Maroon-groge:** Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – review & editing. **Benjamin D. Smith:** Validation, Resources, Writing – review & editing. **Joshua S. Niedzielski:** Supervision, Writing – review & editing. **Simona F. Shaitelman:** Supervision, Validation, Writing – review & editing. **Adam Melancon:** Supervision, Writing – review & editing. **Sanjay Shete:** Supervision, Writing – review & editing. **Thomas J. Whitaker:** Data curation, Writing – review & editing. **Melissa P. Mitchell:** Supervision, Validation, Investigation,



**Fig. 3.** On the left: statistical process control chart for automated breast structure with corrected limits in red. On the right: Axial views of the additional two outlier patients. Both represent clinically acceptable planning variations; The top figure is a wide tangent plan that differs from the tangents/Internal mammary electron matching that was more common for this dataset. The bottom figure was determined to be an acceptable plan, but with higher mean dose to the breast, partly caused by the match line between tangential photon fields and the angled electron field. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Writing – review & editing. **Isidora Yvonne Arzu**: Validation, Writing – review & editing. **Jack Duryea**: Software, Writing – review & editing. **Soleil Hernandez**: Software, Writing – review & editing, Visualization. **Daniel El Basha**: Software, Writing – review & editing. **Raymond Mumme**: Software, Writing – review & editing. **Tucker Netherton**: Conceptualization, Methodology, Supervision, Writing – review & editing. **Karen Hoffman**: Supervision, Resources, Writing – review & editing. **Laurence Court**: Conceptualization, Methodology, Project administration, Funding acquisition, Supervision, Resources, Writing – review & editing.

### Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. Court has grants from Varian, CPRIT, Wellcome Trust, and The Fund for Innovation in Cancer Informatics. Dr. Smith has a royalty and equity interest in Oncora Medical and receives salary support from Varian Medical Systems. Dr. Shaitelman has grants or contracts from Emerson Collective, NIH R21, Artios Pharma. She also has contracted research agreements with Alpha Tau, Exact Sciences and TAE Life Sciences. Dr. Hernandez is supported by a Cancer Prevention and Research Institute (CPRIT) Training Award (RP210028). Daniel el Basha is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2043424.

### Acknowledgments

The authors would like to acknowledge Ms. Sarah J. Bronson of the Research Medical Library at MD Anderson for her editing services. We also acknowledge the support of the High-Performance Computing for Research facility at the University of Texas MD Anderson Cancer Center for providing computational resources that have contributed to this work. This work was supported by the Tumor Measurement Initiative (TMI) through the MD Anderson Strategic Initiative Development Program (STRIDE). Dr. Hernandez is supported by a Cancer Prevention and Research Institute (CPRIT) Training Award (RP210028). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2043424 for Daniel El Basha. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

