Technology in Cancer Research & Treatment
Volume 20: 1-13

DOI: 10.1177/15330338211036325
journals.sagepub.com/home/tct
SAGE

Clinically Oriented Target Contour Evaluation Using Geometric and Dosimetric Indices Based on Simple Geometric Transformations

Lixun Xian, MS[1,2,*], Guangjun Li, MS[1,*], Qing Xiao, MS[1], Zhibin Li, MS[1], Xiangbin Zhang, MS[1], Li Chen, MS[1], Zhenyao Hu, MS[1], and Sen Bai, PhD[1]

Abstract

Purpose: In radiotherapy, geometric indices are often used to evaluate the accuracy of contouring. However, the ability of geometric indices to identify the error of contouring results is limited primarily because they do not consider the clinical background. The purpose of this study is to investigate the relationship between geometric and clinical dosimetric indices. **Methods:** Four different types of targets were selected (C-shaped target, oropharyngeal cancer, metastatic spine cancer, and prostate cancer), and the translation, scaling, rotation, and sine function transformation were performed with the software Python to introduce systematic and random errors. The transformed contours were regarded as reference contours. Dosimetric indices were obtained from the original dose distribution of the radiotherapy plan. The correlations between geometric and dosimetric indices were quantified by linear regression. **Results:** The correlations between the geometric and dosimetric indices were inconsistent. For systematic errors, and with the exception of the sine function transformation (R^2: 0.023-0.04, $P > 0.05$), the geometric transformations of the C-shaped target were correlated with the D98% and D_{mean} (R^2: 0.689-0.988), 80% of which were $P < 0.001$. For the random errors, the correlations obtained by the all targets were $R^2 > 0.384$, $P < 0.05$. The Wilcoxon signed-rank test was used to compare the spatial direction resolution capability of geometric indices in different directions of the C-shaped target (with systematic errors), and the results showed only the volumetric geometric indices with $P < 0.05$. **Conclusions:** Clinically, an assessment of the contour accuracy of the region-of-interest is not feasible based on geometric indices alone. Dosimetric indices should be added to the evaluations of the accuracy of the delineation results, which can be helpful for explaining the clinical dose response relationship of delineation more comprehensively and accurately.

Keywords

contour evaluation, geometric indices, dosimetric indices, geometric transformation, target volume

Abbreviations

OARs, organs-at-risk; IMRT, intensity-modulated radiotherapy; VMAT, volumetric modulated arc radiotherapy; HD, the maximum Hausdorff distances; HD_{mean}, the mean Hausdorff distances; HD95, the 95% Hausdorff distances; DSC, Dice-similarity coefficient; RT, radiotherapy; TG-119 report, the American Association of Physicists in Medicine Task Group No. 119 report; TPS, the treatment planning system; DICOM, Digital Imaging and Communications in Medicine; CT, computer tomography; CI, conformity index; HI, homogeneity index; PTV, planning target volume; ICRU, the International Commission on Radiation Units and Measurements; SDC, Sørensen-Dice similarity coefficient.

Received: January 25, 2021; Revised: June 15, 2021; Accepted: July 08, 2021.

[1] Department of Radiation Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, Sichuan, China
[2] Department of Oncology, Chengdu Second People's Hospital, Chengdu, Sichuan, China
*Lixun Xian and Guangjun Li are contributed equally to this work.

Corresponding Author:
Sen Bai, PhD, Department of Radiation Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China.
Email: baisen@scu.edu.cn

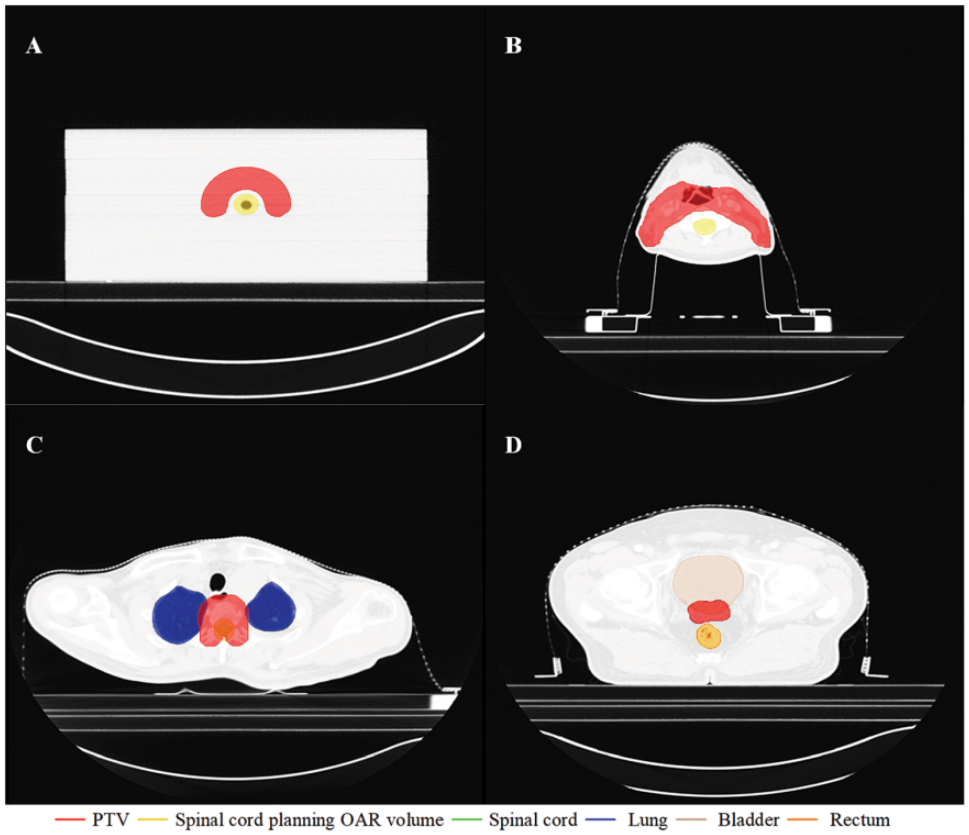

Figure 1. Example of the PTV contours of (A) the C-shaped target, (B) oropharyngeal cancer, (C) metastatic spine cancer, and (D) prostate cancer.

Introduction

Contouring of the target and organs-at-risk (OARs) is a key step in radiotherapy, especially with the highly modulated radiotherapy technologies currently in use, such as intensity-modulated radiotherapy (IMRT) and volumetric modulated arc radiotherapy (VMAT). Inaccurate contouring could cause serious systematic errors to the subsequent radiotherapy work. This error has always existed in the subsequent radiotherapy processes used for patients, such as radiotherapy treatment planning or patient positioning.[1-3] The commonly used slice-by-slice manual approach or interpolation-based semi-automatic contouring approach is time-consuming and labor-intensive, and the corresponding results are susceptible to differences between observers. In other words, the accuracy of contouring depends on the residents' clinical experience and the rational and efficient use of multimodal images, and improper selection of multimodal images may lead to differences in residents' delineation.[4-6] Another emerging contouring approach is the automatic segmentation technique. Compared with the aforementioned manual delineation method, its contouring speed is faster and it is not affected by the subjective factor, but it is necessary to evaluate the accuracy of the contouring results before its use in clinical applications.

Methods used for the assessment of the accuracy of contouring are generally classified in 2 categories, subjective and quantitative evaluations. The subjective evaluations are only based on the experiences and personal preferences of the evaluators. Evaluators are guided to turn off the original contour display and grade all research contours using 3 levels: useful as test contours (= 1), useful with minor edits (= 2), and not useful (= 3). The definition of minor edits is that the test contours would be acceptable after minor modifications.[7] This evaluation method is deeply affected by the individual differences among the evaluators and requires considerable time. At the same time, most of the contour accuracy studies are performed directly by using quantitative evaluations, which involves the employment of geometric indices to characterize the similarity between the test and the reference contours.[8] Geometric indices extensively used in contour evaluations include distance-type geometric indices (e.g., the maximum (HD), mean (HD_{mean}), and 95% Hausdorff distances (HD95)) and volumetric geometric indices (e.g., the Dice-similarity coefficient (DSC) and the Jaccard coefficient).[9] Recent research has suggested that despite the fact that these geometric indices can be easily calculated, they do not consider the clinical effect and may lack clinical relevance.[10-12] Furthermore, different geometric indices have different properties, but

Table 1. Linear Regression Analysis Between Geometric Indices and Dosimetric Endpoints of C-PTV With Systematic Errors.

