

Dosimetric and radiobiological comparison for quality assurance of IMRT and VMAT plans

Nava Raj Paudel^{1,2} | Ganesh Narayanasamy¹ | Eun Young Han³ | Jose Penagaricano¹ | Panayiotis Mavroidis⁴ | Xin Zhang¹ | Anil Pyakuryal⁵ | Dongwook Kim⁶ | Xiaoying Liang⁷ | Steven Morrill¹

¹Department of Radiation Oncology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

²Department of Radiation Oncology, UPMC Susquehanna, Williamsport, PA, USA

³Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

⁴Department of Radiation Oncology, University of North Carolina, Chapel Hill, NC, USA

⁵University of District of Columbia, Washington, DC, USA

⁶Department of Radiation Oncology, Kyung Hee University Hospital at Gangdong, Seoul, Korea

⁷Department of Radiation Oncology, University of Florida, Gainesville, FL, USA

Author to whom correspondence should be addressed. Nava Raj Paudel
E-mail: nrpaudel@yahoo.com;
Telephone: 570 326 8870

Abstract

Introduction: The gamma analysis used for quality assurance of a complex radiotherapy plan examines the dosimetric equivalence between planned and measured dose distributions within some tolerance. This study explores whether the dosimetric difference is correlated with any radiobiological difference between delivered and planned dose.

Methods: VMAT or IMRT plans optimized for 14 cancer patients were calculated and delivered to a QA device. Measured dose was compared against planned dose using 2-D gamma analysis. Dose volume histograms (for various patient structures) obtained by interpolating measured data were compared against the planned ones using a 3-D gamma analysis. Dose volume histograms were used in the Poisson model to calculate tumor control probability for the treatment targets and in the Sigmoid dose–response model to calculate normal tissue complication probability for the organs at risk.

Results: Differences in measured and planned dosimetric data for the patient plans passing at $\geq 94.9\%$ rate at 3%/3 mm criteria are not statistically significant. Average \pm standard deviation tumor control probabilities based on measured and planned data are $65.8 \pm 4.0\%$ and $67.8 \pm 4.1\%$ for head and neck, and $71.9 \pm 2.7\%$ and $73.3 \pm 3.1\%$ for lung plans, respectively. The differences in tumor control probabilities obtained from measured and planned dose are statistically insignificant. However, the differences in normal tissue complication probabilities for larynx, lungs-GTV, heart, and cord are statistically significant for the patient plans meeting $\geq 94.9\%$ passing criterion at 3%/3 mm.

Conclusion: A $\geq 90\%$ gamma passing criterion at 3%/3 mm cannot assure the radiobiological equivalence between planned and delivered dose. These results agree with the published literature demonstrating the inadequacy of the criterion for dosimetric QA and suggest for a tighter tolerance.

PACS
87.45.Qr

KEY WORDS
3DVH, ArcCheck, dosimetric quality assurance, NTCP, radiobiological QA, TCP

1 | INTRODUCTION

Treatment plans used in radiation therapy are generally evaluated on the basis of dose distribution and dose volume parameters. Accuracy of treatment delivery of complex treatment plans is assured through a quality assurance (QA) process using ion chambers, films and more commonly with diode array,^{1,2} and ion chamber array measurements.³ Measured data are generally compared against planned data using two-dimensional (2-D) gamma analysis. In reality, treatment target and organs at risk (OARs) present 3-D geometry, and a 2-D gamma analysis-based dosimetric comparison may not provide information about the criticality of a disagreement. Several studies⁴⁻⁷ have shown that 2-D gamma analysis fails to detect errors in some cases. Even though detailed analysis is still under investigation,⁸ dose volume histogram (DVH)-based dosimetric evaluation can provide structure-by-structure information and a 3-D gamma analysis can be a better option for the QA purpose. Validity of DVH-based evaluation of delivered intensity-modulated radiation therapy (IMRT) plans against the corresponding plans optimized with treatment planning system (TPS) have been demonstrated using film, ion chamber,⁹ and BANG3 gel dosimetry.¹⁰ A number of studies show that a DVH-based 3-D gamma analysis provides more reliable comparison than a point-by-point per-beam 2-D gamma analysis of IMRT plans.^{8,11,12}

