Received: 8 December 2023 | Revised: 14 March 2024 | Accepted: 30 March 2024

DOI: 10.1002/mp.17081

RESEARCH ARTICLE

MEDICAL PHYSICS

Incorporating plan complexity into the statistical process control of volumetric modulated arc therapy pre-treatment verifications

Serenella Russo[1] | **Jordi Saez**[2] | **Marco Esposito**[1,3] | **Andrea Bruschi**[1] | **Alessandro Ghirelli**[1] | **Silvia Pini**[1] | **Silvia Scoccianti**[4] | **Victor Hernandez**[5,6]

[1]Medical Physics Unit, Azienda USL Toscana Centro, Florence, Italy

[2]Department of Radiation Oncology, Hospital Clínic de Barcelona, Barcelona, Spain

[3]Medical Physics Program, The Abdus Salam International Centre for Theoretical Physics Trieste-Italy, Trieste, Italy

[4]Radiotherapy Unit, Azienda USL Toscana Centro, Florence, Italy

[5]Department of Medical Physics, Hospital Sant Joan de Reus, IISPV, Reus, Spain

[6]Universitat Rovira i Virgili (URV), Tarragona, Spain

Correspondence
Victor Hernandez, Department of Medical Physics, Hospital Sant Joan de Reus, IISPV, Reus 43204, Spain.
Email: vhernandezmasgrau@gmail.com

Abstract

Background: Statistical process control (SPC) is a powerful statistical tool for process monitoring that has been highly recommended in healthcare applications, including radiation therapy quality assurance (QA). The AAPM TG-218 report described the clinical implementation of SPC for Volumetric Modulated Arc Therapy (VMAT) pre-treatment verifications, pointing out the need to adjust tolerance limits based on plan complexity. However, the quantification of plan complexity and its integration into SPC remains an unresolved challenge.

Purpose: The primary aim of this study is to investigate the incorporation of plan complexity into the SPC framework for VMAT pre-treatment verifications. The study explores and evaluates various strategies for this incorporation, discussing their merits and limitations, and provides recommendations for clinical application.

Methods: A retrospective analysis was conducted on 309 VMAT plans from diverse anatomical sites using the PTW OCTAVIUS 4D device for QA measurements. Gamma Passing Rates (GPR) were obtained, and lower control limits were computed using both the conventional Shewhart method and three heuristic methods (scaled weighted variance, weighted standard deviations, and skewness correction) to accommodate non-normal data distributions. The 'Identify-Eliminate-Recalculate' method was employed for robust analysis. Eight complexity metrics were analyzed and two distinct strategies for incorporating plan complexity into SPC were assessed. The first strategy focused on establishing control limits for different treatment sites, while the second was based on the determination of control limits as a function of individual plan complexity. The study extensively examines the correlation between control limits and plan complexity and assesses the impact of complexity metrics on the control process.

Results: The control limits established using SPC were strongly influenced by the complexity of treatment plans. In the first strategy, a clear correlation was found between control limits and average plan complexity for each site. The second approach derived control limits based on individual plan complexity metrics, enabling tailored tolerance limits. In both strategies, tolerance limits inversely correlated with plan complexity, resulting in all highly complex plans being

classified as in control. In contrast, when plans were collectively analyzed without considering complexity, all the out-of-control plans were highly complex. **Conclusions:** Incorporating plan complexity into SPC for VMAT verifications requires meticulous and comprehensive analysis. To ensure overall process control, we advocate for stringent control and minimization of plan complexity during treatment planning, especially when control limits are adjusted based on plan complexity.

KEYWORDS
pre-treatment verifications, PSQA, statistical process control

1 | INTRODUCTION

Statistical process control (SPC), developed by Walter A. Shewhart, is a powerful tool for monitoring production processes. Its strength lies in establishing stability and reducing variability using statistical techniques. Applicable in various fields where output measurement is essential, SPC is particularly effective in maintaining process control over time.[1] It identifies random versus systematic errors in time-ordered data, prompting timely corrective actions to elevate quality and drive continuous improvement. Notably, SPC has gained widespread application in healthcare,[2–4] including quality assurance (QA) processes in radiation therapy.[1,5–7]

In radiation therapy, pre-treatment plan-specific quality assurance (PSQA) is essential for validating intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) plans, as well as for monitoring their performance stability over time.[8] Pre-treatment verifications typically involve comparing calculated and delivered doses via gamma index comparisons.[9] Plan acceptability is determined based on these PSQA results against established tolerance limits. The American Association of Physicists in Medicine (AAPM) recommends using SPC to set these limits specifically for each center's PSQA results,[10] with the goal of not only meeting clinical specifications but also refining processes, minimizing variability, and promoting continuous improvement.[1,11] The TG-218 report details the application of SPC in VMAT and IMRT pre-treatment QA,[10] highlighting that results within the control limits indicate that the process is under control, while deviations call for investigation and corrective actions. These control limits, however, are not universal and should be locally determined, considering each institution's specific equipment and procedures.

The agreement between planned and actual dose distributions is influenced by both calculation and delivery accuracy. Plan complexity is closely related to these accuracies, emerging as an important factor in this agreement.[12–17] Consequently, tolerance limits should consider plan complexity, as highly complex plans tend to show greater variability in PSQA results. Various metrics have been proposed to quantify these complexities, with different treatment sites requiring varying degrees of plan complexity for clinically acceptable dose distributions due to differences in target characteristics and proximity to critical structures.[18,19]

Acknowledging this, the AAPM has pointed out the importance of quantifying plan modulation[20] and adapting tolerance limits according to plan complexity.[10] However, the best methods to quantify plan modulation remain subject of discussion and it is not clear how plan complexity should be incorporated into SPC for pre-treatment verifications.

The purpose of this study is to investigate how to incorporate plan complexity into VMAT pre-treatment verifications using SPC. It includes a thorough SPC analysis, computation of multiple complexity metrics, and evaluation of their predictive power in anticipating PSQA failures. The study assesses different strategies for incorporating plan complexity metrics into SPC, evaluating their advantages and limitations, and concludes with recommendations for their clinical implementation.

