Towards Optimizing Patient Specific IMRT QA Techniques in the Accurate Detection of Acceptable and Unacceptable Plans

Elizabeth McKenzie[1], Peter Balter[1], Jimmy Jones[2], David Followill[1], Francesco Stingo[1], Kiley Pulliam[1], Stephen Kry[1]

[1]Dept of Radiation Physics, UT MD Anderson Cancer Center, Houston, TX, [2]The Methodist Hospital, Houston, TX

Introduction

Due to the complexity of intensity modulated radiation therapy (IMRT), verification of patient plans is performed via direct measurement, called patient-specific IMRT Quality Assurance (QA). Despite its widespread practice, IMRT QA is not standardized, and many methods and varieties of equipment exist to accomplish it [1]. This is further complicated by it not only being a question of the detector used, but also of how the data are analyzed. While ion chamber measurements typically rely on a percent dose difference cutoff, gamma analysis for planar QA relies on three parameters: percent dose difference, distance to agreement, and percent of pixels passing a specific criterion[2]. Additionally, multiple types of software exist for gamma analysis which may implement the calculation and analysis differently.

The purpose of this work was to investigate the performance of several patient-specific IMRT QA dosimeters in terms of their ability to correctly label acceptable and unacceptable plans, as determined by a gold standard. Furthermore, a goal of this project was to establish optimal threshold criteria that are consistent and based on the same criteria among different dosimeters.

Materials

Dosimetric systems were selected for their clinical applicability. This study compares the performance of a Wellhofer cc04 ion chamber (CNMC, Nashville, TN), EDR2 radiographic film (Kodak Carestream, Rochester, NY), ArcCheck helical diode array (Sun Nuclear Corporation, Melbourne, FL), and MapCheck diode array (Sun Nuclear Corporation, Melbourne, FL). An in-house designed multiple ion chamber phantom (MIC) (Figure 1) was used as the gold standard for establishing if a plan was dosimetrically acceptable or not. This phantom allowed for a 3D sampling of the plan using 5 ion chambers set in an insert which could rotate to 8 positions.

.

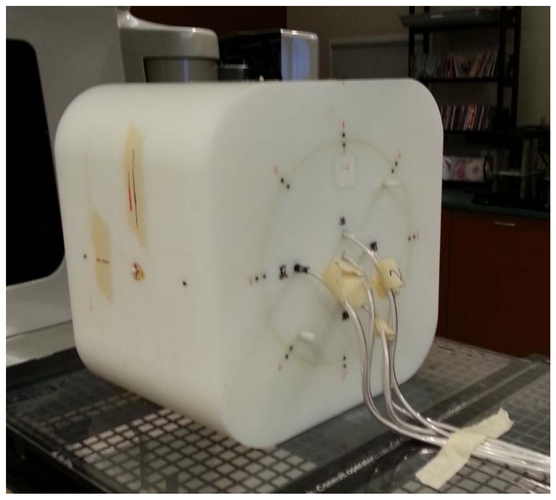

Figure 1:Multiple ion chamber phantom used as gold standard for classifying acceptable and unacceptable plans

Methods

24 IMRT plans were selected from the authors' institutional database. The MIC determined that of these plans, 15 were dosimetrically unacceptable and 9 were acceptable, using a metric that accounts for all the ion chambers in high dose, low gradient regions. The IMRT QA on the same set of plans was measured on the other clinical dosimeters. Additionally, for the MapCheck device, plans were measured in AP field-by-field, AP composite, and planned rotational gantry angle configurations. All planar dosimeters underwent gamma analysis using 2%/2mm, 3%/3mm, and 5%/3mm criteria. The ArcCheck analysis was performed in SNC Patient software, while the radiographic film was analyzed in Omnipro I'mRT software. The MapCheck plans underwent gamma analysis in both Doselab Pro and SNC Patient Software.

