*Journal of Medical Imaging and Radiation Oncology* **54** (2010) 401–410

Journal of Medical Imaging and Radiation Oncology

REVIEW ARTICLE

# A review of methods of analysis in contouring studies for radiation oncology

Michael G Jameson,[1,2] Lois C Holloway,[1,2,3] Philip J Vial,[2,3] Shalini K Vinod[2,4] and Peter E Metcalfe[1]

[1]Centre for Medical Radiation Physics, University of Wollongong, Wollongong, and [2]Department of Radiation Oncology, Liverpool and Macarthur Cancer Therapy Centres, and [3]Institute of Medical Physics, School of Physics, University of Sydney, and [4]University of New South Wales, Sydney, New South Wales, Australia

**MG Jameson** BMedRadPhys; **LC Holloway** PhD; **PJ Vial** PhD; **SK Vinod** MD; **PE Metcalfe** PhD.

**Correspondence**
Mr Michael Jameson, Sydney South West Area Health Service, Cancer Therapy Centre, Liverpool Hospital, Locked Bag 7103, Liverpool BC, NSW 1871, Australia.
Email: michael.jameson@sswahs.nsw.gov.au

Conflict of interest: None.

This work was presented as a poster presentation at Combined Scientific Meeting, Brisbane, QLD, 2009.

Submitted 12 October 2009; accepted 4 July 2010.

doi:10.1111/j.1754-9485.2010.02192.x

## Summary

Inter-observer variability in anatomical contouring is the biggest contributor to uncertainty in radiation treatment planning. Contouring studies are frequently performed to investigate the differences between multiple contours on common datasets. There is, however, no widely accepted method for contour comparisons. The purpose of this study is to review the literature on contouring studies in the context of radiation oncology, with particular consideration of the contouring comparison methods they employ. A literature search, not limited by date, was conducted using Medline and Google Scholar with key words: contour, variation, delineation, inter/intra observer, uncertainty and trial dummy-run. This review includes a description of the contouring processes and contour comparison metrics used. The use of different processes and metrics according to tumour site and other factors were also investigated with limitations described. A total of 69 relevant studies were identified. The most common tumour sites were prostate (26), lung (10), head and neck cancers (8) and breast (7).The most common metric of comparison was volume used 59 times, followed by dimension and shape used 36 times, and centre of volume used 19 times. Of all 69 publications, 67 used a combination of metrics and two used only one metric for comparison. No clear relationships between tumour site or any other factors that may influence the contouring process and the metrics used to compare contours were observed from the literature. Further studies are needed to assess the advantages and disadvantages of each metric in various situations.

**Key words:** contouring; delineation; inter/intra-observer variability; radiotherapy accuracy.

## Introduction

Radiotherapy planning relies on accurate definition of tumour and normal tissues. Variations in contours defining the tumour and normal tissues are analysed using contouring studies where multiple contours are generated either by a number of observers (inter-observer) or a single observer (intra-observer) for a set/s of patient images as shown in Figure 1. Contouring studies are typically used to assess different factors that may influence the contouring process, such as the impact of new imaging technologies on tumour delineation. Imaging technology in radiation therapy is constantly changing with the introduction of new and improved imaging modalities.[1] New treatment planning and delivery techniques have also improved the ability to conform the prescribed doses to the contoured tumour.[2] Thus, the importance of contouring and contouring studies has increased, with an emphasis on defining the tumour and healthy tissues accurately and consistently.[2]

Although contouring studies are commonplace, there is still no widely accepted method of comparing contours. The purpose of this study is to review contouring

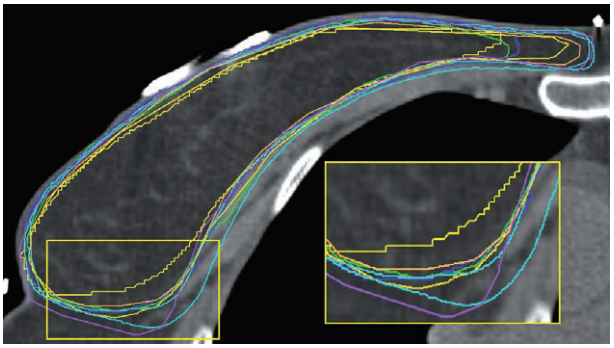


**Fig. 1.** Multiple contours from a number of observers delineating the breast tissue.

studies in the context of radiation oncology, with particular consideration on the contouring comparison methods they employ.

## Methods

A literature search, not limited by date, was conducted using Medline (Ovid Technologies, Inc.) and Google Scholar (Google, Inc.) with key words: contour, variation, delineation, inter/intra observer uncertainty and trial dummy-run. Relevant studies known to the authors and those identified from the reference list of included studies were also analysed. Only those publications relevant to radiation oncology were included. The contouring processes and contour comparison metrics used in these studies were reviewed, and the use of different processes and metrics according to tumour site and other factors was detailed.

## Results

Sixty-nine studies detailing numerous methods of contour analysis covering a range of clinical sites were identified (Table 1). The studies were published between 1993 and 2010.

It is apparent from the literature that the design of a contouring study can be divided into two steps: (i) how the imaging and contouring process will be managed; and (ii) how the contours will be analysed. These steps were described in varying detail in the literature depending on the focus of the contouring study.

### Contouring process

Clinicians were typically given the images, together with imaging and pathology information, and were blinded to other observer's contours.[28,54,56–58,70] In many studies, the clinicians were given guidelines to use for contouring.[3,29,54,70,71] Guidelines usually follow standard contouring protocols[29,54,61] and may specify factors, such as window level settings and software to use, to ensure that contouring conditions are similar for each observer.[3] Very few studies provided detailed guidelines of the protocol used for defining volumes.[4–6,20] For studies utilising different imaging modalities, the images are usually overlaid after they have been spatially registered.[72–74] Image registration is a complex process and will not be considered in this article; readers are referred to Balter and Kessler's review.[75] Most studies did not describe in detail how the image and contouring processes were controlled.

