Best Practice Guidance for Patient-Specific Quality Assurance for IMRT and VMAT Plan Delivery Verification

October 15, 2019

Physics Community of Practice Working Group – Patient-Specific Quality Assurance

Group members:

Gordon Chan (Chair)
Juravinski Cancer Centre

Lee Chin (Co-chair)
Odette Cancer Centre

Ady Abdellatif
Walker Family Cancer Centre

Jean-Pierre Bissonnette
Princess Margaret Cancer Centre

Daria Comsa
Southlake Regional Cancer Centre

Dal Granville
Ottawa Hospital Regional Cancer Centre

Jenna King
Simcoe Muskoka Regional Cancer Centre

Patrick Rapley
Thunder Bay Regional Health Sciences Centre

Aaron Vandermeer
R.S. McLaughlin Durham Regional Cancer Centre

Cancer Care Ontario Lead:

Julie Himmelman

Table of Contents

1. Introduction 4
2. Summaries of Best Practice Recommendations 4
Summary of Key Quality Indicators: Policy and Procedure 5
Summary of Key Quality Indicators: PSQA Instrumentation 5
Summary of Key Quality Indicators: PSQA Measurement Setup and Methodology 6
Summary of Key Quality Indicators: PSQA Evaluation Methodology 6
Summary of Key Quality Indicators: Documentation, Process and Feedback 7
3. Policy and Procedure 8
3.1 Verification of Plan Delivery Parameters 8
3.2 Verification of Plan Deliverability and Dose Accuracy 8
3.3 Prevention of Accidental Alterations to Plan Delivery Parameters 8
3.4 Restrictions/Recommendations on Treatment Optimization and Calculation Parameters 9
4. PSQA Instrumentation 9
4.1 Choice of PSQA Instrument for Technique/Class Solution Development 9
4.2 Choice of Instrument for Routine PSQA 9
4.3 Commissioning of Detectors 10
4.4 QA Program for PSQA tools 10
4.5 Detector Calibration 10
5. PSQA Measurement Setup and Methodology 11
5.1 Placement for Detectors 11
5.2 Detector Acclimatization 11
5.3 Detector Setup 11
5.4 Measurement on Beam-matched Linacs 12
5.5 Density Heterogeneity on CT 12
5.6 Machine Output Variation 12
6. PSQA Evaluation Methodology 13
6.1 Plan Comparison Methodology 13
6.2 Reference and Evaluated Dose Distributions 13
6.3 Dose Normalization 14
6.4 Choice of Normalization Dose 14
6.5 Dose Interpolation 14
6.6 Vender Specific Features 15

6.7 Evaluation and Passing Rate Criteria 15
6.8 Validation of Evaluation and Passing Rate Criteria 15
6.9 Additional Evaluation Parameters 16
7. Documentation, Process and Feedback 16
7.1 Documentation on Acceptance Criteria 16
7.2 Process for PSQA results 17
7.3 Feedback Mechanism 17
7.4 Periodic Reviews 17
Appendix I: Delivery Log 19
Glossary 20
References 22

1. Introduction

The Cancer Care Ontario (CCO) Physics Community of Practice identified Patient-Specific Quality Assurance (PSQA) as a priority quality initiative. The main objective for the initiative was to improve the quality and safety in treatment delivery by supporting coordination and standardization of PSQA practice across the province. A working group was struck to investigate PSQA best practice guidance for IMRT/VMAT delivery verification. It also serves to advise the development of a new provincial funding model for radiation treatment (Radiation Treatment Quality Based Procedure initiative), specific to quality metrics, on the subject of PSQA best practice.

The working group is composed of nine medical physicists from cancer centres across Ontario and a CCO lead. The group conducted a provincial survey on IMRT/VMAT PSQA that illustrated variations in practice across the Ontario cancer centres. It also conducted a literature review on existing guidelines and peer reviewed papers on PSQA to further inform the best practice recommendations outlined in this document. In the absence of evidence, the collective clinical expertise and experience of the group members were utilized.

This document should be considered for review in 3 years unless newly published research or guidance document from a professional society warrants a change in PSQA practice.

2. Summaries of Best Practice Recommendations

This section outlines five summaries of PSQA best practice recommendations for IMRT and VMAT. Each summary compiles a number of Key Quality Indicators (KQIs), which will be explained in details in the subsequent sections.

While VMAT is generally considered to be a subset of IMRT, in this document, VMAT refers to moving gantry delivery, while IMRT refers to fixed gantry delivery. Terminology highlighted by bold face is defined in the Glossary section. The recommendations in the KQIs apply to routine IMRT or VMAT patient-specific delivery quality assurance on regular linacs, including Stereotactic Body Radiation Therapy, but excluding Stereotactic Radiosurgery or PSQA for specialized machines such as Tomotherapy or Cyberknife. “Forward-planned” IMRT or IMRT that primarily generates wedge segmentation (e.g. breast tangents) is also excluded since the bulk of the dose comes from open beams.

