

# Clinically oriented contour evaluation using geometric and dosimetric indices based on simple geometric transformations

**Lixun Xian**

West China Hospital, Sichuan University

**Guangjun Li**

West China Hospital, Sichuan University

**Qing Xiao**

West China Hospital, Sichuan University

**Zhibin Li**

West China Hospital, Sichuan University

**Xiangbin Zhang**

West China Hospital, Sichuan University

**Li Chen**

West China Hospital, Sichuan University

**Zhenyao Hu**

West China Hospital, Sichuan University

**Sen Bai** (✉ [baisen@scu.edu.cn](mailto:baisen@scu.edu.cn))

West China Hospital, Sichuan University

---

## Research

**Keywords:** Contour evaluation, Geometric indices, Dosimetric indices, Geometric transformation

**Posted Date:** October 22nd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-19265/v3>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** 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, mainly because it does not consider any clinical background. To study the relationship between geometric indices and dosimetric indices, four different types of targets were selected to introduce the systematic and random geometric errors in the delineation process.

**Materials and Methods:** The C-shaped target outlined in the American Association of Physicists in Medicine (AAPM) TG-119 report (The report of Task Group 119 of the AAPM) and the targets of three actual cases of oropharyngeal cancer, metastatic spine cancer, and prostate cancer were selected as the test contours that needed to be modified. Python software was used to perform translation, scaling, rotating and sine function transformation to introduce systematic and random errors into the above contours. These geometrically transformed contours were regarded as reference contours corrected by systematic and random errors. The corresponding dosimetric indices were obtained from the original dose distribution of the radiotherapy plan, and correlations ( $R^2$ ) between geometric and dosimetric indices were quantified through linear regression. The Wilcoxon signed rank test was used to compare the ability of spatial-direction discrimination between the geometric indices of different directions of transformations.

**Results:** The correlations between the geometric and dosimetric indices were inconsistent for the targets. For systematic errors, except for the sine function transformation ( $R^2$ : 0.023–0.04,  $p \geq 0.05$ ), the geometric transformations of the C-shaped target's planning target volume (C-PTV) had correlations with the dosimetric indices D98% and Dmean ( $R^2$ : 0.689–0.988), 80% of which were strongly correlated ( $R^2 > 0.8$ ). For the random errors, the correlation coefficients of the actual cases were also high,  $R^2 > 0.384$ ,  $p < 0.05$ . The results of Wilcoxon signed rank test showed that only the p-values of volumetric geometric indices of C-PTV were less than 0.05.

**Conclusions:** Clinically, an assessment of the contour accuracy of the region of interest is not feasible based on geometric indices alone, and should be combined with dosimetric indices. Keywords: Contour evaluation, Geometric indices, Dosimetric indices, Geometric transformation

## Background

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). Inaccuracy in the contouring process 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 plan, patient set-up [1-3]. The commonly used slice-by-slice manual approach or interpolation-based semi-automatic contouring approach is time-consuming and resource-consuming, and the corresponding results are susceptible to differences between observers; furthermore, the accuracy of contouring depends on the residents' clinical experience and the rational and efficient use of imaging techniques. Computed tomography (CT) images provide useful anatomical information and electron density for dose calculation of radiotherapy plans, however, comparing with the results after registration of CT and magnetic resonance imaging (MRI), the contrast of soft tissue in CT images is poor,

and in some cases, the targets and organs at risk can not be clearly displayed. Improper selection of imaging techniques may also lead to differences in residents' delineation [4-6].

Methods for assessing the accuracy of contouring are generally divided into two categories, namely, subjective evaluations and quantitative evaluations. The subjective evaluation is only based on the experiences and personal preferences of the evaluators. Evaluators were 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 was 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 required considerable time. Most contour accuracy studies are performed directly by using quantitative evaluation, which involves the employment of geometric indices to characterize the similarity between the test contour and the reference contour [8]. Geometric indices widely 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., Dice-similarity coefficient (DSC) and Jaccard) [9]. Although these indices are easy to calculate, they do not consider the clinical effect and may lack clinical relevance [10, 11]. Under the assumption of a reference contour, the method for clinically assessing the accuracy of radiotherapy (RT) contours is to determine and predict the deviation of its dosimetric indices based on the dose distribution of the radiation treatment plan [10, 12-14]. However, the relationships between geometric indices and dosimetric indices are yet to be further studied. In addition, different geometric indices have different properties, but some automatic segmentation studies randomly select geometric indices to evaluate the contour results [15-17].

In the present study, we explored the evaluation accuracy of the geometric indices for evaluating the contours under the systematic and random errors. This study artificially introduced contour errors through the following four geometric transformations: translation, scaling, rotation, and sine function transformation. Then, based on these transformations, we investigated the correlations between the distance-type (HD,  $HD_{mean}$ , HD95) and volumetric geometric indices (DSC, Jaccard) and the dosimetric indices ( $D_{98\%}$ ,  $D_{mean}$ ,  $D_{2\%}$ ); explored the ability of geometric indices to distinguish the contours with the same transformation type but different directions.

## Methods

### Contouring

Four different types of targets were selected for this study: C-shaped target, oropharyngeal cancer, metastatic spine cancer, and prostate cancer (Fig. 1). The C-shaped target was delineated on the water phantom, according to the TG-119 report [18], and the remaining three types of targets were outlined by senior residents in our research institution. The structures of the targets were exported from treatment planning system (TPS) Raystation (Raysearch, Stockholm, Sweden) in the form of a DICOM file, and the position information of the contours were read by an in-house developed Python software (version 3.7.3), and then used to perform the geometric transformations. Finally, 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 correction by senior doctors.

In this study, the contour errors were introduced in the form of geometric transformations. The translation transformations were divided into the following three cases: right, anterior, and posterior direction. Based on the location of the original contours, at intervals of 1 mm, the data were moved 10 times to each of the right, anterior, and posterior directions to obtain the reference contours (see Additional file 1: Supplementary Figures S1–S3 for the contours after the right, anterior, and posterior directions translation, respectively). Scaling transformation represented an equidistant expansion or reduction transformation in 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 (see Additional file 1: Supplementary Figures S1–S2 for the contours after the expansion and reduction transformation, respectively). The rotation transformation involved taking the origin of the CT image coordinates as the rotation center point, using 1° as the interval, and rotating clockwise 10 times (see Additional file 1: Supplementary Figures S6 for the contours after the rotation transformation). For the sine function transformation, we extracted the coordinate values ( $x_0, y_0$ ) of the original contour first, and then, we used the function  $y = \sin \omega y_0$  ( $\omega = 3, 4, 5, 6, \dots, 12$ ) to carry out periodic transformations 10 times with a fixed amplitude (see Additional file 1: Supplementary Figures S7 for the contours after the sine function transformation).

