

Real-world examples of sensitivity failures of the 3%/3mm pass rate metric and published action levels when used in IMRT/VMAT system commissioning

B Nelms¹, G Jarry², M Chan³, C Hampton⁴, Y Watanabe⁵ and V Feygelman⁶

¹Canis Lupus LLC, Merrimac, Wisconsin 53561, USA

²L'Hopital Maissonneuve-Rosemont, Montreal, Quebec H1T 2M4, Canada

³Memorial Sloan-Kettering Cancer Center, Basking Ridge, New Jersey 07920, USA

⁴Wake Forest Baptist Medical Center, Winston-Salem, North Carolina 27157, USA

⁵University of Minnesota, Minneapolis, Minnesota 55455, USA

⁶Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL 33612, USA

E-mail: alpha@canislupusllc.com

Abstract. In IMRT/VMAT system commissioning (as with any system), quality is improved by striving for tight tolerances of stringent metrics of accuracy. For 5 cases, passing rates for 3%/3mm gamma analysis were generated following the TG119 instructions. Subsequently, more stringent/sensitive criteria combined with advanced volumetric dose analysis were applied, and in each case significant systematic errors were clearly identified despite the high 3%/3mm passing rates. In 4 of 5 cases, the error was easily remedied. These real-world examples of observed “false negatives” (insensitivities) point towards the inappropriateness of the 3%/3mm gamma passing rate metric as the basis for acceptance testing/commissioning of the IMRT/VMAT delivery chain.

1. Introduction

The AAPM TG-119 report [1] focuses on the commissioning of IMRT systems. The 3%/3mm gamma analysis passing rate metric is used as the basis for proposed action levels of 90% (per-beam) or 88-90% (composite dose) when comparing measured and calculated doses. The same 3%/3mm passing rate metric was found in a 2007 survey to be the metric most frequently used [2]. Recently, the sensitivity and specificity of this ubiquitous 3%/3mm gamma metric have been challenged [3-6]. However, these reports were focused on per-patient dose QA, which was not the topic of TG119. Although it is intuitive that a poor metric for per-patient dose QA would also have limited power in commissioning, this connection has not been directly stated and, as such, 3%/3mm continues to be a metric-of-choice in system commissioning.

Quality leaders such as Deming, Shewhart, and Taguchi stressed that driving out variation and conforming to tight tolerances improves the quality of a system more than voluminous inspection of the system's outputs [7-9]. It is vital to use sensitive metrics that identify and help to root out systematic errors and quantify, and potentially minimize, random errors in the system. An apt question is: as applied to the TPS commissioning, does the popular 3%/3mm passing rate metric meet these needs? In this work, we present real-world examples of important and clinically-relevant systematic errors that fail to be detected by the 3%/3mm passing rate-based action levels described in TG-119.



2. Methods

2.1. Case study summary and conventional 3%/3mm passing rate analyses

Passing rates for 3%/3mm gamma analysis were generated using exactly the TG-119 instructions [10] for five example cases in Table 1 (Cases 1-3: IMRT, 4: VMAT, 5: an open field). For cases 1 and 3, the MapCHECK diode array [11] (Sun Nuclear, USA) was used for per-beam planar measurements, normal to the beam CAX. For Case 2, the measurement method was EPIDose [12-13] (Sun Nuclear) with the same TG-119 criteria as MapCHECK but without “measurement uncertainty” applied. Cases 4 and 5 were measured with the ArcCHECK quasi-3D phantom [14-16] (Sun Nuclear) using the criteria specified for the 2D diode array.

Table 1: Case study summary, including plan details and 3%/3mm gamma passing rates using TG119 instructions [10]. For Cases 1-3, passing rates are the average of per-beam planar passing rates. Cases 4 and 5 are based on composite dose.

