JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

Using an independent dose calculation system to optimize clinical thresholds and reduce the number of physics dose measurements required for patient-specific quality assurance

K.Hasse[1], B. Ziemer[1], Y. Natsuaki[1], G. Valdes[1], E. Hirata[1], T. Solberg[1], J. Scholey[1], and A. Witztum[1]
[1]University of California, San Francisco
Department of Radiation Oncology

UCSF

INTRODUCTION

Patient-specific quality assurance (PSQA) measurements are crucial in order to judge the quality of a delivered treatment plan; however, PSQA requires significant clinical resources. It is common practice to measure point dose or 2D/3D dose distributions prior to treating patients, and then compare these measured dose distributions with those predicted by the planning system. The identification of plans that require measurement has been previously discussed and justified clinically [1-3]. Mobius 3D (M3D) provides a separate beam model used as a secondary check to the clinical treatment planning system (TPS) dose calculation algorithm [4]. The use of M3D was proposed to identify a subset of plans that would fail an ion chamber measurement and thus reduce the number of plans that would require an ion chamber measurement be done while still detecting failing plans and maintaining patient safety. .

AIM

- To assess the use of Mobius 3D (Varian Medical Systems) dose calculation software to reduce the number of physical ion chamber (IC) dose measurements required for patient-specific quality assurance

RESULTS

Table 2: Breakdown of number of passing plans using M3D and IC by machine

Machine	Pass both M3D and IC	Fail both M3D and IC	Fail M3D only
All (n=1114)	1033	8	73
Truebeam (n=544)	525	0	19
Truebeam STx (n=362)	353	0	9
Versa (n=208)	155	8	45

Table 3: Confusion matrix for binary classification problem using a 5% threshold for both M3D and IC results

N = 1114		Actual Values	
		Positive	Negative
Predicted Values	Positive	8	73
	Negative	0	1033

Table 2 shows the breakdown of plans by machine. A total of 1033 plans (92.7%) passed using both the M3D calculation and the IC reading. By machine, 96.5% of Truebeam (planned with RayStation), 97.5% of Truebeam STx (planned with RayStation), and 74.5% of Versa (planned with Pinnacle) plans passed using both the M3D calculation and IC measurement. All 8 plans that failed the IC measurement also failed the M3D calculation. In addition, 73 plans failed the M3D calculation but not the IC measurement. Table 3 shows the confusion matrix that was constructed for this problem.

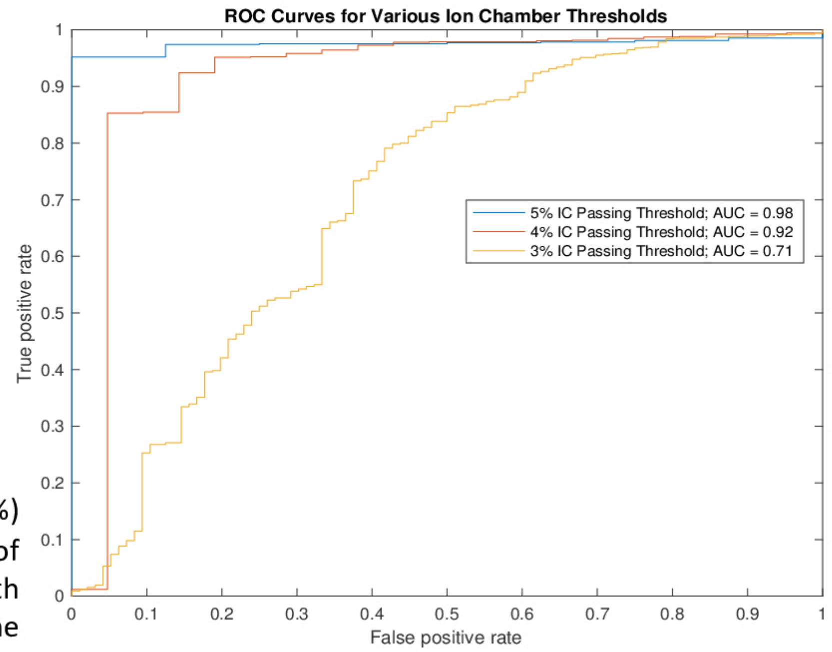

Figure 2: ROC curve for 3%, 4%, and 5% IC thresholds

Table 4: Percent of plans that would need a physical measurement for considered thresholds

IC Threshold	M3D Threshold	Fail M3D only	Fail both	Percent of Current QA Load Still Requiring Physical Measurement
3%	0.5%	885	95	88.0%
4%	3%	259	21	25.1%
5%	4%	133	8	12.7%
5%	5%	73	8	7.3%

To determine whether the hypothesized threshold of 5% difference between M3D and TPS calculations was optimal, an ROC analysis was performed. An ROC analysis was also performed using stricter 4% and 3% ion chamber passing thresholds. The resulting ROC curves are shown in Figure 2.