When a multi-institutional contouring study is performed, the investigators may send images and contouring software as a package with pertinent patient information.[7,8] In the case of Steenbakers *et al*. the software was installed on a number of identical computers (1 GHz with 19-inch monitors) in the institutions participating in the study[7] to ensure that all observers contoured under similar conditions. In other cases the patient imaging and pertinent information was mailed and observers contoured on their own contouring package.[4,5] Some older studies contoured using a marker on CT films.[21,30,59,65] When analysing the contours on film, extra steps are involved in scanning the film or physically measuring the delineations.[21,30,59,65]

**Table 1.** Contouring metrics used for each tumour site as a ratio of the total publications for that site

| Site | # Publications | Volume (%) | CI (%) | Centre of volume (%) | Shape/Dimension (%) | References |
|---|---|---|---|---|---|---|
| Lung | 10 | 8 (80) | 4 (40) | 2 (20) | 5 (50) | 3–12 |
| Breast | 7 | 7 (100) | 3 (42) | 4 (57) | 5 (71) | 13–19 |
| Brain | 8 | 6 (75) | 2 (25) | 4 (50) | 2 (25) | 20–27 |
| Prostate | 26 | 21 (81) | 5 (19) | 4 (15) | 16 (62) | 28–53 |
| Head and neck | 8 | 8 (100) | 3 (37) | 1 (13) | 2 (25) | 25,36,54,55 |
| Pancreas | 1 | 1 | 0 | 0 | 0 | 56–60 |
| Bladder | 3 | 3 | 0 | 2 | 2 | 61–63 |
| Rectum | 1 | 0 | 0 | 0 | 1 | 64 |
| Oesophagus | 2 | 2 | 1 | 0 | 2 | 65,66 |
| Cervix | 3 | 3 | 1 | 2 | 1 | 67–69 |
| Total | 69 | 59 (86) | 18 (26) | 19 (28) | 36 (52) | – |

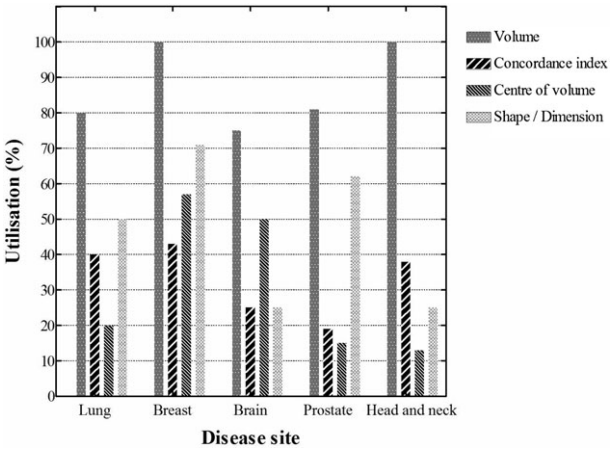



**Fig. 2.** Graphical representation of metrics chosen as a percentage for each clinical site.

## Comparison metrics

A number of metrics that have been used for comparing contours are described next and listed in Table 1. Figure 2 shows a graphical representation of the metrics chosen as a percentage utilisation for the five most common clinical sites studied in this review. Volume was the most frequently used metric across all tumour sites. Shape/dimension was the next most frequently used metric in all tumour sites except for brain and head and neck, where centre of volume (COV) and concordance index (CI) were the next most frequent metrics, respectively.

### *Volume*

Volume has been used for a number of different clinical sites and is the most common metric used in the literature.[3,7,9,13,29,31,32,54,76,77] Although rarely mentioned in the contouring studies reviewed, there are a number of different software methodologies for calculating the volume of a structure which may yield different results.[78] Therefore, when conducting a contouring study, it is important to be aware of the methods of calculation employed by different software packages. This becomes very important if different software packages are to be used for a single contouring study (i.e. a multi-centre clinical trial). One common method of volume calculation takes the number of voxels contained within the contour multiplied by the size of the voxel.[78] The size of the voxel depends on the resolution of the image reconstruction and the image slice thickness. A voxel is usually deemed to be within a contour if the centre of the voxel is within the contour boundary.[29] Studies have used in-house developed software,[7,32] third party applications,[29,31,76] open source,[3,71] and treatment planning systems,[13,54] for calculating volumes. Two structures can have the same volume but different locations, as shown in Figure 3a.

### *Centre of volume*

The COV (sometimes called centre of mass) provides a single point representing the position of the contour. Although the method used to calculate COV was rarely given, there are a number of different methods used by software to calculate the COV of a structure. These methods may give different values of COV for a single structure. One method of COV calculation determines the COV of every polygon contained within a 3D structure and generates a weighted sum.[79] Another calculation method determines the centre of a bounding box around the structure to be the COV.[80] Like the volume metric, the COV is easily calculated and output by treatment planning systems. The variation in the location of the COV can be used as a measure of the accuracy of the contouring process with regard to what is being tested (i.e. imaging modality, auto-contouring algorithm, contouring protocol, etc.). However, the COV has not been used alone for comparing contours. Two vastly different structures can have the same or similar COVs as shown in Figure 3b.