### References

- Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, et al. Failure to Adhere to Protocol Specified Radiation Therapy Guidelines Was Associated With Decreased Survival in RTOG 9704—A Phase III Trial of Adjuvant Chemotherapy and Chemoradiotherapy for Patients With Resected Adenocarcinoma of the Pancreas. *Int J Radiat Oncol* 2012;82:809–16. <https://doi.org/10.1016/j.ijrobp.2010.11.039>.
- Fairchild A, Straube W, Laurie F, Followill D. Does Quality of Radiation Therapy Predict Outcomes of Multicenter Cooperative Group Trials? A Literature Review. *Int J Radiat Oncol* 2013;87:246–60. <https://doi.org/10.1016/j.ijrobp.2013.03.036>.
- Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy. *Radiother Oncol* 2012;105:4–8. <https://doi.org/10.1016/j.radonc.2012.08.008>.
- Zhong H, Men K, Wang J, van Soest J, Rosenthal D, Dekker A, et al. The Impact of Clinical Trial Quality Assurance on Outcome in Head and Neck Radiotherapy Treatment. *Front Oncol* 2019;9:792. <https://doi.org/10.3389/fonc.2019.00792>.
- Clark BG, Brown RJ, Ploquin J, Dunscombe P. Patient safety improvements in radiation treatment through 5 years of incident learning. *Pract Radiat Oncol* 2013; 3:157–63. <https://doi.org/10.1016/j.proro.2012.08.001>.
- Taylor PA, Miles E, Hoffmann L, Kelly SM, Kry SF, Sloth Møller D, et al. Prioritizing clinical trial quality assurance for photons and protons: A failure modes and effects analysis (FMEA) comparison. *Radiother Oncol* 2023;182:109494. <https://doi.org/10.1016/j.radonc.2023.109494>.
- Willett CG, Moughan J, O'Meara E, Galvin JM, Crane CH, Winter K, et al. Compliance with therapeutic guidelines in Radiation Therapy Oncology Group prospective gastrointestinal clinical trials. *Radiother Oncol* 2012;105:9–13. <https://doi.org/10.1016/j.radonc.2012.09.011>.
- Corrigan KL, Kry S, Howell RM, Kouzy R, Jaoude JA, Patel RR, et al. The radiotherapy quality assurance gap among phase III cancer clinical trials. *Radiother Oncol* 2022;166:51–7. <https://doi.org/10.1016/j.radonc.2021.11.018>.
- Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021;18:203–11. <https://doi.org/10.1038/s41592-020-01008-z>.
- Pawlicki T, Whitaker M, Boyer AL. Statistical process control for radiotherapy quality assurance. *Med Phys* 2005;32:2777–86. <https://doi.org/10.1118/1.2001209>.
- Pal B, Pal A, Das S, Palit S, Sarkar P, Mondal S, et al. Retrospective study on performance of constancy check device in Linac beam monitoring using Statistical Process Control. *Rep Pract Oncol Radiother* 2020;25:91–9. <https://doi.org/10.1016/j.rpor.2019.12.004>.
- Zhang H, Lu W, Cui H, Li Y, Yi X. Assessment of Statistical Process Control Based DVH Action Levels for Systematic Multi-Leaf Collimator Errors in Cervical Cancer RapidArc Plans. *Front Oncol* 2022;12:862635. <https://doi.org/10.3389/fonc.2022.862635>.
- Binny D, Aland T, Archibald-Heeren BR, Trapp JV, Kairn T, Crowe SB. A multi-institutional evaluation of machine performance check system on treatment beam output and symmetry using statistical process control. *J Appl Clin Med Phys* 2019; 20:71–80. <https://doi.org/10.1002/acm2.12547>.
- Meyers SM, Balderson MJ, Létourneau D. Evaluation of Elekta Agility multi-leaf collimator performance using statistical process control tools. *J Appl Clin Med Phys* 2019;20:100–8. <https://doi.org/10.1002/acm2.12660>.
- Létourneau D, Wang A, Amin MN, Pearce J, McNiven A, Keller H, et al. Multileaf collimator performance monitoring and improvement using semiautomated quality control testing and statistical process control. *Med Phys* 2014;41:121713. <https://doi.org/10.1118/1.4901520>.
- Breen SL, Moseley DJ, Zhang B, Sharpe MB. Statistical process control for IMRT dosimetric verification. *Med Phys* 2008;35:4417–25. <https://doi.org/10.1118/1.2975144>.
- Gagneur JD, Ezzell GA. An improvement in IMRT QA results and beam matching in linacs using statistical process control. *J Appl Clin Med Phys* 2014;15:190–5. <https://doi.org/10.1120/jacmp.v15i5.4927>.
- Gérard K, Grandhayé J-P, Marchesi V, Kafrouni H, Husson F, Aletti P. A comprehensive analysis of the IMRT dose delivery process using statistical process control (SPC). *Med Phys* 2009;36:1275–85. <https://doi.org/10.1118/1.3089793>.
- Pawlicki T, Yoo S, Court LE, McMillan SK, Rice RK, Russell JD, et al. Moving from IMRT QA measurements toward independent computer calculations using control charts. *Radiother Oncol* 2008;89:330–7. <https://doi.org/10.1016/j.radonc.2008.07.002>.
- Mehrens H, Douglas R, Gronberg M, Nealon K, Zhang J, Court L. Statistical process control to monitor use of a web-based autoplanning tool. *J Appl Clin Med Phys* 2022;23:e13803.
- John S. Oakland, Robert James Oakland. *Statistical Process Control*. 7th Edition. London: Routledge; 2018.
- Talcott WJ, Lincoln H, Kelly JR, Tressell L, Wilson LD, Decker RH, et al. A Blinded, Prospective Study of Error Detection During Physician Chart Rounds in Radiation Oncology. *Pract Radiat Oncol* 2020;10:312–20. <https://doi.org/10.1016/j.proro.2020.05.012>.
- Riegel AC, Polvorosa C, Sharma A, Baker J, Ge W, Lauritano J, et al. Assessing initial plan check efficacy using TG 275 failure modes and incident reporting. *J Appl Clin Med Phys* 2022;23:e13640.
- Gopan O, Zeng J, Novak A, Nyflot M, Ford E. The effectiveness of pretreatment physics plan review for detecting errors in radiation therapy. *Med Phys* 2016;43: 5181. <https://doi.org/10.1118/1.4961010>.
- Ford E, Conroy L, Dong L, de Los Santos LF, Greener A, Gwe-Ya Kim G, et al. Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275. *Med Phys* 2020;47:e236–72. <https://doi.org/10.1002/mp.14030>.
- Rhee DJ, Akinfenwa CPA, Rigaud B, Jhingran A, Cardenas CE, Zhang L, et al. Automatic contouring QA method using a deep learning-based autocontouring system. *J Appl Clin Med Phys* 2022;23:e13647.
- Gronberg MP, Beadle BM, Garden AS, Skinner H, Gay S, Netherton T, et al. Deep Learning-Based Dose Prediction for Automated, Individualized Quality Assurance of Head and Neck Radiation Therapy Plans. *Pract Radiat Oncol* 2023;13(3):e282–e291. doi: 10.1016/j.proro.2022.12.003.
- Jung JW, Mille MM, Ky B, Kenworthy W, Lee C, Yeom YS, et al. Application of an automatic segmentation method for evaluating cardiac structure doses received by breast radiotherapy patients. *Phys Imaging Radiat Oncol* 2021;19:138–44. <https://doi.org/10.1016/j.phro.2021.08.005>.
- Choi MS, Choi BS, Chung SY, Kim N, Chun J, Kim YB, et al. Clinical evaluation of atlas- and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer. *Radiother Oncol* 2020;153:139–45. <https://doi.org/10.1016/j.radonc.2020.09.045>.
- Vaassen F, Boukerroui D, Looney P, Canters R, Verhoeven K, Peeters S, et al. Real-world analysis of manual editing of deep learning contouring in the thorax region. *Phys Imaging Radiat Oncol* 2022;22:104–10. <https://doi.org/10.1016/j.phro.2022.04.008>.
- Simões R, Wortel G, Wiersma TG, Janssen TM, van der Heide UA, Remeijer P. Geometrical and dosimetric evaluation of breast target volume auto-contouring. *Phys Imaging Radiat Oncol* 2019;12:38–43. <https://doi.org/10.1016/j.phro.2019.11.003>.