Most of the DVH-based studies were based on measurement and/or interpolation of 2-D device measured data into 3-D dose distribution.^{7,10,11} The DVH comparison studies have shown differences between planned and measured DVHs as well as differences in mean doses.^{10,13} As per the report of Task Group (TG) 65 of American Association of Physicists in Medicine (AAPM), tumor control probabilities (TCPs) are 2–4 more sensitive with respect to the change in uniform dose and normal tissue complication probabilities (NTCPs) are 4–6 times more sensitive to the change in uniform dose.¹⁴ However, none of the above mentioned studies have evaluated whether radiobiological differences existed in any of the cases. A growing recognition of the limitation of dose volume parameters in correlating with biological response has prompted for the use of radiobiological models for treatment planning¹⁵ but QA of all plans is still performed on the basis of dosimetric comparison alone. Very recently, an attempt for a radiobiological comparison between delivered and planned IMRT treatment plans was made using a 2-D QA device (MapCheck, Sun Nuclear) measured dose.¹⁶

Here, we use a cylindrical QA device, ArcCheck, for the measurement and compare the measured data against TPS-calculated (planned) data using 2-D gamma analysis in SNC PatientTM software. Application of ArcCheck for patient-specific dosimetric QA¹² and DVH-based plan verification using Sun Nuclear's 3DVH[®] software has been experimentally verified elsewhere.¹⁷ In this study, DVHs for various structures were created from the ArcCheck measured data using 3DVH software and compared against the TPS-planned DVHs using 3-D gamma analysis. Radiobiological comparison of ArcCheck measured plans was performed against the corresponding TPS-optimized volumetric-modulated arc therapy (VMAT) or static gantry IMRT plans.

2 | MATERIALS AND METHODS

2.A | Patient selection, treatment planning, and measurement

Fourteen patients treated with IMRT or VMAT were retrospectively selected for this study. Among them, seven were head and neck (H&N) patients and seven were lung patients. Three of H&N patients were planned for VMAT and four were planned for IMRT. Similarly, five of the lung patients were planned for VMAT and two were planned for IMRT. VMAT plans used two full or partial arcs while IMRT plans used 7–9 static fields to get optimal target coverage. Sliding window method was used in all plans. Computed tomography (CT) simulation was performed in a Philips Big Bore CT scanner (Philips Medical Systems, Amsterdam, Netherlands). Vac-Locs and type-S masks (CIVCO Radiotherapy, Coralville, Iowa) were used for immobilizing lung and H&N patients, respectively. Varian couch is modeled in our TPS and was inserted in each treatment plan. H&N mask was included in the body contour (used for dose calculation) while Vac-Loc was not included in the body contour. Even though the beam attenuation due to the immobilization devices is minimal, the field arrangement was optimized to minimize the fraction of the beams passing through them. The lung plans were optimized with Acuros XB algorithm and all other plans were optimized with AAA algorithm in Eclipse TPS (Version 11.0.47, Varian Medical Systems, Palo Alto, CA, USA) using 6 MV beam of a Varian TrueBeam STX linear accelerator. Dose of 66 Gy was prescribed to gross tumor volume (GTV) of all H&N patients while 60 Gy and 54 Gy were prescribed to lymph nodes for a few patients. Five of seven lung patients were prescribed to 66 Gy and other two were prescribed to 61.2 and 72 Gy, respectively. The dose grid resolution used was 1 mm for lung and 2 mm for H&N plans. All the plans were normalized to cover 95% of target volume by 100% of the prescribed dose while minimizing dose to the surrounding OARs. The planning objectives in terms of dose constraints to OARs were the following: maximum dose to the spinal cord 50 Gy, maximum dose to the brainstem 60 Gy, mean dose to parotids <26 Gy (or median dose <30 Gy), mean dose to larynx <40 Gy, mean dose to heart <20 Gy, and mean lung dose <20 Gy and $V_{20\text{ Gy}} <25\%$ to the lungs. In few patients, some of the OAR dose constraints were not met due to the close proximity of infiltration of the tumor to those OARs. The optimized plan doses were calculated on ArcCheck images in Eclipse TPS and delivered to the ArcCheck phantom (Sun Nuclear, Melbourne, FL, USA) with the insert in place. Cumulative dose was recorded using the SNC Patient software version 6.2.1 (Sun Nuclear, Melbourne, FL, USA).