Summary of Key Quality Indicators: Policy and Procedure

KQI		Indicator Measure	Section Reference
Policy and Procedure Quality Indicators			
A1	A procedure must be implemented to verify the consistency of plan delivery parameters prior to first fraction for all IMRT and VMAT treatments.	0 or 1	3.1
A2	A procedure should be in place to verify plan deliverability and dose accuracy prior to first fraction for all IMRT and VMAT treatments.	0 or 1	3.2
A3	A procedure must be implemented to prevent or catch accidental alterations to the plan delivery parameters prior to each treatment.	0 or 1	3.3
A4	Restrictions or recommendations on certain IMRT/VMAT treatment planning system (TPS) optimization and calculation parameters that impact PSQA pass rates should be implemented.	0 or 1	3.4

Summary of Key Quality Indicators: PSQA Instrumentation

KQI		Indicator Measure	Section Reference
PSQA Instrumentation Quality Indicators			
B1	PSQA of new IMRT/VMAT techniques or **class solutions** should be evaluated using measurement devices with high dosimetric accuracy and high spatial resolution.	0 or 1	4.1
B2	Following validation of new IMRT/VMAT techniques or **class solutions**, routine PSQA may employ devices with slightly lower dosimetric accuracy or spatial resolution.	0 or 1	4.2
B3	Detectors must be commissioned prior to clinical use.	0 or 1	4.3
B4	A QA program must be in place to ensure continued and consistent performance of the PSQA tools at the level at which they were commissioned.	0 or 1	4.4
B5	Detector calibration (single or array) must be performed at a frequency dependent on its usage or if it fails regular quality control (QC) tolerance.	0 or 1	4.5

Summary of Key Quality Indicators: PSQA Measurement Setup and Methodology

KQI		Indicator Measure	Section Reference
PSQA Measurement Setup and Methodology Quality Indicators			
C1	The PSQA detector should be set up to maximize the measurement of relevant clinical region.	0 or 1	5.1
C2	Detector and phantom should reach equilibrium temperature with the environment before use or a temperature correction factor is applied.	0 or 1	5.2
C3	**True composite** is the preferred detector setup and delivery method.	0 or 1	5.3
C4	Patient-specific measurements can be done on any **beam-matched linacs** provided consistent inter-machine performance as demonstrated by compliance to appropriate regular linac QC guidelines.	0 or 1	5.4
C5	Density heterogeneity on CT must be considered in relevant software and phantoms in all PSQA measurements.	0 or 1	5.5
C6	Variation of machine output should be accounted for before every measurement session.	0 or 1	5.6

Summary of Key Quality Indicators: PSQA Evaluation Methodology

KQI		Indicator Measure	Section Reference
PSQA Evaluation Methodology Quality Indicators			
D1	If available, reconstructed 3D dose distribution in patient anatomy from measured dose or fluence should be used to provide clinically relevant comparison with the clinical plan. Otherwise, 2D or 3D phantom-based **Gamma** evaluation may be used for plan comparison.	0 or 1	6.1
D2	The **evaluated dose distribution** should have at least the same or higher spatial resolution and dimensionality than the **reference dose distribution**.	0 or 1	6.2
D3	**Global normalization** in absolute dose should be used.	0 or 1	6.3
D4	Normalization dose should be chosen in a consistent manner for each **class solution**, in a low gradient, high dose region.	0 or 1	6.4

D5	Dose interpolation should be done for the **evaluated dose distribution** prior to **Gamma** analysis if the spatial resolution is greater than 1/3 the DTA criterion.	0 or 1	6.5
D6	Optional vendor specific features should be carefully evaluated for their impact on the measurement results before use, and should be enabled in a consistent manner for each **class solution**.	0 or 1	6.6
D7	**Gamma** passing rate tolerance and action levels of 95% and 90%, respectively, using 3%/2mm and 10% low dose threshold in **Gamma** analysis are recommended.	0 or 1	6.7
D8	PSQA program evaluation and passing rate criteria should be validated to ensure the chosen parameters adequately catch known errors.	0 or 1	6.8
D9	If available, clinically relevant parameters in patient anatomy should be evaluated. Otherwise, **Gamma** passing rate, histogram and spatial distribution of **Gamma** index should be evaluated.	0 or 1	6.9

Summary of Key Quality Indicators: Documentation, Process and Feedback

KQI		Indicator Measure	Section Reference
Documentation, Process and Feedback Quality Indicators			
E1	Documentation on PSQA program tolerance levels, action levels, and acceptance criteria should be developed and followed, along with formal procedures on PSQA measurements and analysis.	0 or 1	7.1
E2	A formalized process should be in place to record PSQA results. PSQA results, including failures and the result of any subsequent investigations/decisions, should be clearly documented and approved by a qualified medical physicist for each treatment plan prior to the onset of treatment.	0 or 1	7.2
E3	A feedback mechanism should be in place that allows for discussion and shared learning of issues related to PSQA.	0 or 1	7.3
E4	Periodic programmatic reviews of PSQA results in a database record and overall process should be conducted to identify systematic issues and/or opportunities for process improvement.	0 or 1	7.4

3. Policy and Procedure

3.1 Verification of Plan Delivery Parameters

A procedure must be written and implemented to verify that the plan delivery parameters at the linac are consistent with the approved treatment plan. This may be accomplished using direct or indirect means. A *direct* check could be performed manually by radiation therapists on treatment units or automatically using delivery log. An *indirect* check could be performed by a PSQA measurement or a fluence map comparison.