Through the above seven geometric transformations, systematic errors were introduced into the C-shaped target. In order to introduce random errors, the 5 layers of CT images were randomly selected to keep the position information of the structure unchanged, the remaining CT image layers were divided into three parts, and the above geometric transformations were carried out randomly, and 20 delineation results were obtained for each target.

## Geometric indices

In this study, we chose five widely used geometric indices for the evaluations, including three distance-type indices HD (maximum, mean, 95%) and two volumetric indices (DSC and Jaccard). These five geometric indices were calculated by 3DSlicer version 4.10.2 [19], which is open source software. The calculation of HD was performed on the RT-DICOM structures. The HD indices calculated by the 3DSlicer represent bi-directional distances, and the bi-directional distance is symmetrical; this type of distance is more stable than the unidirectional distance calculated by other methods.

## Dosimetric indices

The original clinical plans for these four targets all used IMRT technology. The C-shaped target met the requirements for a simple version in the TG-119 report, the dose of 5000 cGy received by 90% of the target volume was taken as the prescription, and the dose prescriptions for oropharyngeal cancer, metastatic spine cancer and prostate cancer were 95% of the target volume receiving 5400, 3000 and 5600 cGy, respectively, and the dose grid was 2 mm. After geometric transformation, the RTstructures were imported into the radiotherapy plan of the original clinical plans, and then, on the dose distribution, D98%,  $D_{\text{mean}}$ , and D2% of the PTV were obtained. According to the ICRU-83 report [20], these dosimetric indices represent the minimum dose, mean dose, and maximum dose received by the target, respectively. In this study, the dose differences ( $\Delta D$ ) of three dosimetric indices D98%,  $D_{\text{mean}}$  and D2% were calculated and normalized according to their respective clinical goals [13]. Here,  $\Delta D = (D_{x,\text{reference}} - D_{x,\text{test}}) / D_{x,\text{reference}}$ , where  $x$  represents the type of dosimetric index.

## Analysis

An SPSS version 21.0 software (SPSS, Chicago, IL, USA) was used for linear regression analysis. The correlation coefficient  $R^2$  was used to quantify the correlations between the geometric indices HD (maximum, mean, 95%), DSC, and Jaccard, and the dosimetric indices D98%,  $D_{\text{mean}}$ , and D2%. Two-sided  $p$ -values were obtained, and  $p$ -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, and from scatterplots of geometric indices versus dose difference, the feasibility of assessing the accuracy of contours with geometric indices was analyzed.

valuations. The subjective evaluation is only based on the experiences and personal preferences of the evaluators. Evaluators were 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 was 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 required considerable time. Most contour accuracy studies are performed directly by using quantitative evaluation, which involves the employment of geometric indices to characterize the similarity between the test contour and the reference contour [8]. Geometric indices widely used in contour evaluations include distance-type geometric indices (e.g., the maximum (HD), mean ( $HD_{\text{mean}}$ ), and 95% Hausdorff distances (HD95)) and volumetric geometric indices (e.g., Dice-similarity coefficient (DSC) and Jaccard) [9]. Although these indices are easy to calculate, they do not consider the clinical effect and may lack clinical relevance [10, 11]. Under the assumption of a reference contour, the method for clinically assessing the accuracy of radiotherapy (RT) contours is to determine and predict the deviation of its dosimetric indices based on the dose distribution of the radiation treatment plan [10, 12-14]. However, the relationships between geometric indices and dosimetric indices are yet to be further studied. In addition, different geometric indices have different properties, but some automatic segmentation studies randomly select geometric indices to evaluate the contour results [15-17].

In the present study, we explored the evaluation accuracy of the geometric indices for evaluating the contours under the systematic and random errors. This study artificially introduced contour errors through the following four geometric transformations: translation, scaling, rotation, and sine function transformation. Then, based on these transformations, we investigated the correlations between the distance-type (HD,  $HD_{\text{mean}}$ , HD95) and volumetric geometric indices (DSC, Jaccard) and the dosimetric indices (D98%,  $D_{\text{mean}}$ , D2%); explored the ability of geometric indices to distinguish the contours with the same transformation type but different directions.

## Results

Linear regression analysis method was carried out on the geometric indices and dosimetric indices (Table 1, 2 and 3). Table 1 shows the correlations between geometric indices and dosimetric indices of C-PTV with systematic errors. Except for the sine function transformation ( $R^2$ : 0.023–0.04,  $p > 0.05$ ), the geometric transformations of the C-PTV had strong correlations with the dosimetric indices D98% and  $D_{\text{mean}}$  ( $R^2$ : 0.689–0.988), 80% of which were strongly correlated with a  $p < 0.001$ . Table 2 and 3 show the correlations between geometric indices and dosimetric indices of actual cases with random errors. The geometric indices were correlated with the dosimetric indices D98% and  $D_{\text{mean}}$ , and the correlation coefficients of HD were weaker than

the other four geometric indices. The D2% of targets was not included in the correlation analysis because of its 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 opposite direction. For the analysis results of the five directions of C-PTV, the  $p$ -values of HD, HD<sub>mean</sub> and HD95, were all greater than 0.05, while for DSC and Jaccard, the  $p$ -values were all less than 0.05 and the differences were statistically significant.

Figs. 2–6 show the relationships between the geometric indices of the targets and the dose difference (98%,<sub>mean</sub>). For the targets with random errors 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 non monotonic.

## 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, 13, 14, 17, 21, 22]. In earlier work, van Rooij et al. [14] studied the accuracy of automatic delineation of organs at risk in the head and neck region based on deep learning techniques while using geometric indices and dosimetric indices, and they analyzed the correlation between the geometric index SDC (Sørensen-Dice similarity coefficient) and dose difference. The study found that there was a weak correlation between the SDC and  $\Delta D$  for all of the OARs through automatic segmentation,  $r = -0.24$ ,  $p = 0.002$ , but the correlation was not specific to a certain OAR or a certain patient. It is partly similar with the results described in our study. The correlation was not consistent for the different forms of geometric transformations, we found that the geometric indices obtained by geometric transformation were significantly correlated with the dosimetric indices, but for some specific geometric transformation forms, the situations were different. There was a strong and significant correlation between the geometric indices and dosimetric indices 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. After the periodic transformation of fixed amplitude of sine function, the C-PTV's contour changed very little (see Additional file 1: Supplementary Figures S7 for the contours after the sine function transformation), and the HD values were less than 2 mm (Fig 2b), which led to the small dose difference, and correspondingly, the correlation between them was weak. At the same time, it reflects the importance of contour training for junior residents, when the repeatability of contouring is high and the contour difference is small, the dose difference is also small.