2.B | Dosimetric comparison

Dosimetric comparison between planned and measured data comprised of mean and maximum dose to treatment target and few OARs. For H&N plans, mean dose to GTV, esophagus, larynx,

parotids, and maximum dose to brainstem and cord were compared while for lung plans, mean dose to the GTV, heart, esophagus, normal lung, and maximum dose to the cord were compared.

The measured data were compared against the planned data using a minimum of 90% pass rate with 2-D gamma criteria of 3 mm distance to agreement (DTA), 3% dose difference (DD) global¹ using the SNC Patient. A 10% dose threshold and global normalization was used.

The 3DVH software requires two set of data for comparison. TPS-optimized patient plan, dose distribution, contoured structures and planning CT images of each patient, corresponding TPS-calculated (planned) dose on ArcCheck CT images, and measured data were imported to 3DVH software version 3.2. The ArcCheck measured data were then converted into 3-D dosimetric data with planned dose perturbation (PDP) algorithm^{8,9} and compared against original patient data obtained from TPS using the 3-D gamma analysis (3 mm DTA, 3% DD global). The 3-D gamma analysis results were compared against the corresponding 2-D gamma analysis results. DVH data for target and various structures obtained from 3DVH and TPS (1 cGy bin size) were imported to MATLAB (The MathWorks, Inc., Natick, MA, USA) for radiobiological comparison.

2.C | Radiobiological comparison

DVHs based on measurement and TPS were used in biological models to calculate TCPs and NTCPs using MATLAB. Conventional fractionation scheme (1.8–2 Gy per fraction) was used for the calculations.

2.D | TCPs with Poisson model

Clinical target volume (CTV) is the volume of tumor intended to treat. In our study, GTV included gross tumor and subclinical microscopic disease, and CTV was labeled as GTV. Hence, it is appropriate to calculate TCP for GTV even though dose is prescribed to planning target volume (PTV) to incorporate set up error. In this study, TCPs for H&N and lung GTVs were calculated using the Poisson model,^{18–20} which is expressed in eq. (1).

$$TCP = \left[\frac{1}{2} \sum_i v_i \exp \left\{ 2\gamma_{50} \left(1 - \frac{D_i}{D_{50}} \right) / \ln 2 \right\} \right] \quad (1)$$

Here, D_{50} is dose yielding 50% probability for tumor control and γ_{50} is slope of the dose–response curve at the level of 50% TCP. Similarly, D_i and v_i are the dose and volume elements of DVHs, respectively.

For the calculation purpose, 63.43 Gy and 51.24 Gy were used as D_{50} for H&N GTV and lung GTV, respectively. Similarly, the values of 2.66 and 0.83 were used, respectively, for γ_{50} . These values were derived from clinical data (DVHs and treatment outcomes) using a cohort of 90 patients. A regression analysis was used to fit the TCP model into the clinical data.²¹

2.E | NTCP with sigmoid dose–response model

The values of NTCP for various OARs were calculated using DVHs obtained from planned and measured data in the sigmoid dose–response (SDR) model,^{21–23} expressed in eq. 2.

$$NTCP = \Phi \left(\frac{EUD - D_{50}}{mD_{50}} \right) \quad (2)$$

Here, Φ is the probit function defined by:

$$\begin{aligned} \Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left(-\frac{t^2}{2} \right) dt \\ &= \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \end{aligned} \quad (3)$$

Here, $x = (EUD - D_{50})/mD_{50}$, where D_{50} is the dose yielding 50% NTCP, obtained from dose–response curve, and EUD is equivalent uniform dose, defined as the dose which distributed uniformly over a structure would produce the same effect as the dose specified by the DVH. The parameter m represents the slope of the dose–response curve. EUD is also defined as a generalized equivalent uniform dose (gEUD) calculated using the series of dose volume pairs (D_i, v_i), obtained from the DVH of a structure using the formula expressed in eq. 4.

$$gEUD = \left[\sum_i v_i D_i^{1/n} \right]^n \quad (4)$$

Here, n is a parameter that determines the dose volume dependence of a given OAR.