2 | METHODS

In this study we retrospectively analyzed the PSQA results of 309 VMAT plans treated between 2019 and 2023. All plans were optimized for photon beams with nominal energy 6 or 10 MV, with flattening filter, using the Monaco 5.11 treatment planning system (TPS) with radiobiological cost functions and a Monte Carlo dose calculation algorithm. The voxel size was set to $3 \times 3 \times 2.5\ mm^3$ and the statistical uncertainty for plan dose was 1%. Treatments were delivered using an Elekta Synergy linear accelerator (Elekta, Crawley, UK) equipped with an Agility multileaf collimator (MLC).

The plan sample encompassed various anatomical sites: Head-and-Neck (HN) (60 plans), lung (62), breast (26), brain (53), prostate (46), and pelvis (62). Additional details on the number of PTVs, dose prescriptions, and other plan characteristics are given in the Supporting Material (Table S1).

QA measurements of VMAT plans were performed using the true composite method with the PTW Octavius1500 detector array inserted in the PTW

Octavius4D phantom.[21] This system recorded 2D array measurements as a function of gantry angle at different time intervals, facilitating the reconstruction of a 3D dose volume in the cylindrical phantom.[22,23]

Following the recommendations from AAPM,[10] measurements were compared with calculated doses via 3D absolute γ-analysis using the 3%/2 mm criteria with global normalization (90% of maximum dose) and a dose threshold of 10%. PTW VeriSoft software 6.1 was used and both Gamma Passing Rate (GPR) and average gamma index values were collected.

2.1 | SPC method

The I-chart plot was used to monitor PSQA results and control the process behavior. This chart visually represents time-ordered data and upper and lower control limits (LCL) were computed from the obtained results and the differences between consecutive points.

In this study we followed the definitions of control limits and action limits as given by Miften et al.[10] The control limit was defined as a boundary where the process operates with non-random error and is synonymous to the tolerance limit. As the maximum value of the GPR is bounded to 100%, the only tolerance limit for PSQA results was the LCL. The action limit was defined as the limit value where a negative clinical impact for the patient could occur; in this study the universal 90% value for the action limit recommended in the TG-218 report was used.

Initially, we applied the conventional Shewhart method[10,24–26] for the determination of the LCL values. However, this method was developed under the normality assumption, and several studies have demonstrated that the distributions of GPR values are highly skewed and not normally distributed.[27–29] To address this challenge, we explored the Johnson transformation[30] and the Box–Cox power transformation methods[31] in an effort to normalize the data, thereby rendering the Shewhart method applicable. The normality of the original and transformed datasets were evaluated using the Anderson-Darling test and visually inspected through Quantile-Quantile (Q-Q) plots.

Additionally, we incorporated three different heuristic methods for computing the LCL: the scaled weighted variance (SWV) method,[32] the weighted standard deviations (WSD) method,[33] and the skewness correction (SC) method,[34] beneficial for their independence from specific data distribution assumptions. These methods have been proven effective in determining LCL values for non-normal GPR distributions.[35,36] To improve the robustness of the methodology when retrospectively applying it to unknown processes, the 'Identify-Eliminate-Recalculate' method[36] was used. This method involved initial LCL determination with all data points, followed by iterative elimination of out-of-control points and recalculation until all points were in control. A detailed description on how to apply these heuristic methods and the 'Identify-Eliminate-Recalculate' method can be found in the publication from Xiao et al.[36]

2.2 | Plan complexity analysis

Plans were exported from the TPS in DICOM format and evaluated using the PlanAnalyzer software, coded in MATLAB, which computes multiple complexity metrics and plan parameters.[37] Numerous complexity metrics have been identified in the literature,[18,19] with several metrics targeting analogous aspects of treatment plans and showing strong correlations.[37] To ensure a comprehensive analysis, we chose a variety of complexity metrics that span a broad spectrum of plan characteristics, including the aperture size, aperture irregularity, leaf dynamics, and variations in both dose rate and gantry speed. The following eight complexity metrics were assessed:

1. Plan Irregularity (PI), indicating deviation of the beam aperture shapes from a circle. Its minimum value is 1, which would correspond to a perfect circle.[38]
2. Modulation Complexity Score (MCS), evaluating the MLC aperture variability in both shape and area.[12,13] The lower the MCS value, the higher the plan complexity, which ranges from 1 to 0.
3. Modulation index total (MItotal), quantifying variations in MLC speed and acceleration, gantry speed, and dose rate.[14]
4. Edge Metric (EM), calculating beam aperture complexity based on the ratio of the MLC side length and the aperture area.[39]
5. Leaf travel divided by the arc length (LT/AL), quantifying the average distance travelled by moving leaves divided by the total arc length.[13,37]
6. Mean MLC Gap (meanGap), calculating the average opening of the leaf pairs defining the beam aperture.[40]
7. First quartile of the distribution of leaf pair openings (Q1Gap), providing the 25th percentile of leaf pair openings. This is similar to meanGap but might be more indicative of small gap sizes as it is less affected by the presence of larger gaps.
8. Mean tongue-and-groove index (TGi), determined as the ratio of the average distance between adjacent leaves and the mean MLC gap and indicating the MLC aperture irregularity.[40]

We analyzed the variation and distribution of these metrics and their correlation with GPR values using Spearman's correlation coefficient. Receiver Operating Characteristic (ROC) curves were computed to assess each complexity metric's ability to predict plan acceptability. ROC curves are graphical plots that provide the true positive rate TPR (correct predictions of real failures) as a function of the false positive rate FPR

(wrong predictions of failures). In this study, they indicate the capability of each complexity metric to act as a binary classifier for plan acceptability as its discrimination threshold is varied. For each ROC curve, its area under the curve (AUC) was computed to quantify the performance of each metric as a binary classifier (the ideal value being 1). Further information on the application of ROC curves to the optimization of PSQA verifications can be found in the literature.[41]

2.3 | SPC analysis strategies

Two distinct strategies were implemented to integrate plan complexity into SPC analysis:

2.3.1 | Strategy 1: Analysis per treatment site

This strategy focused on differences between various treatment sites. Separate SPC analyses were conducted for each site, such as HN, lung, breast, brain, prostate, and pelvis. To investigate whether these differences could be attributed to variations in plan complexity, the relationship between the site-specific LCLs and the average complexity for each site was evaluated.