		HD		HD_{mean}		HD95		DSC		Jaccard	
C-PTV	Transformation	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value
D98%	Right	0.931	< 0.001	0.942	< 0.001	0.936	< 0.001	0.936	< 0.001	0.894	< 0.001
	Anterior	0.49	0.017	0.528	0.011	0.502	0.015	0.511	0.013	0.436	0.027
	Posterior	0.972	< 0.001	0.985	< 0.001	0.982	< 0.001	0.981	< 0.001	0.956	< 0.001
	Expansion	0.966	< 0.001	0.976	< 0.001	0.958	< 0.001	0.962	< 0.001	0.937	< 0.001
	Reduction	0.84	< 0.0017	0.724	0.001	0.788	< 0.001	0.689	0.002	0.754	0.001
	Rotation	0.949	< 0.001	0.949	< 0.001	0.944	< 0.001	0.943	< 0.001	0.912	< 0.001
	Sine	0.023	0.778	0.03	0.828	0.027	0.889	0.04	0.859	0.04	0.861
D_{mean}	Right	0.917	< 0.001	0.927	< 0.001	0.923	< 0.001	0.919	< 0.001	0.868	< 0.001
	Anterior	0.651	0.003	0.692	0.002	0.672	0.002	0.675	0.002	0.602	0.005
	Posterior	0.959	< 0.001	0.973	< 0.001	0.964	< 0.001	0.967	< 0.001	0.932	< 0.001
	Expansion	0.971	< 0.001	0.988	< 0.001	0.972	< 0.001	0.977	< 0.001	0.956	< 0.001
	Reduction	0.987	< 0.001	0.94	< 0.001	0.965	< 0.001	0.92	< 0.001	0.955	< 0.001
	Rotation	0.927	<0.001	0.929	<0.001	0.922	<0.001	0.921	<0.001	0.881	<0.001
	Sine	0.024	0.779	0.035	0.72	0.032	0.954	0.027	0.808	0.036	0.708

Table 2. Linear Regression Analysis Between Geometric Indices and D98% of Targets With Random Errors.

	HD		HD_{mean}		HD95		DSC		Jaccard	
Site	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value
C-PTV	0.578	< 0.001	0.876	< 0.001	0.822	< 0.001	0.877	< 0.001	0.836	< 0.001
Oropharyngeal cancer	0.478	0.001	0.813	< 0.001	0.788	< 0.001	0.819	< 0.001	0.759	< 0.001
Metastatic spine cancer	0.384	0.003	0.95	< 0.001	0.87	< 0.001	0.95	< 0.001	0.907	< 0.001
Prostate cancer	0.549	< 0.001	0.754	< 0.001	0.752	< 0.001	0.75	< 0.001	0.725	< 0.001

different automatic segmentation studies were not based on a uniform guide or rules to select geometric indices to evaluate the contour results.[13-15] Subject to the assumption of a reference contour, the method used for the clinical assessment of the accuracy of radiotherapy (RT) contours involves the determination and prediction of the deviation of its dosimetric indices based on the dose distribution of the radiation treatment plan.[10,16-18] However, the relationships between geometric and dosimetric indices still need to be studied in more depth.

This study introduced the systematic and random contour errors based on the following geometric transformations: translation, scaling, rotation, and sine function transformation. Based on these transformations, the specific objectives of this study was a) to investigate the correlations between the geometric and the dosimetric indices, and b) to explore the ability of geometric indices to distinguish the contours with the same transformation type but in different directions.

Materials and Methods

Contouring

Four different types of targets were selected for this study: C-shaped target, oropharyngeal cancer, metastatic spine cancer, and prostate cancer (Figure 1). The C-shaped target was delineated on the water phantom according to the American Association of Physicists in Medicine (AAPM) Task Group No. 119 (TG-119) report,[19] and the remaining 3 types of targets were outlined by senior physicians in our research institution according to institutional clinical protocols. The structures of the targets were exported from the treatment planning system (TPS) Raystation (Raysearch, Stockholm, Sweden) in the form of a Digital Imaging and Communications in Medicine (DICOM) file, and the position information of the contours were read by an in-house software developed in Python (version 3.7.3). The information was used to perform the geometric transformations. Subsequently, the transformed structures were imported back to Raystation system in the form of a DICOM file.

The targets (original contours) before the transformation were regarded as the delineation results by junior residents (test contours), and the transformed targets were regarded as reference contours after systematic and random errors corrections implemented by senior physicians. The contour errors were introduced in the form of geometric transformations. Through the following 7 geometric transformations, systematic errors were only introduced to the C-shaped target. The translational transformations were divided into the following 3 cases: right, anterior, and posterior directions. Based on the location of the original contours at 1 mm intervals, the contours were moved 10 times in each of the right, anterior, and posterior directions to obtain the reference contours. Scaling transformation represented an equidistant expansion or reduction transformation in

Table 3. Linear Regression Analysis Between Geometric Indices and D_{mean} of Targets With Random Errors.

	HD		HD_{mean}		HD95		DSC		Jaccard	
Site	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value
C-PTV	0.555	< 0.001	0.872	< 0.001	0.744	< 0.001	0.875	< 0.001	0.823	< 0.001
Oropharyngeal cancer	0.48	< 0.001	0.788	< 0.001	0.759	< 0.001	0.797	< 0.001	0.734	< 0.001
Metastatic spine cancer	0.418	0.002	0.957	< 0.001	0.877	< 0.001	0.95	< 0.001	0.907	< 0.001
Prostate cancer	0.522	< 0.001	0.794	< 0.001	0.705	< 0.001	0.781	< 0.001	0.744	< 0.001

Table 4. The Wilcoxon Signed Ranks Test Analysis Between the Geometric Indices of Different Directions of Transformations.

Geometric transformation	Parameter	HD	HD_{mean}	HD95	DSC	Jaccard
Right and anterior direction	*Z*-value	−1	−0.475	−0.73	−2.825	−2.848
	P-value	0.317	0.631	0.465	0.005[a]	0.004[a]
Right and posterior direction	*Z*-value	−0.535	−0.547	−1.604	−2.823	−2.842
	P-value	0.593	0.563	0.109	0.005[a]	0.004[a]
Anterior and posterior direction	*Z*-value	−1.604	−0.47	−1.604	−2.598	−2.121
	P-value	0.109	0.572	0.109	0.009[a]	0.034[a]
Scaling transformation	*Z*-value	−1.586	−1.790	−1.792	−2.505	−2.533
	P-value	0.113	0.074	0.069	0.012[a]	0.011[a]

[a]*P*-value = the difference was statistically significant.

reference to the position of the original contour. Considering the fast speed of scaling transformation changes, in the patient modeling module of the Raystation planning system, 10 equidistant transformations were performed at 0.5 mm intervals, excluding the anterior and posterior directions. The rotation transformation was based on the use of the origin of the computer tomography (CT) image coordinates as the rotation center points (at 1° intervals) for 10 clockwise rotations. For the sine function transformation, we extracted the coordinate values (x_0, y_0) of the original contour first, and then used the function $y = \sin\omega y_0$ ($\omega = 3, 4, 5, 6, \ldots, 12$) to conduct periodic transformations 10 times with a fixed amplitude.