In order to study the performance of the dosimeters in terms of their abilities to correctly label both acceptable and unacceptable plans regardless of threshold (% pixels passing for planar dosimeters, or % dose difference for ion chamber), ROC curves were created for each dosimetric system. This led to a total of 25 ROC curves for all combinations of analyses (a selection are shown in Figure 2). The optimal threshold was found by optimizing for both cost and prevalence. In order to compare overall performance, the area under the curve (AUC) was calculated for each ROC curve. A D-test using bootstrapping was performed to determine if any AUC's were statistically different [3]. Then, all AUC's were grouped by device to see if an overall grouping could be determined based on AUC.

Results

Performing pair-wise statistical tests showed that there was no significant difference among the AUC's of the 2%/2mm, 3%/3mm, and 5%/3mm gamma criteria. Also, there wasn't a significant difference between the SNC Patient and Doselab Pro gamma analysis. When the AUC's were grouped by device, two significantly different groups emerged from the results of an ANOVA with a post-hoc Tukey's HSD test

QA System	Average AUC across all analysis systems
cc04 ion chamber	0.94
AP composite MapCheck	0.85
ArcCheck	0.84
EDR2 film	0.82
AP field-by-field MapCheck	0.66
rotationally delivered MapCheck	0.65

Table 1: Average AUC for each device, irrespective of analysis method. The thick line indicates where the devices were significantly grouped based on AUC

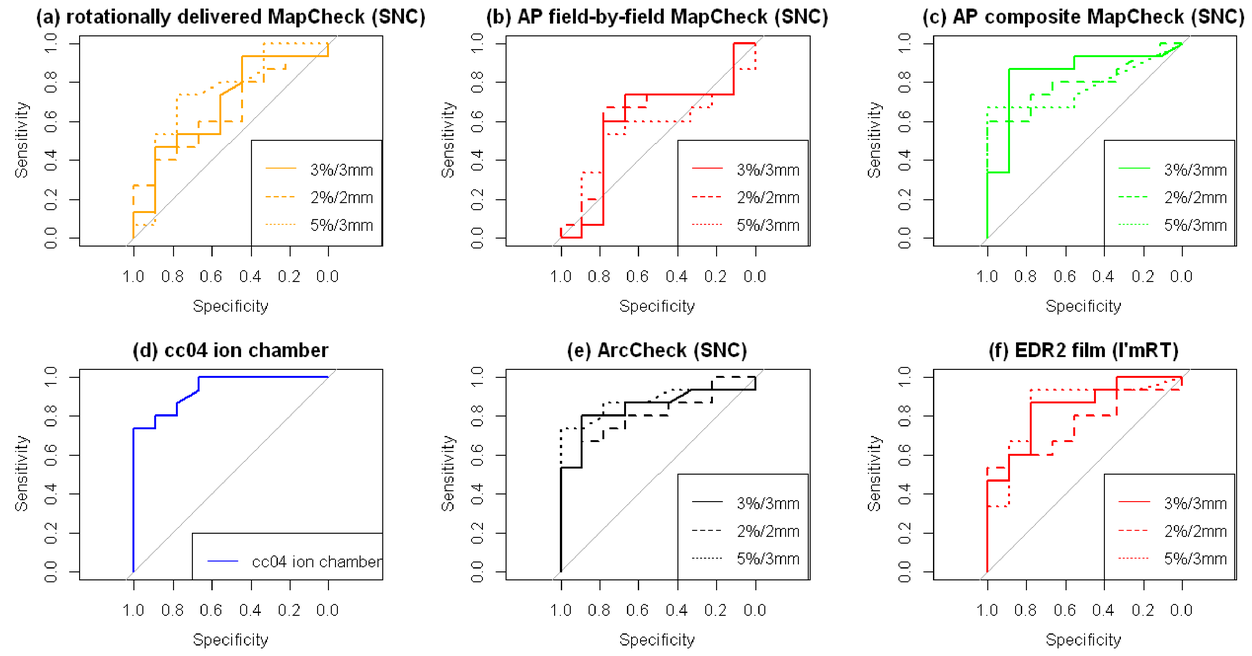

Figure 2: ROC curves generated for each analysis, grouped by dosimetric system. For each planar dosimeter, each panel contains an ROC curve for 2%/2mm, 3%/3mm, and 5%/3mm as the criteria for the gamma analysis. For this figure, all MapCheck gamma analysis was performed using SNC Patient