The correlation coefficient obtained with the C-PTV anterior direction translation was lower than that of the other two translation transformations. In order to avoid high dose radiation to the surrounding organs at risk (Fig.1a), physicians try to keep high-dose areas away from the organs at risk when designing the original radiotherapy plan. The structures in the low-dose region are less likely to have obvious dose changes, even if its contour varies greatly in the spatial extent [23]. When the anterior direction translation occurs in this area, the minimum dose (D98%) of the target in this area is almost unchanged, thus resulting in a weaker correlation coefficient, and the correlation coefficient of D<sub>mean</sub> is higher than D98%. 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 indices and

dosimetric indices can be affected by many factors, such as the method of geometric transformation, the relative positions of the target and organs at risk, and the constraint goals of the radiotherapy plan.

Beasley et al. [24] reported that when measured with a suitable spatial metric, the higher the geometric accuracy of the contour, the smaller the dose difference should therefore be reflected, and vice versa. At the same time, there have been many reports on auto-segmentation technology that randomly select the geometric index to evaluate the acceptability of contouring results. According to the Wilcoxon signed ranks test analysis results 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 three different directions of right, anterior and posterior, and the difference between equidistant scaling. However, there were significant differences between the volumetric geometric indices obtained from the different transformation directions of the irregular shape target. 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 Fig. 2b, when  $HD = 1.547$ , the dose difference was within 5% for equidistant expansion transformation, while for equidistant reduction transformation 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, the corresponding dose differences are different. When the distance-type geometric index value was about 3 mm, the dose difference corresponding to  $HD_{mean}$  was already close to 30% (see Figs 2–6). That's why, despite the highest correlation coefficient of  $HD_{mean}$ , this study believes that  $HD_{mean}$  is not suitable as an index for evaluating contour accuracy alone. Compared with HD and HD95, it is too sensitive and the change gradient is too big to reflect the actual clinical situation. Since there is no independent reference standard for distance-type geometric index values, it is impossible to compare the contour accuracy across different structures. This means that although two different structures have the same geometric index value, they cannot indicate that the quality of the contour is the same. And it can be seen in Figs 2–6 that the quality of the contour is controversial for the two structures (Oropharyngeal cancer and prostate cancer) of the same geometric index value.

For volumetric geometric indices, many studies believe that if the DSC is higher than the normally reported value of 0.7, the agreement between the reference contour and the test contour is considered to be good [22, 25, 26]. Our research showed that when the DSC and Jaccard values of anterior direction translation transformation were between 0.5 and 0.7, the corresponding dose differences were also very small (Fig. 2c), at the same time, there were some cases that the DSC values were greater than 0.7, and the corresponding dose difference values were large. These two 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 transformation, and analyzed the feasibility of clinical evaluation of geometric indices and the ability of geometric indices to identify the direction of transformation. Although geometric indices reflected the geometric difference between a test contour and a reference contour, and the correlation coefficient between the geometric index and the dosimetry parameter in this study was relatively high, the relationship between geometric indices and dosimetric indices were not consistent among different geometric indices, different transformation forms and different targets, so it was illogical to use only geometric indices to evaluate the clinical acceptability of contour results. In addition, our current research is based on the simulation experiment of geometric transformation, we should further explore the relationship between geometric indices and dosimetry indices by making use of the actual cases contouring results of the junior residents.

## Conclusion

At present, there is a lack of guidance for the evaluation of contours by using geometric indices, and therefore, there is a need for a normative framework. We found that the differences between the geometric indices and dosimetric indices were not consistent, which indicates the inaccuracy when using only the geometric indices 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 evaluations of the accuracy of the results of delineation, which can be helpful for explaining the clinical dose response relationship of delineation more comprehensively, accurately.

## Abbreviations

AAPM: the American Association of Physicists in Medicine; TG-119 report: The report of Task Group 119 of the AAPM; PTV: planning target volume; OAR: organ at risk; IMRT: intensity-modulated radiotherapy; VMAT: volumetric modulated arc radiotherapy; CT: computed tomography; MRI: magnetic resonance imaging; HD: the maximum Hausdorff distances;  $HD_{mean}$ : the mean Hausdorff distances; HD95: the 95% Hausdorff distances; DSC: Dice-similarity coefficient; RT: radiotherapy; TPS: the treatment planning system; SDC: Sørensen-Dice similarity coefficient

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

Please contact author for data requests.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 81472807).

### Authors' contributions

LXX and GJL analyzed the data and drafted the manuscript. QX, ZBL and XBZ performed the experiments. LC and ZYH collected data. SB designed this study, interpreted the results and revised the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

Not applicable.