COV analysis has been used frequently in contouring studies for breast and brain Table 1. COV has been used when estimating inter-fractional rigid-organ motion to quantify differences in location between fractions.[29]

### *Concordance index*

The CI is a volume-derived metric[10] that is a measure of the overlap of two or more volumes. It is often defined as the percent ratio of the volume of intersection and the volume of the union of the two volumes. This value attempts to overcome the lack of positional information in the volume metric. A CI of 1 represents two structures perfectly overlapping with identical volume, location and shape. A CI of 0 means there is no overlap. This metric is useful when there is a reference volume with which to compare subsequent volumes. However, the CI does not give any information on how contours may vary quantitatively in size, shape or location in absolute terms; it is a relative measure.

### *Dimension*

Dimension metric has been used when investigating contouring variability with respect to direction.[13] By analysing the contouring variability along a specific axis and using the standard deviation of this variability, it is possible to find the axis for which observers have the most difficulty contouring.[13] Dimension may refer to the encompassing dimension of the structure or the differences in surface dimensions (also called surface variation) of two structures. The encompassing dimension refers to maximal size of the structure along each axis, while the surface dimensions refer to local shape variations between structures.

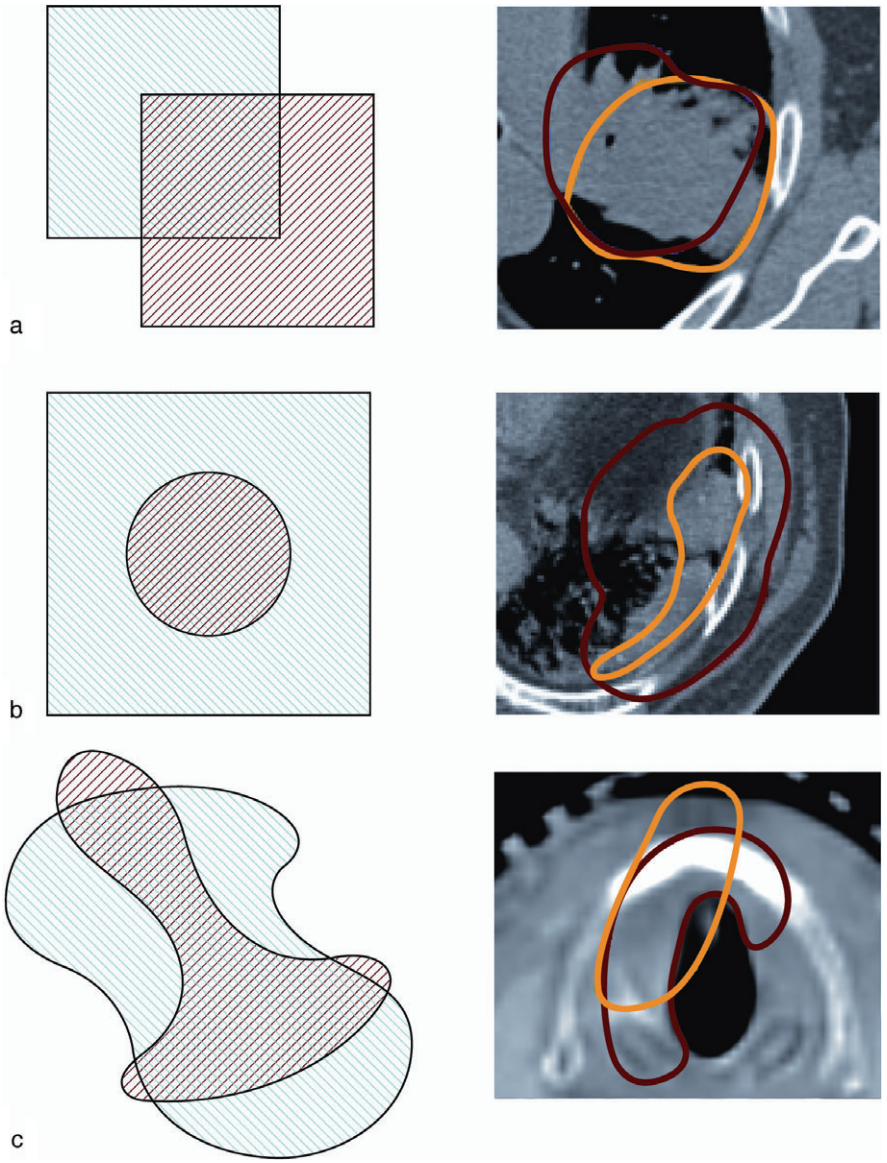



**Fig. 3.** Shows conceptually and clinically (a) two contours of equal volume and shape but different locations; (b) two contours that have the same centre of volume (COV) but different volumes and shapes; and (c) two contours that have similar maximum dimensions but different volumes and COVs.

There are a number of ways to define and calculate dimension. Encompassing dimensions are usually measured in voxels and converted to centimetres.[13,32,64] The encompassing dimensions of the structure may be calculated by defining a regular cuboid around the structure.[13,33,81] The cuboid is defined such that its surfaces 'touch' the outermost surfaces of the structure. Dimensions defined from the COV have also been used by ascertaining the COV as described previously, then defining vectors along each (X, Y, Z) axis.[3,29] The vectors measure the distance from the centre of the volume to the edge of the volume. By summing the vectors that span from the COV to the contour surface in opposite directions, the dimensions are gained.

Structure dimensions provide information on the size and, to some extent, the shape of a structure. The standard deviation of a given dimension can be used as a measure of contouring variability. One of the limitations of using dimensions defined from the COV is that the precision depends on how regular the 3D structure is. For regular structures, the dimensions defined from the COV closely represent the maximum dimensions of the structure, whereas in irregular structures, dimensions defined from the COV may be misleading. Examples of regular structures are breast and prostate; examples of irregular structures are some lung and head and neck tumours. Two structures of different volume could have very similar maximum dimensions as seen in Figure 3c.

