Predicting the likelihood of QA failure using treatment plan accuracy metrics

2014 J. Phys.: Conf. Ser. 489 012051

(http://iopscience.iop.org/1742-6596/489/1/012051)

Journal of Physics: Conference Series **489** (2014) 012051 doi:10.1088/1742-6596/489/1/012051

Predicting the likelihood of QA failure using treatment plan accuracy metrics

T Kairn[1,2,5], S B Crowe[2], J Kenny[3,4], R T Knight[1] and J V Trapp[2]
[1] Premion Cancer Care, Brisbane, Australia
[2] Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
[3] Australian Clinical Dosimetry Service, Australian Radiation Protection and Nuclear Safety Agency, Melbourne, Australia
[4] Radiation Oncology Queensland, Toowoomba, Australia

E-mail: t.kairn@gmail.com

Abstract. This study used automated data processing techniques to calculate a set of novel treatment plan accuracy metrics, and investigate their usefulness as predictors of quality assurance (QA) success and failure. A small sample of 151 beams from 23 prostate and cranial IMRT treatment plans were used in this study. These plans had been evaluated before treatment using measurements with a diode array system. The TADA software suite was adapted to allow automatic batch calculation of several proposed plan accuracy metrics, including mean field area, small-aperture, off-axis and closed-leaf factors. All of these results were compared to the gamma pass rates from the QA measurements and correlations were investigated. The mean field area factor provided a threshold field size (5 cm^2, equivalent to a 2.2 x 2.2 cm^2 square field), below which all beams failed the QA tests. The small aperture score provided a useful predictor of plan failure, when averaged over all beams, despite being weakly correlated with gamma pass rates for individual beams. By contrast, the closed leaf and off-axis factors provided information about the geometric arrangement of the beam segments but were not useful for distinguishing between plans that passed and failed QA. This study has provided some simple tests for plan accuracy, which may help minimise time spent on QA assessments of treatments that are unlikely to pass.

1. Introduction

This study aimed to identify one or more intensity modulated radiotherapy (IMRT) treatment plan dosimetric accuracy metrics, calculable from treatment plan parameters, that could be used to identify plans that were likely to fail routine quality assurance (QA) tests. The use of such metrics to identify failing treatments at the plan completion stage could reduce the linac time required for IMRT QA testing and contribute to the continuous quality improvement of the treatment planning process.

Existing IMRT complexity and deliverability metrics are based on the heterogeneity of the fluence map produced by each beam [1, 2, 3] or on the variations between MLC positions and beam segment aperture geometries [4, 5, 6]. Most of these metrics have been shown to successfully distinguish between IMRT treatment plans with different levels of complexity (for example, distinguishing

[5] To whom any correspondence should be addressed.

between prostate IMRT treatment plans and more-complex head and neck IMRT treatment plans), while showing no clear correlation with increasing dosimetric error for individual treatment sites [3, 6, 7, 8].

This study therefore investigated a set of IMRT treatment plan accuracy metrics that were designed to be sensitive to the treatment plan parameters that are most likely to compromise accurate dose calculations; small field and small segment aperture sizes [9, 10], closed MLC leaves below open linac jaws [11], and small field segments delivered from off-axis positions [12, 13].

2. Method

This study used the results of pre-treatment QA measurements made using a MapCheck2 diode array (Sun Nuclear Corporation, Melbourne, USA), for a small sample of 151 beams from 20 prostate and 3 cranial IMRT treatments, planned and delivered using the BrainLab iPlan treatment planning system (with pencil-beam algorithm) and Brainlab m3 microMLC (Brainlab, Feldkirchen, Germany), with a Varian iX linac (Varian Medical Systems, Palo Alto, USA). The small sample size was determined by the infrequency of microMLC-based IMRT treatments, especially of cranial targets, over the 4 month testing period. 17 of the prostate plans and 2 of the cranial plans were defined as "passed" and approved for treatment, with either 90% of measurement points (above a threshold of 10% of the maximum dose) resulting in $\gamma(2\%, 2mm) < 1.0$, or 95% of measurement points resulting in $\gamma(3\%, 3mm) < 1.0$, when compared with dose to the diode array calculated by the treatment planning system. The remaining plans, which did not meet these criteria, were defined as "failed". This gamma evaluation was performed using MapCheck software (Sun Nuclear Corporation, Melbourne, USA), using the Van Dyk percentage difference evaluation method [14], in absolute mode with no beam smoothing and with a 10% dose threshold.