Case ID	Plan Summary	3%/3mm Passing Rate (%)	Measurement Type/ Dosimeter
Case 1	7-Field IMRT (SMLC) Head/Neck Boost 6 MV [Pinnacle TPS, Elekta 6MV 80-leaf (10 mm)]	99.2	Per-Beam/ 2D Diode Array
Case 2	7-Field IMRT (DMLC) Head/Neck 6 MV [MSKCC TPS, Varian 120-leaf]	99.4	Per-Beam/ EPID-based
Case 3	7-Field IMRT (DMLC) Head/Neck 6 MV [MSKCC TPS, Varian 120-leaf]	99.7	Per-Beam /2D Diode Array
Case 4	1-Arc VMAT Head/Neck 6 MV [Monaco TPS, Elekta 80-leaf (4 mm)]	96.8	Composite/ 3D Diode Array
Case 5	10x10 Open field 10 MV [Pinnacle TPS, Elekta 80-leaf (10 mm)]	97.6	Composite /3D Diode Array

2.2. Advanced diagnostics and identification of errors

In all cases, more stringent/sensitive criteria such as reducing tolerances, turning off “measurement uncertainty” padding, and using local rather than global % dose-difference were applied to probe for error patterns and regions of high local dose error. In addition, in all cases 3D patient measurement-guided dose reconstruction (MGDR) [5,17,18] was performed using 3DVH software (Sun Nuclear) to estimate volumetric dose and DVH differences, accompanied in some cases by film and/or ion chamber measurements for verification. In all cases, the errors were verified as systematic errors of noticeable impact and in 4 of 5 cases (1-3, 5) they were readily remedied.

3. Results and discussion

3.1. Case 1

Despite a high 3%/3mm average passing rate (99.2%), there was a systematic error in the TPS model of the rounded MLC leaf ends. The calculated penumbra at the MLC edges was too wide compared to measurements, resulting in TPS dose being overestimated, in particular along gradients surrounding the target. Using MGDR, the compilation of these errors were reconstructed in the 3D patient dose, highlighting the impact on patient dose and DVH, which turned out to be 3-4% lower than calculated by the TPS (e.g. D95% (Gy) of CTV was in error by 3.8%, TPS being too high). The problem was easily fixed by adjusting the leaf offset table.

3.2. Case 2

Despite a high 3%/3mm average passing rate (99.4%), there was a systematic error in the TPS: the tongue-and-groove (TnG) correction was turned off. The TPS was overestimating dose between the leaves (Figure 1a), but this was essentially hidden in the 3mm distance criterion. The impact on the

patient target volume D95% (Gy) values ranged from 2 to 4% (Figure 1b). Once the TPS TnG model was applied correctly, the differences were essentially eliminated.

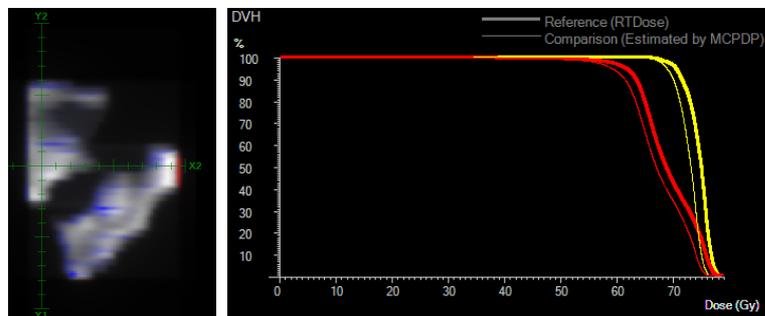


Figure 1a (Left). Measured – Calculated dose difference map, clearly indicating the TPS not correctly accounting for tongue-and-groove effects.

Figure 1b (Right). 99.4% average passing rate hides a systematic error in the TPS tongue-and-groove model which, when properly corrected, results in lower D95% (Gy) dose to the PTV70 (-2.4%) and PTV59.4 (-4.1%).

3.3. Case 3

Despite a high 3%/3mm average passing rate (99.7%), there was a systematic error in the position of the X1 jaw, which was the limiting aperture at the PTV inferior border. The actual X1 position was about 1-2 mm less than indicated, reducing the delivered dose. The MGDR showed a local difference at the inferior target border where the errors accumulated (Figure 2a). Again, the X1 readout error was clearly too small to be detected by a 3mm distance criterion in the gamma index.