## References

1. Nelms B E, Wolfgang A, Tomé, Robinson G, et al. Variations in the Contouring of Organs at Risk: Test Case From a Patient With Oropharyngeal Cancer[J]. *International Journal of Radiation Oncology\*Biology\*Physics*. 2012, 82(1):368-378. doi: 10.1016/j.ijrobp.2010.10.019.
2. Brouwer C L, Steenbakkers R J H M, Heuvel E V D, et al. 3D Variation in delineation of head and neck organs at risk[J]. *Radiation Oncology*. 2012, 7(1): 32. doi: 10.1186/1748-717X-7-32.
3. Brouwer C L, Steenbakkers R J H M, 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[J]. *Radiotherapy and Oncology*. 2015, 117(1): 83-90. doi: 10.1016/j.radonc.2015.07.041.
4. Khoo V S, Adams E J, Saran F, et al. A Comparison of clinical target volumes determined by CT and MRI for the radiotherapy planning of base of skull meningiomas[J]. *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[J]. *Radiation oncology (London, England)*. 2012, 7(1): 160. doi: 10.1186/1748-717X-7-160.
6. Lim J Y, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck[J]. *Acta Oncologica*. 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[J]. *International Journal of Radiation Oncology\*Biology\*Physics*. 2013, 87(4): 809-816. doi: 10.1016/j.ijrobp.2013.08.007.
8. Wittenstein O, Hiepe P, Sowa L H, et al. Automatic image segmentation based on synthetic tissue model for delineating organs at risk in spinal metastasis treatment planning[J]. *Strahlentherapie und Onkologie*. 2019, 195(12): 1094-1103. doi: 10.1007/s00066-019-01463-4.
9. Taha A A, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool[J]. *BMC Medical Imaging*. 2015, 15(1). doi: 10.1186/s12880-015-0068-x.
10. Lim T Y, Gillespie E, Murphy J, et al. Clinically Oriented Contour Evaluation Using Dosimetric Indices Generated From Automated Knowledge-Based Planning[J]. *International Journal of Radiation Oncology\*Biology\*Physics*. 2019, 103(5): 1251-1260. doi: 10.1016/j.ijrobp.2018.11.048.
11. Kaderka R, Gillespie E F, Mundt R C, et al. Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients[J]. *Radiotherapy and Oncology*. 2019, 131: 215-220. doi: 10.1016/j.radonc.2018.07.013.
12. Stuart Y, Tsuji, Andrew Hwang, Vivian Weinberg, et al. Dosimetric Evaluation of Automatic Segmentation for Adaptive IMRT for Head-and-Neck Cancer[J]. *International Journal of Radiation Oncology\*Biology\*Physics*. 2010, 77(3): 707-714. doi: 10.1016/j.ijrobp.2009.06.012.
13. Kieselmann J P, Kamerling C P, Burgos N, et al. Geometric and dosimetric evaluations of atlas-based segmentation methods of MR images in the head and neck region[J]. *Physics in medicine and biology*. 2018,

63(14): 145007. doi: 10.1088/1361-6560/aacb65.

14. 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[J]. *International Journal of Radiation Oncology\*Biology\*Physics*. 2019, 104(3): 677-684. doi: 10.1016/j.ijrobp.2019.02.040.
15. Jameson M G, Holloway L C, Vial P J, et al. A review of methods of analysis in contouring studies for radiation oncology[J]. *J Med Imaging Radiat Oncol*. 2010, 54(5): 401-410. doi: 10.1111/j.1754-9485.2010.02192.x.
16. Hanna G G, Hounsell A R, O Sullivan J M. Geometrical Analysis of Radiotherapy Target Volume Delineation: a Systematic Review of Reported Comparison Methods[J]. *Clinical Oncology*. 2010, 22(7): 515-525. doi: 10.1016/j.clon.2010.05.006.
17. 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[J]. *Radiotherapy and Oncology*. 2016, 121(2): 169-179. doi: 10.1016/j.radonc.2016.09.009.
18. Ezzell G A, Burmeister J W, Dogan N, et al. IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119[J]. *Medical Physics*. 2009, 36(11): 5359-5373. doi: 10.1118/1.3238104.
19. Pinter C, Lasso A, Wang A, et al. SlicerRT: Radiation therapy research toolkit for 3D Slicer[J]. *Medical Physics*. 2012, 39(10): 6332-6338. doi: 10.1118/1.4754659.
20. ICRU (2010) ICRU report 83: prescribing, recording, and reporting photon-beam Intensity Modulated Radiation Therapy (IMRT)[J]. *Journal of the ICRU*. 2010, 10(1): 3. doi: 10.1093/jicru/ndq002.
21. Voet P W J, Dirx M L P, Teguh D N, et al. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis[J]. *Radiotherapy and Oncology*. 2011, 98(3): 373-377. doi: 10.1016/j.radonc.2010.11.017.
22. Eldesoky A R, Yates E S, Nyeng T B, et al. Internal and external validation of an ESTRO delineation guideline – dependent automated segmentation tool for loco-regional radiation therapy of early breast cancer[J]. *Radiotherapy and Oncology*. 2016, 121(3): 424-430. doi: 10.1016/j.radonc.2016.09.005.
23. Feng M, Demiroz C, Vineberg K A, et al. Normal Tissue Anatomy for Oropharyngeal Cancer: Contouring Variability and Its Impact on Optimization[J]. *International Journal of Radiation Oncology\*Biology\*Physics*. 2012, 84(2): e245-e249. doi: 10.1016/j.ijrobp.2012.03.031.
24. Beasley W J, Mcwilliam A, Aitkenhead A, et al. The suitability of common metrics for assessing parotid and larynx autosegmentation accuracy[J]. *Journal of Applied Clinical Medical Physics*. 2016, 17(2): 41-49. doi: 10.1120/jacmp.v17i2.5889.
25. Mattiucci G C, Boldrini L, Chiloiro G, et al. Automatic delineation for replanning in nasopharynx radiotherapy: What is the agreement among experts to be considered as benchmark?[J]. *Acta Oncologica*. 2013, 52(7): 1417-1422. doi: 10.3109/0284186X.2013.813069.
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[J]. *Artificial Intelligence in Medicine*. 2018, 90: 34-41. doi: 10.1016/j.artmed.2018.07.001.

## Table

Table 1 Linear regression analysis between geometric indices and dosimetric indices of C-PTV with systematic errors.

| C-PTV     | Transformation    | HD             |         | HD <sub>mean</sub> |         | HD95           |         | DSC            |         | Jaccard        |         |         |
|-----------|-------------------|----------------|---------|--------------------|---------|----------------|---------|----------------|---------|----------------|---------|---------|
|           |                   | R <sup>2</sup> | p-value | R <sup>2</sup>     | p-value | R <sup>2</sup> | p-value | R <sup>2</sup> | p-value | R <sup>2</sup> | 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           | < 0.001 | 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 <sub>mean</sub> | 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.

| Site                    | HD             |                 | HD <sub>mean</sub> |                 | HD95           |                 | DSC            |                 | Jaccard        |                 |
|-------------------------|----------------|-----------------|--------------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|
|                         | R <sup>2</sup> | <i>p</i> -value | R <sup>2</sup>     | <i>p</i> -value | R <sup>2</sup> | <i>p</i> -value | R <sup>2</sup> | <i>p</i> -value | R <sup>2</sup> | <i>p</i> -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         |

Table 3 Linear regression analysis between geometric indices and D<sub>mean</sub> of targets with random errors.