Key Quality Indicator A1	Indicator Measure
A procedure must be implemented to verify the consistency of plan delivery parameters prior to first fraction for all IMRT and VMAT treatments.	0 or 1

3.2 Verification of Plan Deliverability and Dose Accuracy

Plan deliverability and dose accuracy should be verified prior to first fraction for all IMRT and VMAT treatments according to local policy. Initially with any new **class solutions** or techniques, the verification should be done with PSQA using phantom based measurements. Major modifications to PSQA methodologies such as a change in PSQA frequency, the use of suitable measurement surrogates (e.g. delivery log analysis (see Appendix I) or plan analytics[1,2,3,4,5,6] with clear correlation to measurement results) or replacement of the measurement device warrant studying their impact using data driven methods. Suitable data driven methods include a review of a large number of measurements with statistical control charts[7,8,9,10] and/or sampling theory[11,12].

Key Quality Indicator A2	Indicator Measure
A procedure should be in place to verify plan deliverability and dose accuracy prior to first fraction for all IMRT and VMAT treatments.	0 or 1

3.3 Prevention of Accidental Alterations to Plan Delivery Parameters

A procedure must be implemented to prevent or catch accidental changes of plan delivery parameters at the linac[13] once they have been verified to be consistent with the approved treatment plan. As such, careful consideration must be given to the timing of events in the planning-to-pre-treatment process. This can be achieved by approving or locking the fields in the record and verify (R&V) system immediately after the initial consistency check (see KQI A1). In this way, all subsequent PSQA will be consistent with the fields to be used for patient treatment. Consistency check for subsequent fractions could be done by verifying the time stamp of the beam approval. It is important to consider that any fields for which PSQA is performed prior to the R&V approval or locking could potentially be inadvertently modified before patient treatment.

Key Quality Indicator A3	Indicator Measure
A procedure must be implemented to prevent or catch accidental alterations to the plan delivery parameters prior to each treatment.	0 or 1

3.4 Restrictions/Recommendations on Treatment Optimization and Calculation Parameters

Some TPS optimization and calculation parameters impact PSQA pass rates. Such parameters include: **plan modulation** complexity, number of small segments in a plan, minimum segment size, gantry angle spacing for VMAT delivery, and minimum number of MU per segment. In order to eliminate the effect of these parameters on PSQA pass rates failure, restrictions should be implemented[14].

Key Quality Indicator A4	Indicator Measure
Restrictions or recommendations on certain IMRT/VMAT treatment planning system (TPS) optimization and calculation parameters that impact PSQA pass rates should be implemented.	0 or 1

4. PSQA Instrumentation

4.1 Choice of PSQA Instrument for Technique/Class Solution Development

Devices with the highest level of spatial and dosimetric measurement accuracy allow a full evaluation of plan delivery[13,14,15] and are recommended during verification of new techniques. 3D or quasi-3D devices are preferable than 2D devices if they have similar resolution and dosimetric accuracy.

Key Quality Indicator B1	Indicator Measure
PSQA of new IMRT/VMAT techniques or **class solutions** should be evaluated using measurement devices with high dosimetric accuracy and high spatial resolution.	0 or 1

4.2 Choice of Instrument for Routine PSQA

For routine PSQA, devices with slightly lower dosimetric measurement accuracy or spatial resolution can be employed after increased clinical experience (e.g. using **statistical process control**[16]) with a new technique.

Key Quality Indicator B2	Indicator Measure
Following validation of new IMRT/VMAT techniques or **class solutions**, routine PSQA may employ devices with slightly lower dosimetric accuracy or spatial resolution.	0 or 1

4.3 Commissioning of Detectors

PSQA passing rate may be affected by the detector response. Detector characterization during commissioning should include, if applicable, but not limited to the following: small field response, detector volume averaging, angular response, dose interpolation, reproducibility, precision, accuracy, sensitivity, dose linearity, warm up time, collection efficiency, polarity effect, dependence on temperature, pressure and humidity[17]. If non-coplanar beams or off-axis setup is to be used clinically, the impact of detector response change should be evaluated and, if possible, corrected.

Key Quality Indicator B3	Indicator Measure
Detectors must be commissioned prior to clinical use.	0 or 1

4.4 QA Program for PSQA tools

Detector response may change over time depending on many factors. It is imperative to have a QA program in place to ensure all the PSQA tools are tested at a regular frequency for constancy and accuracy. Depending on the type of detector, its usage, and performance history, this frequency may need to be adjusted over time.

Key Quality Indicator B4	Indicator Measure
A QA program must be in place to ensure continued and consistent performance of the PSQA tools at the level at which they were commissioned.	0 or 1

4.5 Detector Calibration

Calibration accounts for differences in detector sensitivity that will degrade depending on usage[14]. Calibration should follow the manufacturer's methodology, and must be performed at a frequency dependent on its usage or if it fails regular QC tolerance (KQI B4).