2.3.2 | Strategy 2: Analysis based on plan complexity

The second strategy focused on how plan complexity affects control limits in the PSQA process. Rather than categorizing plans by treatment site, they were grouped based on their complexity metrics. A sliding range method was used to form groups of plans centered at specific complexity levels. First, the 60 plans with the lowest complexity were grouped and their LCL was computed. Next, this group of plans was progressively shifted by intervals of 20 plans to include plans with gradually increasing complexity and the LCL was computed for each new group, thus allowing for an assessment of the impact of plan complexity on the SPC control limits. This approach generated overlaps between adjacent groups, which was useful to qualitatively estimate the uncertainties due to the finite number of plans and the sample size used.

3 | RESULTS

3.1 | SPC analysis and predictive power of complexity metrics

The distribution of measured GPRs did not adhere to the normal distribution; the Anderson-Darling test yielded a value of 0.777 at a 5% significance level, thus rejecting the hypothesis of normal distribution. This non-normality persisted after applying Johnson and Box–Cox power transformations, with statistics remaining at 2.0. Q-Q plots further evidencing this non-normality are provided in Figure S1 of the Supporting Material.

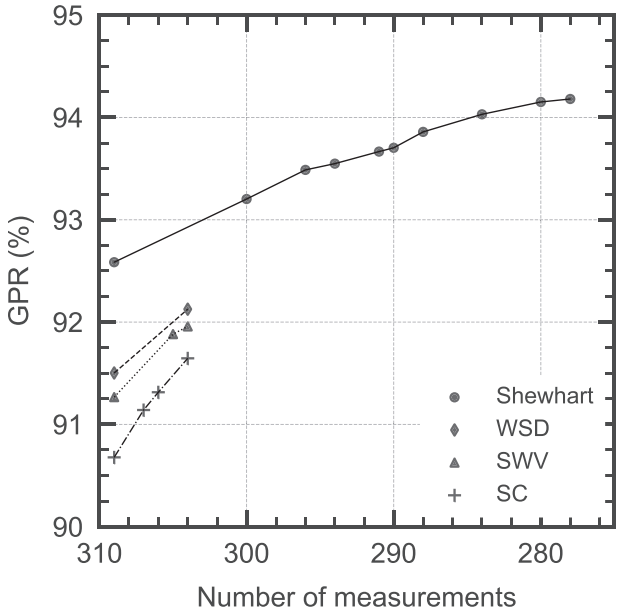

FIGURE 1 LCL values obtained using the three heuristic and the Shewhart procedures. Each symbol denotes an iteration of the 'Identify-Eliminate-Recalculate' method, which progressively eliminated out-of-control points until all data points were in control. LCL, lower control limit; SC, skewness correction method; SWV, scaled weighted variance method; WSD, weighted standard deviations method.

The LCL values determined through heuristic methods (SC, SWV, WSD), showed close similarity, ranging between 91.7% and 92.1% depending on the method applied (see Figure 1). The 'Identify-Eliminate-Recalculate' strategy consistently removed the same five out-of-control data points across all heuristic methods, yielding LCL values derived solely from in-control cases. The number of iterations for convergence varied by method: SC required four, SWV three, and WSD two iterations. In contrast, the Shewhart procedure, based on the assumption of normal data distribution, needed more iterations to converge and resulted in a LCL = 94.2%, approximately 2% higher than those from the heuristic methods.

Due to the similarity of results from the heuristic methods, subsequent results will focus on the WSD method, chosen for its alignment with SWV results (both in between those from the SC and the Shewhart methods) and its faster convergence compared to other methods.

The control chart for PSQA process data, depicted in Figure 2, identified five plans as out-of-control (i.e., below the LCL). These plans comprised three pelvic and two breast cases, constituting 1.6% of the 309 plans evaluated. The Shewart procedure, assuming a

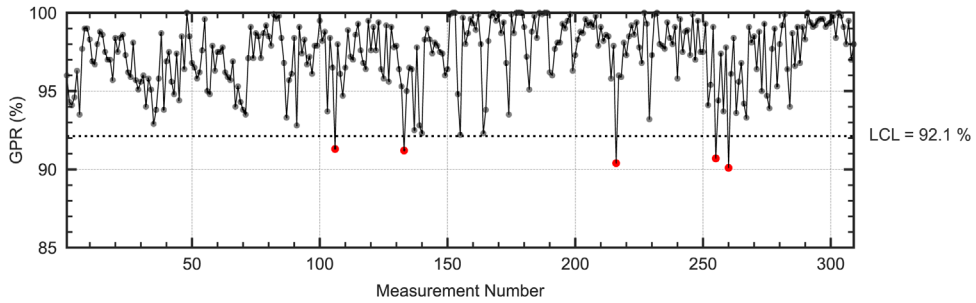

FIGURE 2 I-chart for the pre-treatment verification process, with symbols representing time-ordered results. Black symbols indicate in-control results, while red symbols show out-of-control data points. GPR, gamma passing rate; LCL, lower control limit.