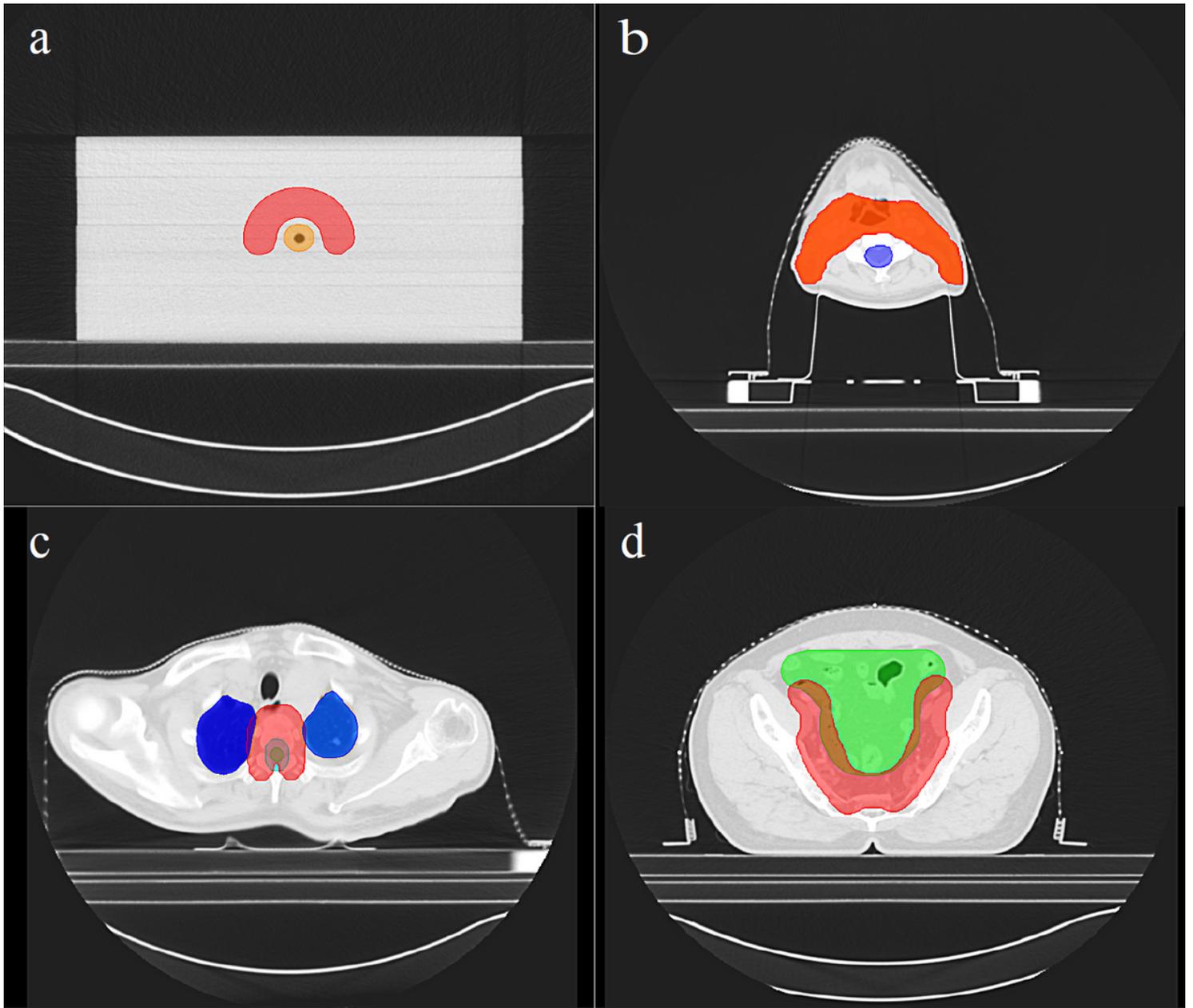
| Site                    | HD             |                 | HD <sub>mean</sub> |                 | HD95           |                 | DSC            |                 | Jaccard        |                 |
|-------------------------|----------------|-----------------|--------------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|
|                         | R <sup>2</sup> | <i>p</i> -value | R <sup>2</sup>     | <i>p</i> -value | R <sup>2</sup> | <i>p</i> -value | R <sup>2</sup> | <i>p</i> -value | R <sup>2</sup> | <i>p</i> -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 <sub>mean</sub> | 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* | 0.004*  |
| 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* | 0.004*  |
| 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* | 0.034*  |
| Scaling transformation           | Z-value   | -1.586 | -1.790             | -1.792 | -2.505 | -2.533  |
|                                  | p-value   | 0.113  | 0.074              | 0.069  | 0.012* | 0.011*  |