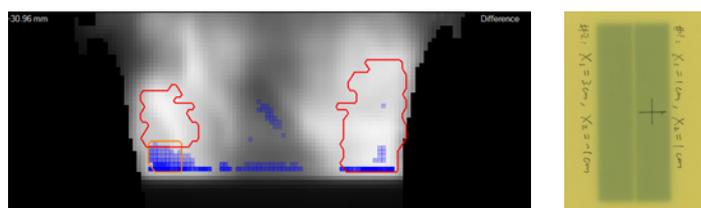


Figure 2a (Left). Measured – Calculated dose difference map using MGDR as measurement shows region of accumulation of low dose at inferior dose border due to X1 impingement.

Figure 2b (Right). Film analysis verifies an error of ~ 1 mm in the X1 jaw position.

3.4. Case 4

Despite a 3%/3mm passing rate of 96.8% (Figure 3a), there were systematic errors in the TPS dose calculation: 1) improper rendering of high gradient dose, and 2) over-estimation of dose under MLC leaves. These errors were hidden by the dose and distance methods prescribed by TG119, but clearly visible as a pattern if more stringent criteria (2% local/2mm) are used (Figure 3b) and when reconstructed in 3D using MGDR (Figure 3c).

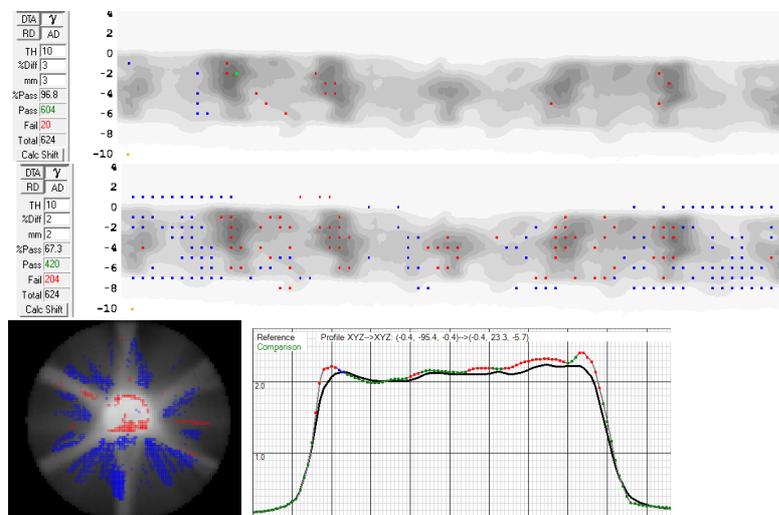


Figure 3a: 96.8% passing rate hides a systematic error in the TPS dose calculation in high gradient regions (peaks, valleys, edges).

Figure 3b: The same data analyzed with more stringent 2% (local)/2mm, clearly showing the TPS dose error patterns in peaks, valleys, and edges.

Figure 3c: MGDR reconstructed dose showing impact of TPS errors in a patient dose axial image (left) and along a sup/inf dose profile (right). Local dose errors in some small regions are in excess of 10%.

3.5. Case 5

Despite a high 3%/3mm passing rate (97.6%), there was a systematic error in how the virtual phantom was modelled in the TPS. The first sign was the 3.8% disagreement between the 0.6cc ion chamber measurement and TPS calculation at isocenter. When examining the exit diodes of the phantom (~23.7 cm deep), the TPS calculation error was almost +7%. 3D MGDR dose confirmed the overestimation of the TPS dose due to improper phantom representation in the TPS. The conventional 3% analysis failed to detect the high TPS error because the global % normalization method uses the maximum measured dose as the denominator, and therefore the exit diode dose, which is less than 0.3 of the entrance, is given an effective absolute dose error tolerance of almost 10% (not 3%). The large local error deep in the phantom was thus completely hidden. See Figures 4a and 4b for the error maps using TG-119 instructions vs. using 2% local/2mm analyses. The source of the error in this case was that instead of appropriately modelling the phantom as a uniform PMMA cylinder [16], a CT-based heterogeneity correction was applied. This resulted in an average electron-density value closer to 1.1 relative to water, instead of the required actual electron density of PMMA, 1.147. Once this was corrected, the TPS calculations became much more accurate, agreeing with both the ion chamber and exit diodes to within 1% (local error).