In addition, to introduce random errors, we performed geometric transformations on each target 20 times, and 3 of the above geometric transformation methods were used in the same target randomly for each transformation. To be specific, we randomly selected a quarter of all the CT image layers to remain the same without applying any geometric transformation, and the remaining layers were randomly divided into 3 parts. A geometric transformation method was randomly selected and used for each part and the transformation amplitude was randomly selected within the aforementioned (corresponding) amplitude range. In the end, 20 delineation results were obtained for each of these 4 targets.

Geometric Indices

In this study, we chose 5 extensively used geometric indices for the evaluations: 3 distance-type indices HD (maximum, mean, 95%) and 2 volumetric indices (DSC and Jaccard). These 5 geometric indices were calculated by the 3DSlicer (version 4.10.2),[20] which is an open-source software. The calculation of HD was performed on the exported DICOM-RT files. The HD indices calculated by the 3DSlicer represent bidirectional distances, and the bidirectional distance was symmetrical; this type of distance is more stable than the unidirectional distance calculated by other methods.

Dosimetric Indices

The original clinical plans for these 4 targets used IMRT technology. The C-shaped target met the requirements for a simple version in the TG-119 report, the dose of 5000 c Gy received by 90% of the target volume was used as the prescription, and the dose prescriptions for oropharyngeal, metastatic spine, and prostate cancers were 95% of the target volume, which received 5400, 3000, and 5600 c Gy, respectively, and the dose grid was 2 mm. After the geometric transformation, the RT structures were imported into the radiotherapy plan of the original clinical plans, and D98%, D_{mean}, D2%, conformity index (CI), and homogeneity index (HI) of the planning target volume (PTV) were obtained. According to the International Commission on Radiation Units and Measurements (ICRU) 83 report,[21] these dosimetric endpoints (D98%, D_{mean}, and D2%) represent the minimum, mean, and maximum doses received by the target. The CI value of target volume is defined according to the following equation,[12,21]

$$CI = \frac{\mathrm{TVPTV}^2}{\mathrm{TV} * \mathrm{PTV}}. \quad (1)$$

where TV is the volume of prescribed isodose line enclosed volume, PTV is the volume of targets, and TVPTV represents the overlap volume between the target volume and the prescribed isodose line enclosed volume.

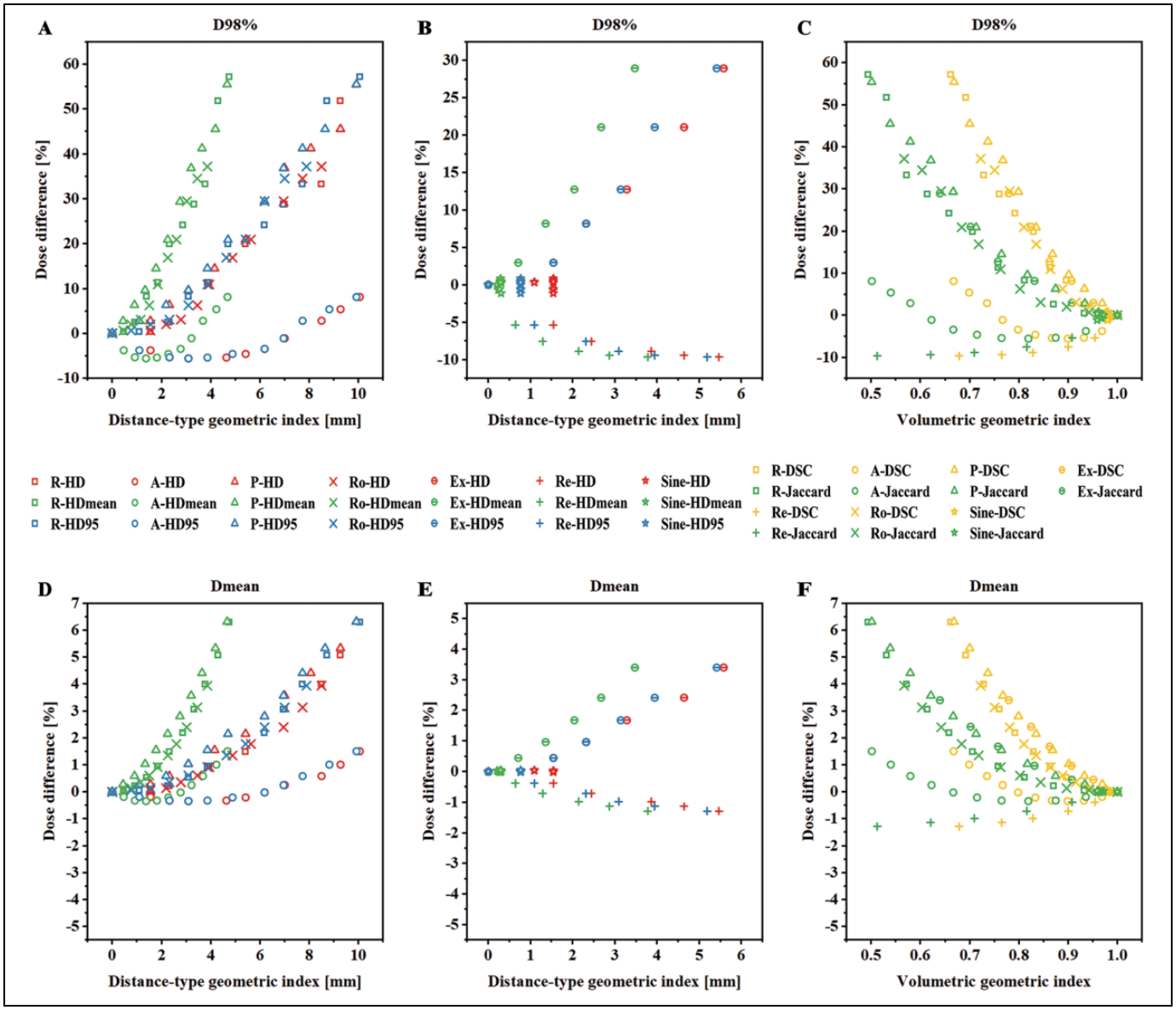

Figure 2. Relationship between the geometric indices of C-PTV and the dose difference. (A-C) and (D-F) show the relationships between the geometric indices of C-PTV after the introduction of systematic errors and the dose difference D98% and D_{mean}, respectively. (R-HD = dose difference corresponding to HD value after translation transformation in the right direction; R-HD_{mean} = dose difference corresponding to HD_{mean} value after translation transformation in the right direction; R-HD95 = dose difference corresponding to HD95 value after translation transformation in the right direction; R-DSC = dose difference corresponding to DSC value after translation transformation in the right direction; R-Jaccard = dose difference corresponding to Jaccard value after translation transformation in the right direction; A = anterior direction translation transformation; P = posterior direction translation transformation; Ex = expansion transformation; Re = reduction transformation; Ro = rotation transformation; Sine = sine function transformation).

The formula of HI suggested by the ICRU 83 report is the following,

$$HI = \frac{D2\% - D98\%}{D50\%}. \quad (2)$$

In this study, the dose differences (ΔD) of 3 dosimetric indices D98%, D_{mean} and D2% were calculated and normalized according to their respective clinical goals.[14] Herein, $\Delta D = (D_{x,reference} - D_{x,test})/D_{x,reference}$, where x represents the type of dosimetric index. Given that the clinical prescription requirements of the 4 targets were different, they were normalized with their respective prescription doses.