Brainstem, spinal cord, esophagus, larynx, and left parotid and right parotid NTCPs were evaluated for H&N plans while bilateral lungs, heart, spinal cord, and esophagus NTCPs were evaluated for lung patients. D_{50} , n , and m values from Burman et al.^{22,24} used for NTCP calculations are tabulated in Table 1.

2.F | Statistical analysis

The Shapiro–Wilk test was used to test the normality of the data. A two-tailed Student t -test (0.05 significance level) was performed for the data following normal distribution to test the significance in difference between (a) planned and measured dose to the studied structures, (b) 2-D and 3-D gamma analysis results, (c) TCPs based on planned and measured data, and (d) NTCPs based on planned and measured data. Wilcoxon signed rank test was performed on the data not following a normal distribution.

2.G | Effect of tighter tolerance on dosimetric analysis

Tighter tolerances of 2.5%/2.5 mm and 2%/2 mm global normalization were used for 2-D gamma analysis to check whether a radiobiological equivalence between planned and delivered dose distributions exists for the patient plans passing at $\geq 90\%$.

TABLE 3 Mean or maximum dose (Gy) to various structures obtained from treatment plan and measurement, and p-values of statistical test for lung patient plans.

Patient No.	Dose (Gy)	GTV _{mean}	Heart _{mean}	Cord _{max}	Esophagus _{mean}	Lung _{mean}
1	Planned	67.6	13.9	40.9	21	18.9
	Measured	69.6	14.6	41.9	21.7	19.6
2	Planned	63.4	3.4	39.1	19.7	16.2
	Measured	63.9	3.5	39.5	19.9	16.4
3	Planned	68.1	0.3	28.1	3.3	3.5
	Measured	67.3	0.3	29	3.4	3.6
4	Planned	67.8	8.7	43.4	29.8	16.3
	Measured	69.3	8.8	44.4	29.8	16.9
5	Planned	70.2	1.7	43.7	6.7	13.3
	Measured	73	1.8	47	6.7	13.8
6	Planned	64.8	4.7	37	15.7	13.2
	Measured	66.3	4.8	37.4	15.8	13.6
7	Planned	69.2	1.2	36.8	16.2	8.5
	Measured	70.6	1.2	37.3	16.6	8.8
Average ± SD	Planned	67.3 ± 2.4	4.8 ± 4.9	38.4 ± 5.3	16.1 ± 8.9	12.8 ± 5.3
	Measured	68.6 ± 3.0	5.0 ± 5.1	39.5 ± 5.9	16.3 ± 9.0	13.2 ± 5.4
P-value		>0.05	>0.05	>0.05	>0.05	>0.05

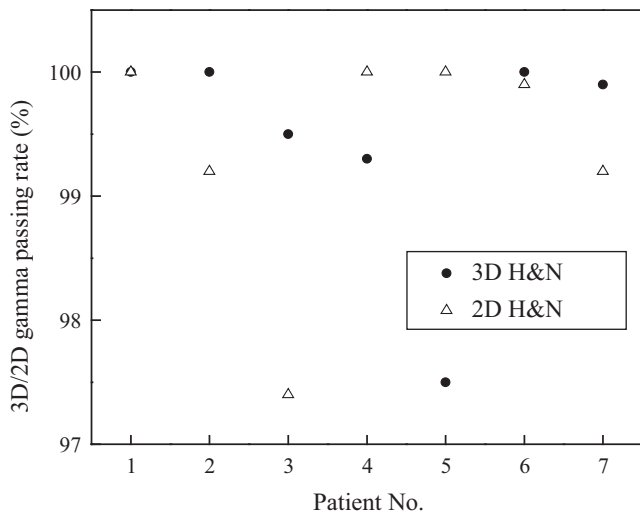


FIG. 1. Two-D and 3-D gamma pass rates for H&N patient plans (patients 4, 5, & 7 are VMAT and all others are IMRT).