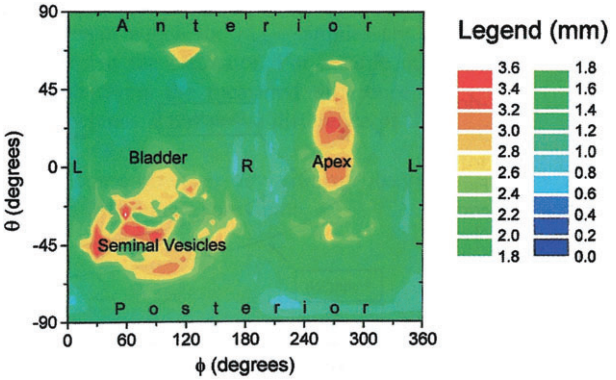



**Fig. 4.** Example of polar maps used to display differences in contouring of the prostate by Remeijer *et al*.[82] (used with permission).

### *Shape/surface variation*

In-house or third party software has frequently been employed for custom analysis of shape or surface dimensions.[7,9,13,31,76,77,82,83] Applications and methods for this type of analysis were recently summarised by van der Put *et al*.[77] They described parameterisation as mapping the surface of a 3D shape to a parametric space such as cylindrical or spherical polar coordinates. Other methods are deformation-based methods and variations of line-based methods.[77] Deformation-based methods use deformable models from image registration algorithms. Line-based methods use a line (local surface normal, closest point or coordinate system-based) between the two contours to determine the difference between them.

The Netherlands Cancer Institute have published a number of papers that utilise a parameterisation method of contour shape analysis.[7,32,61,64,82] The prostate is suited to this type of analysis because of its regular near-spherical shape. In some irregular shapes, the outer surface is not always 'visible' from the origin, leading to incorrect results.[82] Remeijer *et al*.[82] mapped the prostate contours with spherical coordinates and then applied statistical analysis to determine the systematic differences between the contours derived from CT and MRI. The results were displayed as polar maps (Fig. 4) which plot θ on the vertical axis and φ on the horizontal; differences in radius r are then displayed as changes in colour. Remeijer *et al*.[82] concluded that by separating the analysis of the geometric and statistical differences in three-dimensional shapes, the process preserves the geometric information and is non-computationally intensive.

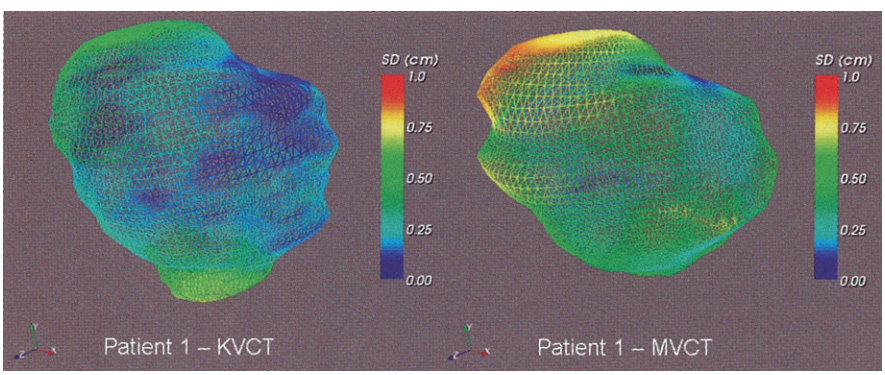
Song *et al*.[29] used a central point line-based method of shape analysis for prostate. They took an average COV and then cast out rays to intersect with the surfaces of the contours. The difference in length of the rays between each contour was then used as a measure of the contouring variability. The observer variability was then mapped onto the average prostate surface in colour to display the results as seen in Figure 5. A limitation of this method is that it only works with regular-shaped volumes like the prostate[21,75]. Both the methods used by Song *et al*.[29] and Remeijer *et al*.[82] can be used in two and three dimensions.

The nearest point line method[77] takes one contour as the reference and one as the target, then measures the distance from the reference to the target. One drawback mentioned is the lack of symmetry.[76] Interchanging the reference contour and the target contour yields different results as demonstrated in Figure 6. This can be overcome by taking the measurement twice, interchanging the contours and then using the mean distance.[76]

In order to overcome inconsistencies (Fig. 6) in shape comparison methods, van der Put *et al*.[77] introduced the ComGrad method. This algorithm works by performing a signed distance transform on both contours. A distance transform is a commonly used image processing technique whereby the pixels in an image are replaced by a scalar value that quantifies the distance to some point or boundary (Fig. 7).

The vector field of the gradient direction for both distance transforms is then computed. This gives a

**Fig. 5.** Example of the display method used by Song *et al*.[29]; the SD of the observer variation is mapped onto the average prostate surface (used with permission).



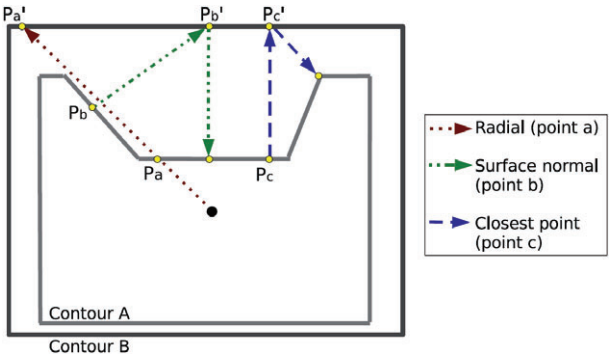



**Fig. 6.** Image shows the limitations of the aforementioned shape analysis methods. The radial and surface normal methods may overestimate the distance. The closest point and surface normal methods are not symmetric, i.e. the distance is different when going from contour A to B than it is from B to A (used with permission from van der Put *et al*.[77]).

directional component to the scalar distance previously calculated for the pixels. The vector fields for both contours are then combined, thus relating the distance and directional information. A line that is parallel to the vector field from some arbitrary point on one contour that corresponds to a point on the other contour is now drawn. This line provides a local distance measure for complex shapes that is not asymmetric. A large number of these lines are drawn and measured, which can then be plotted as a polar map for example.