Analysis

Linear regression analysis was conducted with the software SPSS 21.0 (SPSS, Chicago, IL, USA) and involved the calculation of the coefficient of determination (R^2, where R denotes the correlation coefficient).[13] The R^2 statistic was used to quantify the correlations between the geometric indices (HD

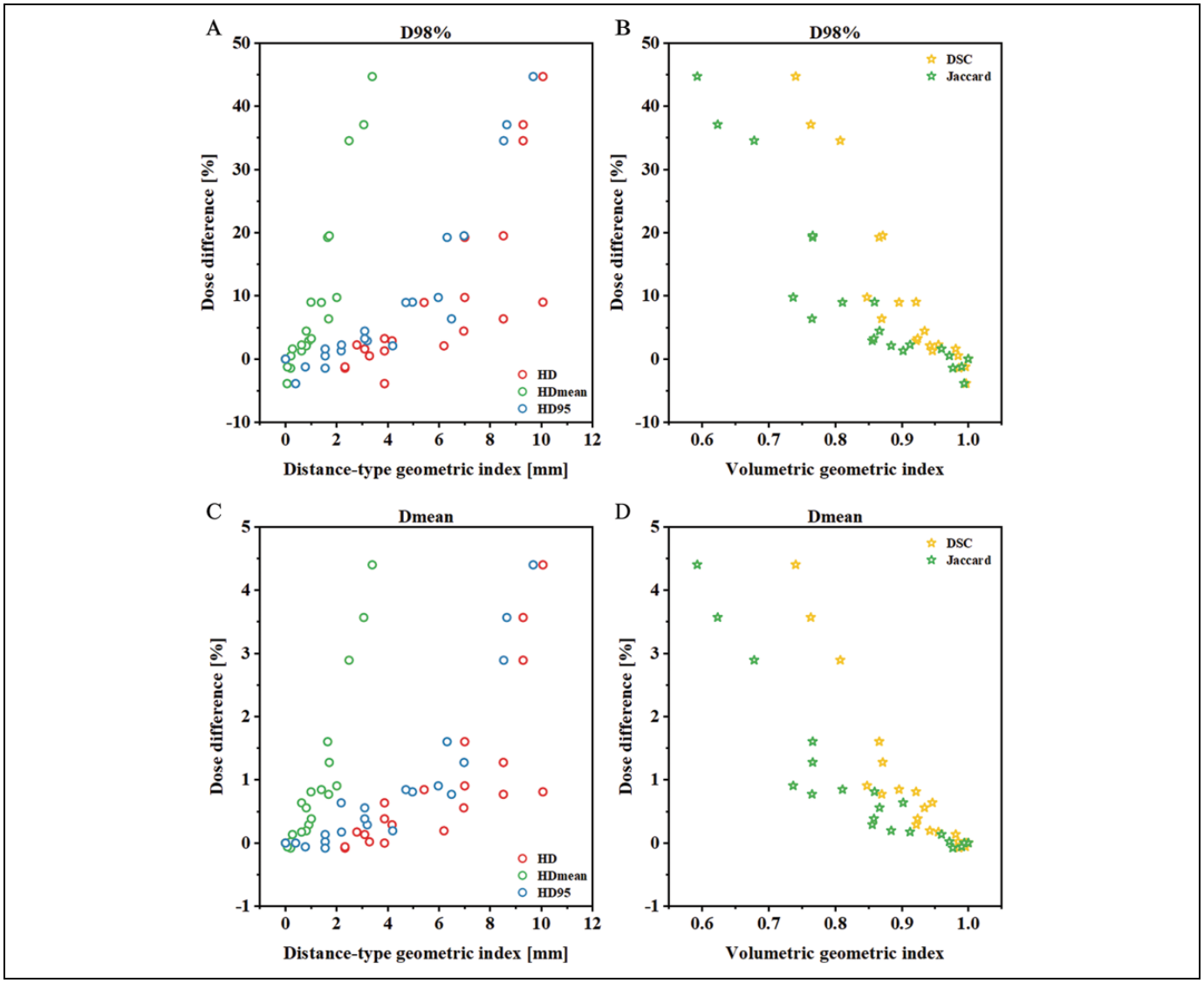

Figure 3. Relationship between the geometric indices of C-PTV and the dose difference. (A-B) and (C-D) are the relationships between the geometric indices of C-PTV after the introduction of random errors and the dose difference D98% and D_{mean}, respectively.

(maximum, mean, 95%), DSC, and Jaccard) and the dosimetric endpoints (D98%, D_{mean}, and D2%). *P*-values < 0.05 were considered significant. In addition, the geometric indices obtained from the equidistant scaling transformations and the right, anterior, and posterior directions of the C-shaped PTV translation transformations were compared, and the difference between them was tested for statistical significance using the Wilcoxon signed-ranks test in SPSS. From the scatterplots of the geometric versus the dosimetric indices, the feasibility of the assessment of the accuracy of contours with geometric indices was analyzed.

Results

Linear regression analysis was conducted on the geometric indices and dosimetric endpoints (Tables 1-3). Table 1 shows the correlations between the geometric indices and dosimetric endpoints of C-PTV with systematic errors. With the exception of the sine function transformation (R^2: 0.023-0.04, $P > 0.05$), the geometric transformations of the C-PTV had links with the dosimetric indices D98% and D_{mean} (R^2: 0.689-0.988), 80% of which were $P < 0.001$. Tables 2 and 3 list the correlations between geometric indices and dosimetric endpoints of actual cases with random errors. The geometric indices were correlated with the dosimetric indices D98% and D_{mean}, and the correlations of HD were weaker than the other 4 geometric indices. The D2% values of the targets were not included in the correlation analysis because of their small variation range.

Table 4 shows the results of the Wilcoxon signed-ranks test analysis between the geometric indices of translation transformation in the right, anterior, and posterior directions, and equidistant scaling transformation in the opposite direction. For the analysis results of the 5 directions of C-PTV, the *P*-values of HD, HD_{mean} and HD95, were all greater than 0.05, while the