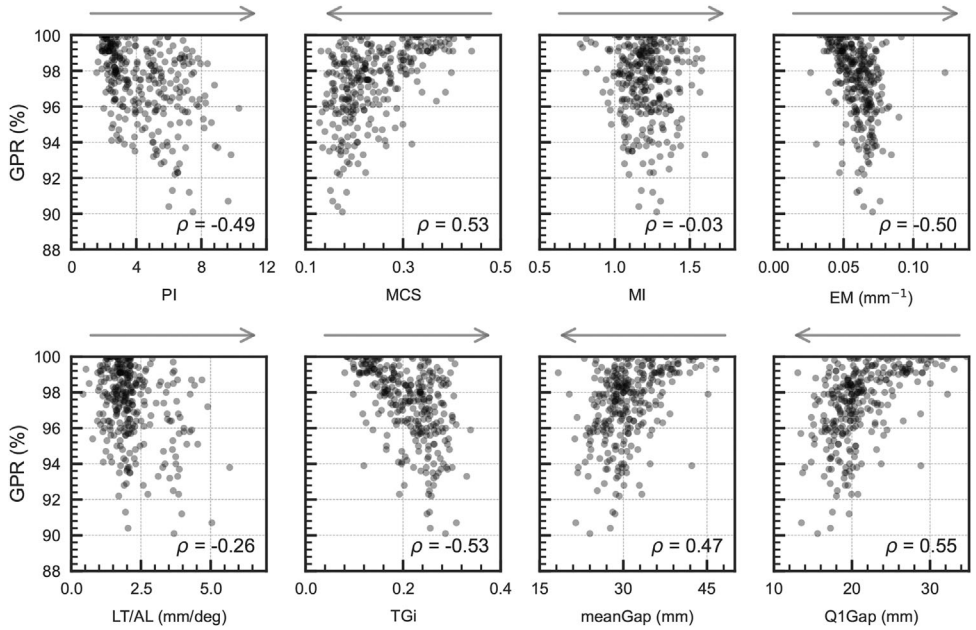

FIGURE 3 Relationship between GPR and evaluated plan complexity metrics with their corresponding Spearman's correlation coefficient *r*. Arrows indicate the direction of increasing plan complexity for each metric. Note that plan complexity increases with the score for PI, MI, EM, LT/AL, and TGi, while for MCS, meanGap, and Q1Gap plan complexity decreases with the metric's value. EM, edge metric; GPR, gamma passing rates; LT/AL, leaf travel divided by the arc length; MI, modulation index; meanGAP, mean MLC gap; PI, plan irregularity; TGi, tongue-and-groove index.

normal distribution, would flag a much higher number of out-of-control plans (27 cases in Figure 2 with GPR below 94.2%). All GPR results were above the lower action limit of 90%, indicating the need for investigation but not for immediate action.

A strong correlation (Spearman's $\rho = -0.87$) was found between GPR and average gamma values, as detailed in Figure S2 of the Supporting Material. Hence, GPR was chosen as the sole criterion for plan acceptability in subsequent SPC analyses.

The analysis of complexity metrics across all treatment plans revealed a broad spectrum of values. The ranges for each metric were 1.5–10.3 for PI, 0.07–0.44 for MCS, 0.77–1.60 for MI total, 0.03–0.12 for EM, 0.4–5.7 for LT/AL, 0.08–0.34 for TGi, 18.3–46.7 mm for meanGap, and 13.1–34.7 mm for Q1Gap.

Figure 3 illustrates the relationship between plan complexity metrics and GPR. Generally, as plan complexity increased, average GPRs tended to decrease, indicating a higher likelihood of PSQA failure in more complex plans. The five out-of-control cases, all with GPRs below 92.1%, demonstrated high plan complexity, particularly as indicated by the metrics PI, MCS, TGi, meanGap, and Q1Gap. However, the correlations between GPRs and complexity metrics were relatively weak, with the highest Spearman's correlation coefficients at 0.55 for Q1Gap and 0.53 for MCS.

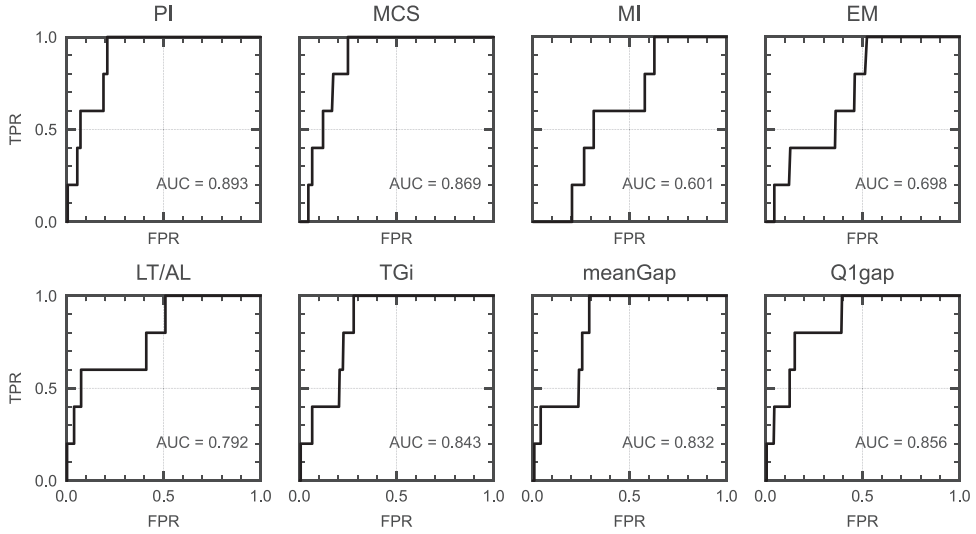

FIGURE 4 ROC curves for each evaluated metric, including their corresponding AUC for each curve. AUC, area under the curve; FPR, false positive rates; ROC, receiver operating characteristic; TPR, true positive rates.

ROC curve analysis, shown in Figure 4, assessed the predictive capability of different complexity metrics in identifying GPR results below the LCL of 92.1% as failures ("positives"). The best-performing metric was PI (AUC = 0.893), followed by MCS (AUC = 0.869), Q1Gap (AUC = 0.856), and TGi (AUC = 0.843). Consequently, these four metrics were the ones selected in subsequent analyses.

3.2 | Introducing plan complexity into the SPC analysis

3.2.1 | Strategy 1: Analysis per treatment site

In this strategy plans were grouped by treatment site and separate analyses were conducted to compute site-specific LCL values from the corresponding GPRs (given in Figure S3 of the Supporting Material). The LCLs obtained for each treatment site were 98.2% for prostate, 93.2% for brain, 92.0% for lung, 91.1% for HN, 88.8% for pelvis, and 88.8% for breast. Complex sites such as HN and pelvis showed lower LCL values compared to simpler sites like prostate. Figure 5 illustrates the correlation between site-specific LCLs and average complexity metrics for each site, focusing on the four metrics with the highest AUC from earlier analysis. A clear correlation is evident, with LCL values inversely related to plan complexity.

Using these site-specific control limits reduced the number of out-of-control plans from five to four. Interestingly, the four out-of-control plans were prostate plans with high GPRs ranging between 96.8% and 98%, only slightly lower than the 98.2% LCL for prostate. In contrast, plans from more complex sites were all classified as in control. This outcome highlights that site-specific LCLs are more forgiving for highly complex sites but stricter for simpler sites, possibly resulting in out-of-control classifications despite high measured GPRs.