## Discussion

A number of different methods have been utilised for comparing contours. The common metrics used are volume, COV and shape/dimension. Studies specifically analysing contouring variation, such as evaluation of auto-contour algorithms, tend to have more in-depth analysis of the contours generated, i.e. dimension analysis and shape variation,[3] as dimension analysis and shape variation can give more detailed insight into any systematic errors in the algorithms.

The accuracy and consistency of contouring may be affected by several factors, including: (i) advances in medical imaging technology;[84] (ii) developments in software capabilities, such as auto-contour algorithms;[3,9,31,64,76,85] (iii) the design and use of imaging and contouring protocols; (iv) inconsistent levels of imaging expertise; and (v) access to different imaging modalities.[71,84] Contouring studies reviewed in this paper used a variety of metrics to assess the influence of these factors. There are no metrics used consistently for assessing the influence of each of these factors on contouring.

Each of the comparison metrics has limitations and thus it is desirable to use multiple metrics where possible. Volume gives no indication of the location or shape of the contour. Conversely, COV and shape/dimension give no indication of the volume of the contour (Fig. 3). CI attempts to overcome this with a measure of overlap which gives some insight into the volumetric and spatial relationships of two or more contours. Limitations of the different shape/surface analysis techniques were observed, most notably the lack of symmetry of the nearest point method.

The absence of a gold standard contour that outlines the true extent of the object being contoured makes it impossible to make conclusions about the absolute accuracy of contours. Contouring comparisons were generally limited to measuring the variation of contour differences. A reference contour from which to measure the other contours against is commonly used, giving a common frame of reference for the statistical variation of each of

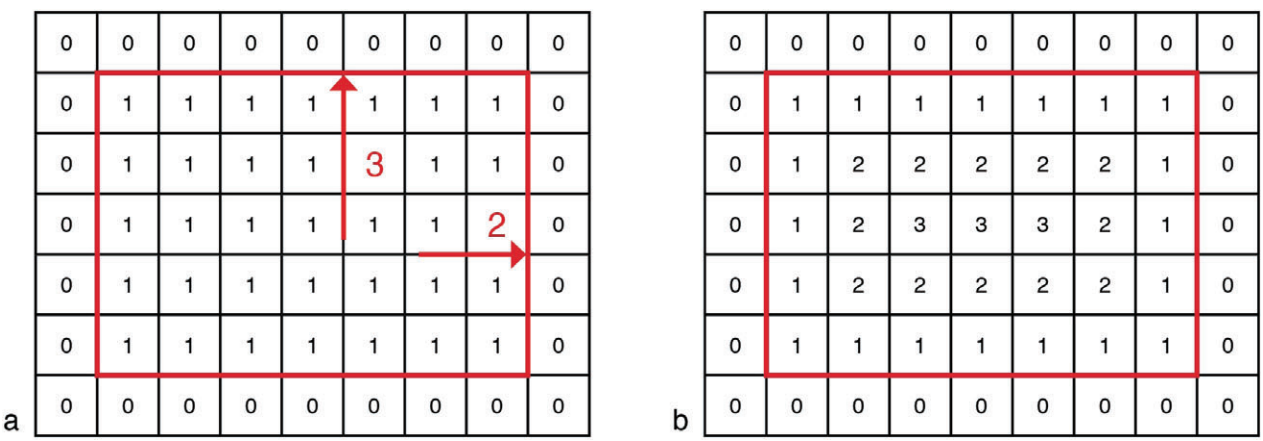



**Fig. 7.** Two-dimensional graphical representation of a chess board distance transform on a binary image; the pixels are replaced with a scalar indicating the distance to the boundary, where the boundary is displayed in red (a) shows the original field where pixels with 1 are within the boundary, and (b) shows the transformed image where the numbers represent the distance to the boundary.

the contouring metrics to be compared. The choice of this gold standard or reference contour varies in the literature from a mathematical average contour, a radiologist-defined contour, an experienced oncologist-defined contour or a consensus contour that is decided upon by a panel of experts.[10,13,29,54] One study by Gao *et al*. compared prostate gland delineations to images from the Visible Human Project.[28] This study demonstrated that inter-observer contour variations persist even when the structure being contoured is well visualised.

The clinical impact of variation in contouring is unknown. The geometric uncertainty as a result of contouring variation is larger than that of set-up errors and organ motion for some tumour sites.[84] Weiss and Hess[84] describe the uncertainty because of contour variation as systematic and random for the individual but random for a population. An individual may consistently define larger or smaller volumes with some intra-observer variation, but for large numbers of patients and observers, these contribute to an overall random error. This becomes important in multi-centre clinical trials where a large variation in contouring processes between centres may impact on trial outcomes. While the 'true' tumour is unknown, it is intuitive that variation in contouring increases the probability of some geometric miss of the tumour, which will have a clinical impact.[86]

No clear relationship between the choice of metrics used and tumour site or any other factors was observed from the literature. This suggests that the choice of metrics in many studies is somewhat arbitrary and not determined on any established clinical basis. Further studies are needed to assess the impact of contouring variation. Inconsistencies in methods of contour comparison may be addressed through the implementation of consensus guidelines for analysis. These could be included as a software package not dissimilar to current clinical trials packages like SWAN.[87] Given the clear uncertainty in contouring, without established methods of contour comparison the impact of variation will be difficult to investigate on a large scale.

