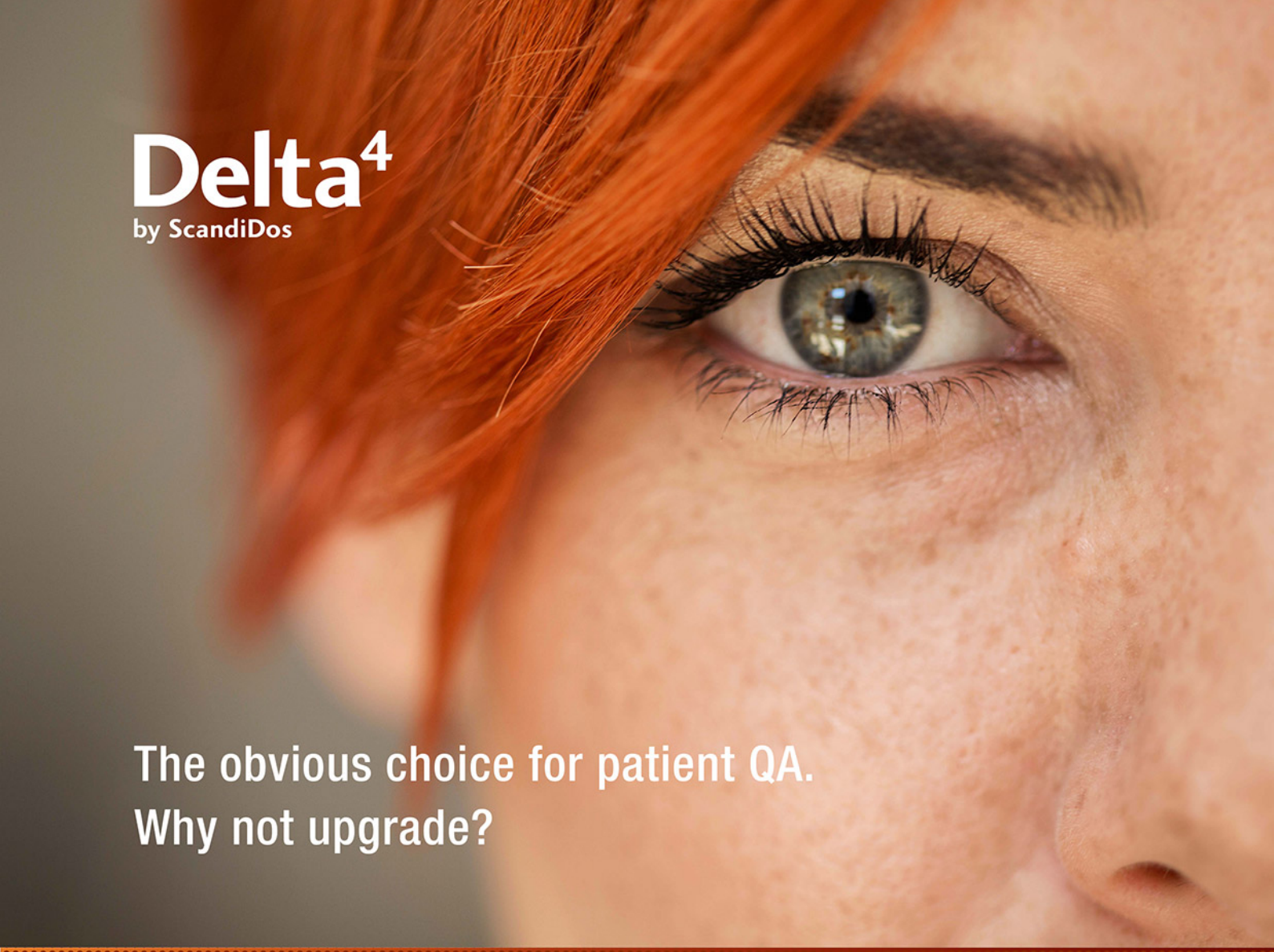


DELIVERY DOSAGE ENSURED

INDEPENDENT VERIFICATION DURING TREATMENT

OPTIMIZED CLINICAL WORKFLOW

A complete solution for all your patient QA needs

The Delta4 Family of products delivers Quality Assurance from prescription to final fraction.

Not only can you expect to increase your workflow efficiency but, most importantly, you can also carry out your work with full confidence that the treatment dose delivered to your patient is safe.

The Delta4 Family covers all modern treatment technologies, including VMAT, IMRT, SRS/SBRT, 4D-RT, MRgRT, bore-based linacs, Radixact and TomoTherapy.

Delta4
by ScandiDos

Innovative and Efficient QA
Delta4family.com

Received: 19 April 2021 | Revised: 25 May 2021 | Accepted: 21 June 2021

DOI: 10.1002/mp.15069

MEDICAL PHYSICS

AAPM SCIENTIFIC REPORT

Report of AAPM Task Group 219 on independent calculation-based dose/MU verification for IMRT

Timothy C. Zhu[1] | **Sotiris Stathakis**[2] | **Jennifer R. Clark**[3] | **Wenzheng Feng**[4] | **Dietmar Georg**[5] | **Shannon M. Holmes**[6] | **Stephen F. Kry**[7] | **Chang-Ming Charlie Ma**[8] | **Moyed Miften**[9] | **Dimitris Mihailidis**[1] | **Jean M. Moran**[10] | **Niko Papanikolaou**[2] | **Bjorn Poppe**[11] | **Ying Xiao**[1]

[1]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA, USA

[2]Department of Radiation Oncology, UTHSCSA, San Antonio, TX, USA

[3]Sun Nuclear Corporation, Melbourne, FL, USA

[4]Department of Radiation Oncology, Columbia University, New York, NY, USA

[5]Department of Radiation Oncology, Medical University Vienna, Vienna, Austria

[6]Standard Imaging, Middleton, WI, USA

[7]IROC, UT MD Anderson Cancer Center, Houston, TX, USA

[8]Department of Radiation Oncology, Fox Chase Cancer Center, Philadelphia, PA, USA

[9]Department of Radiation Oncology, University of Colorado Denver, Aurora, CO, USA

[10]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[11]Pius Hospital & Carl von Ossietzky University, Oldenburg, Germany

Correspondence
Timothy C. Zhu, University of Pennsylvania, Philadelphia, PA, USA.
Email: Timothy.Zhu@pennmedicine.upenn.edu

Sotiris Stathakis, Mays Cancer Center - MD Anderson Cancer Center, San Antonio, TX, USA.
Email: Stathakis@uthscsa.edu

Abstract

Independent verification of the dose per monitor unit (MU) to deliver the prescribed dose to a patient has been a mainstay of radiation oncology quality assurance (QA). We discuss the role of secondary dose/MU calculation programs as part of a comprehensive QA program. This report provides guidelines on calculation-based dose/MU verification for intensity modulated radiation therapy (IMRT) or volumetric modulated arc therapy (VMAT) provided by various modalities. We provide a review of various algorithms for "independent/second check" of monitor unit calculations for IMRT/VMAT. The report makes recommendations on the clinical implementation of secondary dose/MU calculation programs; on commissioning and acceptance of various commercially available secondary dose/MU calculation programs; on benchmark QA and periodic QA; and on clinically reasonable action levels for agreement of secondary dose/MU calculation programs.

Abbreviations: A(d), Attenuation function as a function of depth d; BEV, Beam's eye view; C/S, Convolution/Superposition; D, Absorbed dose; d_{max}, Depth of maximum dose; D_p, Primary dose, that is, dose from charged particles released from the photon's first interaction in the patient; D_s, Scatter dose, that is, dose from charged particles released from the photon's second or later interactions in the patient; DLG, Dosimetric leaf gap; DTA, Distance-to-agreement; DVH, Dose–volume histogram; GBBS, Grid-Based Boltzmann Solver; IMRT, Intensity modulated radiation therapy; f, Source-to-detector distance; fMU, Fractional MU; K, Collision kerma (K_d for direct beam, K_{air} for kerma in air, K_h for headscatter component); ks, Scatter polyenergetic point-spread kernel; MCI, Modified Clarkson Integration; MU, Monitor unit; OAR, Off-axis ratio; POAR, Primary off-axis ratio; PDD, Percentage depth dose; ROI, Region of interest; s, Projected field size at point of interest; SAD, Source-to-axial distance, usually 100 cm; SPD, Source-to-point distance; SSD, Source-to-skin distance; SDD, Source-to-detector distance; S_c, In-air output ratio; S_{cp}, (In-water) output ratio; SPR, Scatter-to-primary dose ratio; SF, Dose scatter factor, equals 1+SPR; S_p, Phantom scatter factor; TPS, Treatment planning system; TMR, Tissue-maximum ratio; TPR, Tissue-phantom ratio; x, y, Lateral positions relative to central axis of radiation source; VMAT, Volumetric arc therapy; z, Distance from the source to the point of interest.

1 | STATEMENT OF THE PROBLEM AND TG CHARGES

An independent check of dose/monitor units has been and continue to be an important part of quality assurance (QA) for patient treatment plans. AAPM Report Task Group 71 reports on the formalism for calculating monitor units.[1] AAPM Task Group 114 reports on methods and requirements for verification of data for conformal external beam plans.[2] The need for monitor unit verification programs was identified early in the adoption of IMRT treatment planning and delivery techniques.[3] Several different types of programs were developed ranging from confirmation of dose at a single point in a simple phantom geometry to calculation of dose at a single point while taking patient anatomy and geometry into consideration. While these programs have been in use for some time, guidance was lacking in how to commission such technologies as well as their role as part of an IMRT QA program as was noted in the ASTRO white paper entitled "Safety Considerations for IMRT".[4,5]

This task group was charged with: (a) Reviewing and evaluating the algorithms for "independent/second check" of monitor unit calculations for IMRT; (b) Making recommendations on the clinical implementation of calculation programs (e.g., number of points, locations, accuracy, evaluation methods, and heterogeneities); (3) Describing commissioning and benchmark QA of secondary MU calculation programs, proposing additional measurements, if necessary; and (4) Describing clinical testing and periodic QA of secondary MU calculation programs and recommendations on test tolerance.

The terminology used in this report follows that used in other AAPM Task Group reports including in particular:

1. *Shall* indicates a procedure that is essential for either (a) establishment of uniform practices, or (b) the most safe and effective result and/or maintaining established standards of practice to ensure the accuracy of dose/MU determination.
2. *Should* indicates an advisory recommendation that is to be applied when practicable. The task group favors the indicated procedure but understands that there are other procedures which can accomplish the same goal. Deviations from the recommended procedure should only be carried out after careful analysis demonstrates that an equivalent result will be produced.
3. *May* indicates a statement that is likely (or probably) to be correct but the task group does not make any recommendations.

As part of this work, the AAPM community was surveyed in 2012 regarding the type of software being used. According to the responses, a dose/MU verification program was generally used for the majority of IMRT/VMAT treatment plans although approximately 31% of responders did not use dose/MU verification software for VMAT treatment plans at that time. The most common commercial system reported in the survey was RadCalc (Lifeline, Tyler, TX). The most common treatment planning system (TPS) reported in the survey was Eclipse (Varian, Palo Alto, CA). This is not to be considered as an endorsement of these products. The most common passing rate criteria for dose/MU verification software was 5% for IMRT (51%) and "None Specified" for VMAT (34%), although 30% of VMAT responders used 5% as passing rate. More than 50% of users used a single point for their calculations and only 6% used three-dimensional (3D) volumetric dose in the 2012 survey. Additional measurements (typically the MLC dosimetric leaf gap (DLG) measurements) were also required during commissioning. The most common IMRT dose/MU verification calculation algorithm represented by software in use at the time was a "factor-based calculation algorithm." Note that the clinical practice has changed, with more widespread use of VMAT as well as the introduction and adoption of new software tools (such as 3D volumetric calculation systems) since 2012, so users should consider the age and context of the data when interpreting the survey.

2 | ROLE OF DOSE/MU VERIFICATIONS IN A COMPREHENSIVE QA PROGRAM

2.1 | Review of the problem

The implementation of new treatment techniques such as IMRT/VMAT in a radiotherapy department increases the complexities in planning and delivery and thus, the potential for serious errors in the planning and delivery of radiotherapy. An effective set of QA procedures is therefore essential. The goal of a routine pretreatment verification procedure is to identify and resolve any errors before patient treatment. For IMRT, verification measurements are commonly used to verify correct delivery of treatment plans, for example with ionization chambers, films, or multidimensional detector arrays. Experimental methods for patient-specific QA in advanced radiotherapy are, however, time-consuming in both manpower and accelerator time and have been shown to be unable to detect some unacceptable plans.[6-11] Recent studies have demonstrated a sensitivity of only 5% to detect IMRT plan errors using IMRT pretreatment measurements.[12,13] Moreover, as treatment planning becomes more efficient and the number of patients treated with advanced radiotherapy techniques steadily increases, measurement-based verification may result in a continued increase in workload.