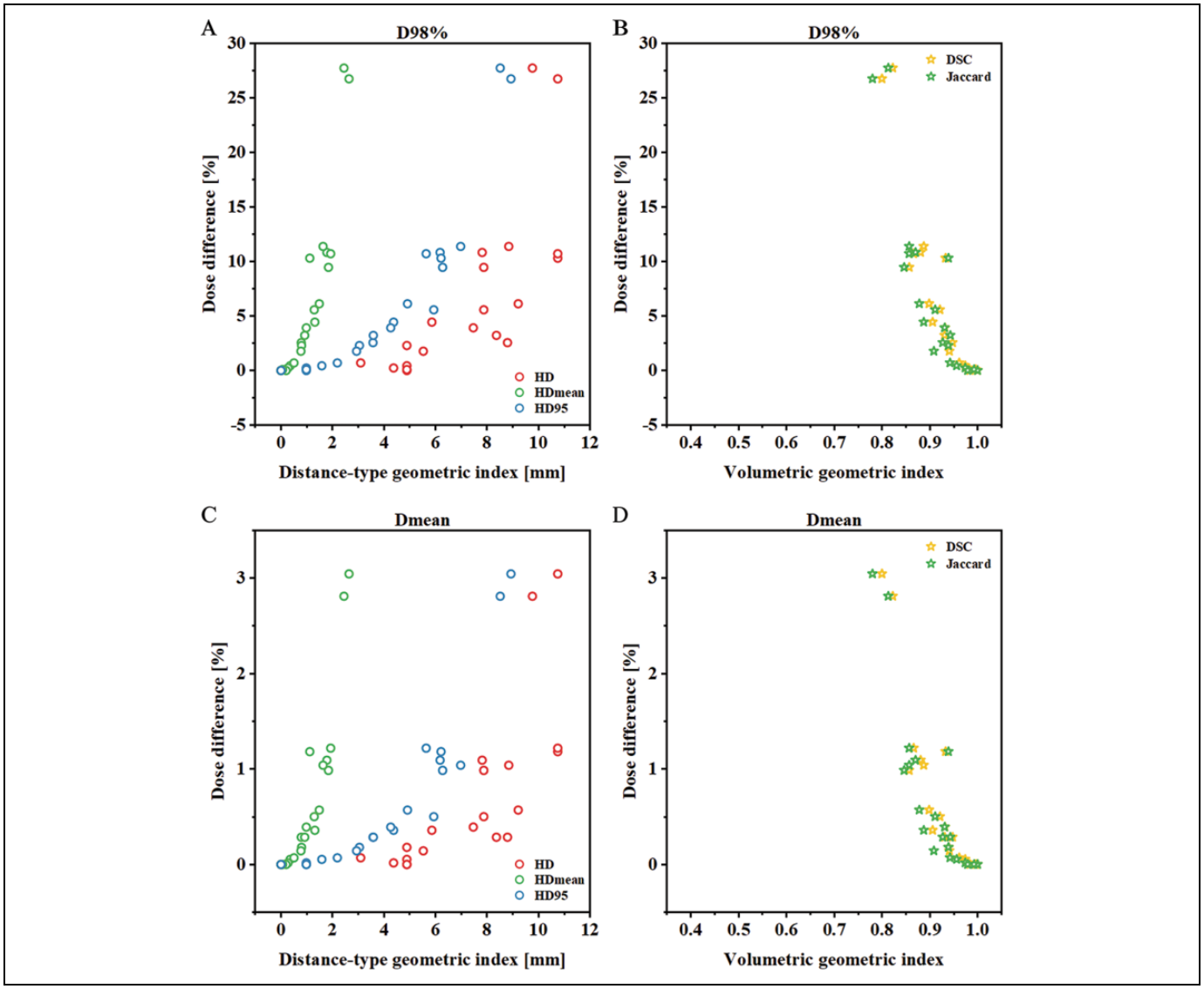

Figure 4. Relationship between the geometric indices of oropharyngeal cancer and dose difference. (A-B) and (C-D) are the relationships between the geometric indices of oropharyngeal cancer after the introduction of random errors and the dose difference D98% and D_{mean}, respectively.

P-values for DSC and Jaccard were all less than 0.05, and the differences were statistically significant.

Figures 2-6 show the relationships between the geometric indices of the targets and the dose differences (98%, $_{mean}$). Figures 7 and 8 show the relationship between the geometric indices and CI, HI. For the targets with random or systematic errors, the relationships were inconsistent in different geometric indices. In addition, for the water phantom target (C-PTV) with systematic errors, the relationships were nonmonotonic.

Discussion

Many studies have shown that although it is important to quantify the degree of variation or uncertainty of the contouring, it is more important to determine the dose difference and clinical impact.[10,11,15,17,18,22,23] In an earlier work, Ward van Rooij *et al*[18] studied the accuracy of automatic delineation of organs at risk in the head and neck regions based on deep learning techniques, while the use of geometric and dosimetric indices allowed them to analyze the correlation between the geometric index SDC (Sørensen-Dice similarity coefficient) and dose difference. They found that there was a weak correlation between the SDC and ΔD for all of the OARs based on automatic segmentation, $r = -0.24$, $P = 0.002$, but the correlation was not specific to a certain OAR or a certain patient. This is partly similar with the results described in our study. Ward van Rooij *et al* evaluated the OARs, and our study evaluated the results of the targets. Dosimetrically, this is a major difference. For the targets, high doses need to be achieved, while for the OARs,

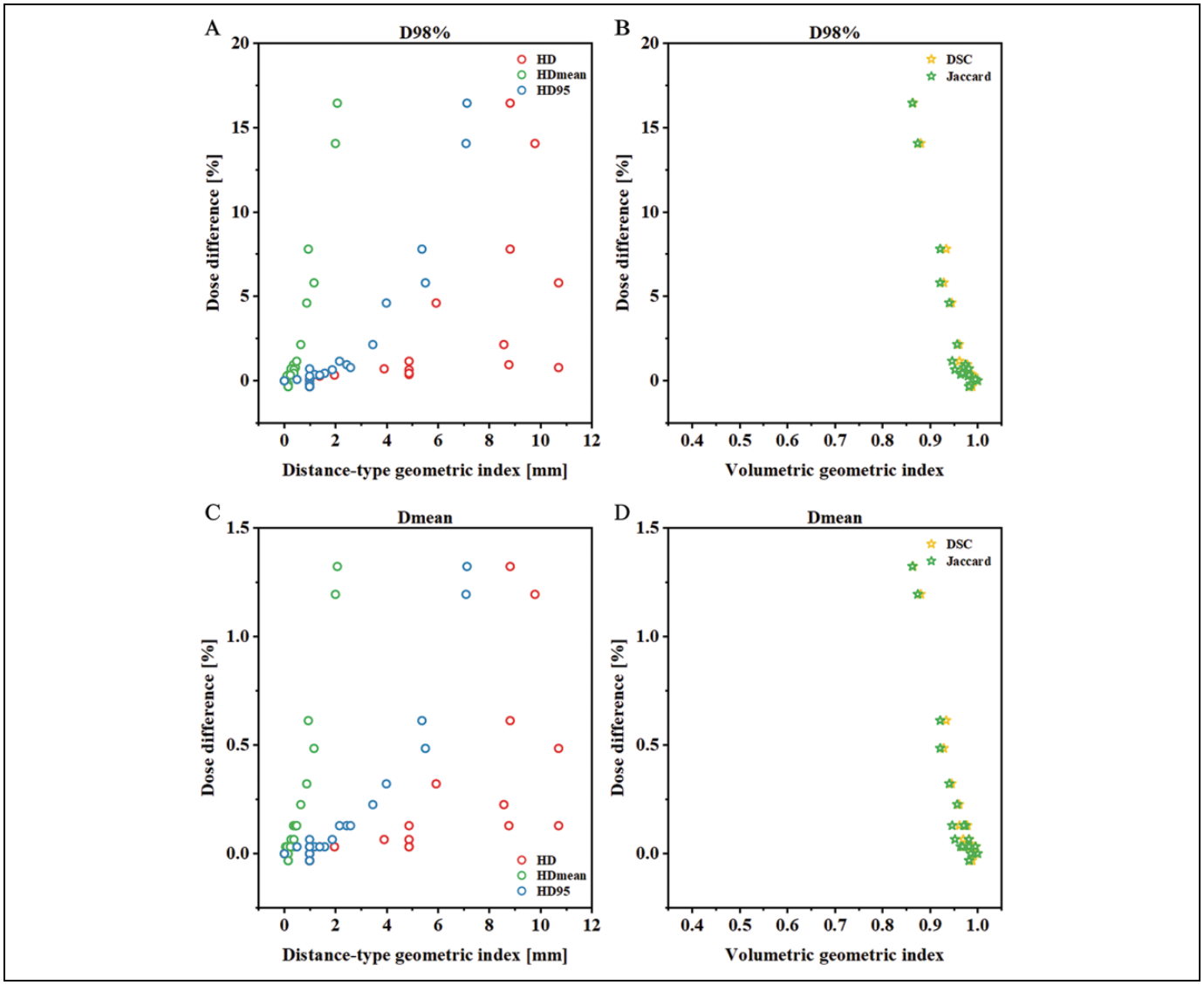

Figure 5. Relationship between the geometric indices of metastatic spine cancer and the dose difference. (A-B) and (C-D) are the relationships between the geometric indices of metastatic spine cancer after the introduction of random errors and the dose difference D98% and D_{mean}, respectively.

high doses need to be avoided. Given that all the beams are directed toward the targets, these volumes are much more sensitive to dosimetric changes in cases where there are contour errors; this effect is thus considerably less prominent in OARs. We found that the geometric indices obtained by geometric transformation were correlated with the dosimetric indices, but for some specific geometric transformation forms, the situations were different, and the correlation was not consistent for the different forms of the geometric transformations. A correlation existed between the geometric indices and dosimetric endpoints in the translation, scaling, and rotation transformations of the C-PTV with systematic errors, but the results for the sine function transformation were not significant or weak. Following the periodic transformation of sine function with fixed amplitude, the C-PTV's contour changed very little, and the HD values were less than 2 mm (Figure 2B). These outcomes led to small-dose differences, and correspondingly, the correlation between them was weak. At the same time, this re-emphasizes the importance of contour training for junior residents, whereby the repeatability of contouring is high, the contour difference is small, and the dose difference is also small.