3.2.2 | Strategy 2: Analysis based on plan complexity

The analysis of the complexity of individual plans within each treatment site, shown in Figure 6, highlighted that individual plans within the same site exhibit significant variability in complexity. Furthermore, overlaps in complexity metrics across different sites show that some plans may exhibit higher complexity than others from sites with lower average complexity, and vice versa.

To further explore the relationship between SPC limits and plan complexity, in strategy #2 plans were grouped by their individual complexity degree, regardless of their treatment site, and the shifting range method was used to compute LCLs through the full spectrum of complexities. Figure 7 summarizes the results of applying this method to the four best-performing metrics from the earlier ROC analysis, showing LCLs decreasing with increasing plan complexity. Clear differences were observed across metrics. For PI, LCLs ranged from GPR = 88% to 95%, with all data points in control. For other metrics, higher LCLs were found (e.g., up to 98.2% for TGi), leading to some data points being flagged as out-of-control. The number of out-of-control points varied depending on the metric used, with five for MCS, two for Q1 Gap, and four for TGi, considering the linear fits in Figure 7. Notably, most out-of-control data points

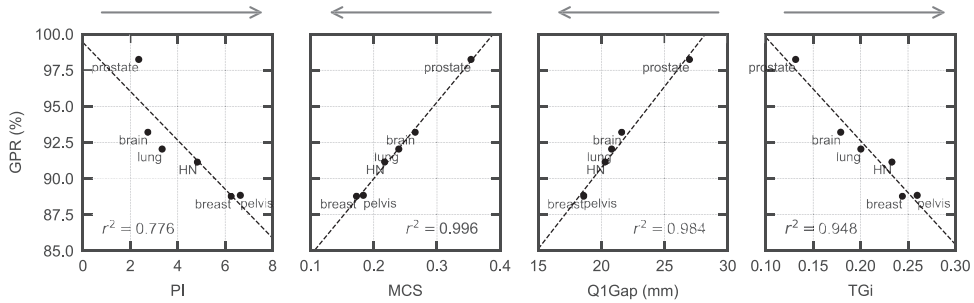

FIGURE 5 Relationship between LCL values and average complexity metrics for each treatment site, including a linear fit with its correlation coefficient r^2. Arrows indicate the direction of increasing plan complexity for each metric. LCL, lower control limit.

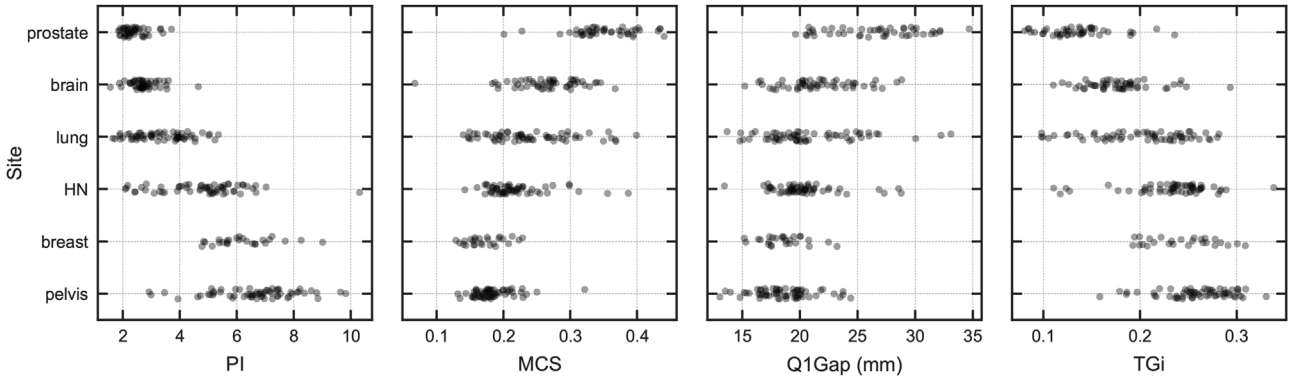

FIGURE 6 Complexity metrics as a function of the treatment site.

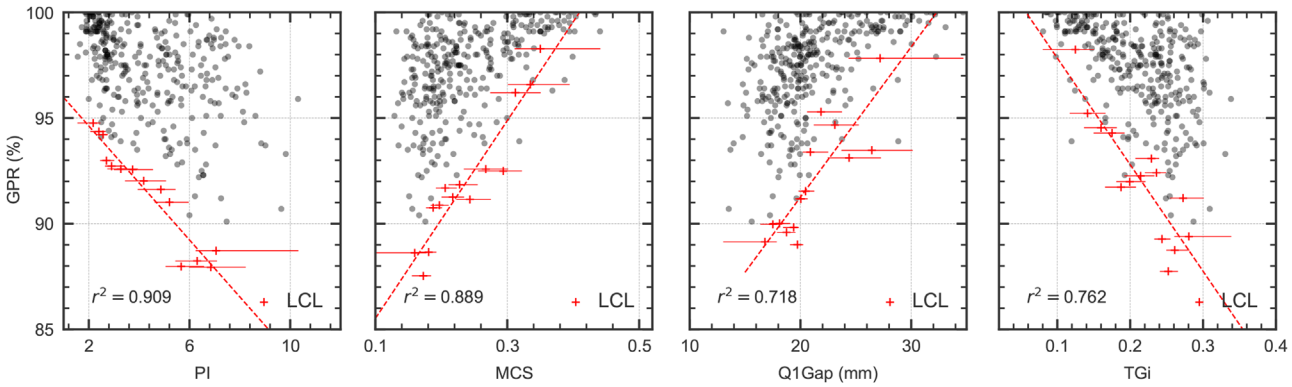

FIGURE 7 GPR results and LCL values relative to plan complexity. Circles represent GPR results. Red crosses indicate LCLs for similar complexity plans, positioned at each group's median complexity, with the dashed line indicating the linear fit. GPR, gamma passing rate; LCL, lower control limit.

corresponded to plans with low complexity and high GPRs (even higher than 98% for MCS), a result of the stricter LCLs for lower complexity plans, while all highly complex plans were consistently classified as in control.