As a result, it is important to seek additional complemental QA methods to increase the chance of catching any errors. Independent dose calculations can serve such a role and are complementary to measurement-based patient-specific QA. Such methods have been used at least since the 1950s[14] as routine QA in conventional radiotherapy, that is, radiotherapy with uniform intensity beams, using empirical algorithms in a manual calculation procedure, or utilizing software based on simple dose calculation algorithms.[15,16] These historic empirical dose calculation models are of very limited applicability for advanced treatment techniques and do not cover IMRT with static or rotating gantry.[17]

2.1.1 | Planning and dose delivery errors for IMRT

Compared to three-dimensional conformal radiotherapy (3D CRT), IMRT planning and delivery is a more complex process. End-to-end QA verification tests for the IMRT TPS and IMRT delivery equipment, along with patient-specific verification QA, are required to ensure the accuracy of the radiation delivery to patients.[4] Since the design and delivery of the IMRT treatment plans for patients involve both treatment planning software and delivery equipment, errors and/or uncertainties in the planning and delivery process can result in erroneous dose distributions delivered to the patient. Table 1 reports a consensus from TG 219 members on sources of error in secondary calculations, and considerations from AAPM Task Group 114.[2]

The main sources of error for IMRT plans can be attributed to input data, users, and planning-related features. The first two sources of error, data-related and user-related, are similar between IMRT and conventional plans, and are described in detail in TG 114.[2] These include errors in beam data (PDD, TMR, S_{cp}, etc.) and user errors (wrong plan, wrong point of calculation, wrong prescription, and wrong images for dose calculation). Table 1 lists our estimates of the relative probability of error occurrence for different algorithm dimensionality (one-dimensional [1D], two-dimensional [2D], and 3D as described in Section 3.1). The third category, planning-related errors, is more specific to IMRT. This includes dose errors due to high-gradient regions, small field sizes, small MUs, split fields, and low-density regions (e.g., lung) impacted by secondary electron disequilibrium of the IMRT plan.

Accurate IMRT beam modeling in the TPS is essential to reduce the uncertainties associated with the modeling process and consequently ensure good patient-specific verification QA agreement (a) between treatment planning calculations and measurements and (b) between treatment planning calculations and independent dose or MU calculations.[18] This includes accurate dosimetric data collection & entry especially for small fields, accurate MLC characterization, and verification of a full range of parameters in the TPS. These include leaf positions, effective leaf gap for non-divergent MLCs, leaf transmission values (representing an accurate average for a range of clinical depths and field sizes), output factors, percent depth doses (PDDs), and dose profiles for square and rectangular field sizes less than 4 × 4 cm^2 (including asymmetric fields).[19] Factors that should be part of validation include severe tongue-and-groove effects, small static step-and-shoot field segments and/or narrow sliding-window shapes, large leaf speed and dose rate interplay effects, and dose distributions that are calculated for a range of grid sizes and anatomical sites (eg prostate, lung, head & neck (H&N)). These tests are performed as part of commissioning to improve the overall accuracy and agreement of the IMRT treatment planning and delivery process with measurements, and establish baseline estimates for the expected accuracy and agreement

TABLE 1 Sources of error for the secondary MU calculation program as related to IMRT. H = high, M = medium, L = low

	Probability of error			
Common sources of error	**(Algorithm "dimensions")**			
Data related	**1D**	**2D**	**3D**	**Comment**
S_{cp}	L	L	L	Errors are more likely for small fields
PDD	L	L	L	
TMR	M	M	M	Usually calculated from PDD
Fit	-	-	M	1D and 2D algorithms do not usually use a fit to the data
User related				
Wrong plan	M	M	L	
Wrong points	M	M	M	
Wrong Rx	L	L	L	High if manually input
Wrong images	M	M	M	
Plan related				
Low dose region	H	M	M	
High gradient region	H	H	H	
Small field	M	M	M	
Small MU	L	L	L	
Dynamic beams	H	M	M	
Split fields (large angle scattering)	M	M	M	
Lateral electron disequilibrium	H	M	M	

between planning calculations and independent dose or MU calculations.[20-22]

The accuracy of IMRT delivery is impacted by linac output variations, field output variations (especially for small static step-and-shoot field segments and/or narrow sliding-window shapes with low MUs), MLC leaf position accuracy, leaf-end design, leaf position, and leaf gap reproducibility, inter- and intraleaf transmission, leaf acceleration/deceleration speed, tongue-and-groove effect, and MLC system sag with gantry rotation.[8,23-29] Further considerations include the possible failure of individual leaves and MLC carriages (if present) to sustain their calibration and any weaknesses in the torque of individual leaves' motors during IMRT delivery which also affect the accuracy of IMRT delivery.[24,26,30-32]

Patient-specific IMRT QA, including both measurements and independent IMRT dose/MU verification, plays a critical role in discovering these errors and ensuring patient safety.[29,33,34] For instance, MLC modeling errors ultimately results in dose errors especially in a high gradient region for small fields and/or dynamic fields. Importantly, IROC-Houston has reported that during IMRT delivery the vast majority of errors involve dose calculation errors from the TPS.[35,36]

2.1.2 | IMRT delivery techniques

With fixed gantry IMRT, the spatial intensity distribution of each treatment field is modulated to produce the desired dose distribution. This modulation is determined from the treatment plan, which can be generated in a forward manner (as with field-in-field planning), or more commonly using inverse planning. MLC leaf sequences generated for IMRT can either be step-and-shoot (the beam is turned off while the leaves move between static positions) or dynamic (MLC motion occurs while the beam is on).

Another approach to IMRT delivery involves the use of rotational treatments, such as volumetric modulated arc therapy (VMAT).[37-39] In addition, certain VMAT planning and delivery systems allow for modulation of the dose rate and/or gantry speed during arc delivery.

The TomoTherapy Hi-Art system is an image-guided radiation therapy device that utilizes a 6-MV linear accelerator mounted on a CT ring gantry to deliver intensity modulated fan-beams of radiation while the patient is translated through the gantry on a moving table. The vendor has also introduced TomoDirect (also referred to in the literature as Topotherapy[40-42]), which uses the same delivery platform but enables fixed beam treatments by moving the patient through the machine while keeping the linac fixed at user-specified gantry angles. However, the latter technique is considered a 3D CRT technique and not IMRT.

The CyberKnife Robotic Radiosurgery System collimates the radiation beam with tungsten collimators ranging in size from 5 mm to 60 mm, the IRIS variable aperture collimator, or an optional MLC. The robotic mounting allows repositioning of the source, which enables the system to deliver radiation from many different non-coplanar directions without the need to move the patient as required by current gantry configurations.

Many of the independent dose or monitor unit calculation software products for IMRT are designed to calculate the dose and/or MUs from the intensity pattern and/or MLC pattern for these different techniques and provide the user with tools to compare the calculations from IMRT/VMAT, TomoTherapy, and CyberKnife TPSs. Given the complexity of the IMRT delivery techniques described above, independent dose or monitor unit calculation software plays an increasing role in the IMRT QA process.

2.2 | Limitations of dose/MU verification programs

Reports and scientific publications that describe the accuracy of independent IMRT dose or monitor unit calculation software, whether developed in-house or commercial products, are scarce. Most publications deal with static (step-and shoot) or dynamic IMRT delivery based on a multileaf collimator or VMAT,[43-45] and only a few of them with advanced treatment techniques performed with a robotic linac[46] and TomoTherapy.[47] Nevertheless, the commercial vendor list (Table 2) reveals that these tools are clinically applied for verifying IMRT or VMAT.

For such calculation-based techniques, it was common practice to verify the dose at a single point by projecting the treatment geometry onto a flat, semi-infinite water phantom or "slab geometry" with the heterogeneity often modeled by the radiological path-length.[48] When using this technique, procedures are restricted to a limited number of points, even if the procedure is repeated several times for multiple points, and usually do not provide a full 3D dose verification. This is an important limitation for this type of independent dose calculation method. As long as patient anatomy is not included in verification calculations, the accuracy of independent dose calculations is influenced by treatment site-specific factors which might affect the analyses or definition of acceptance criteria. Moreover, when heterogeneities are not taken into account, fairly loose acceptance levels of about ±5–7% compared to the dose calculations carried out with a TPS and/or measurements have been reported.[49-51] The ability to perform a 2D or full 3D verification calculation based on the patient CT data set is largely dependent on the availability of appropriate calculation algorithms in the independent dose calculation software.

It is important to note that secondary calculations cannot catch hardware delivery errors, such as MLC

TABLE 2 Commercially available 2nd MU verification software

Software	Algorithm	Supported	Input	Output
RadCalc (LifeLine Software, Inc.)	Modified Clarkson	IMRT VMAT TomoTherapy CyberKnife Halcyon	Effective depth Patient external contour Plan parameters	One point/2D
MUCheck (Oncology Data Systems, Inc.)	Modified Clarkson	IMRT VMAT TomoTherapy CyberKnife	Effective depth Average depth Average SSD Plan parameters	One point
IMSure (Standard Imaging, Inc.)	Three Source Model	IMRT VMAT	Effective depth Plan parameters	Multiple points
Diamond (PTW Freiburg GmbH)	Modified Clarkson	IMRT VMAT	Effective depth Plan parameters	One point
DoseCHECK (Sun Nuclear, Corp)	Collapsed Cone Convolution/ Superposition	IMRT VMAT TomoTherapy Halcyon	Patient geometry Plan parameters	3D dose calculation
DosimetryCheck (Math Resolutions LLC)	Collapsed Cone Convolution/ Superposition	IMRT VMAT	Plan parameters EPID measurements	3D dose calculation
Mobius 3D (Varian Medical Systems, Inc)	Collapsed Cone Convolution/ Superposition	IMRT VMAT TomoTherapy	Plan parameters EPID measurements	3D dose calculation

leaf motor slippage, etc. For this reason, independent MU calculation for IMRT is unlikely to completely replace measurement-based methods for patient-specific QA. Furthermore, tools for independent IMRT dose/MU calculations cannot replace measurements for commissioning IMRT equipment. On the contrary, independent MU calculation tools for IMRT require verification of the software and shall be commissioned prior to their clinical use.

In 3D CRT, the total number of MUs per treatment plan or beam can be intuitively estimated for a given standard technique, such as a four-field pelvic box. Based on experience, outliers can be detected by a physicist during plan review. For IMRT, it is difficult to estimate intuitively whether a given number of MUs makes sense for a given patient and treatment setup. When used as a general verification tool without the expectation that small deviations will be detected, independent dose/MU verification using simple/standard algorithms can be of value in a department for detecting gross calculation errors.

2.3 | Dose/MU verification programs as a complement to measurement-based IMRT QA Methods

ESTRO[15] and AAPM[1,2] have both produced documents on dose/MU verification, although they are not specific to IMRT/VMAT. These projects were motivated by the desire for a tool that supports a detailed and effective investigation of the dosimetric accuracy of the primary TPS to catch systematic errors of those systems, due to limitations in algorithms, or the basic beam input data which might have systematic measurement errors.[52]

Currently, IMRT guidance[4,53,54] recommends experimental verification for IMRT patient- specific QA. A major advantage of this approach is the breadth of failure modes that can be checked: the treatment plan exports from the TPS, the plan imports into the treatment management system (TMS) system, and the plan is delivered as approved. A major disadvantage of this procedure is the substantial workload required. Furthermore, verification measurements are usually not performed in an anthropomorphic geometry but rather in simple geometric phantoms neglecting heterogeneities. Perhaps most concerning, many common methods are largely ineffectual in catching errors.[6,8-11]

Dose verification systems provide an opportunity for complementing the measurement-based approach, and hence TG 219 recommends that secondary dose/MU calculation should be performed for every IMRT/VMAT plan, at least in 1D but preferably in 2D/3D, regardless of the method of measurement-based verification utilized. A main advantage of an independent dose calculation method for IMRT is that it is far less time consuming than experimental methods for patient-specific QA. Furthermore, such calculational procedures do not require machine time or additional effort to perform the measurements. These advantages cannot be neglected when evaluating the overall IMRT QA strategy in a department; as personnel feel increasingly

confident about the reliability of techniques such as IMRT, it is reasonable to revise efforts in order to reduce the overall workload. However, an evaluation of the trade-off and opportunity is beyond the scope of this report and more studies[55] are required to make concrete recommendations. In addition to efficiency, independent dose/MU calculations also have an advantage over measurement-based QA because they are performed in the individual patient geometry including heterogeneities. Calculation-based techniques also have an advantage in terms of calculation flexibility, not only in terms of patient geometry, but also in terms of comprehensiveness of evaluation (full 3D) and direct determination of dose-volume metrics. Finally, there is evidence that, despite the subset of failure modes tested by an independent dose/MU calculation framework, such a system may actually be more robust at detecting unacceptable plans than traditional measurement-based approaches.[55] IROC observed the surprising result that an independent recalculation dramatically outperformed institutional measurement-based QA in predicting whether a plan would pass the IROC phantom.[55]

The above result from IROC is surprising because a major limitation of dose/MU calculations is that they are insensitive to many potential failure modes, such as a leaf calibration error, patient setup and/or organ motion variations, or errors in the beam calibration. Other errors that are not necessarily caught by calculation are related to issues with the patient CT or selection of the wrong treatment plan or wrong treatment unit. Furthermore, measured 2D fluence or dose map measurements may provide additional information about other potential sources of errors that may be difficult to determine when an independent dose calculation is limited to a single point.