The correlation obtained by the C-PTV anterior direction translation was lower than the correlation obtained by the other 2 translation transformations. To avoid high-dose radiation to the surrounding OARs (Figure 1A), physicians try to maintain high-dose areas away from the OARs when designing the original radiotherapy plan. The structures in the low-dose region are less likely to have noticeable dose changes even if its contour varies greatly in the spatial domain.[24] When the translation occurred in this area along the anterior direction, the minimum dose (D98%) of the target in this area was almost unchanged.

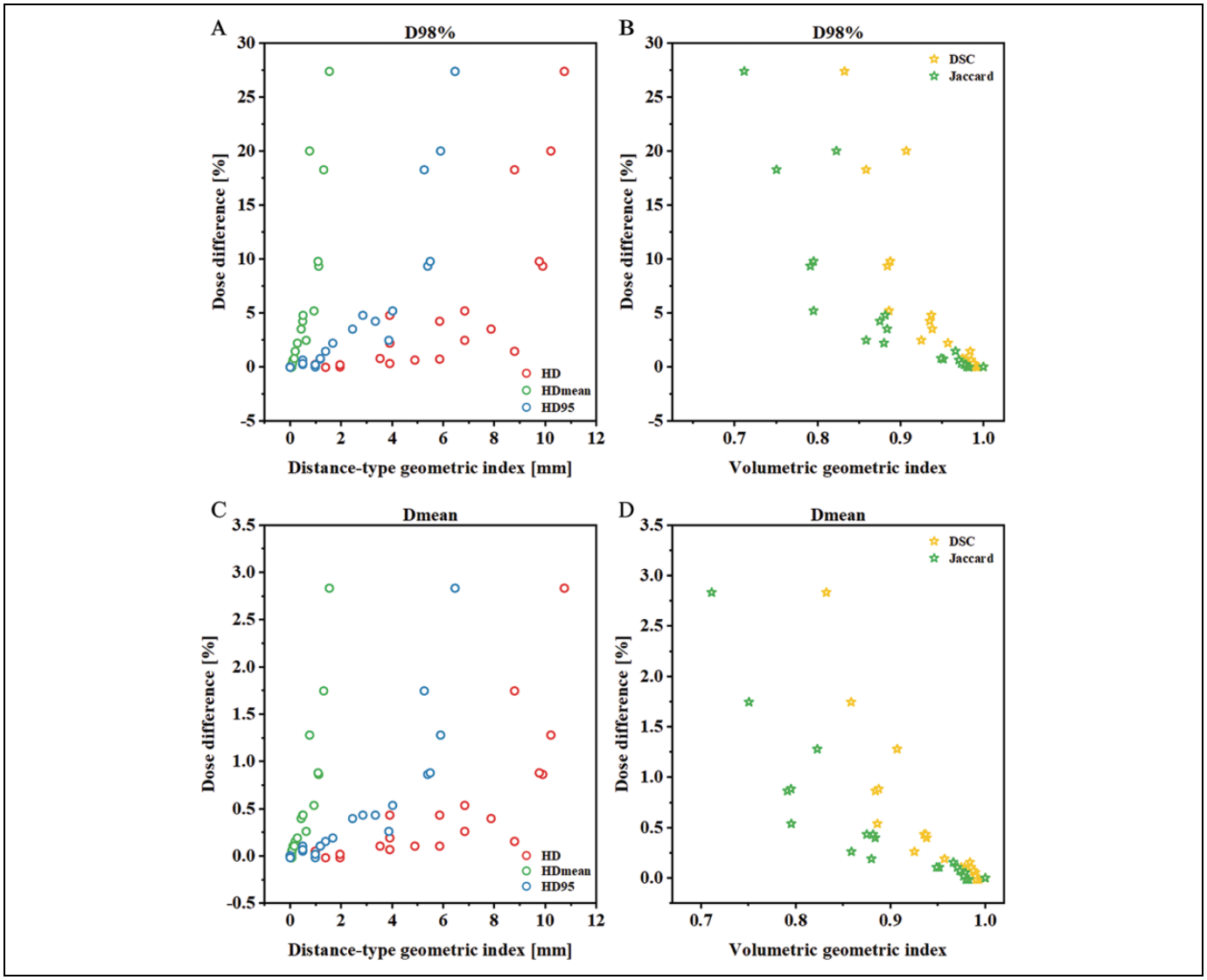

Figure 6. Relationship between the geometric indices of prostate cancer and the dose difference. (A-B) and (C-D) are the relationships between the geometric indices of prostate cancer after the introduction of random errors and the dose difference D98% and D_{mean}, respectively.

This resulted in a weaker correlation, and the correlation of D_{mean} was higher than D98%. In addition, as it can be observed in Figure 7, the HIs are almost unchanged during the anterior direction translation of the C-PTV; additionally, the relationships between the geometric indices and CIs, and the HIs in the cases of other transformation methods are also different. This is consistent with the study by Lim *et al*,[10] which found that the correlation between geometric indices and dosimetric indices was affected by the goals of the treatment plan. From these studies, it can be shown that the correlation between geometric and dosimetric indices can be affected by many factors, such as the geometric transformation method, the relative positions of the target and OARs, and the constraint goals of the radiotherapy plan.

According to the analysis outcomes of the Wilcoxon signed-ranks in Table 4, for the geometric transformation results of the C-PTV, the distance-type geometric indices HD, HD_{mean} and HD95 cannot express the difference of the translation transformations in the right, anterior, and posterior directions, and the difference at different equidistant scalings. However, there were significant differences between the volumetric geometric indices obtained from the different transformation directions in irregular shape targets. The HD values of the equidistant scaling transformation were the same for C-PTV, but the clinical effect on them were different. As shown in Figure 2B, when HD = 1.547 mm, the dose difference was within 5% for equidistant expansion transformation, while for equidistant reduction transformation it was beyond −5%. The geometric indices obtained from the same type of geometric transformation in different directions were not distinguishable. For a target contour, different types of geometric indices have the same value, and the corresponding dose differences are different. Beasley *et al*[25] reported that when measured with a suitable spatial metric, the higher the geometric