## Conclusion

There is no consistent or widely accepted method of systematic contour comparison. A number of contouring metrics exist; some of which are available in treatment planning systems and others require specialised software. Volume was the most frequently used metric across all tumour sites. Shape/dimension was the next most frequently used metric in all tumour sites except for brain and head and neck, where COV and CI were the next most frequent metrics, respectively. A number of different methods of calculating the same metrics were reported but many studies did not provide details on calculation methods. No clear relationships between tumour sites or any other factors that may influence the contouring process and the metrics used to compare contours were observed from the literature. This suggests that the choice of metrics in many studies is somewhat arbitrary and not determined on any established clinical basis. Further studies are needed to assess the advantages and disadvantages of each metric in various situations.

### References


1. Schnall M, Rosen M. Primer on imaging technologies for cancer. *J Clin Oncol* 2006; **24**: 3225.
2. Bernier J, Hall EJ, Giaccia A. Radiation oncology: a century of achievements. *Nat Rev Cancer* 2004; **4**: 737–47.
3. Wijesooriya K, Weiss E, Dill V *et al*. Qantifying the accuracy of automated structure segmentation in 4D CT images using a deformable image registration algorithm. *Med Phys* 2008; **35**: 1251–60.
4. Fitton I, Steenbakkers R, Gilhuijs K *et al*. Impact of anatomical location on value of CT-PET co-registration for delineation of lung tumors. *Int J Radiat Oncol Biol Phys* 2008; **70**: 1403–7.
5. Giraud P, Elles S, Helfre S *et al*. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother Oncol* 2002; **62**: 27–36.
6. Vorwerk H, Beckmann G, Bremer M *et al*. The delineation of target volumes for radiotherapy of lung cancer patients. *Radiother Oncol* 2009; **91**: 455–60.
7. Steenbakkers RJHM, Duppen JC, Fitton I *et al*. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis.[see comment]. *Int J Radiat Oncol Biol Phys* 2006; **64**: 435–48.
8. Spoelstra FOB, Senan S, Le Péchoux C *et al*. Variations in target volume definition for postoperative radiotherapy in stage III non-small-cell lung cancer: analysis of an international contouring study. *Int J Radiat Oncol Biol Phys* 2009; **76**: 1106–13.
9. Pevsner A, Davis B, Joshi S *et al*. Evaluation of an automated deformable image matching method for quantifying lung motion in respiration-correlated CT images. *Med Phys* 2006; **33**: 369–76.
10. Kepka L, Bujko K, Garmol D *et al*. Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy. *Radiother Oncol* 2007; **85**: 450–5.
11. Bowden P, Fisher R, Mac Manus M *et al*. Measurement of lung tumor volumes using three-dimensional computer planning software. *Int J Radiat Oncol Biol Phys* 2002; **53**: 566–73.
12. Song Y, Chan M, Burman C, Cann D. Inter-modality variation in gross tumor volume delineation in 18FDG-PET guided IMRT treatment planning for lung cancer. *Conference Proceedings:* Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2006.