4 | DISCUSSION

We applied the SPC methodology to determine LCLs for measurement-based pre-treatment verifications and explored the incorporation of plan complexity into SPC analysis. In line with previous studies,[27–29] we confirmed the non-normal distribution of measured GPRs, which persisted even after attempting to normalize the data through statistical transformations. The heuristic methods (SC, SWV, WSD), combined with the iterative 'Identify-Eliminate-Recalculate' procedure, proved effective and robust for SPC analysis of retrospective GPRs, corroborating findings by Xiao et al.[36]

The LCL value of 92.1% for all VMAT plans, irrespective of treatment site or plan complexity, was lower than the 95% universal value proposed in the TG-218 report.[10] This variance can be attributed to the equipment and procedures specific to each institution, the distribution of anatomical sites treated, and the methodology employed for calculating the LCL. For instance, employing the Shewhart procedure in our analysis yielded a LCL of 94.2%. Moreover, our LCL of 92.1% aligns well with findings from similar studies[25,35] and is slightly higher than the example values TG-218 provided for VMAT plans.

We also want to highlight that, despite weak correlations between complexity metrics and PSQA results, several individual complexity metrics notably succeeded in classifying plan acceptability. PI emerged as the most predictive metric (AUC = 0.893), followed by MCS, Q1Gap, and TGi. PI's simplicity in computation, compared to other metrics such as MCS and MItotal, made it a practical choice, though the four metrics with the highest AUC were included in the subsequent analyses for completeness.

In this study, we selected eight complexity metrics covering a broad spectrum of complexity aspects and plan characteristics, aiming to identify the most critical factors influencing plan acceptability. The metrics demonstrating the highest predictive power were those associated with MLC aperture characteristics, such as aperture irregularity (PI and TGi), aperture size (Q1Gap), and the variability in aperture shapes and sizes (MCS). These findings are in line with those of the wider community, highlighting the significance of MLC aperture-related factors in determining plan acceptability.[42] However, it is important to note that the relevance of these metrics might vary depending on local equipment and procedures. Therefore, we recommend conducting a local evaluation to ascertain the predictive power of different complexity metrics, rather than applying these findings universally.

The two strategies evaluated for incorporating plan complexity into SPC analysis used a different approach. The first strategy, focusing on LCLs per treatment site, revealed large differences across sites, in agreement with the findings by Xiao et al.[36] Interestingly, this strategy yielded strong correlations between LCLs and the average plan complexity at each site (see Figure 5). Hence, this site-specific approach serves as an indirect way to account for plan complexity in the management of pre-treatment verifications. However, this strategy also altered the outcome of the SPC analysis radically, identifying a completely different set of out-of-control plans and classifying all highly complex plans as in control.

Note that the 'per-treatment site' strategy does not require computing complexity metrics. Although we computed several metrics to explore the relationship between control limits and average plan complexity for each site, this strategy can be applied by determining control limits for each anatomical site independently of complexity metrics. Nonetheless, we believe that assessing complexity for individual plans is useful to control plan complexity, thus keeping it within a safe range and minimizing the associated uncertainties. Tailoring this range to each treatment site is crucial, as it facilitates the identification of plans with excessive complexity that might benefit from being replanned. While commercial TPSs offer various tools to manage and reduce plan complexity, such as limiting the number of monitor units (MUs) or other specific features,[43] these tools vary across different TPS platforms. This approach aligns with class solution concepts, where plan parameters are controlled to maximize consistency and quality.[44,45] Ideally, complexity metrics should be computed and handled directly into the TPS for better management during the planning stage.[16–19]

The second strategy, focusing on LCLs based on individual plan complexity, enabled the tailoring of LCLs for each plan. This strategy was justified by the large variability in plan complexity within each treatment site, which is in line with the difficulty to differentiate between treatment sites using complexity metrics.[46] Like the site-specific approach, strategy #2 effectively classified most high-complexity plans as in control, with most out-of-control plans exhibiting low complexity. The number of out-of-control plans varied depending on the complexity metric used, and this approach required complexity quantification for all plans, a challenging task with current commercial TPSs.[47]

Both strategies pose risks. Excessively low LCLs for complex plans may compromise treatment accuracy; in such instances, the lower action limit must be considered, and clinical acceptability should be carefully evaluated. Conversely, high LCLs for low-complexity plans can be overly sensitive to minor uncertainties that hold little clinical relevance. To mitigate this, an upper threshold for the LCL could be set, for instance around GPR = 97% -98%. Finally, measurement-based pre-treatment verifications are often carried out with the purpose of controlling plan complexity and the impact of its associated uncertainties.[17] However, this goal could be undermined if control limits are adjusted based on the degree of plan complexity. While the clinical impact of adjusting SPC limits based on plan complexity is expected to be small, managing PSQA failures can be resource-intensive and potentially lead to treatment delays. By refining these limits to take into account plan complexity, we can assign lower LCL values to plans where higher complexity is needed, which can be useful in reducing the incidence of PSQA failures and can contribute to optimizing the clinical workflow.

In our opinion, complexity metrics should ideally be available at the TPS to facilitate the minimization of

plan complexity during treatment planning and reduce uncertainties in dose calculation and delivery. Control limits can then be adjusted based on the individual plan complexity (strategy #2), which can be useful to prevent out-of-control instances in challenging clinical cases in which highly complex plans are needed. For this approach, the complexity metric with the highest predictive power for each center should be used. In situations where tools for computing complexity metrics are unavailable, categorizing plans by treatment site (strategy #1) remains a valid alternative. Regardless of the chosen strategy, manual adjustments to clinically used tolerance limits may be necessary to keep them within an appropriate range. Furthermore, action limits for plan acceptability should be based on clinical criteria.

A limitation of our study is the reliance on GPR as the metric for determining plan acceptability. It has been observed that GPR may not be highly sensitive to discrepancies between calculated and delivered dose distributions, nor does it necessarily correlate well with their clinical impact.[48] Despite these shortcomings, we chose GPR for its endorsement by current guidelines[10] and its widespread use in clinical settings. Importantly, the methodologies and strategies we have introduced are adaptable and can be applied irrespective of the specific metrics utilized to assess plan acceptability.[41] While our research concentrated on measurement-based verifications, the approach to adjusting SPC limits that we propose can also be extended to secondary dose calculations or other software-based verification methods.[49–51]

Another limitation is that it was conducted within a single institution. As indicated in the TG218 report,[10] tolerance limits for PSQA are influenced by local equipment, processes, case types, and physicist's expertise. Therefore, our results should not be directly extrapolated to other institutions. However, we believe that our strategies and findings will assist centers to incorporate plan complexity into their handling of pre-treatment verifications.