Device	Empirical Cutoffs: Prevalence is 50%, and Cost of FN is equal to the Cost of FP			Empirical Cutoffs: Prevalence is 3%, and Cost of FN is 0.06 times the cost of FP		
cc04 ion chamber	1.6	±	1.1	3.0	±	0.7
AP field-by-field MapCheck at 3%/3mm	97.0	±	2.6	90.8	±	4.2
rotationally delivered MapCheck at 3%/3mm	90.0	±	8.6	69.3	±	10.2
AP composite MapCheck at 3%/3mm	97.9	±	1.6	82.0	±	9.1
ArcCheck at 3%/3mm	92.0	±	7.1	69.2	±	14.0
EDR2 film at 3%/3mm	97.0	±	9.7	76.3	±	8.2
AP field-by-field MapCheck at 2%/2mm	87.3	±	4.3	74.8	±	8.8
rotationally delivered MapCheck at 2%/2mm	74.4	±	13.9	51.0	±	13.4
AP composite MapCheck at 2%/2mm	85.8	±	5.8	63.8	±	10.7
ArcCheck at 2%/2mm	74.4	±	14.2	49.2	±	11.4
EDR2 film at 2%/2mm	68.1	±	15.0	59.8	±	6.4
AP field-by-field MapCheck at 5%/3mm	99.0	±	1.1	97.0	±	1.5
rotationally delivered MapCheck at 5%/3mm	98.3	±	4.4	83.8	±	7.6
AP composite MapCheck at 5%/3mm	99.7	±	0.3	97.4	±	1.3
ArcCheck at 5%/3mm	96.3	±	1.7	92.1	±	6.1
EDR2 film at 5%/3mm	99.8	±	1.5	91.2	±	5.7

Table 2: Optimal thresholds calculated without weighting (left) and with weighting (right). Included are 95% confidence intervals

Results (continued)

In the better performing group were the cc04 ion chamber, radiographic film, AP composite MapCheck, and ArcCheck. In the poorer performing group were the AP field-by-field and original gantry angle MapCheck.(Table 1)

Adjusting the parameters to set the optimized ion chamber threshold to 3%, it was found that the relative weight of passing an unacceptable plan (false negative) was 0.06 times that of failing an acceptable plan (false positive). Using this weighting combined with an assumed unacceptable plan prevalence of 3% [4], the optimal thresholds were calculated for each dosimetric system. The thresholds were also calculated without weighting (Youden Index). (Table 2)

Conclusion

Differences were noted among the IMRT QA techniques' ability to correctly identify acceptable and unacceptable patient plans, based on their AUC. However, it is interesting to note that no

Conclusion (continued)

statistical difference was found between different distance to agreement and dose difference gamma criteria for any planar device, as well as gamma software used.

This analysis offered a way to calculate optimal thresholds based on prevalence of an unacceptable plan and the relative costs of misclassifying a plan. Using this method could allow a clinic to develop quantitatively justifiable thresholds for their IMRT QA.

References

1) Nelms, B.E. and J.A. Simon, *A survey on planar IMRT QA analysis.* J Appl Clin Med Phys, 2007. **8**(3): p. 2448.
2) Low, D.A., et al., *A technique for the quantitative evaluation of dose distributions.* Med Phys, 1998. **25**(5): p. 656-61.
3) Robin, X., et al., *pROC: an open-source package for R and S plus to analyze and compare ROC curves.* Bmc Bioinformatics, 2011. **12**.
4)Dong, L., et al., *Patient-specific point dose measurement for IMRT monitor unit verification.* International journal of radiation oncology, biology, physics, 2003. **56**(3): p. 867-877.

Support

This investigation was supported by PHS grant CA10953 awarded by the NCI, DHHS.