13. Petersen RP, Truong PT, Kader HA *et al*. Target volume delineation for partial breast radiotherapy planning: clinical characteristics associated with low interobserver concordance. *Int J Radiat Oncol Biol Phys* 2007; **69**: 41–8.
14. Hurkmans CW, Borger JH, Pieters BR, Russell NS, Jansen EP, Mijnheer BJ. Variability in target volume delineation on CT scans of the breast. *Int J Radiat Oncol Biol Phys* 2001; **50**: 1366–72.
15. Landis DM, Luo W, Song J *et al*. Variability among breast radiation oncologists in delineation of the postsurgical lumpectomy cavity. *Int J Radiat Oncol Biol Phys* 2007; **67**: 1299–308.
16. Pitkanen MA, Holli KA, Ojala AT, Laippala P. Quality assurance in radiotherapy of breast cancer – variability in planning target volume delineation. *Acta Oncol* 2001; **40**: 50–5.
17. Struikmans H, Warlam-Rodenhuis C, Stam T *et al*. Interobserver variability of clinical target volume delineation of glandular breast tissue and of boost volume in tangential breast irradiation. *Radiother Oncol* 2005; **76**: 293–9.
18. Li XA, Tai A, Arthur DW *et al*. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study. *Int J Radiat Oncol Biol Phys* 2009; **73**: 944.
19. van Mourik AM, Elkhuizen PHM, Minkema D, Duppen JC, van Vliet-Vroegindeweij C. Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines. *Radiother Oncol* 2010; **94**: 286–91 (March 2010).
20. Weltens C, Menten J, Feron M *et al*. Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. *Radiother Oncol* 2001; **60**: 49–59.
21. Leunens G, Menten J, Weltens C, Verstraete J, Van der Schueren E. Quality assessment of medical decision making in radiation oncology: variability in target volume delineation for brain tumours. *Radiother Oncol* 1993; **29**: 169–75.
22. Buis DR, Lagerwaard FJ, Dirven CM *et al*. Delineation of brain AVMs on MR-Angiography for the purpose of stereotactic radiosurgery. *Int J Radiat Oncol Biol Phys* 2007; **67**: 308–16.
23. Cattaneo GM, Reni M, Rizzo G *et al*. Target delineation in post-operative radiotherapy of brain gliomas: interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans. *Radiother Oncol* 2005; **75**: 217–23.
24. Coles CE, Hoole AC, Harden SV *et al*. Quantitative assessment of inter-clinician variability of target volume delineation for medulloblastoma: quality assurance for the SIOP PNET 4 trial protocol. *Radiother Oncol* 2003; **69**: 189–94.
25. Horan G, Roques TW, Curtin J, Barrett A. 'Two are better than one': a pilot study of how radiologist and oncologists can collaborate in target volume definition. *Cancer Imaging* 2006; **6**: 16–9.
26. Mazzara GP, Velthuizen RP, Pearlman JL, Greenberg HM, Wagner H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *Int J Radiat Oncol Biol Phys* 2004; **59**: 300–12.
27. Yamamoto M, Nagata Y, Okajima K *et al*. Differences in target outline delineation from CT scans of brain tumours using different methods and different observers. *Radiother Oncol* 1999; **50**: 151–6.
28. Gao Z, Wilkins D, Eapen L, Morash C, Wassef Y, Gerig L. A study of prostate delineation referenced against a gold standard created from the visible human data. *Radiother Oncol* 2007; **85**: 239–46.
29. Song WY, Chiu B, Bauman GS *et al*. Prostate contouring uncertainty in megavoltage computed tomography images acquired with a helical tomotherapy unit during image-guided radiation therapy. *Int J Radiat Oncol Biol Phys* 2006; **65**: 595–607.
30. Seddon B, Bidmead M, Wilson J, Khoo V, Dearnaley D. Target volume definition in conformal radiotherapy for prostate cancer: quality assurance in the MRC RT-01 trial. *Radiother Oncol* 2000; **56**: 73–83.
31. Pasquier D, Lacornerie T, Vermandel M, Rousseau J, Lartigau E, Betrouni N. Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *Int J Radiat Oncol Biol Phys* 2007; **68**: 592–600.
32. Deurloo KEI, Steenbakkers RJHM, Zijp LJ *et al*. Quantification of shape variation of prostate and seminal vesicles during external beam radiotherapy.[erratum appears in Int J Radiat Oncol Biol Phys. 2005 Apr 1;61(5):1609]. *Int J Radiat Oncol Biol Phys* 2005; **61**: 228–38.
33. Nichol AM, Brock KK, Lockwood GA *et al*. A magnetic resonance imaging study of prostate deformation relative to implanted gold fiducial markers. *Int J Radiat Oncol Biol Phys* 2007; **67**: 48–56.
34. Court LE, Dong L, Taylor N *et al*. Evaluation of a contour-alignment technique for CT-guided prostate radiotherapy: an intra- and interobserver study. *Int J Radiat Oncol Biol Phys* 2004; **59**: 412–8.
35. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiother Oncol* 1998; **47**: 285–92.
36. Jeanneret-Sozzi W, Moeckli R, Valley JF *et al*. The reasons for discrepancies in target volume delineation: a SASRO study on head-and-neck and prostate cancers. *Strahlenther Onkol* 2006; **182**: 450–7.
37. Livsey JE, Wylie JP, Swindell R, Khoo VS, Cowan RA, Logue JP. Do differences in target volume definition in prostate cancer lead to clinically relevant differences in normal tissue toxicity? *Int J Radiat Oncol Biol Phys* 2004; **60**: 1076–81.