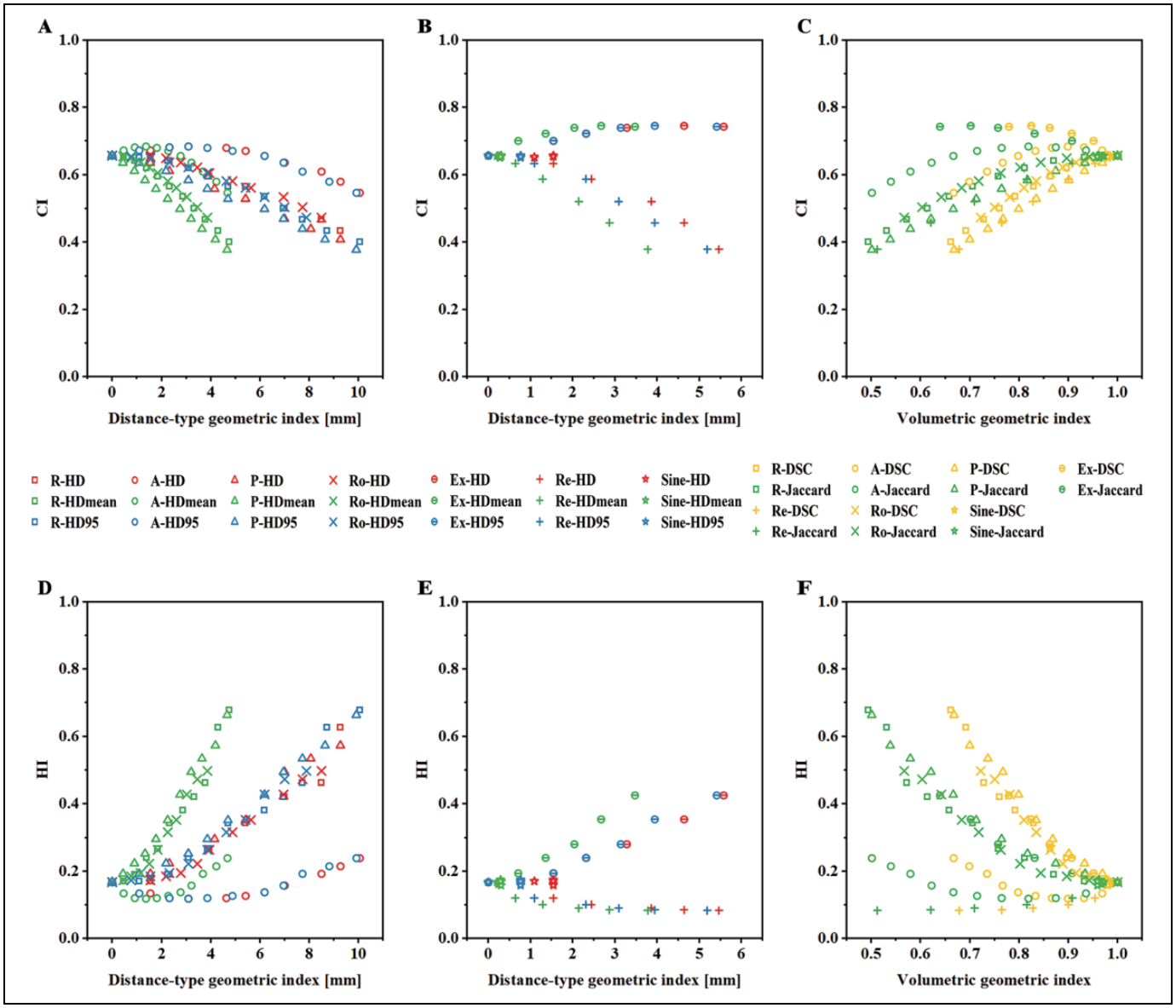

Figure 7. Relationships between the CI (A-C), HI (D-F) and geometric indices of C-PTV after the introduction of systematic errors.

accuracy of the contour is, the smaller the dose difference should be, and vice versa. When the distance-type geometric index value was approximately 3 mm, the dose difference corresponding to HD_{mean} was already close to 30% (see Figures 2–6). That is why this study asserts that HD_{mean} is not suitable as an index for evaluating contour accuracy alone despite the highest correlation coefficient of HD_{mean}. Compared with HD and HD95, it is too sensitive and the change gradient is too large to reflect the actual clinical situation. Given that there is no independent reference standard for distance-type geometric index values, it is impossible to compare the contour accuracy across different structures. This implies that although 2 different structures have the same geometric index value, they cannot indicate whether the quality of the contour is the same. As shown in Figures 2–8, the quality of the contour is controversial for the 2 structures (such as oropharyngeal and prostate cancers) with the same geometric index value. Additionally, it is difficult to compare the contour accuracy among different automatic segmentation studies by only using the distance-type geometric index. The article on the study of automatic segmentation of OARs showed that the HD (Liver) $= 15.770 \pm 1.0$ mm had a high accuracy,[26] while another article indicated that an HD $= 37.7 \pm 13.8$ mm also yielded a high accuracy.[8] Although the former value is smaller than the latter, the identification of the value with the higher accuracy cannot be clearly stated owing to the lack of a standard reference.

For volumetric geometric indices, many studies indicated that if the DSC was higher than the normally reported value of 0.7, the agreement between the reference contour and the test contour was considered to be good.[23,26,27] Our research showed that when the DSC and Jaccard values of the anterior direction

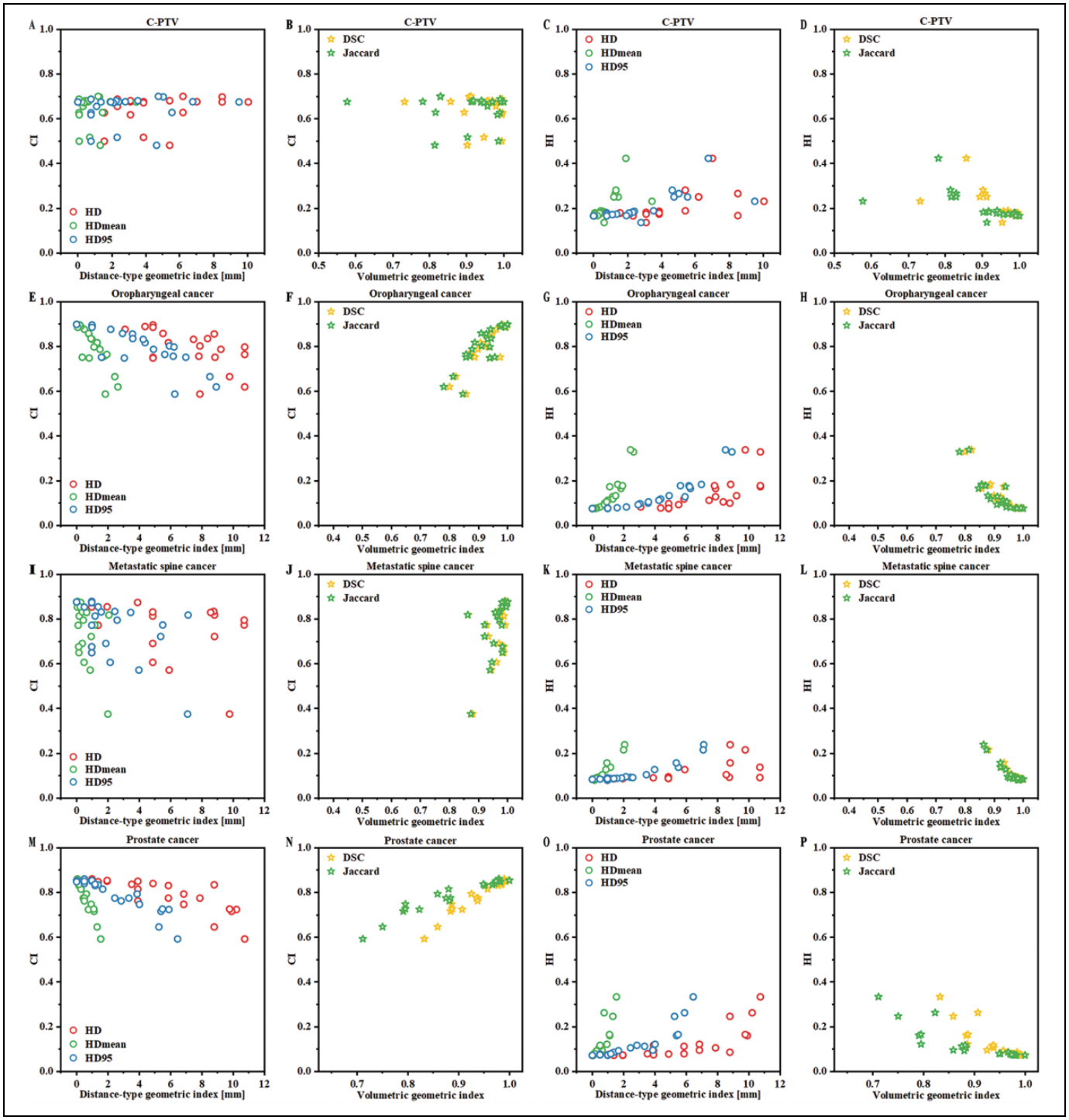

Figure 8. Relationships between the CI, HI, and the geometric indices, after the introduction of systematic errors. (A-D) C-PTV, (E-H) oropharyngeal cancer, (I-L) metastatic spine cancer, and (M-P) prostate cancer.

translation transformation were between 0.5 and 0.7, the corresponding dose differences were also very small (Figure 2C). At the same time, there were some cases in which the DSC values were greater than 0.7, and the corresponding dose difference values were large. These 2 contradictory situations show that it is not reliable to set an acceptable threshold for DSC.