Hybrid approaches, between traditional measurements and pure secondary MU checks are also appearing. For example, linac log files can be used as input for independent dose verification.[56] Alternatively, measured fluences with an EPID or a transmission detector can be used as input data for independent dose calculation.[57-59]

3 | ALGORITHMS FOR MU CALCULATION VERIFICATION METHODS

3.1 | Introduction

The algorithms for dose per MU verification programs can be based on factors derived from measurements, kernel-based convolution/superposition models, or Monte Carlo (MC) simulations. We will here review all three groups of formalisms but will focus on the first two because these are commercially implemented: a factor-based formalism tailored for "hand" calculations, and a model-based energy fluence formalism typical of modern TPSs. As general requirements, an ideal verification dose calculation model should be independent of the TPS, should be based on physical effects which are accurately described, and should be based on an independent set of algorithm input data.[60] A brief description of the available dose calculation algorithms is discussed below. A summary of these algorithms is listed in Tables 3 and 4. Table 3 focuses on 2D algorithms and Table 4 shows the 3D dose calculation algorithms. All factor-based algorithms are included with the 2D dose calculation algorithms. These two tables (Tables 3 and 4) demonstrate the evaluation options available for each type of algorithm. Table 5 summarizes the known algorithms for all commercial MU calculation programs.

Traditionally the formalisms used for independent MU and dose verification have been designed for verification calculations to a single point in the target (1D). However, for IMRT in which dose distributions are nonuniform, dosimetric agreement throughout the PTV(s) and organs at risk are often of interest.[50] While a multidimensional or volumetric verification including dose assessment is generally desirable, such an attempt might be considered overkill for dose verification in a QA procedure unless it can be performed with a minimum of additional effort. The institutional demand regarding the category of dimensionality for independent MU and dose verification in IMRT will certainly impact the algorithm or method used and the tolerance criteria. As the desired dimensionality increases so does the need for more sophistication in the algorithm. For example, a 2D approach requires penumbra modeling and also preferably modeling of the tongue-and-groove effect. In a full 3D approach, scatter and heterogeneity corrections become more important particularly when comparison between the TPS dose calculations performed on the patient's CT and independent volumetric calculations are evaluated.

3.2 | Factor- and model-based approaches

Independent dose/MU verification is most often performed with either a factor-based dose calculation algorithm or a model-based dose calculation algorithm.

3.2.1 | Factor-based approaches

With a factor-based method,[1,15] the parameters (e.g., *TPR*, S_c, S_p, etc.) are obtained either from direct measurement in a water phantom or extrapolated from such measurements. The dose/MU is determined by using the product of standardized dose ratio measurements.[61]

TABLE 3 2D algorithms and evaluation methods available in various second dose/MU calculation system and the specifics of various algorithm types

Alg. types	Hetero. Corr. Methods	Head Scatter Models	Pat. Geom.	# Calc. points	Eval. Method
1. Factor based	A. RTAR[1]	a. HS central axis meas.	2D contour/CT	α. one point	(a). % err.
2. Model based	B. Batho power[2]	b. HS off-axis meas.[3]		β. 2 – 10 points	(b). Gamma Index (or DTA)
3. Monte Carlo (MC)	C. ETAR[4]	c. Model: flattening filter[3]		γ. Planar dose	(c). DVH
4. Deterministic (GBBS)	D. FFT[5-7]	d. Model: ff+cs+ps[a]			
	E. Material Z				

[a]This refers to three source headscatter model composed of flattening filter (ff), collimator scattering (cs), and primary-collimator scattering (ps)

TABLE 4 3D algorithms and evaluation methods available in various second MU calculation system and the specifics of various algorithm types

Alg. types	Hetero. Corr. Methods	Head Scatter Models	Pat. Geom.	# Calc. points	Eval. Method
1. Factor based	A. FFT[5]	a. HS off-axis meas.[3]		β. 2 – 10 points	(a). % err.
2. Model based	B. Collapsed cone[8,9]	b. Model: flattening filter[3]	3D contour/CT	γ. Planar dose	(b). Gamma Index (or DTA)
3. Monte Carlo (MC)	C. Material Z	c. Model: ff+cs+ps[a]		η 3D dose cloud	(c). DVH
4. Deterministic (GBBS)	D. Secondary electron transport	d. Model: source obscuring[3]			
		e. Model: monitor backscattering[3]			

[a]This refers to three-source head scatter model composed of flattening filter (ff), collimator scattering (cs), and primary-collimator scattering (ps).

TABLE 5 A summary of dose calculation algorithms used in commercial dose/MU verification software. The number and letters listed represent algorithm type, heterogeneity correction method, and head scatter model for 2D (Table 3) and 3D (Table 4), respectively

Products (Vendors)	2D	3D
RadCalc[b,c] (LifeLine Software, Inc.)	1Aa[c]	
MUCheck[b,c] (Oncology Data Systems, Inc.)	1Aa	
IMSure[c] (Standard Imaging, Inc.)	1Ad	
DoseCHECK[b] (Sun Nuclear, Corp)		2Bb[e]
Dosimetry Checka (Math Resolutions LLC)		2Bb
DIAMOND (PTW Freiburg GmbH)	1Aa	
Mobius 3D (Varian Medical Systems, Inc)		2Bb

[a]Dosimetry Check is owned by Lifeline Software Inc.
[b]Supports TomoTherapy treatments
[c]Supports CyberKnife treatments
[d]1Aa means: **1** Factor based, **A** RTAR heterogeneity correction, **a** head scatter (HS) central axis measured Sc only as shown in Table 3 for 2D.
[e]2Bb means: **2** model-based, **B** collapsed cone algorithm heterogeneity correction, **b** flattening filter-based HS model as shown in Table 4 for 3D.

Successive dose ratio factors are multiplied for a chain of geometries, and thus the dose ratio factors are varied one by one until the geometry of interest is linked back to the reference geometry.[1,15] Most of the existing commercial systems specifically designed for IMRT verification use factor-based empirical methods.

3.2.2 | Model-based approaches

Model-based algorithms are usually more versatile and powerful than factor-based empirical models. Model-based dose calculations use a two-step procedure: first, modeling the energy fluence exiting the treatment head with a multi-source approach and second, determining the dose deposition in the patient with energy deposition kernels. Such model-based approaches generally require only a few easily obtainable input data for model tuning. Several examples of model-based algorithms have been developed and published.[61] For example, the AAPM TG 74 developed a multisource model that included direct radiation from the target; scattered radiation from the flattening filter, the collimator edges, wedges, backscatter to the monitor chamber; and electron contamination.[61] Other pencil-beam model-based algorithms as well as the results of a multi-institutional test have been described in the literature.[50,60,62,63]

In general, the dose contribution (D) from photons is separated into the primary (P) and scatter (S) components, such as $D = P + S = P*SF$, where $SF = 1 + S/P$ is defined as scatter factor.[64-66] This method is adequate for regions of transient charge particle equilibrium (TCPE). The primary component (which usually accounts for 70–80% of the total dose in a megavoltage photon beam and is dominant for small fields) can be obtained directly from measured quantities

between the source of the radiation and the point of measurement, that is, attenuation function through the local medium $A(d)$,[67,68] inverse square law ($1/f^2$),[65] and the in-air output ratio (S_c).[61] These quantities can be determined easily from the in-air measurements.[61] For IMRT, it is also important to take into account the head scatter outside beam collimation, the leaf ends, the leakage through the leaves and the tongue-and-groove effect, which are difficult to model in an accurate way. The scatter factor, *SF*, can be determined from phantom measurement, for example, *PDD* and S_p.[66] In advanced Clarkson's methods for scatter calculations, the fluence or MU map is first obtained, and then convolved with a scatter kernel.[69] For an IMRT field with modulation, a convolution with a scatter kernel can be used to calculate dose for an arbitrary field. Scatter dose in an IMRT field is often the cause of the "gradient effect," which refers to the secondary electron disequilibrium created by the (lateral) gradients introduced by intensity modulation, so that the depth dose is affected due to the strong intensity modulation of the IMRT field.

3.2.3 | Heterogeneity corrections

Heterogeneity corrections are necessary for accurate independent dose calculations for IMRT/VMAT. If patient-specific anatomic information is not included in verification calculations, the accuracy of independent dose checks can be seriously compromised.

For both IMRT and 3D CRT in the thoracic or head-and-neck regions, simple calculation conditions based on a semi-infinite homogenous phantom are insufficient to obtain accurate verification results. However, once simple radiological depth corrections are incorporated for head-and-neck treatments, a reasonable accuracy can be achieved.[50] For regions with large heterogeneities, accuracy will heavily depend on scatter modeling by the different dose calculation systems, that is, the TPS and the independent method.

3.2.4 | Use of a second TPS as secondary dose/MU check software

The use of a second TPS is an alternative model-based approach for independent dose verification. The utilization of a second TPS especially for IMRT and VMAT has the advantage that small and systematic uncertainties of the primary TPS potentially can be traced,[70] but has the disadvantage of high cost. Similarly, a new generation of calculation-based QA tools are coming into clinical practice that are based on more advanced dose calculation algorithms, for example, superposition/convolution algorithms similar to the ones used for treatment planning. These tools enable efficient and accurate fully 3D independent dose calculations based on the patient's CT dataset, with analysis tools such as DVH verification, etc. similar to a second TPS. Results have been published on using a second TPS or an advanced model-based independent dose calculation software, in which the robustness and time efficiency of such an approach were emphasized.[71,72] The relationship between calculation accuracy and different algorithms is summarized in Table 6 and discussed in detail in Section 3.4.

3.3 | Monte Carlo-based approaches

At the time this report was written, there were no commercial MC-based secondary MU check systems available. MC dose computation methods are appealing because they allow for highly accurate calculation of radiation transport through the patient. This has high value, particularly for heterogeneous anatomy. However, it should be noted that commercial MC TPSs are based on source models that are similar to those in conventional dose algorithms and subject to similar inaccuracies.[73] The potential for MC methods as an independent tool for radiotherapy treatment planning dose calculation and QA has been reviewed thoroughly in the AAPM TG 105 report.[74] This report highlighted that MC solutions provided substantially different results, particularly for calculation of dose in highly heterogeneous anatomy (lung), when

TABLE 6 Error ranges between secondary MU calculation algorithm types and measurement or Monte Carlo; see Section 3.4 for details

	Typical error range (local % difference from measurement or MC)				
	Center of lung tumor	In or downstream of lung	In bone	At surface	High Z (e.g., dental)
Factor-based	4.9	3–10	3–10	>40	20–40
AAA[a]	3.7	2–5	2–3	20	10–15
Collapsed cone (C/S)	3.7	2–5	1–2	20	10–15
Deterministic (GBBS)	1.5	1–2	<1	-	5
MC	<1	-	-	-	5

Numeric values are from Refs 71–87 as detailed in the text of Section 3.4.

[a]AAA stands for Analytical Anisotropic Algorithm, used in Varian Eclipse treatment planning system.

compared to simple pencil-beam algorithms.[65-68,72] MC results have uncertainties associated with them that depend on the number of histories used, which, if larger than the uncertainty between the secondary calculation and primary calculation could render the comparison less useful. High accuracy results without the uncertainty problem can be obtained with grid-based Boltzmann solver (GBBS) algorithms (i.e., Acuros),[73-75] although no secondary MU verification systems have been developed with this algorithm to date.

3.4 | Summary regarding the accuracy of calculation algorithms

With improved calculation algorithms, dose calculation accuracy is improved. The calculation accuracy is dependent on the location and type of heterogeneities; typical accuracy seen clinically around different heterogeneities is summarized in Table 6. These accuracies should help guide physicists in selection of calculation point placement, and more generally with expectations of agreement with the TPS.

3.4.1 | Lung GTV

Even in the center of a lung target, dosimetric accuracy is dependent on dose algorithm. In the IROC-Houston lung phantom, pencil-beam type algorithms systematically overestimate the dose by 4.9% on average compared to measurement.[75] More surprisingly, superposition/convolution algorithms also overestimate the dose to the center of the target by 3.7%; while some MC algorithms show agreement within 1% compared to measurement (as do GBBS results) compared to other MC algorithms showing systematic differences of several percent.[76,77]

3.4.2 | Lung and Bone

Within the lung, convolution/superposition (C/S) and AAA algorithms typically achieve accuracy within 2–5% compared with MC.[77-79] Agreement is poorer in low-density lung, which exacerbates the effect of the heterogeneity, resulting in dose errors of 5–10%.[78,79] GBBS is within 1–2% in lung[77,79] and 3% in light-density lung compared to MC.[79]

In bone, the accuracy of C/S is typically within 1–3%,[77,78,80] and within 1% for GBBS.[77]

3.4.3 | Skin/surface dose

At the patient surface, most algorithm calculations are generally poor. While the surface dose (and therefore the accuracy of the TPS calculation) can vary dramatically depending on treatment parameters,[81] studies have typically found that at the surface the TPS overestimates the dose by ~20% compared to measurements or MC calculations.[82-84] As the depth increases, the accuracy of C/S calculations increases and agreement within a few percent is reached within 4 mm.[82-84] Recently, Zhang *et al* demonstrated that C/S dose calculations can be accurate within 2% of measurements at the surface (and throughout the build-up region), but only with careful modeling.[85]

3.4.4 | High-Z heterogeneities

Dose accuracy is particularly challenging to achieve near high-*Z* interfaces and is exacerbated for higher energy beams and higher Z materials.[77,86] At the vicinity of both upstream and downstream metal interfaces, dose errors in C/S are often in the range of 10–15% compared to measurement (underestimating dose at the upstream interface and overestimating it at the downstream), but these errors can easily exceed 20%.[87-89] GBBS has been shown to agree within 1%–2% with MC,[86] which in turn agrees reasonably well (within ~5%) with measurement.[86,90] These effects can extend several cm from the implant; even 2 cm away, C/S algorithms can still show residual dose error of 6%–12%.[87-89,91]

4 | ACCEPTANCE OF AVAILABLE DOSE/MU VERIFICATION SOFTWARE

It is not the purpose of this document to provide detailed instruction for acceptance of various commercial or in-house developed dose/MU verification software for IMRT. The physicist should consult the manufacturer's documentation of the respective software for detailed instructions. However, a conceptual overview of elements that should be considered is presented in Table 7, including both tests of the hardware and software. Completion of such a set of tests will ensure that the user knows how the system works under various situations.