To avoid missing any IMRT failures, we mandated that our sensitivity must remain 100%. Achieving a sensitivity of 100% with 4% and 3% IC failing criteria requires M3D thresholds of 3% and 0.5%, respectively. With the goal of maximizing specificity while maintaining 100% sensitivity, the 5% M3D threshold is optimal. 3%, 4%, and 5% IC thresholds translate to performing PSQA for 88.0%, 25.1%, and 7.3% of the current PSQA load respectively. This information is shown in Table 4.

METHODS

- A total of 1114 inversely-planned treatment using Pinnacle or RayStation TPS were considered.
- These plans were delivered using Elekta Versa HD, Varian Truebeam, and Varian Truebeam STx linear accelerators between June 2018 and November 2019.
- For each plan, an independent dose calculation was performed using M3D, and an absolute dose measurement was performed using an ion chamber (IC) inside the Mobius phantom. Measured and recorded doses were obtained as illustrated in Figure 1 below.
- Agreement between TPS and IC was used to define ground-truth plan failure. The classification system is illustrated in Table 1.
- ROC analysis was used to determine the M3D agreement threshold for IC plans failing at 3,4, and 5%.

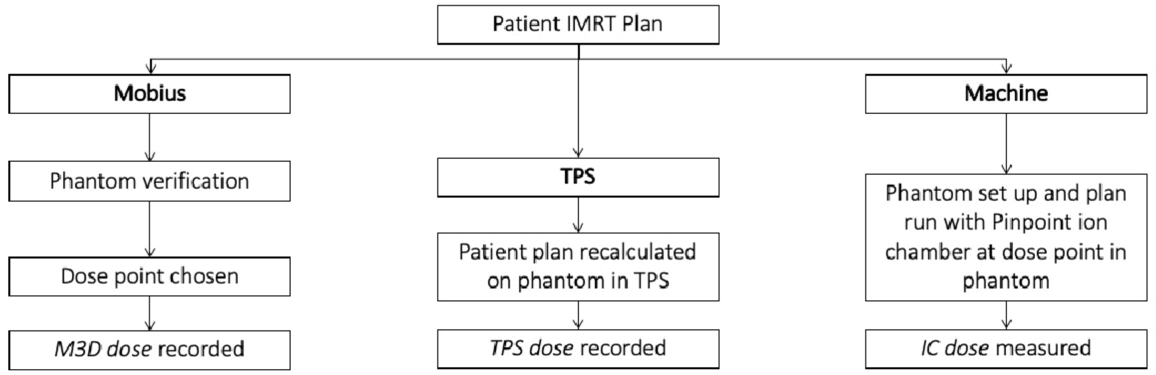

Figure 1: Flowchart of procedure for obtaining measured IC and recorded M3D and TPS doses

Table 1: Classification system for plans using M3D and IC

Classification	Description
True Positive (TP)	Failing plans were correctly identified as failing (Failed both M3D and IC)
False Positive (FP)	Passing plans were incorrectly identified as failing (Failed M3D but passed IC)
True Negative (TN)	Passing plans were correctly identified as passing (Passed both M3D and IC)
False Negative (FN)	Failing plans were incorrectly identified as passing (Passed M3D but failed IC)

CONCLUSIONS

This is the first study to demonstrate that the Mobius 3D system (M3D, Varian Inc) can be used to inform PSQA performance decisions, and to evaluate the thresholds required to maintain 100% sensitivity. Our results indicate that a pre-delivery M3D calculation can identify plans failing an IC measurement using a 5% threshold with 100% sensitivity. By only taking IC measurements on this subset of plans, 1033 fewer physical measurements would have been taken over a period of 15 months, freeing up significant clinical resources while still detecting potentially failing plans and maintaining a high level of confidence in the safety and accuracy of treatment.

REFERENCES

[1] Valdes, G., et al. "A Mathematical Framework for Virtual IMRT QA Using Machine Learning." *Medical Physics*, vol. 43, no. 7, 2016, pp. 4323–4334., doi:10.1118/1.4953835.
[2] Valdes, G., Chan, M.F., Lim, S.B., Scheuermann, R., Deasy, J.O. and Solberg, T.D. (2017), IMRT QA using machine learning: A multi-institutional validation. J Appl Clin Med Phys, 18: 279-284. doi:10.1002/acm2.12161
[3] Interian, Y., Rideout, V., Kearney, V.P., Gennatas, E., Morin, O., Cheung, J., Solberg, T. and Valdes, G. (2018), Deep nets vs expert designed features in medical physics: An IMRT QA case study. Med. Phys., 45: 2672-2680. doi:10.1002/mp.12890
[4] Y. Hillman, et.al., " Refinement of MLC modeling improves commercial QA dosimetry system for SRS and SBRT patient-specific QA," Med Phys **45**, 1351-59 (2018).

CONTACT INFORMATION

Katelyn Hasse, Ph.D.
katelyn.hasse@ucsf.edu