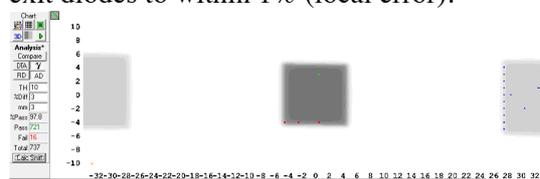


Figure 4a: 97.6% passing rate hides a systematic error, causing large TPS calculation errors go unnoticed.

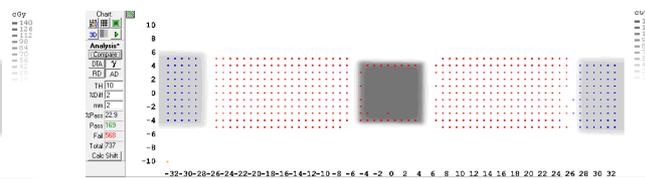


Figure 4b: The same data analyzed with more stringent 2% (local)/2mm criteria, where the results clearly show the TPS dose errors. The blue dots indicate the exit dose errors.

4. Conclusions

In IMRT/VMAT system commissioning (as with any system), quality is improved by the proven strategy of striving for tight tolerances of stringent metrics of accuracy. However, the common 3%/3mm gamma passing rate is insensitive (at any action level) when it comes to detecting systematic errors. A collection of observed “false negatives” are presented here as real-world examples of the inappropriateness of the metric if used as the basis for acceptance testing/commissioning. Overreliance on this *de-facto* current standard inhibits detection and mitigation of systematic errors that can become pervasive in the system, and thus imposes direct and indirect costs (Taguchi Loss Function).

5. References

- [1] Ezzell G A *et al* 2009 *Med. Phys.* **36** 5359-73
- [2] Nelms B E and Simon J A 2007 *J. Appl. Clin. Med. Phys.* **8** 76-90
- [3] Kruse J J 2010 *Med. Phys.* **37** 2516-24
- [4] Nelms B E *et al* 2011 *Med. Phys.* **38** 1037-44
- [5] Zhen H *et al* 2011 *Med. Phys.* **38** 5477-89
- [6] Carrasco P *et al* 2012 *Med. Phys.* (In press)
- [7] Deming W E 1986 *Out of the Crisis (The MIT Press)*
- [8] Aguayo R 1991 *Dr. Deming: The American who taught the Japanese about quality (Simon and Schuster)*
- [9] Shewhart W A 1939 *Statistical method from the viewpoint of quality control (The Graduate School, Department of Agriculture. Washington DC)*
- [10] Ezzell G A *et al* 2009 *Med. Phys.* **36** 5359-73
- [11] Jursinic P A and Nelms B E 2003 *Med. Phys.* **30** 870-79
- [12] Nelms B E *et al* 2010 *J. Appl. Clin. Med. Phys.* **11** 140-57
- [13] Bailey D W *et al* 2012 *J. Appl. Clin. Med. Phys.* **4** 82-99
- [14] Letourneau D *et al* 2009 *Med. Phys.* **36** 1813-21
- [15] Feygelman V *et al* 2011 *J. Appl. Clin. Med. Phys.* **12** 146-67

- [16] Kozelka J *et al* 2011 *Med. Phys.* **38** 5021-32
- [17] Olch A J 2012 *Med. Phys.* **39** 81-6
- [18] Nelms B E *et al* 2012 *Med. Phys.* **39** 4228-38