In recent years, various documents have been published for acceptance testing of TPS by the International Atomic Energy Agency (IAEA)[92,93] and AAPM reports TG 53[94] and MPPG 5A[95] to differentiate between tests for which the manufacturer is primarily responsible, and those which are the responsibility of the user. Tests to be performed by the manufacturer are related to the specific design of a TPS or dose calculation system to establish compliance with specified criteria. Tests to be performed after installation of an individual device or equipment establish compliance with the specified criteria at a particular site. This distinction of acceptance test procedures between general software functionalities that are site independent (e.g., protection against unauthorized use,

TABLE 7 Key tasks for dose algorithm check, acceptance, and commissioning for the secondary MU calculation program

Tasks	Data required
Dose algorithm check	
Linac Physics Model	Energy, SAD, Dmax, size/angle range (Jaw, gantry, collimator, couch)
Linac Dosimetry Model/ Beam Data	PDD/TMR(open, wedge), Profile(open, wedge), Output Factor (open Sc/Sp, wedge), transmission factors (Jaw, block tray, comp tray, couch, immobilization, etc.), reference MU definition
MLC Physics Model	MLC type, leaf number, size, etc.
MLC Dosimetry Model	Attenuation (inter and intra leaf), dosimetric leaf gap, etc.
Tasks	**Test required**
Acceptance*	
Software	Software running Import-export PDD and profile comparisons Test cases
Hardware	Printing
Tasks	**Test required**
Commissioning	
Open beam Homogenous phantom	SSD setup, various Jaw size and depth SAD setup, various Jaw size and depth SAD setup, various Off axis point with representative jaw size and depth
Static field Homogenous phantom	Blocked field (Block/MLC) Compensator field Wedge field (CAX and Off axis) Field edge Skin Flash Surface slope
Dynamic field Homogenous phantom	Dynamic wedge (CAX and Off axis) Step and shoot Sliding window VMAT
Heterogeneous phantom	Different density tissue internal (lung/bone, etc) Different density tissues interface Different density field edge
Real patient plan	Statistic evaluation between real patient plans and MU calculation program results.
Criteria	Percentage, Gamma index or DVH (based on plan type, site, etc.)
Benchmark points	Dose/MU points, see Table 8

*We recommend following the manufacturer's recommendation for acceptance tests.

principles of anatomic modeling) and those which depend on the installation site (e.g., treatment unit at the facility) can in principal also be applied to software for independent dose/MU calculation. For more detail, refer to recommendations for acceptance testing of TPS.[92-96]

5 | COMMISSIONING OF DOSE/MU VERIFICATION SOFTWARE

5.1 | Current commercial dose/MU verification systems

Methods to calculate monitor units independently for IMRT have been available since the late 1990s.[22,97-99] Several commercial systems exist which support verification of IMRT techniques. Important common features of these commercial software tools are support for both step-and-shoot and dynamic MLC delivery, ability to import leaf sequence patterns including the number of monitor units directly from the TPS, and performance of calculations at the dose specification point located on or off the central axis in a flat homogenous semi-infinite water phantom. A few products support calculations in the patient CT anatomy. Table 2 gives a brief description of the available software, their dose calculation methods, as well as the input parameters and expected output. Readers are referred to Appendix A for further details for each software. Since software products are updated continuously, the description below serves primarily as a starting point only.

5.2 | Commissioning

The commissioning requirements for dose/MU verification software depends on the type of algorithm. It is important to note that measurement errors in acquired beam data will propagate as systematic uncertainties in the secondary dose/MU calculation values. Table 7 provides an overview of the components of commissioning that should be completed.

As with the primary TPS, the data input into the secondary dose calculation system shall match the actual machine output characteristics. Accurate matching between these calculational systems and the photon characteristics of the linac is crucial for dose calculations.

As with any system used in the clinical treatment of patients, the secondary dose/MU verification system requires commissioning and ongoing quality control monitoring to ensure the accuracy and efficacy of the system as recommended by AAPM TG 53[94] and AAPM MPPG 5A.[95] That is, during commissioning the secondary dose/MU verification system should undergo a testing procedure similar to the one for a conventional TPS. The independent verification of the dose/MU should be validated and compared against institutional data and other calculation systems, if available. Although the main focus of such testing procedures is on the dosimetric tests, the non-dosimetric functions (e.g., import/export) and the treatment geometry (field size, gantry, and collimator rotation, etc.) should also be commissioned and documented. Full documentation of the commissioning tests and results should be kept in place at the institution and serve as guidelines for the ongoing QA program. Action levels can be established during the commissioning testing phase, provided that a sufficient sampling of clinical situations have been investigated to explore the limitations of the secondary dose/MU verification system. Users should also test the performance of the secondary dose/MU verification system against published data.

5.3 | Validation and benchmark guidance

Benchmark tests should be carefully designed to examine the system's dose accuracy under specific conditions and to establish baselines. Table 8 lists 19 points for Dose/MU evaluations in a homogeneous phantom that examine the dose calculation algorithm at off-axis points and on-axis points for various SSDs and field sizes under transient electron equilibrium conditions. For completeness, dose/MU in electron disequilibrium regions (points 17–19) can also be examined. The secondary MU verification program should be commissioned as thoroughly as the primary TPS. Additional guidance for TPS benchmarks can be found in AAPM TG 53.[94]

After commissioning of the secondary MU software per the recommendations of AAPM TG 114[2] and AAPM MPPG 5A,[95] calculation of a set of IMRT beams that exercise the MLC to produce simple patterns that can be verified manually with the use of look up tables will establish a baseline of the secondary MU software accuracy in a homogenous geometry. Such fields could be a “pyramid,” a “step wedge,” and/or a “checker board” field. Second, a set of IMRT fields based on a test plan similar to the AAPM TG 119[100] recommendation should be verified using the MU software. Finally, fields for IMRT plans on a phantom that incorporate inhomogeneities such as the lung or head and neck phantom from IROC-Houston[101] or a phantom that can accommodate detectors for absolute measurements should be used as an additional validation of the secondary dose/MU software.

The dosimetric results from the validation fields described above should provide the physicist with benchmark data for the accuracy of the secondary dose/MU software. Furthermore, this approach will validate the correct network communication between the TPS and the secondary MU software and the correct data transfer. The results can be then analyzed on a beam-by-beam or composite plan basis. The general consensus is to develop and use confidence limits from clinical data points. The values listed in Table 9 depends on whether the validation is for a single IMRT beam or a composite plan such as VMAT. These values are based on consensus and corresponding recommendations from TG 218.[53] This validation should serve as the baseline for any additional analysis and should be carried out again after any software updates.

6 | CLINICAL IMPLEMENTATION OF MU CALCULATION VERIFICATION METHODS

The goal of any routine pretreatment IMRT verification procedure is to catch errors before the actual patient treatment begins. Aspects of workload efficiency have a major impact on the clinical usability and utilization of tools for independent dose verification by calculations. To perform the secondary MU calculation efficiently, one should be able to import approved treatment plan data (e.g., MLC settings) directly from the TPS, the oncology information system (OIS) or the TMS. Such an automated data transfer can be realized utilizing the DICOM RT data exchange protocol. For any methods based on an automated computerized approach, single beam and multiple beam verification procedures do not differ significantly from a workload perspective.

TABLE 8 Benchmark points for independent verification of Dose/MU of photon beams using conditions different from those used for commissioning beam data collection. The table gives an example of 19 comparison points between measurements and secondary dose calculation results for Dose/MU in a 6 MV beam from a Varian TrueBeam. Benchmark measurements should be repeated for all photon energies. The first and second columns give the index and energy specification; the third to sixth columns give the conditions for measurements, for example, SSD, collimator settings, Collimator angle (CA) and Gantry angle (GA), and locations of the measurement point in the water phantom (x, y, z) that matched the coordinates used in a TPS, x and y are parallel to the X and Y jaw, note z is in the depth direction; the seventh to eighth columns give example measured and calculated values for Dose/MU; the ninth column gives the percentage difference between the calculation and the measurement

Index	Beam	SSD	Jaws (X,Y)	CA/GA	Location (x, y, z)	Meas. [cGy/MU]	2nd Calc [cGy/MU]	Difference [%]
1	6 MV	90	10 × 10	0/0	(0,0,10)	0.801	0.801	0.0%
2	6 MV	90	10 × [−10,20][a]	0/0	(0,15,10)	0.829	0.827	−0.3%
3	6 MV	90	10 × [−10,20][a]	0/0	(0,16.5,20)	0.445	0.444	−0.1%
4	6 MV	110	36 × 36	0/0	(0,0,5)	0.810	0.804	−0.7%
5	6 MV	110	36 × 36	0/0	(0,12,5)	0.836	0.825	−1.3%
6	6 MV	110	36 × 36	0/0	(12,0,5)	0.838	0.826	−1.4%
7	6 MV	80	5 × 20	0/0	(0,0,5)	1.280	1.281	0.1%
8	6 MV	80	5 × 20	0/0	(0,5,5)	1.313	1.303	−0.8%
9	6 MV	80	5 × 20	0/0	(0,0,20)	0.511	0.512	0.3%
10	6 MV	80	20 × 5	0/0	(0,0,5)	1.266	1.262	−0.3%
11	6 MV	80	20 × 5	0/0	(5,0,5)	1.300	1.284	−1.2%
12	6 MV	80	20 × 5	0/0	(0,0,20)	0.503	0.505	0.3%
13	6 MV	80	4 × 4	0/0	(0,0,5)	1.185	1.183	−0.2%
14	6 MV	80	4 × 4	0/0	(0,0,20)	0.439	0.439	0.0%
15	6 MV	80	36 × 36	0/0	(0,0,20)	0.702	0.701	−0.1%
16	6 MV	80	36 × 36	0/0	(0,12,20)	0.670	0.662	−1.3%
17	6 MV	100	1 × 1	0/0	(0,0,10)	0.606	0.600	−1.0%
18	6 MV	100	3 × 3	0/0	(0,0,0.5)	0.993	0.988	−0.5%
19	6 MV	100	40 × 40	0/0	(0,0,0.5)	1.291	1.319	2.2%

[a][−10, 20] is for independent collimator jaw setting, Y1 = −10 cm and Y2 = 20 cm so that an offset 10 x 10 cm^2 field is formed.

TABLE 9 Action levels of the secondary MU calculation compared to TP calculations for various clinical situations for a single point. All percentage differences are defined as local relative difference. Note: the action level is larger than the tolerance level as described in the AAPM TG218 report.[10] For 2D or 3D action levels, use the specification as described by TG218 given the lack of literature specific to second MU calculations.[10] A general guidance is to use 90% for 3%/2 mm as action level.[10] However, the users are allowed to tighten their criteria as they wish

	Homogeneous		Heterogeneous	
	Single beam	Composite plan	Single beam	Composite plan
High Dose/ Low Gradient	5%	3%	7%	5%
Low Dose/ High Gradient[a]	7%	5%	10%	7%

[a]Low-dose region is defined as dose <5% of maximum dose, where head scatter and leakage dominate. High gradient region is defined as dose gradient >5%/mm or in regions of electron disequilibrium (e.g., 4x4 cm field in the lung for 15 MV photons).

6.1 | Considerations regarding independence of the verification software

For a dose calculation of individual IMRT beams or a composite IMRT treatment plan method to be considered independent, it should be in principle a completely independent commercial or in-house developed software solution. This independence from the primary TPS should consider both the dose calculation algorithm and the dosimetric input data.[2]

To avoid introducing systematic errors in two dose calculation systems, it is highly recommended to use two different sets of experimentally determined beam data. This can be data determined in different reference conditions, that is, isocentric versus at fixed source-to-surface distance (SSD), data determined in the same setup but acquired for different field sizes, or similar data determined by different personnel. Some commercially available calculation applications provide independent beam models with their software.

Golden beam data set can be used for secondary dose per MU verification program to check beams produced by the same model type from the same manufacturer. However, users are cautioned to use benchmark points (e.g., Table 8) to validate the agreement of dose/MU between calculations and measurements.

6.2 | Components of the secondary dose/MU verification check

6.2.1 | Dose/MU

In general, the independent secondary dose/MU verification system should calculate the MU required for the delivery of the prescribed dose or confirm the resultant dose from a given MU value. For such a dose calculation, there are several input data that are required for the system to proceed. These data include: prescription information, jaw settings, MLC shapes, *SSD*, and calculation depth. This information is provided from the treatment plan data that are transferred electronically to the secondary dose/MU verification system either directly or through the TMS. The data should be verified prior to dose calculation. It should be noted that electronic transfer is preferable to reduce human error. In the case of IMRT and VMAT dose calculations, manual entry is almost impossible due to the amount of data required. After secondary dose/MU verification system computations are completed, the dose to point(s) or 2D/3D dose is reported or the number of MU for a given field is given.