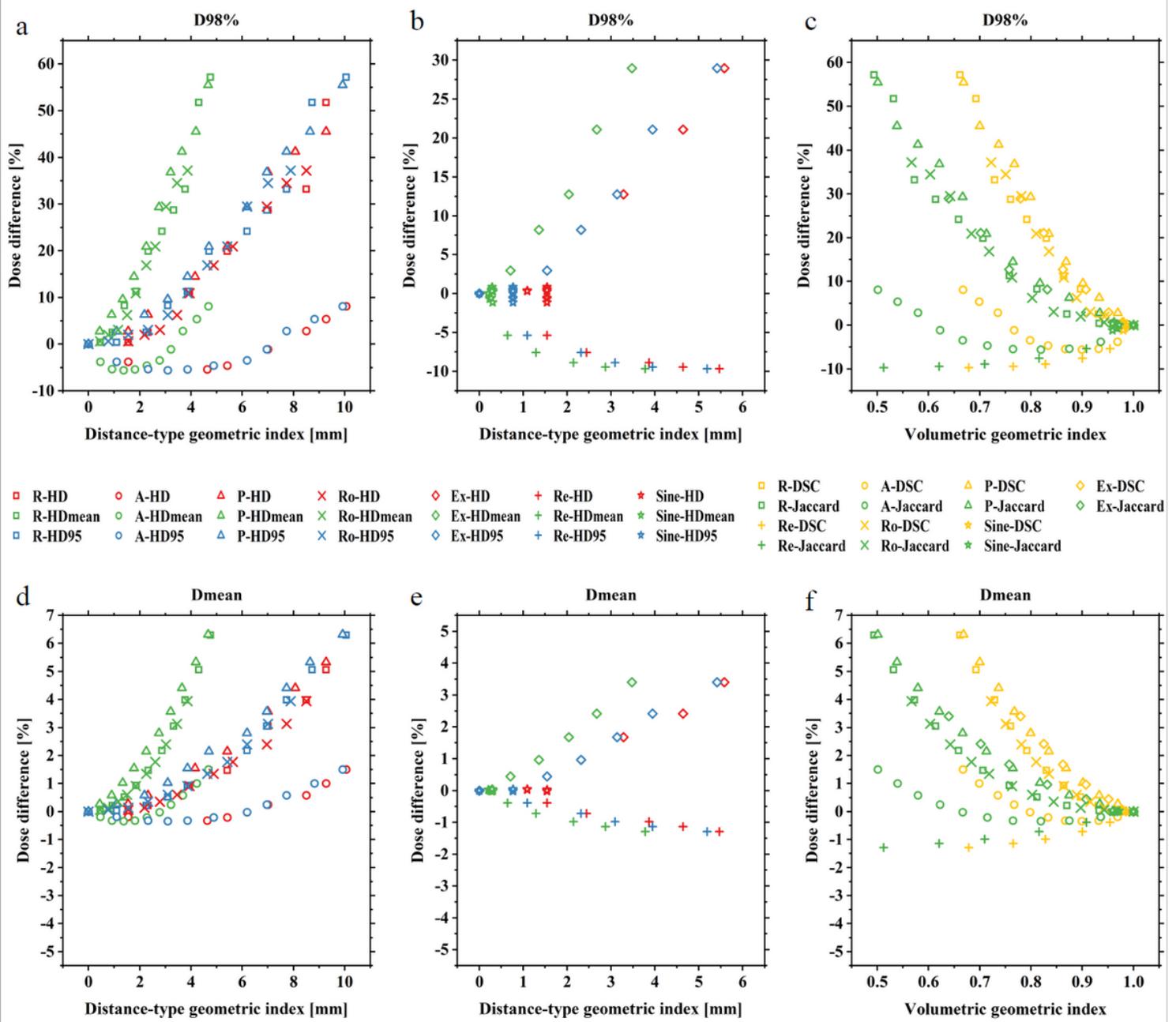
\* *p*-value = the difference was statistically significant.

## Figures



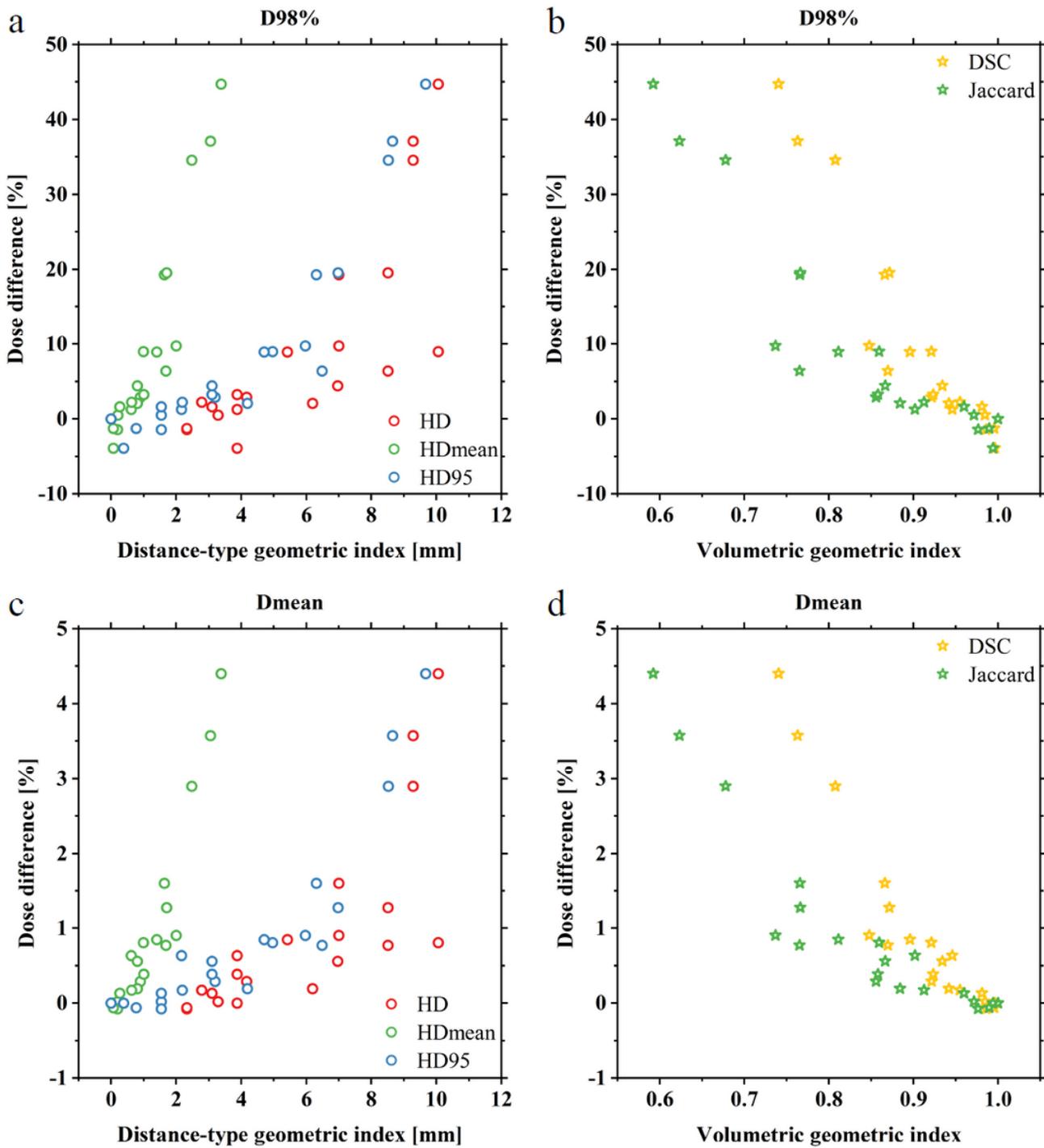
**Figure 1**

An example of the PTV contours (orange red) of (a) the C-shaped target, (b) oropharyngeal cancer, (c) metastatic spine cancer and (d) prostate cancer.