The treatment plans were exported from the planning system in DICOM format for analysis. Automatic batch calculation of a set of novel treatment accuracy metrics was achieved using the TADA (Treatment And Dose Assessor) code [15], an extension to the MCDTK software suite [16, 17]. The TADA code was adapted to allow automatic batch calculation of metrics including:

- Mean field area (MFA), the average open area of field segments in each beam
- Mean aperture displacement (MAD), the average distance between the midway point between each pair of leaves and the central axis
- Cross-axis score (CAS), the proportion of open leaf pairs where one leaf crossed the central axis
- Closed leaf score (CLS), the proportion of leaf pairs that were closed, downstream of open jaws
- Small aperture score (SAS), the proportion of open leaf pairs that were separated by less than a given threshold distance.

These metrics were calculated separately for each individual beam and then combined in monitor unit weighted averages for each treatment plan.

The treatment plan analysis results obtained using TADA were compared with the pass rates from the QA tests, in order to identify which, if any, of the metrics could be related to the MapCheck2 pass rate or the likelihood of QA failure. The relationship between beam accuracy metrics and individual gamma pass rates was analysed using Excel 2007 (Microsoft, Redmond, USA), with linear regression used for trend estimation and p-value calculation, using two-tailed T-tests. Two-tailed Welch's T-tests were also used to obtain p values expressing the statistical significance of observed differences between the mean values of the accuracy metrics for the prostate treatment beams that passed their QA tests, the prostate treatment beams that failed their QA tests, the cranial treatment beams that passed their QA tests and the cranial treatment beams that failed their QA tests.

3. Results and Discussion

Figures 1(a)-(d) show the results of evaluating the metrics relating to the mean area of the IMRT beam segment apertures and the displacement of those apertures across the linac's central axis. The MFA results shown in figure 1(a) are correlated with gamma pass rates ($p \approx 0.02$) for both the prostate and the cranial treatment fields. Figure 1(a) also suggests that there is a threshold mean aperture size of 5

cm^2 (equivalent to a 2.2 x 2.2 cm^2 square) below which all beams fail their QA tests. Above this threshold, the MFA metric fails to distinguish between plans that pass and fail their QA tests, with 17% of beams with mean field areas between 5 cm^2 and 20 cm^2 failing their QA tests. Results for the MAD, CAS and CLS metrics, shown in figures 1(b), (c) and (d), respectively, provide no useful relationships with gamma pass rates, although they provide information about the geometric arrangement of the IMRT beam segments. The results for MAD and CAS are qualitatively similar, since the number of segments per beam that are centred away from the central axis is correlated with the number of MLC leaves per beam that cross the central axis. The CLS results in figure 1(d) are inversely related to the MFA results in figure 1(a) for similar, geometric reasons.

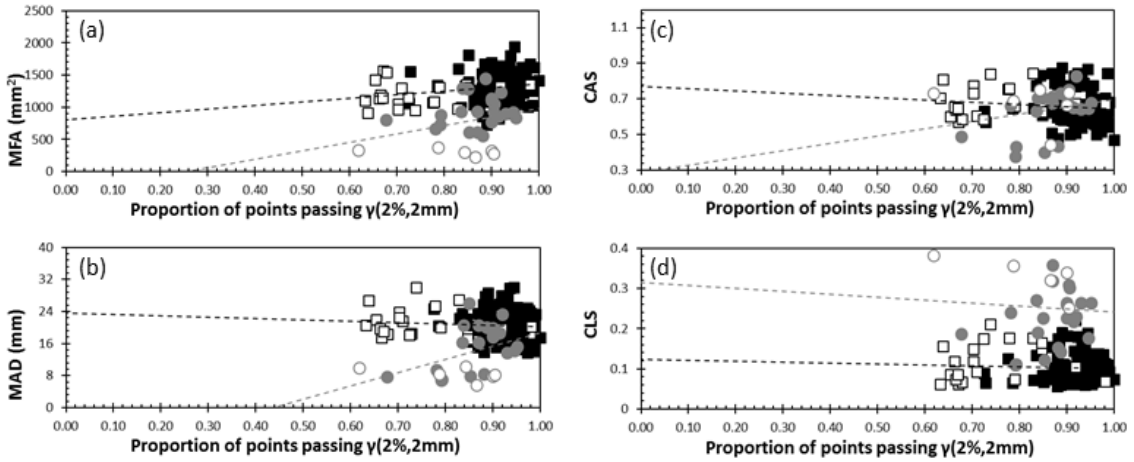

Figure 1. (a) Mean field area, (b) mean aperture displacement (c) cross-axis score and (d) closed leaf score, plotted against the proportion of measurement points in each field that passed a (2%, 2mm) gamma comparison with the planned dose. Filled data points represent fields from plans that were defined as "passed". Open data points represent fields from plans that were defined as "failed". Black squares represent beams from prostate treatments and grey circles represent beams from cranial treatments. Linear trend lines are included as a visual guide only.