Key Quality Indicator B5	Indicator Measure

Detector calibration (single or array) must be performed at a frequency dependent on its usage or if it fails regular quality control (QC) tolerance.	0 or 1

5. PSQA Measurement Setup and Methodology

5.1 Placement for Detectors

Placement for 2D detectors should be chosen so that the dose plane for evaluation has the largest area or volume of clinically relevant high dose region and the least amount of large dose gradient region. 3D or quasi-3D devices should be set up so that the high dose region is at the device centre, provided any changes in detector response are accounted for (see KQI B3). However, the detector placement should also be chosen to minimize the change in measurement results (e.g. **Gamma** passing rate) due to a small setup uncertainty.

Key Quality Indicator C1	Indicator Measure
The PSQA detector should be set up to maximize the measurement of relevant clinical region.	0 or 1

5.2 Detector Acclimatization

If moving detector or phantom from one room to the treatment unit, measurement should begin after temperature equilibrium has been reached unless a temperature correction factor is applied.

Key Quality Indicator C2	Indicator Measure
Detector and phantom should reach equilibrium temperature with the environment before use or a temperature correction factor is applied.	0 or 1

5.3 Detector Setup

True composite provides a more direct measurement of dose summation that most closely mimics patient treatment delivery[14]. Angular dependence of the detector must be negligible or accounted for. If **true composite** is not suitable for the detector (e.g. EPID for portal dosimetry), **perpendicular field-by-field** method may be used.

Key Quality Indicator C3	Indicator Measure
True composite is the preferred detector setup and delivery method.	0 or 1

5.4 Measurement on Beam-matched Linacs

Patient-specific measurements can be performed on any **beam-matched linacs** provided a set of appropriate IMRT/VMAT specific tests are done on a regular interval in addition to compliance to linac QC guidelines[18,19,20,21,22]. Appropriate IMRT/VMAT specific QC may include measuring a set of identical IMRT/VMAT plans delivered on multiple **beam-matched linacs** and determining if the PSQA results meet a passing rate threshold[23]. Applying the principles of **statistical process control**[16] to PSQA results also allows small differences between closely matched linacs to be identified and reduced[24].

Key Quality Indicator C4	Indicator Measure
Patient-specific measurements can be done on any **beam-matched linacs** provided consistent inter-machine performance as demonstrated by compliance to appropriate regular linac QC guidelines.	0 or 1

5.5 Density Heterogeneity on CT

Systematic dosimetric errors can occur in phantom based PSQA due to incorrect handling of heterogeneity settings in the TPS. Consideration should be made to account for:

1. CT artifacts caused by high Z detectors/electronics, and
2. Incorrect densities for high Z detectors/electronics/phantom[25].

Decision to override the densities on CT with either the correct ones or a homogeneous phantom density should be made after literature review (e.g. Chaswal et al[26]), manufacturer's recommendations, and phantom measurement verification. For example, the measurement verification may be done by delivering simple phantom plans based on static fields[27].

Key Quality Indicator C5	Indicator Measure
Density heterogeneity on CT must be considered in relevant software and phantoms in all PSQA measurements.	0 or 1

5.6 Machine Output Variation

Correcting for linac output variation allows for more meaningful comparison between PSQA measured and planned dose distribution, and provides better long term statistical tracking and analysis across different machines. It is also acceptable to analyze the results without output variation correction if the goal is simply to provide results that are more representative of patient treatment on a particular linac. Variation in accelerator output can be accounted for by measuring

the dose of a reference beam using the PSQA detector and comparing with the expected value, or by directly calibrating the detector using the reference beam. If the expected value is to be determined using an absolute dosimetry protocol, care must be taken to account for potential changes in dose due to the phantom/detector's non-flat surface, non-water medium, and heterogeneous density, if applicable.

Key Quality Indicator C6	Indicator Measure
Variation of machine output should be accounted for before every measurement session.	0 or 1

6. PSQA Evaluation Methodology

6.1 Plan Comparison Methodology

If available, 3D dose distribution in patient anatomy allows for clinically relevant comparison with treatment plan, such as DVH of targets and organs at risk (OARs). Numerous studies demonstrated phantom-based PSQA techniques may not be sensitive to clinically meaningful errors[28,29,30,31,32,33,34,35,36,37]. However, the reconstructed 3D dose distribution must be commissioned and thoroughly investigated for potential pitfalls (e.g. heterogeneous correction), before clinical use. If 3D dose distribution in patient anatomy is not available, 2D or 3D **Gamma** evaluation (for dose difference and distance-to-agreement (DTA)) in phantom, ideally similar in size as the patient, may be adequate. Other comparable evaluation methods are acceptable as long as they have been thoroughly tested. However, **Composite evaluation** does not provide histogram distribution and may omit important information such as the magnitude of failure.

Key Quality Indicator D1	Indicator Measure
If available, reconstructed 3D dose distribution in patient anatomy from measured dose or fluence should be used to provide clinically relevant comparison with the clinical plan. Otherwise, 2D or 3D phantom-based **Gamma** evaluation may be used for plan comparison.	0 or 1

6.2 Reference and Evaluated Dose Distributions

In **Gamma** (or comparable evaluation methods) evaluation, the **reference dose distribution** is the one against which the **evaluated dose distribution** is compared. The **reference distribution** can have any spatial resolution and dimensionality. The **evaluated distribution** should have at least the same or higher spatial resolution and dimensionality[14]. An example of the **reference** and **evaluated distributions** is 2D array detector measurement and 3D treatment plan dose distribution, respectively. The **reference** and **evaluated distributions** should not be reversed because the DTA analysis is not invariant. However, the choice of the distributions may subject to vendor's software limitation.