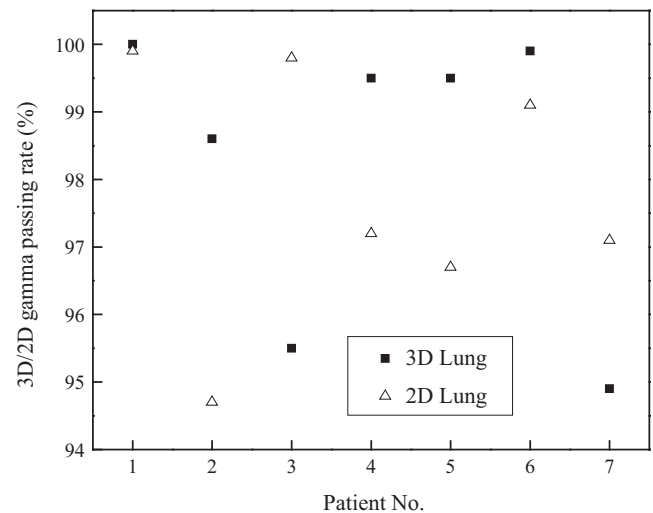


FIG. 2. Two-D and 3-D gamma pass rates for lung patient plans (patients 1 & 6 are IMRT and all others are VMAT).

and measured data as well as P-values of the statistical test for the studied OARs are tabulated in Tables 6 and 7, respectively. The structures with significant difference in NTCPs are marked bold.

NTCPs for the other OARs in most of the studied patients were negligible, ranging between 10⁻⁷% and less than a percent. Only the values greater than or equal to 0.5% have been tabulated to one significant figure in Tables 6 and 7, and all values smaller than 0.5% have been indexed as <0.5%.

As evident from Tables 6 and 7, differences in NTCPs based on measurement and calculation are statistically significant for larynx, lungs-GTV, heart, and cord. Even though the values are a few

percent, the statistically significant difference in NTCPs may indicate a possibility of clinically significant difference between delivered and intended outcome.

3.C | Effect of tighter tolerance on dosimetric gamma analysis

A 2-D gamma analysis performed at 2.5%/2.5 mm reduced the pass rate slightly for most of the patient plans but not below 92.7% for any of the plans. The pass rate ranged between 92.7% and 100.0% with median pass rate of 99.4% for H&N plans and between 91.0%

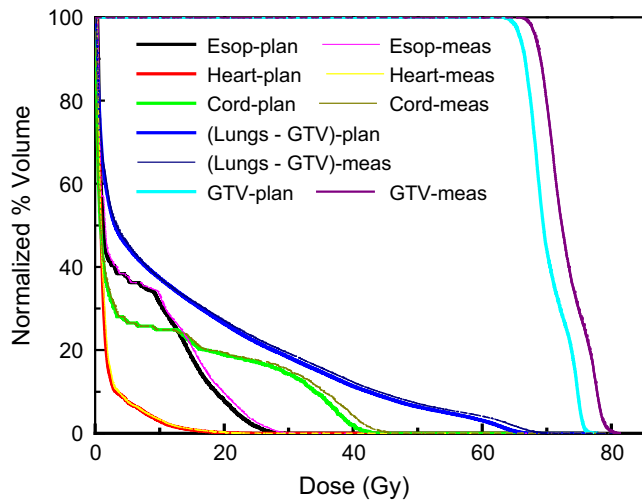


FIG. 3. DVHs obtained from treatment plan and ArcCheck measured data for a lung patient plan.

and 99.5% with median pass rate of 96.6% for lung plans. However, for a 2-D gamma analysis based on 2%/2 mm criterion, pass rate for H&N patients ranged between 85.3% and 99.6% with the median

pass rate of 98.3%. For lung patient plans, it ranged between 83.3% and 97.1% with the median pass rate of 90.2%. Only five H&N patient plans and three lung patient plans met the passing criterion of $\geq 90\%$ at 2%/2 mm. The statistical test on the plans passing by $\geq 90\%$ at 2%/2 mm criterion did not show any radiobiological difference for any of the structures studied.