This study introduced the systematic and random geometric errors through translation, scaling, rotation, and sine function geometric transformations; analyzed the feasibility of the clinical evaluation of geometric indices, and determined the ability of geometric indices to identify the direction of transformation. Although geometric indices reflected the geometric differences

between test and reference contours, and the correlation between the geometric index and the dosimetric endpoints in this study were relatively high, the relationship between geometric and dosimetric indices was not consistent among different geometric indices, different transformation forms, and different targets. Thus, it was illogical to use only geometric indices to evaluate the clinical acceptability of contour results. In addition, our current research was based on the simulation experiment of geometric transformation. Accordingly, we should explore further the relationship between geometric and dosimetric indices using the actual contouring results cases of the junior residents.

Conclusion

At present, there is a lack of guidance for the evaluation of contours using geometric indices. Therefore, there is a need for a normative framework. We found that the differences between the geometric and dosimetric indices were not consistent. This justifies the inaccuracy arising in instances where only the geometric indices are used to evaluate the results of contouring. The clinical acceptability of contouring results cannot be judged by geometric indices alone. Therefore, we suggest that dosimetric indices should be added to the evaluations of the accuracy of the results of automatic segmentation or contouring training of the residents, which can be helpful in explaining the clinical dose–response relationship of delineation more comprehensively and accurately.

Authors' Note

Lixun Xian and Guangjun Li are contributed equally to this work. This study was approved by the ethics committee of West China Hospital (No.2021017). The consent of the patients has been waived by the ethics committee because this is a retrospective study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (grant numbers 81472807 and 81972848), and the Sichuan Science and Technology Program (grant number 2021YFS0143).

ORCID iD

Guangjun Li, MS https://orcid.org/0000-0003-2054-1771

References

1. Nelms BE, Tomé WA, Robinson G, et al. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. *Int J Radiat Oncol Biol Phys*. 2012;82(1):368-378. doi:10.1016/j.ijrobp.2010.10.019
2. Brouwer CL, Steenbakkers RJHM, Heuvel EVD, et al. 3D Variation in delineation of head and neck organs at risk. *Radiat Oncol*. 2012;7(1):32. doi:10.1186/1748-717X-7-32
3. Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol*. 2015; 117(1):83-90. doi:10.1016/j.radonc.2015.07.041
4. Khoo VS, Adams EJ, Saran F, et al. A comparison of clinical target volumes determined by CT and MRI for the radiotherapy planning of base of skull meningiomas. *Int J Radiat Oncol Biol Phys*. 2000; 46(5):1309-1317. doi:10.1016/s0360-3016(99)00541-6
5. La Macchia M, Fellin F, Amichetti M, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiat Oncol (London, England)*. 2012;7(1):160. doi:10.1186/1748-717X-7-160
6. Lim JY, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. *Acta Oncol*. 2016;55(7):799-806. doi:10.3109/0284186X.2016.1173723
7. Zhu MY, Bzdusek K, Brink C, et al. Multi-institutional quantitative evaluation and clinical validation of smart probabilistic image contouring engine (Spice) autosegmentation of target structures and normal tissues on computer tomography images in the head and neck, thorax, liver, and male pelvis areas. *Int J Radiat Oncol Biol Phys*. 2013;87(4):809-816. doi:10.1016/j.ijrobp.2013.08.007
8. Wittenstein O, Hiepe P, Sowa LH, Karsten E, Fandrich I, Dunst J. Automatic image segmentation based on synthetic tissue model for delineating organs at risk in spinal metastasis treatment planning. *Strahlenther Onkol*. 2019;195(12):1094-1103. doi:10.1007/s00066-019-01463-4
9. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imag*. 2015;15(1):29. doi:10.1186/s12880-015-0068-x
10. Lim TY, Gillespie E, Murphy J, Moore KL. Clinically oriented contour evaluation using dosimetric indices generated from automated knowledge-based planning. *Int J Radiat Oncol Biol Phys*. 2019;103(5):251-1260. doi:10.1016/j.ijrobp.2018.11.048
11. Kaderka R, Gillespie EF, Mundt RC, et al. Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients. *Radiother Oncol*. 2019;131: 215-220. doi:10.1016/j.radonc.2018.07.013
12. Zhu J, Chen X, Yang B, et al. Evaluation of automatic segmentation model with dosimetric metrics for radiotherapy of esophageal cancer. *Front Oncol*. 2020;10:564737. doi:10.3389/fonc.2020.564737
13. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. *J Med Imaging Radiat Oncol*. 2010;54(5):401-410. doi:10.1111/j.1754-9485.2010.02192.x
14. Hanna GG, Hounsell AR, O Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. *Clin Oncol*. 2010;22(7):515-525. doi:10.1016/j.clon.2010.05.006
15. Vinod S K, Jameson M G, Min M, et al. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. *Radiother Oncol*. 2016; 121(2):169-179. doi:10.1016/j.radonc.2016.09.009
16. Tsuji SY, Hwang A, Weinberg V, Yom SS, Quivey JM, Xia P. Dosimetric evaluation of automatic segmentation for adaptive

IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2010;77(3):707-714. doi:10.1016/j.ijrobp.2009.06.012

17. Kieselmann JP, Kamerling CP, Burgos N, et al. Geometric and dosimetric evaluations of atlas-based segmentation methods of MR images in the head and neck region. *Phys Med Biol*. 2018; 63(14):145007. doi:10.1088/1361-6560/aacb65
18. van Rooij W, Dahele M, Ribeiro Brandao H, et al. Deep learning-based delineation of head and neck organs at risk: geometric and dosimetric evaluation. *Int J Radiat Oncol Biol Phys*. 2019;104(3): 677-684. doi:10.1016/j.ijrobp.2019.02.040
19. Ezzell GA, Burmeister JW, Dogan N, et al. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM task group 119. *Med Phys*. 2009;36(11): 5359-5373. doi:10.1118/1.3238104
20. Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: radiation therapy research toolkit for 3D slicer. *Med Phys*. 2012; 39(10):6332-6338. doi:10.1118/1.4754659
21. ICRU. ICRU report 83: prescribing, recording, and reporting photon-beam Intensity Modulated Radiation Therapy (IMRT). *J ICRU. 2010*;10(1):3. doi:10.1093/jicru/ndq002
22. Voet PWJ, Dirkx MLP, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJ. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. *Radiother Oncol*. 2011;98(3):373-377. doi:10.1016/j.radonc.2010.11.017
23. Eldesoky AR, Yates ES, Nyeng TB, et al. Internal and external validation of an ESTRO delineation guideline—dependent automated segmentation tool for loco-regional radiation therapy of early breast cancer. *Radiother Oncol*. 2016;121(3):424-430. doi:10.1016/j.radonc.2016.09.005
24. Feng M, Demiroz C, Vineberg KA, Eisbruch A, Balter JM. Normal tissue anatomy for oropharyngeal cancer: contouring variability and its impact on optimization. *Int J Radiat Oncol Biol Phys*. 2012;84(2):e245-e249. doi:10.1016/j.ijrobp.2012.03.031
25. Beasley WJ, Mcwilliam A, Aitkenhead A, Mackay RI, Rowbottom CG. The suitability of common metrics for assessing parotid and larynx autosegmentation accuracy. *J Appl Clin Med Phys*. 2016;17(2):41-49. doi:10.1120/jacmp.v17i2.5889
26. Liang F, Qian P, Su K, et al. Abdominal, multi-organ, auto-contouring method for online adaptive magnetic resonance guided radiotherapy: an intelligent, multi-level fusion approach. *Artif Intell Med*. 2018;90:34-41. doi:10.1016/j.artmed.2018.07.001
27. Mattiucci GC, Boldrini L, Chiloiro G, et al. Automatic delineation for replanning in nasopharynx radiotherapy: what is the agreement among experts to be considered as benchmark? *Acta Oncol*. 2013;52(7):1417-1422. doi:10.3109/0284186X.2013.813069