Key Quality Indicator D2	Indicator Measure
The **evaluated dose distribution** should have at least the same or higher spatial resolution and dimensionality than the **reference dose distribution**.	0 or 1

6.3 Dose Normalization

In **Gamma** (or comparable evaluation methods) analysis, **global normalization** in absolute dose, where the absolute dose distributions in both the **reference** and **evaluated dose distributions** are normalized to the same dose value should be used. **Local normalization** in absolute dose may also be used since it provides a more stringent analysis, useful for commissioning a new technique or troubleshooting. But the recommendation remains **global normalization** in absolute dose.

Key Quality Indicator D3	Indicator Measure
Global normalization in absolute dose should be used.	0 or 1

6.4 Choice of Normalization Dose

Normalization dose should be chosen in a low gradient, high dose region such as the maximum dose or the prescription dose. The dose point should be chosen in a consistent manner for each **class solution**, and not solely to increase passing rate.

Key Quality Indicator D4	Indicator Measure
Normalization dose should be chosen in a consistent manner for each **class solution**, in a low gradient, high dose region.	0 or 1

6.5 Dose Interpolation

The spatial resolution of the **evaluated dose distribution** has been shown to affect the **Gamma** calculation accuracy[38]. If the distribution has a spatial resolution greater than 1/3 the DTA criterion[14], dose interpolation can improve the **Gamma** (or comparable evaluation methods) calculation accuracy. Interpolation algorithms must be investigated for appropriateness and accuracy before use.

Key Quality Indicator D5	Indicator Measure
Dose interpolation should be done for the **evaluated dose distribution** prior to **Gamma** analysis if the spatial resolution is greater than 1/3 the DTA criterion.	0 or 1

6.6 Vender Specific Features

Vendor specific features such as **auto grid shift**, **measurement uncertainty**, and fast DTA algorithm to save computing time can affect passing rates in a way that can mask clinically significant problems. They should be enabled in a consistent manner and not solely to increase passing rate. Rather, systematic error in the measurements should be eliminated.

Key Quality Indicator D6	Indicator Measure
Optional vendor specific features should be carefully evaluated for their impact on the measurement results before use, and should be enabled in a consistent manner for each **class solution**.	0 or 1

6.7 Evaluation and Passing Rate Criteria

Assuming using **global normalization** in absolute dose, evaluation criteria of 3%/2mm in **Gamma** analysis should be used if the detector is capable of measuring with high precision and spatial resolution. Larger dose difference and/or DTA criterion may not be sensitive to detect significant leaf position and clinically significant errors[27,28,32,33,39]. Tighter tolerance may reveal the need to improve the accuracy in treatment planning beam models and delivery[40]. Low dose threshold should be used in **Gamma** analysis to exclude low dose, clinically insignificant regions. A value of 10% of the normalization point dose is recommended. For plans where very low dose to OARs is important, centres may select a different threshold value on an ad hoc basis. However, the selection must not be based solely on increasing the passing rate.

Universal passing rate tolerance and action levels of 95% and 90% (3%/2mm and 10% low dose threshold), respectively, should be used[14]. **Statistical process control**[16] may be used to establish equipment and **class solution** limits if universal limits cannot be met. However, steps should be taken to improve the passing rates.

In-house evaluation criteria may continue to be used in addition to recommended criteria for historical comparison. Other criteria might be used if the type of treatment or **class solutions** requires higher dose or targeting accuracy. However, the criteria should be standardized.

Key Quality Indicator D7	Indicator Measure
Gamma passing rate tolerance and action levels of 95% and 90%, respectively, using 3%/2mm and 10% low dose threshold in **Gamma** analysis are recommended.	0 or 1

6.8 Validation of Evaluation and Passing Rate Criteria

Determining evaluation criteria and passing rate tolerance and action levels for PSQA should be part of implementing a new IMRT or VMAT technique, **class solution** or measurement procedure. PSQA results may depend upon details within the TPS, linac, type of measurement device and setup, analysis methods, treatment technique and treatment site, and therefore, should be verified when a major change is made in any of these.

Regardless of how the PSQA evaluation criteria and tolerance/action levels are determined (whether equipment- or site-specific), all should be validated to ensure the chosen parameters adequately catch relevant errors of interest[14,41]. One option is by creating and measuring plans with known errors of the type one would expect to see a failed PSQA test.

Key Quality Indicator D8	Indicator Measure
PSQA program evaluation and passing rate criteria should be validated to ensure the chosen parameters adequately catch known errors.	0 or 1

6.9 Additional Evaluation Parameters

If available, clinically relevant parameters such as DVH, min, max, mean dose, and **Gamma** statistics of target(s) and OARs should be evaluated in 3D reconstructed dose distribution in patient anatomy. Otherwise, **Gamma** passing rate, histogram distribution (min, max, median, and mean **Gamma**) and spatial distribution of **Gamma** index should be evaluated as well. Potential clinical impact should be evaluated if any pixels/voxels with **Gamma** index >> 1, or if a large number of failed pixels/voxels are in the target or in the OARs.