6.2.2 | Patient geometry

Patient geometry (e.g., entry point, *SSD*, depth, heterogeneity) will have a large effect on the calculated dose per MU. The secondary dose/MU verification will be calculated based on the data received from the TPS. Modeling of patient geometry in the verification software may differ from that in the TPS and is often much simpler. In the case of discrepancies in the parameters that define the patient geometry, a difference between the values calculated by the TPS and the secondary software may arise. It should be noted that there are secondary calculation applications that will import patient contours and/or CT images from the TPS. In such cases, a more rigorous commissioning of the software that includes testing of geometrical parameters should be conducted using the applicable guidelines of the AAPM Task Group 53.[94]

6.2.3 | Dose–volume constraints

DVH metrics have been shown to be more sensitive to critical dose errors[9] than single point dose verification or gamma passing rates. Newer 3D-based verification applications allow evaluation of the dose volume via some combination of a DVH graph, dose distribution visualization, and/or dose–volume constraint sampling. The dose volume is evaluated by importing the RT Structure Set file and sampling the dose distribution relative to the defined contours. Metrics (such as D95 for PTV, V20 for lung) can then be compared to clinically accepted dose protocols and/or customized limits defined by the physician during treatment planning.

6.3 | Considerations regarding calculation points

Since many of the current secondary dose/MU verification programs utilize only a single point for dose/MU verification (though some do allow for multiple points of interest), it is critical that the chosen point is in a *high dose* and *low dose gradient* region so that the comparison is clinically significant. However, this is often not feasible in practice. Some commercial secondary dose/MU verification programs have implemented 2D calculations to one plane or 3D calculations to the entire volume and may additionally provide gamma maps for dose evaluation. This method is more broadly meaningful and is thus recommended for general use for IMRT secondary calculations.

6.4 | Comparison between TPS and dose/MU verification calculation

It is important to consider the dose and dose per monitor unit deviations in absolute as well as relative terms. Large relative deviations might result in small overall dose uncertainties. For IMRT, deviations that are large in relative terms but acceptable in absolute terms have been reported,[49,97,102] predominantly in areas outside the high dose region.

6.4.1 | Acceptable difference criteria

The comparisons between the TPS and the secondary dose/MU verification system can be done for each beam and for the composite plan dose contribution to a selected point(s). ESTRO booklet 9[103] summarizes the experience of several European institutions and discusses the use of confidence limits. They recommend tolerance limits of 3% for ion chamber measurements in target areas and action limits of 5% for point dose verification. The AAPM TG 218 report[53] summarizes a more up-to-date experience including that from North America and comes to a similar conclusion. The physicist should evaluate the dose(s) computed by the dose/MU verification program carefully, especially for the use of a single beam since it is not guaranteed that in an IMRT plan the dose point used is in a low gradient

and high dose area. There will be beams that deliver low dose to that point but contribute higher dose to the PTV.

Figure 1 shows the performance of 4 commercially available secondary MU verification software tested in 2012 using 206 IMRT and VMAT treatment fields.[43] All data in the figure represent single dose calculation points. A slab geometry was used for measurements without any inhomogeneities. The results in Figure 1 demonstrate the agreement between the TPS (Pinnacle[3]) and each verification algorithm for 6 MV. In Figure 1a, V1 represents a Varian linac with HD120 MLC, V5 a Varian 600C linac with 120 Millennium MLC and V8 a Varian 21EX linac with 120 Millennium MLC. Figure 1b shows how each software performed against the TPS. The overall agreement for all software is within 10%, the average is 0.8% and the standard deviation is 2.9% for all evaluated fields. Further analysis of the results shown in Figure 1 was performed to determine if there was a correlation between the differences observed and the MLC/linac model used (see Figure 2). The results in Figure 2 show that there was not a significant correlation between the MLC model and the algorithm model used by the secondary verification software.

Evaluation of the differences between TPS calculations and MU verification software for the composite plan (Figure 3) shows that all algorithms can predict the total dose to the point within 5%. The average from all plans is 0.2% with a standard deviation of 2.1%.[43]

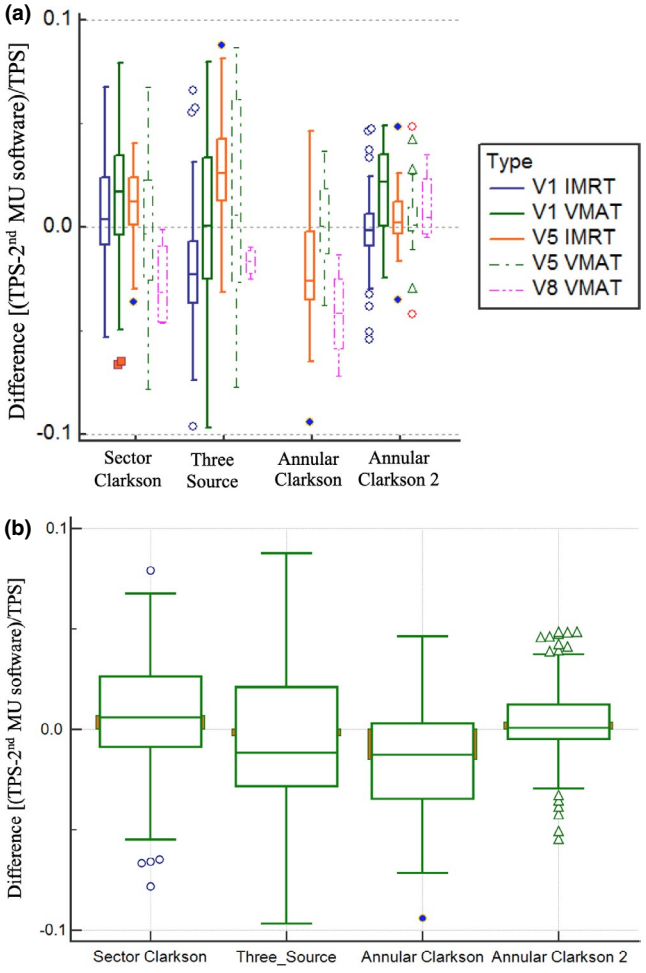


FIGURE 1 Comparisons of IMRT/VMAT secondary MU calculation uncertainty from several commercial MU calculation algorithms: (a) results separated for IMRT/VMAT. (b) Combined results alone with one standard deviation (box) and 90% confidence level (bars). Label "V1" represents a Varian linac with HD120 MLC, "V5" a 600C linac with 120 Millennium MLC, and "V8" a Varian 21EX linac with 120 Millennium MLC. All data represent single dose calculation points

6.4.2 | Action levels for unacceptable differences

Action levels for unacceptable difference have always been a topic of much discussion.[53] TG 219 recommends that users develop and use confidence limits from clinical data points. The values are listed in Table 9 depending on whether it is a single IMRT beam or a composite plan such as VMAT. These values are based on consensus and corresponding recommendation from TG 218.[53]

The AAPM Task Group report on IMRT commissioning (TG 119)[100] also used confidence limits to assist in judging the adequacy of IMRT commissioning. TG 219 recommends that measurements of a suite of IMRT tests be performed, mimicking the range of cases that will be encountered in clinical practice. The average and standard deviation of the results can be compared with those obtained by this group. A 1.96 multiplier is used in the confidence limit calculation which only strictly applies when a very large number of samples are available. The confidence limit thus provides a mechanism for determining reasonable action levels for per-patient IMRT verification studies for this group. The confidence limit for ion chamber measurements in the target region was 4.5% and for the low dose region was 4.7%, consistent with the recommendations of Palta *et al.*[104] and ESTRO guidelines.[103] TG 219 provides additional support for action levels expressed in terms of percentage of points passing gamma criteria of 3%/2 mm: 90% for per-field measurements and 88%–90% for composite irradiations.

The gamma passing rate, even if calculated based on patient dose grids, has generally weak correlation to critical patient DVH errors.[9] A commercial "planned dose perturbation" (PDP) algorithm was shown to predict the DVH impact using conventional planar QA results.[105] Using IMRT QA with patient DVH-based metrics allows per-patient dose QA to be based on metrics that are both sensitive and specific. Further studies are required to analyze new processes and action levels associated with DVH-based metrics to ensure effectiveness and practicality in the clinical setting.

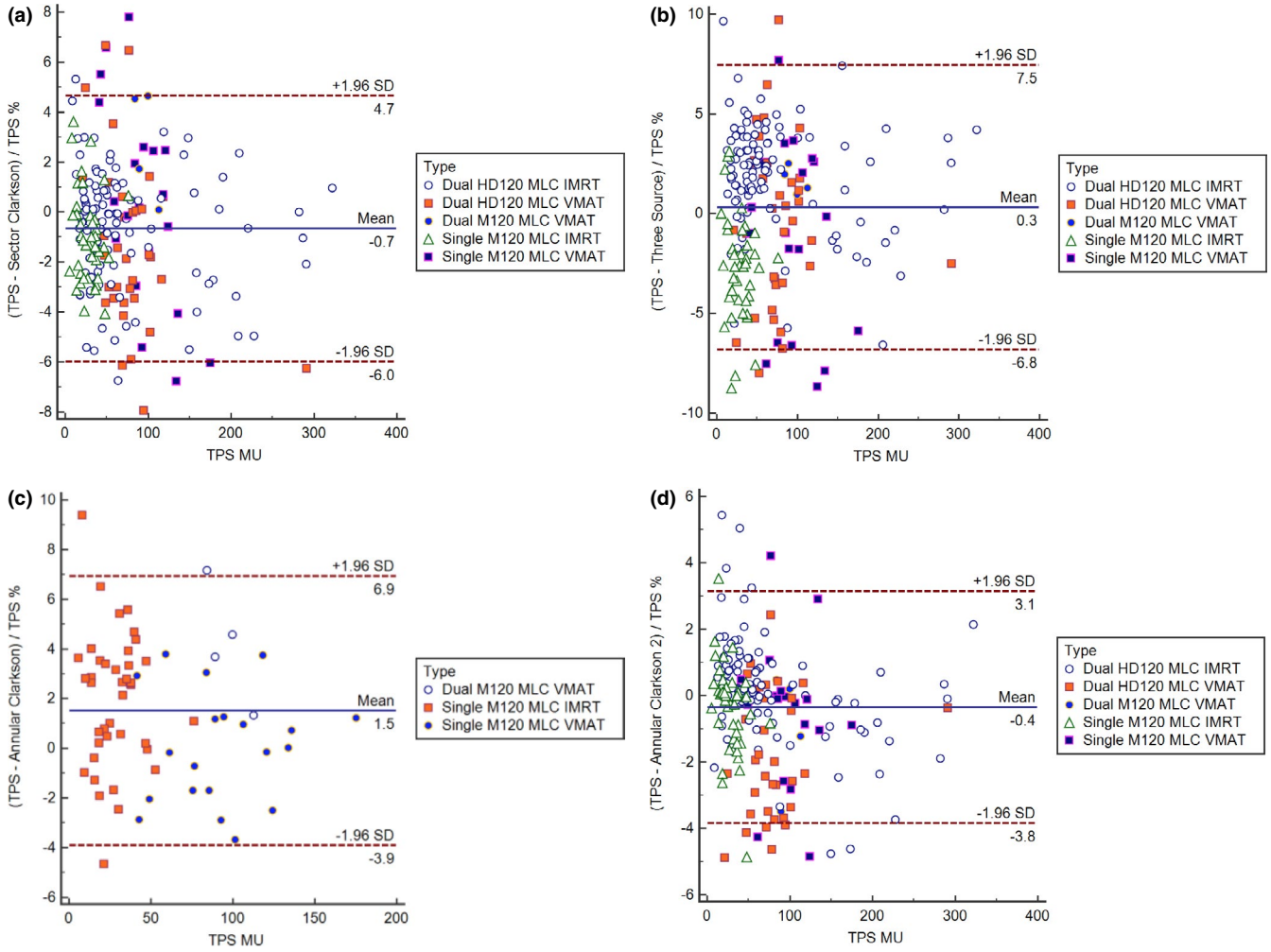
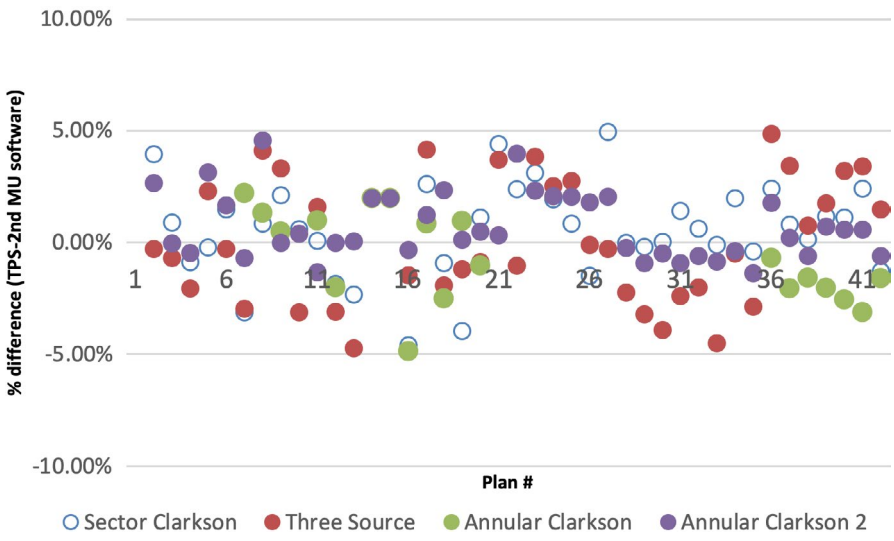