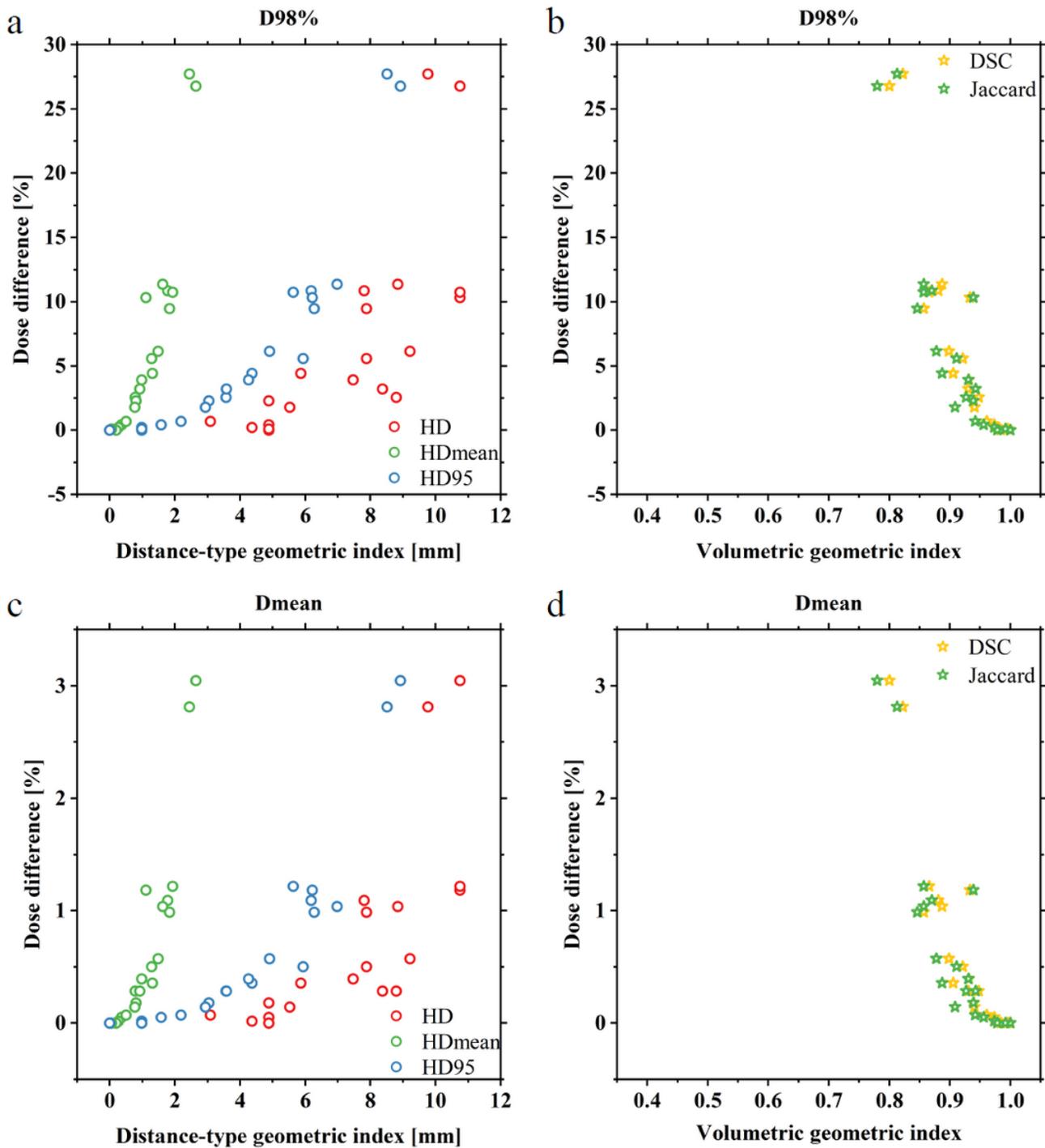
**Figure 2**

The relationship between the geometric indices of C-PTV and the dose difference. (a-c) and (d-f) are the relationships between the geometric indices of C-PTV after introducing systematic errors and the dose difference D98% and Dmean, respectively. R-HD = dose difference corresponding to HD value after right direction translation transformation; R-HDmean = dose difference corresponding to HDmean value after right direction translation transformation; R-HD95 = dose difference corresponding to HD95 value after right direction translation transformation; R-DSC = dose difference corresponding to DSC value after right direction translation transformation; R-Jaccard = dose difference corresponding to Jaccard value after right direction translation transformation; A = anterior direction translation transformation; P = posterior direction translation transformation; Ex = expansion transformation; Re = reduction transformation; Ro = rotation transformation; Sine = sine function transformation.