Key Quality Indicator D9	Indicator Measure
If available, clinically relevant parameters in patient anatomy should be evaluated. Otherwise, **Gamma** passing rate, histogram and spatial distribution of **Gamma** index should be evaluated.	0 or 1

7. Documentation, Process and Feedback

7.1 Documentation on Acceptance Criteria

Tolerance and action levels for PSQA tests as well as acceptance criteria (i.e., pass/fail conditions, plan approval process) should be clearly defined, reviewed periodically and documented along with formal procedures on measurements and analysis, such as data preparation in the TPS, detector/phantom setup, and software settings for analysis.

Key Quality Indicator E1	Indicator Measure

Documentation on PSQA program tolerance levels, action levels, and acceptance criteria should be developed and followed, along with formal procedures on PSQA measurements and analysis.	0 or 1

7.2 Process for PSQA results

A formalized process should be in place to make sure all staff involved in treating the patient is aware that the PSQA results have been completed, documented, evaluated and approved by a qualified medical physicist so that treatment can proceed[41,42]. In case of PSQA warnings/failures, the formalized process, which should be followed for all patients, should include how to troubleshoot for a suboptimal delivery, estimate the clinical impact, determine the appropriate course of action, mechanisms of communication to all staff involved in the treatment of possible delay, and the investigation/decision results documented for each plan prior to the onset of treatment[14,42].

Key Quality Indicator E2	Indicator Measure
A formalized process should be in place to record PSQA results. PSQA results, including failures and the result of any subsequent investigations/decisions, should be clearly documented and approved by a qualified medical physicist for each treatment plan prior to the onset of treatment.	0 or 1

7.3 Feedback Mechanism

Issues with individual plans should be discussed with the dosimetrist, physicist, physics associate/assistant, and/or radiation oncologist involved. Systematic issues or frequent individual plan issues should also be communicated with a wider group for continuous learning and improvement. Examples of formal feedback mechanisms that could be implemented include discussion at staff meetings (standing agenda item) or within the QA committee.

Key Quality Indicator E3	Indicator Measure
A feedback mechanism should be in place that allows for discussion and shared learning of issues related to PSQA.	0 or 1

7.4 Periodic Reviews

Periodic programmatic reviews of PSQA results in a database record and the overall PSQA process can identify systematic issues and lead to improvement in plan/delivery accuracy. Records should include enough information, such as patient identifier, plan/trial name, treatment site/technique/**class solution**, device used, linac used, and PSQA result etc., to be able to identify trends across treatment sites/techniques/**class solutions**, equipment, or treatment machines.

Key Quality Indicator E4	Indicator Measure
Periodic programmatic reviews of PSQA results in a database record and overall process should be conducted to identify systematic issues and/or opportunities for process improvement.	0 or 1

Appendix I: Delivery Log

Delivery log-based PSQA can be an acceptable alternative surrogate to phantom based PSQA provided it is commissioned, potential pitfalls identified, and suitable linac specific QC performed on a regular basis to independently confirm log file integrity. Several authors have shown that log based and phantom based PSQA give similar results[43,44,45]. However, in two studies[46,47], it was found that MLC leaf positions in the log files could differ from measured or planned positions by a clinically significant amount. Therefore, delivery-log based software must be fully evaluated before clinical use. For example,

1. Does the delivery log contain all relevant information in reconstructing accurate dose distribution, such as dose rate and MU?
2. Does the delivery log contain accurate information, such as MLC leaf positions, collimator/gantry angles (if they are mis-calibrated)?
3. Does the delivery log (or the software reading the delivery log) have adequate sampling frequency?
4. If applicable, is the dose reconstruction algorithm sufficiently independent and accurate?
5. If applicable, do the delivery log based dose distribution calculations of a phantom/detector provide consistent results as measured dose distribution of the same phantom/detector?

Glossary

Auto grid shift: vendor specific tool that automatically calculates the best alignment (to produce the highest pass rate) between **reference** and **evaluated dose distributions**. It is intended to compensate for potential mis-alignment of the detector in field.

Beam-matched linacs: linacs commissioned to have identical or nearly identical dosimetric characteristics such that a patient can be moved between machines using the same treatment plan. A unique beam model for multiple beam-matched linacs exists in the TPS.

Class solution: pre-defined beam geometry and/or optimization objectives for a given treatment site that usually results in a consistent solution for the achieved planning objectives and PSQA.

Composite evaluation: a tool to compare **reference** and **evaluated dose distributions** by calculating the dose difference and the distance-to-agreement between the two distributions for each point in the **evaluated distribution**. The calculation is deemed to be a pass if either dose difference or the distance-to-agreement passes an evaluation criterion.

Evaluated (dose) distribution: the evaluated dose distribution is compared against the **reference dose distribution.** The evaluated distribution should have at least the same or higher spatial resolution and dimensionality than the **reference distribution**.