FIGURE 2 Comparison of IMRT/VMAT secondary MU calculations for different MLC types and secondary MU calculation algorithms: (a) sector Clarkson; (b) three-source; (c) annular Clarkson; and (d) annular Clarkson 2

FIGURE 3 Percent difference for each secondary MU software dose calculation algorithm and TPS (pinnacle) for the composite dose from all beams to the calculation point



6.4.3 | Investigating cases with poor agreement

In cases where there is poor agreement (e.g., difference is beyond the action level listed in Table 9) between the primary TPS and the secondary dose/MU verification system, it is important to examine the case to understand the nature of the disagreement. A disagreement should not be dismissed offhand; it is important to make use of all tools available to detect plan errors, including the secondary MU verification results. When a disagreement happens, there are multiple levels of investigation that can be used. First, check whether the disagreement may be simply a matter of poor point placement, such as a high gradient region (e.g., dose gradient >10%/mm), in regions of electron disequilibrium (e.g., 4 × 4 cm field in lung for 15 MV photons), or in regions of low dose (e.g., <10% of maximum dose) where head scatter and leakage dominate. While this possibility should be examined, moving the point routinely for numerous patients or repeatedly for a given patient to ensure acceptability is not an appropriate solution. Second, the physicist should understand the algorithm limitations of the secondary dose/MU verification system: for example, some secondary dose/MU calculation systems cannot provide the correct answer for a point inside lung medium with small fields (e.g., 4 × 4 cm^2 field inside lung). The secondary MU verification system should be well commissioned, and users should be well educated on issues such as point placement to ensure that good agreement is routinely observed between the secondary dose/MU system and the TPS. When a radius of average, defined as a sphere where dose to all points are averaged to represent an average value to a point, is used for a second dose/MU calculation, a 3-mm radius is a reasonable choice. Particularly if there are consistent disagreements between the two computational systems, the input data and models of the computational systems should be evaluated. This does not mean forcing the secondary system to match the TPS; the secondary system should match the delivery/linac output. Both the secondary system and the primary TPS should be reviewed to ensure that there are no errors in either system. Third, if no known cause is found for the disagreement, check if the IMRT/VMAT plan needs to be revised to ensure safe delivery of the plan. Decisions at this stage can be greatly aided by measurement-based IMRT QA or 2D or 3D comparisons between the TPS and secondary calculation software (e.g., gamma index analysis). Finally, an option is to contact the manufacturer for possible software clarification where there is a possibility that the specific scenario has been reported previously and possibly resolved in a later version of the software.

7 | ESTABLISHING BENCHMARK TESTS FOR COMMISSIONING AND PERIODIC QA

After commissioning of the secondary dose/MU verification system, it is imperative to verify the accuracy of this system. For a secondary dose/MU verification system, the use of test/benchmark cases is appropriate. This is similar to the process outlined in TG 53[94] and MPPG 5A[95] for the verification and QA of TPS. Per TG 53,[94] benchmark cases should have been previously established in the TPS during its commissioning, and these test cases should be both comprehensive and clinic-specific. A subset of these test cases is likely include those cases specified in TG 119.[100] The validity of the dose calculations in the TPS should be verified through phantom-based IMRT QA measurements of the test cases. These benchmark cases can then be applied to the commissioning of the secondary dose/MU verification system. During commissioning of the secondary dose/MU verification system, test cases should be transferred to the secondary dose/MU verification system and the agreement with the TPS should be evaluated and documented. The scope and extent of the test cases for the secondary dose/MU verification system need not be as comprehensive as the set used for commissioning of the TPS. Nevertheless, the test suite should cover the range of plans encountered clinically, including routine cases (e.g., head and neck and prostate), as well as a comprehensive range of field sizes. Benchmark plans should also include the different techniques used clinically: step-and-shoot, dynamic MLC, and/or VMAT. If the secondary calculation system includes heterogeneity corrections, a heterogeneous benchmark should also be evaluated. Assuming all the benchmark plans calculated in the planning system agree with measurements, the secondary calculation software should agree with the TPS. Reasonable agreement is within 5% (for both field-by-field or composite), and this should be achieved for the benchmark cases. Failure to achieve this level of agreement should result in either (a) improved commissioning of the secondary calculation system such that appropriate agreement is achieved, or (b) identification of the limitations of the secondary system, particularly in the case of challenging benchmarks, and establishment of alternate criteria for treatment plans of a similar nature.

In addition to verifying the commissioning of the secondary calculation system, use of benchmark cases with the secondary dose/MU verification software during commissioning is also important to establish a baseline for the performance of the system on specific test cases. This then serves as a baseline for QA of the secondary dose/MU calculation software in the case of a software update. When software updates occur, benchmark plans should be recalculated and

any change in the MU calculation should be noted. For any substantial change in the calculation from the secondary MU system, the new result can be compared to the original, hopefully yielding improved agreement. If agreement deteriorates for the benchmark cases, the issue should be explored before patient treatment.

It is reasonable for the benchmark plans used for periodic QA of the second calculation system to be a trimmed version of the commissioning set, consisting of a few cases. Potential errors in the secondary dose/MU verification software will be evaluated routinely with patient-specific calculations. Therefore, while concerning and potentially leading to delays in the plan check process, errors in updates to the secondary dose/MU verification system are unlikely to harm a patient due to other QA measures in place such as pretreatment measurements. QA of this secondary calculation system should be commensurate with this low risk.

If a new modality is added to clinical use, existing benchmark cases can be re-planned using this new technique to provide new benchmark cases. Both the old and new plans should be maintained and used in QA if both techniques remain in clinical use.

QA of the secondary dose/MU verification system is also warranted when the TPS is updated. A single test case of each delivery type (step-and-shoot, dynamic MLC, VMAT) should be exported to the second dose/MU verification system to verify the integrity of the data transfer and to confirm the secondary dose/MU verification system is able to process the data.

8 | SUMMARY RECOMMENDATIONS FOR PATIENT-SPECIFIC DOSE/MU VERIFICATION

1. The goal of independent dose/MU calculation in IMRT is to catch errors before the actual treatment begins. It should be used in such a way that the frequency of direct dose measurements may be limited or optimized (Section 2.3). Physicists should not rely solely on independent dose/MU calculation tools for IMRT QA. Such software currently cannot detect errors in dose calibration, MLC errors, collimator or gantry discrepancies, or patient setup inaccuracies (Section 2.2).
2. Secondary dose/MU calculations should be performed for every IMRT/VMAT plan, at least in 1D as predominantly available at the present time but preferably in 2D/3D, regardless of the method of measurement-based verification (Section 2.3).
3. "Independence" for secondary dose/MU software follows the same definition as outlined in TG 114[2], that is, it can be comprised of independent algorithms and/or independent beam data. It is acceptable to use the same beam data used for the TPS commissioning provided the algorithm for dose calculation is different, but it is preferred that both the algorithm and beam data are independent (Section 6.1).
4. Commissioning of the secondary dose/MU software should be performed based on the recommendations of AAPM TG 53[94] and MPPG 5A.[95] Routine disagreement between the secondary dose/MU verification system and the primary TPS should prompt thorough review of the commissioning and QA of the systems. (Section 5.2)
5. The software validation and benchmarking should use the benchmark points suggested in Table 8 and follow the recommendations of AAPM TG 119[100] (Section 5.3).
6. Ongoing QA for the secondary dose/MU software should be carried out both annually and anytime a TPS or secondary dose/MU software upgrade occurs, consistent with MPPG 5A[95] (Section 5.3).
7. For each individual IMRT/VMAT field, the agreement between the TPS and secondary dose/MU verification should be within the recommended action levels shown in Table 9. Plan acceptability should be based on the composite plan. Single beam agreement may be used to better understand discrepancies (Section 6.4).
8. Plans failing to meet acceptability criteria should be evaluated to understand the cause of the disagreement and manage it appropriately. (Section 6.4.3).
9. 2D/3D verification for IMRT and VMAT is recommended. Vendors should move away from systems that offer only a single comparison point and should develop secondary dose/MU verification systems that compute the dose distribution throughout the high dose volume (Section 6.3). Action levels suggested by TG 218[53] should be used for 2D/3D comparisons.

ACKNOWLEDGMENTS

The authors thank the reviewers from AAPM Therapy Physics Committee (TPC), Relative Dosimetry Subcommittee (RDSC) and Work Group on Treatment Planning (WGTP) for their valuable input.

CONFLICT OF INTEREST

The Co-Chairs of the "AAPM Task Group 219 on independent calculation-based dose/MU verification for IMRT" have reviewed the required Conflict of Interest statement on file for each member of "AAPM Task Group 219 on independent calculation-based dose/MU verification for IMRT" and determined that disclosure of potential Conflicts of Interest is an adequate management plan. Disclosures of potential Conflicts of Interest for each member of "AAPM Task Group 219 on independent calculation-based dose/MU verification for IMRT" are found at the close of this document.

DISCLOSURE STATEMENT

1. The members of "AAPM Task Group 219 on independent calculation-based dose/MU verification for IMRT" listed below attest that they have no potential Conflicts of Interest related to the subject matter or materials presented in this document.

Timothy C. Zhu, Sotiris Stathakis, Jennifer R. Clark, Wenzheng Feng, Dietmar Georg, Shannon M Holmes, Stephen F Kry, Chang-Ming Charlie Ma, Moyed Miften, Dimitris Mihailidis, Niko Papanikolaou, Bjorn Poppe, Ying Xiao.

2. The members of "AAPM Task Group 219 on independent calculation-based dose/MU verification for IMRT" listed below disclose the following potential Conflict(s) of Interest related to subject matter or materials presented in this document.

Jean M. Moran reports funding from NIH, Varian Medical Systems, and Blue Cross Blue Shield of Michigan unrelated to this work.