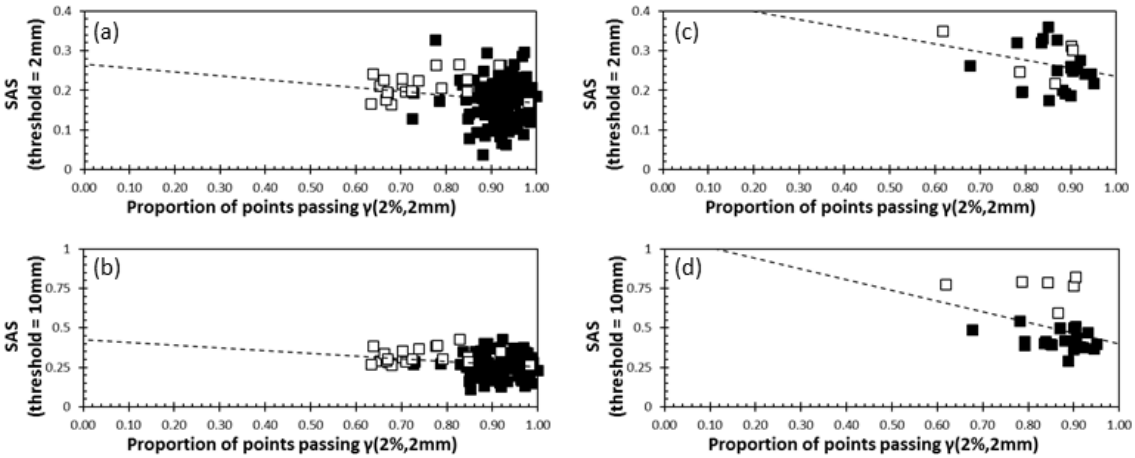

Figure 2. (a) Small aperture scores for prostate treatment beams, calculated using a threshold value of 2 mm, (b) small aperture scores for prostate treatment beams, calculated using a threshold value of 10 mm, (c) small aperture scores for cranial treatment beams, calculated using a threshold value of 2 mm, and (d) small aperture scores for cranial treatment beams, calculated using a threshold value of 10 mm, plotted against the proportion of measurement points in each

field that passed a (2%, 2mm) gamma comparison with the planned dose. Filled data points represent fields from plans that were defined as "passed". Open data points represent fields from plans that were defined as "failed". Linear trend lines are included as a visual guide only.

Figures 2(a) and (b) show the SAS values for the prostate IMRT treatment beams, respectively evaluated using thresholds of 2 mm and 10 mm to define the "small" aperture. Correlations between these metrics and the gamma pass rates for these individual beams are either weak ($p \approx 0.05$ when using the 10 mm threshold) or insignificant ($p \approx 0.2$ when using the 2 mm threshold), however it is apparent from figures 2(a) and (b) that when fewer than 18% of open MLC leaf pairs are open by less than 2 mm, or when fewer than 25% of open MLC leaf pairs are open by less than 10 mm, all beams pass their QA tests. When these small aperture scores are evaluated as monitor unit weighted averages over all beams in each treatment, the small aperture score (defined at 10 mm) has a useful threshold value at 0.32, above which all of the prostate plans that failed their QA tests, plus two of the plans that passed their QA tests, would be identified as likely to fail. All prostate plans with mean SAS values less than 0.32 passed their QA tests.

Results shown in figures 2(c) and (d) suggest that the SAS is similarly able to distinguish between cranial treatment beams that passed and failed their QA tests, when the "small" aperture is defined as 10 mm. Given that a small sample of beams from only one failing and two passing cranial treatments were available for use in this analysis, it is important that this result be confirmed by further study.

Tables 1 and 2 summarise the suitability of the metrics defined in this study for predicting the likely QA outcomes of the IMRT treatment plans examined here. Table 1 lists the mean treatment plan accuracy metrics for the prostate and cranial plans that passed and failed their QA tests. Table 2 provides an indication of the significance of differences between the calculated metrics for the different plan cohorts.

Table 1. Summary of plan accuracy metrics (mean and one standard deviation) for prostate and cranial plans that passed and failed their QA tests. The small aperture thresholds used in the SAS calculations are indicated in parentheses.