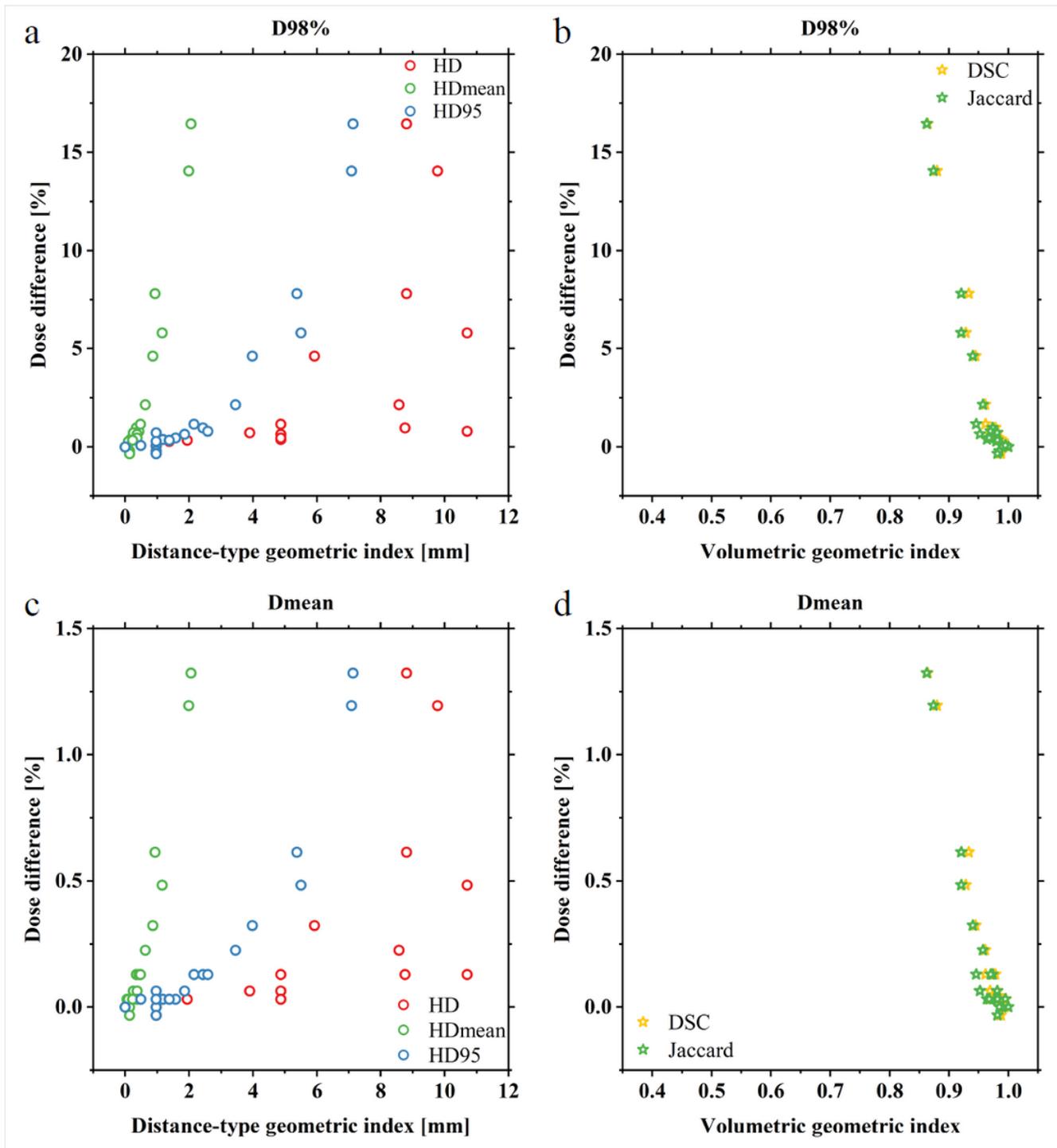
**Figure 3**

The 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 introducing random errors and the dose difference D98% and Dmean, respectively.



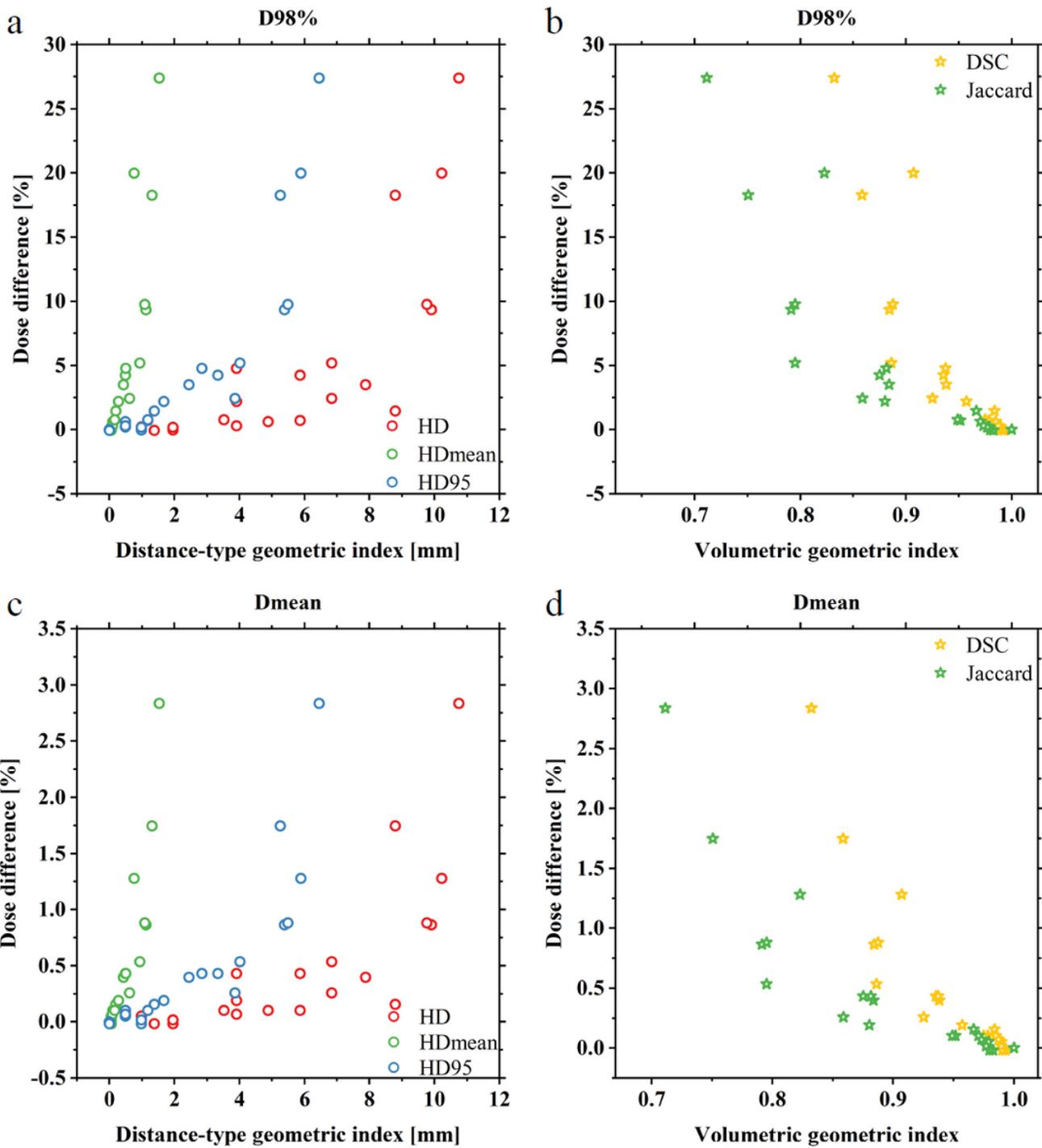
**Figure 4**

The relationship between the geometric indices of oropharyngeal cancer and the dose difference. (a-b) and (c-d) are the relationships between the geometric indices of oropharyngeal cancer after introducing random errors and the dose difference D98% and Dmean, respectively.



**Figure 5**

The 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 introducing random errors and the dose difference D98% and Dmean, respectively.



**Figure 6**

The 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 introducing random errors and the dose difference D98% and Dmean, respectively.

## Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [Additionalfile1.pdf](#)