REFERENCES

1. Gibbons JP, Antolak JA, Followill DS, et al. Report of the AAPM Therapy Physics Committee Task Group. No. 71. *Med Phys*. 2014;41(3):031501.
2. Stern RL, Heaton R, Fraser MW, et al. Verification of monitor unit calculations for non-IMRT clinical radiotherapy: report of AAPM Task Group 114. *Med Phys*. 2011;38(1):504-530.
3. Boyer AL, Butler EB, DiPetrillo TA, et al. Intensity-modulated radiotherapy: current status and issues of interest. *Int J Rad Onc Biol Phys*. 2001;51(4):880-914.
4. Moran JM, Dempsey M, Eisbruch A, et al. Safety considerations for IMRT: executive summary. *Med Phys*. 2011;38(9):5067-5072.
5. Fraass BA, Marks LB, Pawlicki T. Commentary: safety considerations in contemporary radiation oncology: introduction to a series of ASTRO safety white papers. *Prac Rad Onc*. 2011;1(3):188-189.
6. Kruse JJ. On the insensitivity of single field planar dosimetry to IMRT inaccuracies. *Med Phys*. 2010;37(6):2516-2524.
7. Nelms BE, Zhen H, Wolfgang AT. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys*. 2011;38(2):1037-1044.
8. Nelms BE, Chan MF, Jarry G, 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.
9. Zhen H, Nelms BE, Tome WA. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Med Phys*. 2011;38(10):5477-5489.
10. McKenzie EM, Balter PA, Stingo FC, Jones J, Followill DS, Fry SF. Toward optimizing patient-specific IMRT QA techniques in the accurate detection of dosimetrically acceptable and unacceptable patient plans. *Med Phys*. 2014;41(12):121702.
11. Kry SF, Molineu A, Kerns JR, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. *Int J Radiat Oncol Biol Phys*. 2014;90(5):1195-1201.
12. Ford EC, Terezakis S, Souranis A, Harris K, Gay H, Mutic S. Quality Control Quantification (QCQ): a tool to measure the value of quality control checks in radiation oncology. *Int J Rad Onc Biol Phys*. 2012;84(3):e263-e269.
13. Bojechko C, Phillps M, Ford EC. A quantification of the effectiveness of EPID dosimetry and software-based plan verification systems in detecting incidens in radiotherapy. *Med Phys*. 2015;42(9):5363-5369.
14. Day MJ. A note on the calculation of dose in x-ray fields. *Br J Radiol*. 1950;23:368-369.
15. Dutreix A, Bjangard B, Bridier A, Mijnheer B, Shaw J, Svensson H. Monitor Unit calculation for high energy photon beams. ESTRO. 1997;Booklet 3.
16. Huyskens D, Bogaerts R, Verstraete J, et al. Practical guidelines for the impletation of in vivo dosimetry with diodes in external RT with photon beams. ESTRO. 2001;Booklet 5.
17. Georg D, Olofsson J, Kunzler T, Aiginger H, Karlsson M. A practical method to calculate head scatter factors in wedged rectangular and irregular MLC shaped beams for external and internal wedges. *Phys Med Biol*. 2004;49(20):4689-4700.
18. Webster GJ, Rowbottom CG, Mackay RI. Development of an optimum photon beam model for head and-neck intensity-modulated radiotherapy. *JACMP*. 2007;8(4):2711.
19. Followill DS, Kry SF, Qin L, et al. The Radiological Physics Center's standard dataset for small field size output factors. *J Appl Clin Med Phys*. 2012;13(5):3962.
20. Dong L, Antolak J, Salehpour M, et al. Patient-specific point dose measurement for IMRT monitor unit verification. *Int J Radiat Oncol Biol Phys*. 2003;56(3):867-877.
21. Fan J, Li J, Chen L, et al. A practical Monte Carlo MU verification tool for IMRT quality assurance. *Phys Med Biol*. 2006;51(10):2503-2515.
22. Linthout N, Verellen D, Van Acker S, Storme G. A simple theoretical verification of monitor unit calculation for intensity modulated beams using dynamic mini-multileaf collimation. *Radiother Oncol*. 2004;71(2):235-241.
23. Buckey CR, Stathakis S, Papanikolaou N. The inter- and intrafraction reproducibilities of three common IMRT delivery techniques. *Med Phys*. 2010;37(9):4854-4860.
24. Agnew A, Agnew CE, 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.
25. Furlow B. IMRT guidelines aim to prevent "catastrophic" errors. *Lancet Oncol*. 2011;12(9):836.
26. Ung NM, Harper CS, Wee L. Dosimetric impact of systematic MLC positional errors on step and shoot IMRT for prostate cancer: a planning study. *Australas Phys Eng Sci Med*. 2011;34(2):291-298.
27. 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):2915.
28. Sastre-Padro M, Welleweerd J, Malinen E, Eilertsen K, Olsen DR, van der Heide UA. Consequences of leaf calibration errors on IMRT delivery. *Phys Med Biol*. 2007;52(4):1147-1156.
29. Childress NL, Bloch C, White RA, Salehpour M, Rosen II. Detection of IMRT delivery errors using a quantitative 2D dosimetric verification system. *Med Phys*. 2005;32(1):153-162.
30. Bai S, Li G, Wang M, Jiang Q, Zhang Y, Wei Y. Effect of MLC leaf position, collimator rotation angle, and gantry rotation angle errors on intensity-modulated radiotherapy plans for nasopharyngeal carcinoma. *Med Dosim*. 2013;38(2):143-147.
31. Moiseenko V, Lapointe V, James K, Yin L, Liu M, Pawlicki T. Biological consequences of MLC calibration errors in IMRT delivery and QA. *Med Phys*. 2012;39(4):1917-1924.
32. 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.
33. Stapleton S, Zavgorodni S, Popescu IA, Beckham WA. Implementation of random set-up errors in Monte Carlo calculated dynamic IMRT treatment plans. *Phys Med Biol*. 2005;50(3):429-439.
34. Zygmanski P, Kung JH, Jiang SB, Chin L. Dependence of fluence errors in dynamic IMRT on leaf-positional errors varying with time and leaf number. *Med Phys*. 2003;30(10):2736-2749.

35. Carson ME, Molineu A, Taylor P, Followill DS, Stingo FS, Kry SF. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropormorphic head and neck phantom. *Med Phys*. 2016;43(12):6491-6496.
36. Kerns JR, Stingo F, Followill DS, Howell RM, Melancon A, Fry SF. Treatment planning system calculation errors are present in the majority of IROC Houston phantom failures. *Int J Radiat Oncol Biol Phys*. 2017;98(5):1197-1203.
37. Otto K. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Med Phys*. 2008;35(1):310-317.
38. Yu CX. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Phys Med Biol*. 1995;40(9):1435-1449.
39. Korreman S, Medin J, Kjaer-Kristoffersen F. Dosimetric verification of RapidArc treatment delivery. *Acta Oncol*. 2009;48(2):185-191.
40. Schubert LK, Gondi V, Sengbusch E, et al. Dosimetric comparison of left-sided whole breast irradiation with 3DCRT, forward-planned IMRT, inverse-planned IMRT, helical tomotherapy, and topotherapy. *Radiother Oncol*. 2011;100(2):241-246.
41. Lu W. Real-time motion-adaptive delivery (MAD) using binary MLC: I. Static beam (topotherapy) delivery. *Phys Med Biol*. 2008;53(22):6491-6511.
42. McIntosh A, Read PW, Khandelwal SR, et al. Evaluation of coplanar partial left breast irradiation using tomotherapy-based topotherapy. *Int J Radiat Oncol Biol Phys*. 2008;71(2):603-610.
43. Tuazon B, Narayanasamy G, Papaniolaou N, Kirby N, Mavroidis P, Stathakis S. Evaluation and comparison of second-check monitor unit calculation software with Pinnacle3 treatment planning system. *Physica Med*. 2018;45:186-191.
44. Au IW, Ciurlionis L, Compbell N, Goodwin D. Validation of Mobius system for patient-specific quality insurance using introduced intentional errors. *Australas Phy Eng Sci Med*. 2017;40(1):181-189.
45. McDonald DG, Jacqmin DJ, Mart CJ, et al. Validation of a modern second-check dosimetry system using a novel verification phantom. *J Appl Clin Med Phys*. 2017;18(1):170-177.
46. Dieterich S, Pawlicki T. Cyberknife image-guided delivery and quality assurance. *Int J Rad Onc Biol Phys*. 2008;71(1):S126 -S130.
47. Dzintars E, Papanikolaou N, Mavroidis P, Sadeghi A, Stathakis S. Application of an independent dose calculation software for estimating the impact of inter-fractional setup shifts in Helical Tomotherapy treatments. *Phys Med*. 2013;29(6):615-623.
48. Colodro FM, Berna AS, Puchades VP. Dosimetric validation of a redundant independent calculation software for VMAT fields. *Phys Med*. 2013;29:341-349.
49. Linthout N, Verellen D, Van Acker S, Storme G. A simple theoretical verification of monitor unit calculation for intensity modulated beams using dynamic mini-multileaf collimation [published online ahead of print 2004/04/28]. *Radiother Oncol*. 2004;71(2):235-241.
50. Georg D, Stock M, Kroupa B, et al. Patient-specific IMRT verification using independent fluence-based dose calculation software: experimental benchmarking and initial clinical experience. *Phys Med Biol*. 2007;52:4981-4992.
51. Wilcox EE, Daskalov GM, Pavlonnis G 3rd, Shumway R, Kaplan B, VanRooy E. Dosimetric verification of intensity modulated radiation therapy of 172 patients treated for various disease sites: comparison of EBT film dosimetry, ion chamber measurements, and independent MU calculations. *Med Dosim*. 2008;33(4):303-309.
52. Karlsson M, Ahnesjö A, Georg D, Nyholm T, Olofsson J. Independent dose calculations concepts and models. ESTRO. 2010;Booklet 10.
53. Miften M, Olch A, Mihailidis DN, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: a report of the AAPM TG 218. *Med Phys*. 2018;45(4):e53-e83.
54. Hartford AC, Galvin JM, Beyer DC, et al. American College of Radiology (ACR) and American Society for Radiation Oncology (ASTRO) Practice Guideline for intensity-modulated Radiation Therapy (IMRT). *Am J Clin Onc*. 2012;35(6):612-617.
55. Kry SF, Glenn MC, Peterson CB, et al. Independent recalculation outperforms traditional measurement-based IMRT QA methods in detecting unacceptable plans. *Med Phys*. 2019;46(8):3700-3708.
56. Litzenberg DW, Moran J, Fraass B. Verification of dynamic and segmental IMRT delivery by dynamic log file analysis. *J Appl Clin Med Phys*. 2002;3(2):63-72.
57. Sumida I, Yamaguchi H, Das IJ, et al. Intensity-modulated radiation therapy dose verification using fluence and portal imaging device. *J Appl Clin Med Phys*. 2016;17(1):259-271.
58. Bresciani S, Poli M, Miranti A, et al. Comparison of two different EPID-based solutions performing pretreatment quality assurance: 2D portal dosimetry versus 3D forward projection method. *Phys Med*. 2018;52:65-71.
59. Sait AA, Figueredo J, Jones GW, et al. Validation of three-dimensional electronic portal imaging device-based PerFRACTIONTM software for patient-specific quality assurance. *J Med Phys*. 2019;44(1):16-20.
60. Olofsson J, Nyholm T, Georg D, Ahnesjo A, Karlsson M. Evaluation of uncertainty predictions and dose output for model-based dose calculations for megavoltage photon beams. *Med Phys*. 2006;33(7):2548-2556.
61. Zhu TC, Ahnesjö A, Lam KL, et al. Report of AAPM Therapy Physics Committee Task Group 74: in-air output ratio, Sc, for megavoltage photon beams. *Med Phys*. 2009;36(11): 5262-5291.
62. Olofsson J, Georg D, Karlsson M. A widely tested model for head scatter influcence on photon beam output. *Radiother Oncol*. 2003;67:225-238.
63. Nyholm T, Olofsson J, Ahnesjo A, Georg D, Karlsson M. Pencil kernel correction and residual error estimation for quality-index-based dose calculations. *Phys Med Biol*. 2006;51:6245-6262.
64. Bjarngard BE, Cunningham JR. Comments on "Validity of the concept of separating primary and scatter dose". *Med Phys*. 1986;13:760-762.
65. Bjarngard BE, Vadash P. Analysis of central-axis doses for high-energy x rays. *Med Phys*. 1995;22(7):1191-1195.
66. Bjarngard BE. Scatter factor for a 25-MV x-ray beam. *Med Phys*. 1993;20(2):357-361.
67. Bjarngard BE, Shackford H. Attenuation in high-energy x-ray beams. *Med Phys*. 1994;21(7):1069-1073.
68. Siddon RL. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med Phys*. 1985;12(2):252-255.
69. Ahnesjo A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Med Phys*. 1989;16(4):577-592.
70. Sun B, Rangaraj D, Boddu S, et al. Evaluation of the efficiency and effectiveness of independent dose calculation followed by machine log file analysis against conventional measurement based IMRT QA. *J Appl Clin Med Phys*. 2012;13(5):3837.
71. Anjum MN, Parker W, Ruo R, Aldahlawi I, Afzal M. IMRT quality assurance using a second treatment planning system. *Med Dosim*. 2010;35(4):274-279.
72. Fontenot JD. Evaluation of a novel secondary check tool for intensity-modulated radiotherapy treatment planning. *J Appl Clin Med Phys*. 2014;15(5):4990.
73. Ma C-M, Chetty I, Deng J, et al. Beam modeling and beam model commissioning for MC dose calculation based radiation therapy treatment planning - The AAPM Task Group 157 Report. *Med Phys*. 2020;47(1):e1-e18.
74. Chetty I, Curren B, Cygler J, et al. Report fo the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning. *Med Phys*. 2007;34:4818-4853.