- [32] Jagsi R, Chadha M, Moni J, Ballman K, Laurie F, Buchholz TA, et al. Radiation field design in the ACOSOG Z0011 (Alliance) Trial. *J Clin Oncol* 2014;32:3600–6. <https://doi.org/10.1200/JCO.2014.56.5838>.
- [33] Yu C, Anakwenze CP, Zhao Y, Martin RM, Ludmir EB, Niedzielski S, et al. Multi-organ segmentation of abdominal structures from non-contrast and contrast enhanced CT images. *Sci Rep* 2022;12:19093. <https://doi.org/10.1038/s41598-022-21206-3>.
- [34] Weissmann T, Huang Y, Fischer S, Roesch J, Mansoorian S, Ayala Gaona H, et al. Deep learning for automatic head and neck lymph node level delineation provides expert-level accuracy. *Front Oncol* 2023;13:1115258. <https://doi.org/10.3389/fonc.2023.1115258>.
- [35] Chung SY, Chang JS, Choi MS, Chang Y, Choi BS, Chun J, et al. Clinical feasibility of deep learning-based auto-segmentation of target volumes and organs-at-risk in breast cancer patients after breast-conserving surgery. *Radiat Oncol* 2021;16:44. <https://doi.org/10.1186/s13014-021-01771-z>.
- [36] Li XA, Tai A, Arthur DW, Buchholz TA, Macdonald S, Marks LB, et al. Variability of Target and Normal Structure Delineation for Breast Cancer Radiotherapy: An RTOG Multi-Institutional and Multiobserver Study. *Int J Radiat Oncol Biol Phys* 2009;73:944–51. <https://doi.org/10.1016/j.ijrobp.2008.10.034>.