Metric	Prostate - pass	Prostate - fail	Cranial - pass	Cranial – fail
MFA (cm^2)	13 ± 2	12 ± 2	9 ± 2	3.0 ± 0.5
MAD (cm)	20 ± 4	21 ± 4	16 ± 6	8 ± 2
CAS	0.65 ± 0.09	0.68 ± 0.08	0.6 ± 0.1	0.7 ± 0.1
CLS	0.10 ± 0.04	0.12 ± 0.05	0.22 ± 0.07	0.35 ± 0.07
SAS (2mm)	0.17 ± 0.06	0.21 ± 0.03	0.25 ± 0.05	0.31 ± 0.08
SAS (10mm)	0.26 ± 0.07	0.32 ± 0.05	0.42 ± 0.06	0.75 ± 0.08

Table 2. Calculated p values (Welch's T-test), for comparisons between metrics listed in Table 1, for prostate plans that passed (PP) and failed (PF) their QA tests and cranial plans that passed (CP) and failed (CF) their QA tests. Small p-values (< 0.001) suggest that the differences between the datasets are unlikely to have arisen by random fluctuations.

PP vs PF	CP vs CF	PP vs CP	PF vs CF
0.02	< 0.001	< 0.001	< 0.001
0.39	< 0.001	< 0.001	< 0.001
0.21	0.52	0.43	0.81
0.13	0.005	< 0.001	< 0.001
< 0.001	0.14	< 0.001	0.03
< 0.001	< 0.001	< 0.001	< 0.001

4. Conclusion

This study evaluated a set of metrics designed to be sensitive to the treatment plan parameters that are most likely to compromise accurate dose calculations. Based upon our small sample our results suggest that some of these metrics may be used to identify plans that are unlikely to pass routine pre-treatment QA tests.The MFA factor provided a threshold field size (5 cm^2), below which all beams failed their QA tests. The SAS provided a useful predictor of plan failure, when averaged over all beams, despite being weakly correlated with gamma pass rates for individual beams. By contrast, the MAD, CAS and CLS factors provided information about the geometric arrangement of the beam segments but were not useful for distinguishing between plans that passed and failed QA. This study has provided some simple tests for plan accuracy, which may help minimise time spent on QA assessments of treatments that are unlikely to pass.

Acknowledgements

This study was supported by the Australian Research Council, the Wesley Research Institute, Premion and the Queensland University of Technology (QUT), through linkage grant number LP110100401**.**

References

[1] Webb S 2003 *Phys. Med. Biol.* **48**(14) 2051–2062
[2] Llacer J, Solberg T D and Promberger C 2001 *Phys. Med. Biol*. **46**(10) 2637–2663
[3] Nicolini G, Vanetti E, Clivio A, Fogliata A, Korreman S, Bocanek J and Cozzi L 2008 *Radiat. Oncol*. **3** 24
[4] McNiven A L, Sharpe M B and Purdie T G 2010 *Med. Phys.* **37** 505–515
[5] Lee M T, Purdie T G, Eccles C L, Sharpe M B, Dawson L A et al. 2010 *Radiat. Oncol***.5** 115–115.
[6] McGarry C K, Chinneck C D, O'Toole M M, O'Sullivan J M, Prise K M and Hounsell A R 2011 *Med. Phys*. **38** 2027–2034.
[7] Tonigan J, Kry S, Dong L, Purdie T, White R, Ibbott G and Followill D 2011 *Med. Phys*. **38** 3804.
[8] Tonigan J R 2011 Evaluation *of intensity modulated radiation therapy (IMRT) delivery error due to IMRT treatment plan complexity and improperly matched dosimetry data* Master's thesis University of Texas
[9] Kairn T, Crowe S, Kenny J and Trapp J V 2011 *Radiat. Meas.* **46**(12) 1985–1988
[10] Kairn T, Hardcastle N, Kenny J, Meldrum R, Tome W A and Aland T 2011 *Australas. Phys. Eng. Sci. Med.* **34**(3) 333–343
[11] Fenoglietto P, Lalibert´e B, Ailleres N, Riou O, Dubois J B and Azria D 2011 *Radiat. Oncol.* **6** 85
[12] Ahnesjo A and Aspradakis M M 1999 *Phys. Med. Biol*. **44**(11) R99–R155
[13] Brainlab AG 2010 *Brainlab Physics Technical Reference Guide* Revision 1.2
[14] Van Dyk J, Barnett R B, Cygler J E and Shragge P C 1993 *Int. J. Radiat. Oncol. Phys*. **26** 261–273
[15] Crowe S B, Kairn T, Middlebrook N, Hill B, Knight R T, Kenny J, Langton C M and Trapp J V 2013 *Australas. Phys. .Eng. Sci. Med.* **36**(1): 74-74
[16] Crowe S, Kairn T and Fielding A L 2009 *Radiother. Oncol*. **92** S71
[17] Crowe S B, Kairn T, Trapp J V and Fielding A L 2013 Experimental evaluation of MCDTK, the Monte Carlo DICOM ToolKit *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China* (Berlin: Springer) pp. 1807–1810