Gamma (evaluation): a tool to compare **reference** and **evaluated dose distributions** by calculating the Euclidian "distance" as a function of their dose difference and distance-to-agreement between the two distributions for each point in the evaluated distribution. Gamma between 0 and 1 is deemed to be a pass, and greater than 1 is considered a fail[48].

Global/local normalization: in relation to gamma analysis normalization methods: with "global normalization" for all evaluation points, the dose difference between the **reference** and **evaluated dose distributions** is normalized to a single value (typically the max point dose), whereas for "local normalization", the dose difference is normalized to the reference dose at the local point.

Measurement uncertainty: vendor specific tool that allows for an additional 1% of the measured dose at each detector to be added to the dose difference criterion before performing dose difference analysis.

Perpendicular field-by-field: dose distribution is measured for each beam separately, with the detector exposed always perpendicular to the beam.

Plan modulation: a measure of the variability of aperture area and leaf sequence from control point to control point.

Reference (dose) distribution: the reference dose distribution is the one against which the **evaluated dose distribution** is compared. The reference distribution can have any spatial resolution and dimensionality.

Statistical process control: a method of quality control which employs statistical methods to monitor and control a process; key tools used include run charts, control charts, and a focus on continuous improvement of the process.

True composite: summation of dose distribution by delivering all beams to a stationary detector/phantom at actual planned positions. This method best reproduces the dose distribution delivered to the patient.

References

1. Crowe SB, Kairn T, Kenny J, Knight RT, Hill B, Langton CM, et al. Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results. *Australas Phys Eng Sci Med.* 2014;37(3):475-82.
2. 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(2):021716.
3. Li G, Wu K, Peng G, Zhang Y, Bai S. A retrospective analysis for patient-specific quality assurance of volumetric-modulated arc therapy plans. *Med Dosim.* 2014;39(4):309-313.
4. McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. *Med Phys.* 2010;37(2):505-15.
5. 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(7):071718.
6. Valdes G, Chan MF, Lim SB, Scheuermann R, Deasy JO, Solberg TD. IMRT QA using machine learning: A multi-institutional validation. *J Appl Clin Med Phys.* 2017;18(5):279-284.
7. Chung JB, Kim JS, Ha SW, Ye SJ. Statistical analysis of IMRT dosimetry quality assurance measurements for local delivery guideline. *Radiat Oncol.* 2011;6(1):27.
8. Gérard K, Grandhaye JP, 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(4):1275-85.
9. Breen SL, Moseley DJ, Zhang B, Sharpe MB. Statistical process control for IMRT dosimetric verification. *Med Phys.* 2008;35(10):4417-25.
10. Nakamura S, Okamoto H, Wakita A, Umezawa R, Takahashi K, Inaba K, et al. A management method for the statistical results of patient-specific quality assurance for intensity-modulated radiation therapy. *J Radiat Res.* 2016;58(4):572-578.
11. Schilling ED, Neubauer DV. *Acceptance sampling in quality control, 2nd ed. Statistics: Textbooks and Monographs, ed. Balakrishnan N, Schilling ED, Schucany WR*. Boca Raton: Taylor & Francis Group; 2009. 683 p.
12. Montgomery DC, Runger GC. *Applied statistics and probability for engineers, 5th ed.* Hoboken, NJ: Wiley; 2011. Xv, 768 p.
13. Van der Wal E, Wiersma J, Ausma AH, Cuijpers JP, Tomsej M, Bos LJ, et al. *NCS Report 22: Code of practice for the quality assurance and control for intensity modulated radiotherapy.* Delft: Netherlands Commission on Radiation Dosimetry; 2013.
14. Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys.* 2018;45(4):e53-e83.
15. Mans A, Schuring D, Arends MP, Vugts CA, Wolthaus JW, Lotz HT, et al. The NCS code of practice for the quality assurance and control for volumetric modulated arc therapy. *Phys Med Biol.* 2016;61(19):7221-7235.
16. Montgomery DC. *Introduction to statistical quality control. 6th ed.* Hoboken, N.J.: John Wiley & Sons; 2009. Xiv, 734 p.

17. Low DA, Moran JM, Dempsey JF, Dong L, Oldham M. Dosimetry tools and techniques for IMRT. *Med Phys.* 2011;38(3):1313-1338.

18. Kirkby C, Ghasroddashti E, Angers C, Zeng G, Barnett E. COMP report: CPQR technical quality control guideline for medical linear accelerators and multileaf collimators. *J Appl Clin Med Phys.* 2018;19(2):22-28.

19. Klein EE, Hanley J, Bayouth J, Yin FF, Simon W, Dresser S, et al. Task Group 142 report: Quality assurance of medical accelerators. *Med Phys.* 2009;36(9):4197-4212.

20. Smith K, Balter P, Duhon J, White Jr. GA, Vassy Jr. DL, Miller RA, et al. AAPM Medical Physics Practice Guideline 8.a.: Linear accelerator performance tests. *J Appl Clin Med Phys.* 2017;18(4):23-39.

21. Bedford JL, Warrington AP. Commissioning of volumetric modulated arc therapy (VMAT). *Int J Radiat Oncol Biol Phys.* 2009;73(2):537-545.