Finally, to simplify the SPC analysis and interpretation we focused on single complexity metrics. AI models utilizing multiple plan parameters and complexity metrics show promise as more effective classifiers for plan acceptability[52–54] and it is reasonable to believe that such AI models could outperform single-metric classifiers and further improve the approach's efficiency. Future research is needed to investigate how these AI models can be incorporated into the management of pre-treatment verifications and SPC methodologies.

5 | CONCLUSIONS

SPC proves effective in reducing variability in VMAT pre-treatment verifications. Our investigation of two distinct strategies for integrating plan complexity into SPC has revealed a strong dependency of control limits on the complexity of treatment plans. The first strategy, focusing on different treatment sites, showed a direct correlation between control limits and the average plan complexity at each site. The second strategy, taking a more individualized approach, derived control limits directly from the complexity of each treatment plan. In both strategies, tolerance limits decreased as plan complexity increased, leading to higher complexity plans consistently meeting control standards, while some less complex plans with high GPRs did not. This outcome contrasts markedly with scenarios where plan complexity was not considered, where all plans deemed out-of-control were of high complexity. The incorporation of plan complexity into SPC requires meticulous analysis and possibly manual adjustments to tolerance limits. Our recommendation is to control and minimize plan complexity during treatment planning for reducing the uncertainties in clinical plans and for maintaining control throughout the entire process, especially when control limits are tailored to plan complexity.

ACKNOWLEDGMENTS

The authors have nothing to report.

CONFLICT OF INTEREST STATEMENT

The authors have no relevant conflicts of interest to disclose.

ORCID

Serenella Russo https://orcid.org/0000-0002-8740-1039
Jordi Saez https://orcid.org/0000-0003-4888-9323
Marco Esposito https://orcid.org/0000-0002-9332-5351
Andrea Bruschi https://orcid.org/0000-0002-0717-4981
Silvia Scoccianti https://orcid.org/0000-0002-8457-2123
Victor Hernandez https://orcid.org/0000-0003-3770-8486

REFERENCES

1. Pawlicki T, Whitaker M, Boyer AL. Statistical process control for radiotherapy quality assurance. *Med Phys*. 2005;32:2777-2786.
2. Suman G, Prajapati DR. Control chart applications in healthcare: a literature review. *Int J Metrol Qual Eng*. 2018;9:1-21.
3. Tennant R, Mohammed MA, Coleman JJ, Martin U. Monitoring patients using control charts: a systematic review. *Int J Qual Health Care*. 2007;19(4):187-194.
4. Noyez L. Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interact Cardiovasc Thorac Surg*. 2009;9(3):494-499.
5. Pawlicki T, Whitaker M. Variation and control of process behavior. *Int J Radiat Oncol Biol Phys*. 2008;71:S210-S214.
6. Pawlicki T, Yoo S, Court LE, et al. Process control analysis of IMRT QA: Implications for clinical trials. *Phys Med Biol*. 2008;53:5193-5205.