4 | DISCUSSION

Our study showed small dosimetric differences between 2-D and 3-D gamma analysis results, which are in line with the results obtained by Infusino et al.¹⁷ using ArcCheck for the measurement. However, our radiobiological comparisons do not agree with the results from Sumida et al.¹⁶ where the TCPs based on measured data were found to be significantly smaller and NTCPs to be significantly higher than the ones based on planned data. Possible differences could be because of differences in device type, geometry, differences in measurement and analysis techniques, as well as the different radiobiological models used to calculate TCPs and NTCPs. While Sumida et al. had used per-beam analysis using MapCheck measured data, we have used cumulative dose analysis using ArcCheck measured data.

TABLE 4 TCPs and statistical values from planned and measured data for H&N patients.

Patient No.		1	2	3	4	5	6	7	Mean	SD	P-value
TCP (%)	Planned	57.9	64.7	65.0	66.5	67.0	69.2	70.1	65.8	4.0	>0.05
	Measured	60.2	63.8	70.3	71.2	69.5	69.6	70.0	67.8	4.1	

TABLE 5 TCPs and statistical values from planned and measured data for lung patients.

Patient No.		1	2	3	4	5	6	7	Mean	SD	P-value
TCP (%)	Planned	67.5	69.1	72.4	72.7	72.9	74.1	74.9	71.9	2.7	>0.05
	Measured	68.2	70.8	74.5	74.2	72.1	75.5	77.6	73.3	3.1	

TABLE 6 NTCPs from planned and measured data for H&N patient OARs and P-values of statistical test.

Structure	NTCP (%)	Patient No.							Average \pm SD	P-value
		1	2	3	4	5	6	7		
Brainstem	Planned	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5 \pm <0.5	>0.05
	Measured	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5 \pm <0.5	
Cord	Planned	1.6	1.1	0.7	<0.5	<0.5	<0.5	<0.5	0.5 \pm 0.6	>0.05
	Measured	1.7	1.3	0.8	1.1	<0.5	<0.5	<0.5	0.7 \pm 0.7	
Esophag.	Planned	0.7	1.5	<0.5	<0.5	<0.5	<0.5	<0.5	0.5 \pm 0.6	>0.05
	Measured	0.8	1.8	<0.5	<0.5	<0.5	<0.5	<0.5	0.5 \pm 0.6	
Larynx	Planned	2.0	3.8	37.1	0.9	<0.5	<0.5	14.7	8.4 \pm 13.6	<0.05
	Measured	2.4	4.4	44.4	1.2	<0.5	<0.5	15.8	9.8 \pm 16.2	
Lt Parotid	Planned	1.1	<0.5	<0.5	<0.5	1.2	<0.5	56.7	8.5 \pm 21.3	>0.05
	Measured	0.9	<0.5	<0.5	<0.5	1.4	<0.5	52.0	7.8 \pm 19.5	
Rt Parotid	Planned	9.2	<0.5	<0.5	<0.5	2.7	<0.5	<0.5	1.7 \pm 3.4	>0.05
	Measured	8.6	<0.5	<0.5	<0.5	3.1	<0.5	<0.5	2.5 \pm 4.0	

TABLE 7 NTCPs from planned and measured data for lung patient OARs and *P*-values of statistical test.

Structure	NTCP (%)	Patient No.							Average ± SD	P-value
		1	2	3	4	5	6	7		
Lungs-GTV	Planned	15.6	1.3	1.4	<0.5	4.9	6.7	<0.5	4.3 ± 5.6	<0.05
	Measured	20.0	1.5	2.0	<0.5	6.4	7.4	<0.5	5.3 ± 7.1	
Heart	Planned	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5 ± <0.5	<0.05
	Measured	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5 ± <0.5	
Cord	Planned	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5 ± <0.5	<0.05
	Measured	<0.5	<0.5	0.5	<0.5	0.5	<0.5	<0.5	2.3 ± 5.5	
Esophag	Planned	2.5	<0.5	<0.5	<0.5	28.4	1.3	<0.5	4.6 ± 10.5	>0.05
	Measured	4.9	<0.5	<0.5	<0.5	34.5	1.6	<0.5	5.9 ± 12.8	