22. Ling CC, Zhang P, Archambault Y, Bocanek J, Tang G, LoSasso T. Commissioning and quality assurance of RapidArc radiotherapy delivery system. *Int J Radiat Oncol Biol Phys.* 2008;72(2):575-581.

23. Leung R, Lee V, Cheung S, Lee K, Law G, Wong M, et al. SU-F-T-643: Feasibility of performing patient specific VMAT QA on single linac for plans treated in beam-matched Elekta Agility linacs. *Med Phys.* 2016;43(6):3612.

24. 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(5):190-195.

25. Coolens C, Childs PJ. Calibration of CT Hounsfield units for radiotherapy treatment planning of patients with metallic hip prostheses: the use of the extended CT-scale. *Phys Med Biol.* 2003;48(11):1591-1603.

26. Chaswal V, Weldon M, Gupta N, Chakravarti A, Rong Y. Commissioning and comprehensive evaluation of the ArcCHECK cylindrical diode array for VMAT pretreatment delivery QA. *J Appl Clin Med Phys.* 2014;15(4):212-225.

27. Nelms BE, Chan MF, Jarry G, Lemire M, Lowden J, Hampton C, et al. Evaluating IMRT and VMAT dose accuracy: practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. *Med Phys.* 2013;40(11):111722.

28. Yan G, Liu C, Simon TA, Peng LC, Fox C, Li JG. On the sensitivity of patient-specific IMRT QA to MLC positioning errors. *J Appl Clin Med Phys.* 2009;10(1): 120-128.

29. Kruse JJ. On the insensitivity of single field planar dosimetry to IMRT inaccuracies. *Med Phys.* 2010;37(6);2516-2524.

30. Nelms BE, Zhen H, Tomé WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys.* 2011;38(2):1037-1044.

31. Zhen H, Nelms BE, Tomé WA. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Med Phys.* 2011;38(10):5477-5489.

32. Carrasco P, Jornet N, Latorre A, Eudaldo T, Ruiz A, Ribas M. 3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification. *Med Phys.* 2012;39(8):5040-5049.

33. 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.

34. Coleman L, Skourou C. Sensitivity of volumetric modulated arc therapy patient specific QA results to multileaf collimator errors and correlation to dose volume histogram based metrics. *Med Phys.* 2013;40(11):111715.

35. Fredh A, Scherman JB, Fog LS, Munck af Rosenschöld P. Patient QA systems for rotational radiation therapy: A comparative experimental study with intentional errors. *Med Phys.* 2013;40(3):031716.

36. Chan MF, Li J, Schupak K, Burman C. Using a novel dose QA tool to quantify the impact of systematic errors otherwise undetected by conventional QA methods: clinical head and neck case studies. *Technol Cancer Res Treat.* 2014;13(1):57-67.

37. Jin X, Yan H, Han C, Zhou Y, Yi J, Xie C. Correlation between gamma index passing rate and clinical dosimetric difference for pre-treatment 2D and 3D volumetric modulated arc therapy dosimetric verification. *Br J Radiol.* 2015;88(1047):20140577.

38. Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med Phys.* 2003;30(9):2455-2464.

39. Mu G, Ludlum E, Xia P. Impact of MLC leaf position errors on simple and complex IMRT plans for head and neck cancer. *Phys Med Biol.* 2008;53(1):77-88.

40. Stambaugh C, Gagneur J, Uejo A, Clouser E, Ezzell G. Improvements in treatment planning calculations motivated by tightening IMRT QA tolerances. *J Appl Clin Med Phys.* 2019;20(1):250-257.

41. Moran JM, Dempsey M, Eisbruch A, Fraass BA, Galvin JM, Ibbott GS, et al. Safety considerations for IMRT: Executive summary. *Pract Radiat Oncol.* 2011;1(3):190-195.

42. McNiven A. Canadian Partnership for Quality Radiotherapy Technical Quality Control Guidelines for Patient-Specific Dosimetric Measurements for Intensity Modulated Radiation Therapies. http://www.cpqr.ca/wp-content/uploads/2017/01/PDM-2016-07-01.pdf. 2016.

43. Stanhope CW, Drake DG, Liang J, Alber M, Söhn M, Habib C, et al. Evaluation of machine log files/MC-based treatment planning and delivery QA as compared to ArcCHECK QA. *Med Phys.* 2018;45(7):2864-2874.

44. Agnew CE, Irvine DM, McGarry CK. Correlation of phantom-based and log file patient-specific QA with complexity scores for VMAT. *J Appl Clin Med Phys.* 2014;15(6):204-216.

45. Vazquez-Quino LA, Huerta-Hernandez CI, Rangaraj D. Clinical experience with Mobius FX software for 3D dose verification for prostate VMAT plans and comparison with physical measurements. *Journal of Physics: Conf. Series.* 2017;847(1):012060.

46. Neal B, Ahmed M, Kathuria K, Watkins T, Wijesooriya K, Siebers J. A clinically observed discrepancy between image-based and log-based MLC positions. *Med Phys.* 2016;43(6):2933-2935.

47. Agnew A, Agnew C, Grattan MW, Hounsell AR, McGarry CK. Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. *Phys Med Biol.* 2014;59(9):N49-63.

48. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys.* 1998;25(5):656-661.