7. Arjomandy B, Taylor P, Ainsley C, et al. AAPM task group 224: Comprehensive proton therapy machine quality assurance. *Med Phys*. 2019;46:e678-e705.
8. Moran JM, Dempsey M, Eisbruch A, et al. Safety considerations for IMRT: executive summary. *Pract Radiat Oncol*. 2011;1:190-195.
9. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys*. 1998;25:656-661.
10. Miften M, Olch A, Mihailidis D, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM task group No. 218. *Med Phys*. 2018;45:e53-e83.
11. Breen SL, Moseley DJ, Zhang B, Sharpe MB. Statistical process control for IMRT dosimetric verification. *Med Phys*. 2008;35:4417-4425.
12. McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. *Med Phys*. 2010;37:505-515.
13. Masi L, Doro R, Favuzza V, Cipressi S, Livi L. Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy. *Med Phys*. 2013;40:071718.
14. Park JM, Park SY, Kim H, Kim JH, Carlson J, Ye SJ. Modulation indices for volumetric modulated arc therapy. *Phys Med Biol*. 2014;59:7315-7340.
15. Götstedt J, Hauer AK, Bäck A. Development and evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings. *Med Phys*. 2015;42:3911-3921.
16. Younge KC, Roberts D, Janes LA, Anderson C, Moran JM, Matuszak MM. Predicting deliverability of volumetric-modulated arc therapy (VMAT) plans using aperture complexity analysis. *J Appl Clin Med Phys*. 2016;17:124-131.
17. Hernandez V, Hansen CR, Widesott L, et al. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans. *Radiother Oncol*. 2020;153:26-33.
18. Chiavassa S, Bessieres I, Edouard M, Mathot M, Moignier A. Complexity metrics for IMRT and VMAT plans: a review of current literature and applications. *Br J Radiol*. 2019;92:20190270.
19. Antoine M, Ralite F, Soustiel C, et al. Use of metrics to quantify IMRT and VMAT treatment plan complexity: a systematic review and perspectives. *Physica Med*. 2019;64:98-108.
20. 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;36:5359-5373.
21. Stelljes TS, Harmeyer A, Reuter J, et al. Dosimetric characteristics of the novel 2D ionization chamber array OCTAVIUS Detector 1500. *Med Phys*. 2015;42:1528-1537.
22. Stathakis S, Myers P, Esquivel C, Mavroidis P, Papanikolaou N. Characterization of a novel 2D array dosimeter for patient-specific quality assurance with volumetric arc therapy. *Med Phys*. 2013;40:071731.
23. Esch AV, Basta K, Evrard M, Ghislain M, Sergent F, Huyskens DP. The Octavius1500 2D ion chamber array and its associated phantoms: dosimetric characterization of a new prototype. *Med Phys*. 2014;41:091708.
24. Sanghangthum T, Suriyapee S, Srisatit S, Pawlicki T. Statistical process control analysis for patient-specific IMRT and VMAT QA. *J Radiat Res*. 2013;54:546-552.
25. Bellec J, Delaby N, Jouyaux F, et al. Plan delivery quality assurance for CyberKnife: statistical process control analysis of 350 film-based patient-specific QAs. *Physica Med*. 2017;39:50-58.
26. Esposito M, Ghirelli A, Pini S, et al. Clinical implementation of 3D in vivo dosimetry for abdominal and pelvic stereotactic treatments. *Radiother Oncol*. 2021;154:14-20.
27. Chou YM, Polansky AM, Mason RL. Transforming non-normal data to normality in statistical process control. *J Qual Technol*. 1998;30:133-141.
28. Riaz M, Mehmood R, Iqbal MR, Abbasi SA. On efficient skewness correction charts under contamination and non-normality. *Qual Reliab Eng Int*. 2016;32:837-854.
29. Aako OL, Adewara JA, Adekeye KS, Nkemnole EB. X and R control charts based on marshall-olkin inverse log-logistic distribution for positive skewed data. *African Journal of Pure and Applied Sciences*. 2020;1:9-16.
30. Johnson NL. Systems of frequency curves generated by methods of translation. *Biometrika*. 1949;36:149-176.
31. Box GEP, Cox DR. An analysis of transformations. *J Roy Statist Soc Ser B (Methodological)*. 1964;26:211-252.
32. Choobineh F, Ballard JL. Control-limits of QC charts for skewed distributions using weighted-variance. *IEEE Trans Reliab*. 1987;R-36:473-477.
33. Chang YS, Bai DS. Control charts for positively-skewed populations with weighted standard deviations. *Qual Reliab Eng Int*. 2001;17:397-406.
34. Chan LK, Cui HJ. Skewness correction X and R charts for skewed distributions. *Nav Res Logist*. 2003;50:555-573.
35. Xiao Q, Bai S, Li G, et al. Statistical process control and process capability analysis for non-normal volumetric modulated arc therapy patient-specific quality assurance processes. *Med Phys*. 2020;47:4694-4702.
36. Xiao Q, Bai L, Li G, et al. A robust approach to establish tolerance limits for the gamma passing rate-based patient-specific quality assurance using the heuristic control charts. *Med Phys*. 2022;49:1312-1330.
37. Hernandez V, Saez J, Pasler M, Jurado-Bruggeman D, Jornet N. Comparison of complexity metrics for multi-institutional evaluations of treatment plans in radiotherapy. *Phys Imaging Radiat Oncol*. 2018;5:37-43.
38. Du W, Cho SH, Zhang X, Hoffman KE, Kudchadker RJ. Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. *Med Phys*. 2014;41:021716.
39. Younge KC, Matuszak MM, Moran JM, McShan DL, Fraass BA, Roberts DA. Penalization of aperture complexity in inversely planned volumetric modulated arc therapy. *Med Phys*. 2012;39:7160-7170.
40. Vieillevigne L, Khamphan C, Saez J, Hernandez V. On the need for tuning the dosimetric leaf gap for stereotactic treatment plans in the eclipse treatment planning system. *J Appl Clin Med Phys*. 2019;20:68-77.
41. McKenzie EM, Balter PA, Stingo FC, Jones J, Followill DS, Kry SF. Toward optimizing patient-specific IMRT QA techniques in the accurate detection of dosimetrically acceptable and unacceptable patient plans. *Med Phys*. 2014;41:121702.
42. Brooks FMD, Glenn MC, Hernandez V, et al. A radiotherapy community data-driven approach to determine which complexity metrics best predict the impact of atypical TPS beam modeling on clinical dose calculation accuracy. *J Appl Clin Med Phys*. Published online March 1, 2024. doi:10.1002/acm2.14318
43. Scaggion A, Fusella M, Agnello G, et al. Limiting treatment plan complexity by applying a novel commercial tool. *J Appl Clin Med Phys*. 2020;21(8):27-34.
44. Van der Wal E, Wiersma J, Ausma AH, et al. Code of practice for the quality assurance and control for intensity modulated radiotherapy report 22 of the Netherlands commission on radiation dosimetry. 2013.
45. Mans A, Schuring D, Arends MP, et al. The NCS code of practice for the quality assurance and control for volumetric modulated arc therapy. *Phys Med Biol*. 2016;61:7221-7235.
46. Desai VK, Labby ZE, DeWerd LA, Culberson WS. On the implementation of the plan-class specific reference field using multidimensional clustering of plan features and alternative strategies for improved dosimetry in modulated clinical linear accelerator treatments. *Med Phys*. 2020;47(8):3621-3635.

47. Kamperis E, Kodona C, Hatziioannou K, Giannouzakos V. Complexity in radiation therapy: it's complicated. *Int J Radiat Oncol Biol Phys*. 2020;106:182-184.
48. Stasi M, Bresciani S, Miranti A, Maggio A, Sapino V, Gabriele P. Pretreatment patient-specific IMRT quality assurance: a correlation study between gamma index and patient clinical dose volume histogram. *Med Phys*. 2012;39(12):7626-7634.
49. Zhu TC, Stathakis S, Clark JR, et al. Report of AAPM task group 219 on independent calculation-based dose/MU verification for IMRT. *Med Phys*. 2021;48(10):e808-e829.
50. Wang L, Li J, Zhang S, et al. Multi-task autoencoder based classification-regression model for patient-specific VMAT QA. *Phys Med Biol*. 2020;65:235023.
51. Huang Y, Pi Y, Ma K, et al. Virtual patient-specific quality assurance of IMRT using UNet++: classification, gamma passing rates prediction, and dose difference prediction. *Front Oncol*. 2021;11:700343.
52. Valdes G, Scheuermann R, Hung CY, Olszanski A, Bellerive M, Solberg TD. A mathematical framework for virtual IMRT QA using machine learning. *Med Phys*. 2016;43:4323-4334.
53. Li J, Wang L, Zhang X, et al. Machine learning for patient-specific quality assurance of VMAT: prediction and classification accuracy. *Int J Radiat Oncol Biol Phys*. 2019;105:893-902.
54. Mehrens H, Molineu A, Hernandez N, et al. Characterizing the interplay of treatment parameters and complexity and their impact on performance on an IROC IMRT phantom using machine learning. *Radiother Oncol*. 2023;182:109577.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Russo S, Saez J, Esposito M, et al. Incorporating plan complexity into the statistical process control of volumetric modulated arc therapy pre-treatment verifications. *Med Phys*. 2024;51:3961–3971. https://doi.org/10.1002/mp.17081