75. Kry SF, Alvarez P, Molineu A, Amador C, Galvin J, Followill DS. Algorithms used in heterogeneous dose calculations show systematic differences as measured with the radiological physics center's anthropomorphic thorax phantom used for RTOG credentialing. *Int J Radiat Oncol Biol Phys*. 2013;85(1):e96-e99.
76. Clark CH, Hurkmans CW, Kry SF. The role of dosimetry audit in lung SBRT multi-centre clinical trials. *Med Phys*. 2017;44:171-176.
77. Han T, Mikell JK, Salehpour M, Mourtada F. Dosimetric comparison of Acuros XB deterministic radiation transport method with Monte Carlo and model-based convolution methods in heterogeneous media. *Med Phys*. 2011;38(5):2651-2664.
78. Fogliata A, Vanetti E, Albers D, et al. On the dosimetric behaviour of photon dose calculation algorithms in the presence of simple geometric heterogeneities: comparison with Monte Carlo calculations. *Phys Med Biol*. 2007;52:1363-1385.
79. Bush K, Gagne IM, Zavgorodni S, Ansbacher W, Beckham W. Dosimetric validation of Acuros XB with Monte Carlo methods for photon dose calculations. *Med Phys*. 2011;38(4):2208-2221.
80. Kan MWK, Leung LHT, So RWK, Yu PKN. Experimental verificaation of the Acuros XB and AAA dose calculation adjacent to heterogeneous media for IMRT and RapidArc of nasopharygeal carcinoma. *Med Phys*. 2013;40:031714.
81. Kry SF, Smith SA, Weathers R, Stovall M. Skin dose during radiotherapy: a summary and general estimation technique. *J Appl Clin Med Phys*. 2012;13(3):20-34.
82. Court LE, Tishler RB, Allen AM, Xiang H, Makrigiorgos M, Chin L. Experimental evaluation of the accuracy of skin dose calculation for a commercial treatment planning system. *J Appl Clin Med Phys*. 2008;9(1):29-35.
83. Chung H, Jin H, Dempsey JF, et al. Evaluation of surface and build-up region dose for intensity-modulated radiation therapy in head and neck cancer. *Med Phys*. 2005;32(8):2682-2689.
84. Panettieri V, Barsoum P, Westermark M, Brualla L, Lax I. AAA and PBC calculation accuracy in the surface build-up region in tangential beam treatments. Phantom and breast case study with the Monte Carlo code penelope. *Radiother Oncol*. 2009;93(1):94-101.
85. Zhuang AH, Olch AJ. Validation of OSLD and a treatment planning system for surface dose determination in IMRT treatments. *Med Phys*. 2014;41(8):081720.
86. Lloyd SAM, Ansbacher W. Evaluation of an analytical linear Boltzmann transport equation solver for high-density inhomogeneities. *Med Phys*. 2013;40(1):011707.
87. Lin M-H, Li J, Price RA Jr, Wang L, Lee C-C, Ma C-M. The dosimetric impact of dental implants on head-and-neck volumetric modulated arc therapy. *Phys Med Biol*. 2013;58:1027-1040.
88. Wieslander E, Knoos T. Dose perturbation in the presence of metallic implants: treatment planning system versus Monte Carlo simulations. *Phys Med Biol*. 2003;48:3295-3305.
89. Wang X, Yang JN, Li X, et al. Effect of spine herdware on small spinal stereotactic radiosurgery dosimetry. *Phys Med Biol*. 2013;58:6733-6747.
90. Spirydovich S, Papiez L, Langer M, Sandison G, Thai V. High density dental materials and radiotherapy planning: Comparison of the dose predictions using superposition algorithm and fluence map Monte Carlo method with radiochromic film measurements. *Radiother Oncol*. 2006;81:309-314.
91. Huang JY, Followill DS, Howell RM, et al. Approaches to reducing photon dose calculation errors near metal implants. *Med Phys*. 2016;43(9):5117-5130.
92. Ibbott G, Izewska J, Schmidt R, et al. Specification and acceptance testing of Radiotherapy Treatment Planning Systems. IAEA. 2007;TECDOC 1540.
93. Andreo P, Cramb J, Fraass BA, et al. Commissioning and quality assurance of computerized planning systems for radiation treatment of cancer. IAEA. 2004;Technical Reports Series (430).
94. Fraass B, Doppke K, Hunt M, et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. *Med Phys*. 1998;25(10):1773-1829.
95. Smilowitz JB, Das IJ, Feygelman V, et al. AAPM Medical Physics Practice Guideline 5.a.: commissioning and QA of treatment planning dose calculations - megavoltage photon and electron beams. *J Appl Clin Med Phys*. 2015;16(5):14-34.
96. Mijnheer B, Olszewska A, Fiorino C, et al. Quality assurance of treatment planning systems practical examples for non-imrt photon beams. ESTRO. 2004;Booklet 7.
97. Baker CR, Clements R, Gately A, Budgell GJ. A separated primary and scatter model for independent dose calculation of intensity modulated radiotherapy. *Radiother Oncol*. 2006;80(3):385-390.
98. Kung JH, Chen GT, Kuchnir FK. A monitor unit verification calculation in intensity modulated radiotherapy as a dosimetry quality assurance. *Med Phys*. 2000;27(10):2226-2230.
99. Xing L, Chen Y, Luxton G, Li JG, Boyer AL. Monitor unit calculation for an intensity modulated photon field by a simple scatter-summation algorithm. *Phys Med Biol*. 2000;45(3):N1-7.
100. 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(11):5359-5373.
101. Followill DS, Evans DR, Cherry C, et al. Design, development, and implementation of the Radiological Physics Center's pelvis and thorax anthropomorphic quality assurance phantoms. *Med Phys*. 2007;34:2070-2076.
102. Chen Z, Xing L, Nath R. Independent monitor unit calculation for intensity modulated radiotherapy using the MIMiC multileaf collimator. *Med Phys*. 2002;29(9):2041-2051.
103. Alber M, Mijnheer B, Georg D. *Guidelines for the Verification of IMRT*. Brussels, Belgium: ESTRO booket 9; 2008.
104. Palta JR, Kim S, Li J, Liu C. Tolerance limits and action levels for planning and delivery of IMRT. In: Palta JR, Mackie TR, eds. *Intensity-Modulated Radiation Therapy: The State of Art*. Madison, WI: Medical Physics Publishing; 2003:593-612.
105. Olch AJ. Evaluation of the accuracy of 3DVH software estimates of dose to virtual ion chamber and film in composite IMRT QA. *Med Phys*. 2012;39(1):81-86.
106. Yang Y, Xing L, Li JG, et al. Independent dosimetric calculation with inclusion of head scatter and MLC transmission for IMRT. *Med Phys*. 2003;30(11):2937-2947.
107. Kahn FM. *The Physics of Radiation Therapy*. 4th, edn. New York: Lippincott Williams & Wilkins; 2010.
108. Jacques R, Wong J, Taylor R, McNutt T. Real-time dose computation: GPU-accelerated source modeling and superposition/convolution. *Med Phys*. 2011;38(1):294-305.
109. Jacques R, Taylor R, Wong J, McNutt T. Towards real-time radiation therapy: GPU accelerated superposition/convolution. *Comput Meth Prog Biomed*. 2010;98:285-292.
110. Ahmed S, Hunt D, Kapatoes J, et al. Validation of a GPU-based 3D dose calculator for IMRT/VMAT QA applications. *J Appl Clin Med Phys*. 2017;18(3):73-82.
111. Renner WD, Sarfaraz M, Earl MA, Yu CX. A dose delivery verification method for conventional and intensity-modulated radiation therapy using measured field fluence distributions. *Med Phys*. 2003;30(11):2996-3005.

How to cite this article: 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;00:1–22. https://doi.org/10.1002/mp.15069

APPENDIX A

Current Commercial MU Verification Systems

RadCalc (Lifeline Software Inc., Austin, TX) is a point dose Monitor Unit secondary check software which can support IMRT, CyberKnife, Volumetric modulated arc therapy (VMAT), TomoTherapy, compensator-based IMRT, and enhanced dynamic wedges (EDW).

IMRT dose/MU calculations in RadCalc use the modified Clarkson integration (MCI) algorithm developed by Kung *et al*.[98] It first performs an azimuthal fluence average along a CAX. Second, the Clarkson integration is performed over annular sectors instead of pie sectors. RadCalc incorporates three inhomogeneity correction methods: fixed inhomogeneity correction value, equivalent path length, and equivalent path length with field size scaling (equivalent square will be scaled by ratio of effective depth to geometric depth). The region of interest (ROI) module can compute the depth and effective depth to any defined point for conventional, IMRT and VMAT calculations, to allow for a more robust and accurate VMAT calculation. The Volume Average Dose Tool can be used to analyze the variation in dose around the primary calculation point, which is effective when calculation point is in a high dose gradient area. MLC parameters such as MLC transmission and dosimetric leaf gap (DLG) are needed for IMRT calculations.

MUCheck (Oncology Data Systems, Inc., Oklahoma City, OK) is an independent Monitor Unit/dose validation software for IMRT/VMAT, CyberKnife, Gamma Knife, and brachytherapy. It can support both step-and-shoot and dynamic IMRT plans. The IMRT dose calculation algorithm is a modified Clarkson's integration developed by J.H. Kung using annular sectors instead of pie sectors.[98] To calculate a VMAT plan, the input required includes: the average depth, average effective depth, and average SSD for the arc. Dose Volumetric Averaging provides a useful tool that can be helpful in analyzing the dose in areas of greater fluence intensity; it will calculate 9 additional points within a defined radius of the chosen calculation point. Average dose can also be calculated.

IMSure (Standard Imaging, Inc., Middleton, WI) is a 2D fluence and point dose calculation comparison software that uses the three-source model developed by Yang *et al*.[106] Fluence is calculated based on machine parameters input by the user and segment shapes defined by the treatment plan. Primary dose is computed using a TMR-OCR method, and scatter contributions from each fluence modulated beamlet are computed using a modified Clarkson integration.[99] Effective depths are calculated based on the CT image, including any density overrides. IMSure is used for the MU verification of 3D, IMRT, and VMAT treatment plans.

IMSure IMRT QA input requires a DICOM format patient plan file, with the option of including CT images for effective depth calculation. In addition to the reference calculation point for MU verification, numerous dose calculation points can be included at arbitrary locations within the patient volume. The input fluence map is compared with a fluence map calculated using the three-source model, and gamma calculations have user-defined percent difference and DTA criteria. Global dose normalization is applied.

DIAMOND (PTW, Freiburg, Germany) is a measurement-based monitor unit verification calculation system based generally on the formalism of Kahn in which scatter and tissue heterogeneity effects are separated from each other.[107] The dose calculation follows a modified Clarkson integration method and also uses a proprietary head scatter algorithm. Diamond can be used for MU verification of 3D, IMRT, and VMAT radiation therapy treatments.

The conventional calculation in DIAMOND uses sector integration to calculate effective TPR (TPR_{eff}), head scatter, and phantom scatter components for the given MLC pattern. Field parameters such as collimator settings, gantry angles, and field sizes as well as MLC patterns are extracted from imported DICOM data. User input includes the calculation point location and the radiological depth.

The IMRT and VMAT routines use the same method as the conventional calculation except that instead of calculating monitor units for the entire field they calculate dose for each control point of an IMRT or VMAT MLC sequence and sum the results.

DoseCHECK (Sun Nuclear Corporation, Melbourne, FL) is an automated, independent secondary calculation software that performs a full 3D dose calculation on the patient's planning CT and provides analysis via composite and per-beam point dose(s), dose volume constraints (Clinical Goals), and gamma on a whole volume and structure-by-structure basis. It is compatible with 3D CRT, varian enhanced dynamic wedge (EDW) and Elekta Universal Wedge, IMRT, VMAT, FFF, Halcyon, TomoTherapy, and SRS/SBRT. Heterogeneity correction is automatically applied. Couch tops, bolus, and density overrides from the RT Structure Set are supported. The photon dose calculation algorithm is a GPU-accelerated collapsed cone superposition/convolution initially developed at Johns Hopkins University,[108,109] and for which accuracy was benchmarked by Moffitt Cancer Center.[110]

Independent standardized photon beam models provided with the DoseCHECK software are based on an aggregate of measured beam data across several systems of the same type (e.g., TrueBeam, Synergy, etc.). Users assign the appropriate beam model to each configured linac by selecting from a dropdown. DoseCHECK data commissioning requirements do include the measured absolute Reference Dose for each photon energy under the following conditions: 100 SSD, 10x10 cm field, 10 cm depth, 100 MU. A CT-to-ED

table shall also be defined for each CT scanner used for treatment planning. (Additional TPS beam data may be required to produce a customized beam model, if deemed necessary.)

The DICOM dataset (RT Plan, RT Structures, CT images, and RT Dose) is received via export from the TPS to a dedicated DICOM listener built into the product. Dose calculation and analysis are performed automatically upon receipt of the DICOM dataset. The Point Dose analysis task evaluates absolute and relative difference between planned and calculated doses to the calculation point of each beam, as well as composite doses from all beams to all Points of Interest (POI) defined for the plan.

Additionally, following the recommendations by Zhen, *et al*.[9] that DVH-based metrics are more sensitive to dose errors, DoseCHECK automatically analyses dose volume constraints for targets and organs at risk, referred to as "Clinical Goals," and reports them against the ideal and/or acceptable values defined in a template by the user.

Dosimetry Check (Lifeline Software Inc., Austin, TX) is a 3D dose calculation software that uses the patient planning CT for both IMRT and VMAT pretreatment patient QA.[111] Dosimetry Check can also perform in vivo QA by measuring the beam transmitted through the patient, but this capability is outside the scope of the current investigation and therefore will not be addressed here.

Dosimetry Check pretreatment IMRT and VMAT QA input requirements include the DICOM format patient plan including the patient CT image, structure files, and 3D dose matrix. For IMRT QA, integrated EPID images are taken at each treatment angle, and for VMAT QA a cine image set is required. If gantry angle information is not included in the cine image file, time-stamped data from an inclinometer installed on the back of the gantry shall also be imported.

After calculating dose within the planning CT, Dosimetry Check will then interactively display the planned and calculated dose in 1D profiles, 2D isodose curves, or 3D volumes. Dose difference curves, DVH plots, gamma maps or volumes, and gamma volume histograms can also be displayed. Users can define distance, percent difference, and global dose normalization for the 3D gamma calculation.

Mobius 3D (Mobius Medical Systems, LLC, Houston, TX) is a full 3D recalculation of the patient treatment using independent beam data and an independent superposition/convolution algorithm. The patient CT, structure, and plan DICOM set are exported from the primary TPS to the Mobius system. Mobius 3D then recalculates the treatment plan on the patient CT image. This can be done for any IMRT or VMAT type delivery based on a linear accelerator. The recalculation is done with an independent and superposition/convolution algorithm. In this algorithm, TERMA is calculated from a three-source model (primary photons, scattered photons, and electron contamination); the TERMA is convolved with energy-spread kernels to determine the dose.

Input data for Mobius 3D depends on user preference. Mobius 3D comes with predefined stock models of Varian, Elekta, Siemens, and TomoTherapy linear accelerators for all common beam energies including FFF. Individual clinics can fine-tune these models to match their linac if desired. Modeling data necessary for tuning a Mobius 3D beam model is minimal, consisting of 9 percent depth dose values, output factors, and less than 10 off-axis ratios. The built-in automodeling tool adjusts beam models to match the entered data. Mobius 3D sensitivity for IMRT QA is benchmarked recently by introducing artificial errors.[44].