Due to nonzero detecting threshold of detectors used in the measurement, per-beam analysis suppresses low dose for every beam data while with the cumulative analysis, low doses from multiple beams can add up to be detectable enough by the detectors. Also ArcCheck measures both entrance and exit doses essentially doubling the detector density while Mapcheck measures only the entrance dose. Sumida et al. had used voxel-based Niemierko's model to calculate TCPs and NTCPs but we have used Poisson model for TCP calculation and SDR model for NTCP calculation. The other difference is that the study is based on structure-by-structure gamma analysis but we have performed overall gamma analysis. However, the D_{50} and γ_{50} values used to calculate TCPs for H&N patients in our study match with the values used in the study.

The accuracy of ArcCheck measurement and PDP algorithm is out of scope of this study. Study by Sun Nuclear Corporation shows an excellent accuracy of PDP²⁵ and the algorithm has been applied in a number of independent studies^{8–12} before and validated against ion chamber, diode array, and gel dosimetry.

AAPM TG 119 recommends a passing criteria of 90% with 3%/3 mm DTA for per-beam analysis and 88%–90% for composite dose analyzed with radiographic films. The minimum pass rate in our study was 94.9% at 3%/3 mm DTA for composite analysis. In spite of 94.9% minimum pass rate, the differences in NTCPs based on planned and measured data for few of the structures studied were statistically significant. However, for the patients with passing rate of $\geq 90\%$ at 2%/2 mm, no statistically significant radiobiological difference was observed for any of the structures. As the number of patient plans passing by $\geq 90\%$ at 2%/2 mm is small, it may not be a wise idea to generalize these results and hence a study on a large number of patients is required to determine whether 90% gamma passing rate at 2%/2 mm is sufficient for radiobiological assurance. However, our results demonstrating a statistically significant radiobiological difference for the patient plans meeting $\geq 94.9\%$ pass rate at 3%/3 mm cohere with the inadequacy of 3%/3 mm gamma analysis criterion as discussed by Cadman,²⁶ demonstrated by Nelms et al.¹² and outlined by AAPM Medical Physics Practice Guideline 5.a.²⁷ Our future study will focus on finding the tolerance criteria for the radiobiological quality assurance.

Another point worth mentioning is that the dosimetric or biological pass rate was not favored by any of the IMRT or VMAT

modality. Although this finding is clear in our dataset, generalizing this argument is not in the scope of this study. The investigation of this topic would require a study of a different design, where other elements of the plans such as the level of beam modulation, etc., would be thoroughly studied.

5 | CONCLUSIONS

Differences between 2-D and 3-D gamma analysis results for H&N and lung patients are small and statistically insignificant. The differences between TCPs obtained from the planned and measured data are also small and insignificant. However, the differences in NTCPs based on planned and measured data for a few of the structures studied are statistically significant even though the dosimetric agreements are $\geq 94.9\%$ at 3%/3 mm DTA. Our study based on 14 patients suggests that $\leq 94.9\%$ pass rate at 3%/3 mm DTA used for 2-D or 3-D gamma analysis cannot assure the radiobiological equivalence between a delivered and the corresponding planned dose. Hence, radiobiological analysis in addition to dosimetric comparison may have to be considered for the QA of complex radiotherapy plans.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Ezzell GA, Burmeister JW, Dogan N, Losasso TJ, Mechalakos JG, Mihailidis D. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. *Med Phys*. 2009;36:5359–5373.
- Sadagopan R, Bencomo JA, Martin RL, Nilsson G, Matzen T, Balter PA. Characterization and clinical evaluation of a novel IMRT quality assurance system. *J Appl Clin Med Phys*. 2009;10:104–119.
- Spezi E, Angelini AL, Romani F, Ferri A. Characterization of a 2D ion chamber array for the verification of radiotherapy treatments. *Phys Med Biol*. 2005;50:3361–3373.
- Nelms BE, Zhen H, Tome WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys*. 2011;38:1037–1044.

5. 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;90:1195–1201.
6. Coleman L, Skourou C. Sensitivity of volumetric modulated arc therapy patient specific QA results to multileaf collimator errors and correlation to dose volume histogram based metrics. *Med Phys*. 2013;40:111715-1–111715-7.
7. Heilemann G, Poppe B, Laub W. On the sensitivity of the common gamma – Index evaluation methods to MLC misalignments in Rapidarc quality assurance. *Med Phys*. 2013;40:031702-1–031702-12.
8. Zhen H, Nelms BE, Tome WA. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Med Phys*. 2011;38:5477–5489.
9. 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;39:81–86.
10. Watanabe Y, Nakaguchi Y. 3D evaluation of 3DVH program using BANG3 polymer gel dosimeter. *Med Phys*. 2013;40:821011–821011.
11. Carrasco P, Jornet N, Latorre A, Eudaldo T, Ruiz A, Ribas M. 3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification. *Med Phys*. 2012;39:5040–5049.
12. 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;40:111722-1–111722-15.
13. Létourneau D, Publicover J, Kozelka J, Moseley DJ, Jaffray DA. Novel dosimetric phantom for quality assurance of volumetric modulated arc therapy. *Med Phys*. 2009;36:1813–1821.
14. Papanikolaou N, Battista JJ, Boyer AL, et al. *Tissue inhomogeneity corrections for megavoltage photon beams: report of Task Group No. 65 of the radiation therapy committee of the american association of physicists in medicine*. Medical Physics Publishing; 2004: 1–135. <https://aapm.org/pubs/reports/detail.asp?docid=86>.
15. Allen LX, Alber M, Deasy JO, et al. The use and QA of biologically related models for treatment planning. *Med Phys*. 2012;39:1386–1409.
16. Sumida I, Yamaguchi H, Kizaki H, et al. Novel radiobiological gamma index for evaluation of 3-dimensional predicted dose distribution. *Int J Radiat Oncol Biol Phys*. 2015;92:779–786.
17. Infusino E, Mameli A, Conti R, et al. Initial experience of ArcCheck and 3DVH software for rapidarc treatment plan verification. *Med Dosim*. 2014;39:276–281.
18. Kallman P, Agren A, Brame A. Tumour and normal tissue responses to fractionated non-uniform dose delivery. *Int J Radiat Biol*. 1992;62:249–262.
19. Warkentin B, Stavrev P, Stavreva N, Field C, Fallone BG. A TCP-NTCP estimation module using DVHs and known radiobiological models and parameter sets. *J Appl Clin Med Phys*. 2004;5:50–63.
20. Narayanasamy G, Pyakuryal AP, Pandit S, et al. Radiobiological evaluation of intensity modulated radiation therapy treatments of patients with head and neck cancer: a dual-institutional study. *J Med Phys*. 2015;40:165–169.
21. Okunieff P, Morgan D, Niemierko A, Suit HD. Radiation dose-response of human tumors. *Int J Radiat Oncol Biol Phys*. 1995;32:1227–1237.
22. Lyman JT. Complication probability as assessed from dose volume histograms. *Radiat Res*. 1985;104:S13–S19.
23. Luxton G, Keall PJ, King CR. A new formula for normal tissue complication probability (NTCP) as a function of equivalent uniform dose (EUD). *Phys Med Biol*. 2008;53:23–36.
24. Burman C, Kutcher GJ, Emami B, Goitein M. Fitting of normal tissue tolerance data to an analytic function. *Int J Radiat Oncol Biol Phys*. 1991;21:123–135.
25. Sun Nuclear Corporation. 3DVH: On the accuracy of the planned dose perturbation algorithm. Last accessed Nov 9, 2015: <http://www.sunnuclear.com/documents/1212W05062010.pdf>
26. Cadman PF. Comment on 'IMRT commissioning: some causes for concern. *Med Phys*. 2011;38:4464.
27. 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* 2016;17:14–34.