38. Nakamura K, Shioyama Y, Tokumaru S *et al*. Variation of clinical target volume definition among Japanese radiation oncologists in external beam radiotherapy for prostate cancer. *Jpn J Clin Oncol* 2008; **38**: 275–80.
39. Rasch C, Remeijer P, Koper PC *et al*. Comparison of prostate cancer treatment in two institutions: a quality control study. *Int J Radiat Oncol Biol Phys* 1999; **45**: 1055–62.
40. Valicenti RK, Sweet JW, Hauck WW *et al*. Variation of clinical target volume definition in three-dimensional conformal radiation therapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 1999; **44**: 931–5.
41. van der Wielen GJ, Mutanga TF, Incrocci L *et al*. Deformation of prostate and seminal vesicles relative to intraprostatic fiducial markers. *Int J Radiat Oncol Biol Phys* 2008; **72**: 1604–11.*e3*.
42. Villeirs GM, Van Vaerenbergh K, Vakaet L *et al*. Interobserver delineation variation using CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer.[see comment]. *Strahlenther Onkol* 2005; **181**: 424–30.
43. Gual-Arnau X, Ibanez-Gual MV, Lliso F, Roldan S. Organ contouring for prostate cancer: interobserver and internal organ motion variability. *Comput Med Imaging Graph* 2005; **29**: 639–47.
44. Cazzaniga LF, Marinoni MA, Bossi A *et al*. Interphysician variability in defining the planning target volume in the irradiation of prostate and seminal vesicles. *Radiother Oncol* 1998; **47**: 293–6.
45. Davis JB, Reiner B, Dusserre A, Giraud JY, Bolla M. Quality assurance of the EORTC trial 22911. A phase III study of post-operative external radiotherapy in pathological stage T3N0 prostatic carcinoma: the dummy run* 1. *Radiother Oncol* 2002; **64**: 65–73.
46. Debois M, Oyen R, Maes F *et al*. The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer. *Int J Radiat Oncol Biol Phys* 1999; **45**: 857–65.
47. Dubois DF, Prestidge BR, Hotchkiss LA, Prete JJ, Bice WS. Intraobserver and interobserver variability of MR imaging-and CT-derived prostate volumes after transperineal interstitial permanent prostate brachytherapy. *Radiology* 1998; **207**: 785.
48. Foppiano F, Fiorino C, Frezza G, Greco C, Valdagni R. The impact of contouring uncertainty on rectal 3D dose–volume data: results of a dummy run in a multicenter trial (AIROPROS01–02). *Int J Radiat Oncol Biol Phys* 2003; **57**: 573–9.
49. Matzinger O, Poortmans P, Giraud JY *et al*. Quality assurance in the 22991 EORTC ROG trial in localized prostate cancer: dummy run and individual case review. *Radiother Oncol* 2009; **90**: 285–90.
50. Mitchell DM, Perry L, Smith S *et al*. Assessing the effect of a contouring protocol on postprostatectomy radiotherapy clinical target volumes and interphysician variation. *Int J Radiat Oncol Biol Phys* 2009; **75**: 990–3.
51. Rasch C, Steenbakkers R, Fitton I *et al*. Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer. *Radiat Oncol* 2010; **5**: 21.
52. Rosewall T, Kong V, Vesprini D *et al*. Prostate delineation using CT and MRI for radiotherapy patients with bilateral hip prostheses. *Radiother Oncol* 2009; **90**: 325–30.
53. White EA, Brock KK, Jaffray DA, Catton CN. Inter-observer variability of prostate delineation on cone beam computerised tomography images. *Clin Oncol* 2009; **21**: 32–8.
54. Breen SL, Publicover J, De Silva S *et al*. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. *Int J Radiat Oncol Biol Phys* 2007; **68**: 763–70.
55. Rasch C, Eisbruch A, Remeijer P *et al*. Irradiation of paranasal sinus tumors, a delineation and dose comparison study. *Int J Radiat Oncol Biol Phys* 2002; **52**: 120–7.
56. O'Daniel JC, Rosenthal DI, Garden AS *et al*. The effect of dental artifacts, contrast media, and experience on interobserver contouring variations in head and neck anatomy. *Am J Clin Oncol* 2007; **30**: 191.
57. Rasch C, Keus R, Pameijer FA *et al*. The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer. *Int J Radiat Oncol Biol Phys* 1997; **39**: 841–8.
58. Riegel AC, Berson AM, Destian S *et al*. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. *Int J Radiat Oncol Biol Phys* 2006; **65**: 726–32.
59. Valley JF, Bernier J, Tercier PA *et al*. Quality assurance of the EORTC radiotherapy trial 22931 for head and neck carcinomas: the dummy run. *Radiother Oncol* 1998; **47**: 37–44.
60. Yamazaki H, Nishiyama K, Tanaka E *et al*. Dummy run for a phase II multi-institute trial of chemoradiotherapy for unresectable pancreatic cancer: inter-observer variance in contour delineation. *Anticancer Res* 2007; **27**: 2965–71.
61. Lotz HT, Pos FJ, Hulshof MCCM *et al*. Tumor motion and deformation during external radiotherapy of bladder cancer. *Int J Radiat Oncol Biol Phys* 2006; **64**: 1551–8.
62. Meijer GJ, Rasch C, Remeijer P, Lebesque JV. Three-dimensional analysis of delineation errors, setup errors, and organ motion during radiotherapy of bladder cancer. *Int J Radiat Oncol Biol Phys* 2003; **55**: 1277–87.
63. Logue JP, Sharrock CL, Cowan RA, Read G, Marrs J, Mott D. Clinical variability of target volume description in conformal radiotherapy planning. *Int J Radiat Oncol Biol Phys* 1998; **41**: 929–32.
64. Hoogeman MS, van Herk M, de Bois J, Muller-Timmermans P, Koper PCM, Lebesque JV. Quantification of local rectal wall displacements by

virtual rectum unfolding. *Radiother Oncol* 2004; **70**: 21–30.
65. Tai P, Van Dyk J, Yu E, Battista J, Stitt L, Coad T. Variability of target volume delineation in cervical esophageal cancer1. *Int J Radiat Oncol Biol Phys* 1998; **42**: 277–88.
66. Vesprini D, Ung Y, Dinniwell R *et al*. Improving observer variability in target delineation for gastro-oesophageal cancer – the role of 18ffluoro-2-deoxy-d-glucose positron emission tomography/computed tomography. *Clin Oncol* 2008; **20**: 631–8.
67. Dimopoulos JCA, Vos VD, Berger D *et al*. Inter-observer comparison of target delineation for MRI-assisted cervical cancer brachytherapy: application of the GYN GEC-ESTRO recommendations. *Radiother Oncol* 2009; **91**: 166–72.
68. Saarnak AE, Boersma M, van Bunningen B, Wolterink R, Steggerda MJ. Inter-observer variation in delineation of bladder and rectum contours for brachytherapy of cervical cancer. *Radiother Oncol* 2000; **56**: 37–42.
69. Weiss E, Richter S, Krauss T *et al*. Conformal radiotherapy planning of cervix carcinoma: differences in the delineation of the clinical target volume: a comparison between gynaecologic and radiation oncologists. *Radiother Oncol* 2003; **67**: 87–95.
70. McJury M, Dyker K, Nakielny R, Conway J, Robinson MH. Optimizing localization accuracy in head and neck, and brain radiotherapy. *Br J Radiol* 2006; **79**: 672–80.
71. van Sornsen de Koste JR, Senan S, Underberg RWM *et al*. Use of CD-ROM-based tool for analyzing contouring variations in involved-field radiotherapy for Stage III NSCLC. *Int J Radiat Oncol Biol Phys* 2005; **63**: 334–9.
72. Ashamalla H, Rafla S, Parikh K *et al*. The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer. *Int J Radiat Oncol Biol Phys* 2005; **63**: 1016–23.
73. Blodgett TM, Meltzer CC, Townsend DW. PET/CT: form and function. [Review] [267 refs. *Radiology* 2007; **242**: 360–85.
74. Greco C, Rosenzweig K, Cascini GL, Tamburrini O. Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). [Review] [45 refs]. *Lung Cancer* 2007; **57**: 125–34.
75. Balter JM, Kessler ML. Imaging and alignment for image-guided radiation therapy. *J Clin Oncol* 2007; **25**: 931–7.
76. Rao M, Stough J, Chi YY *et al*. Comparison of human and automatic segmentations of kidneys from CT images. *Int J Radiat Oncol Biol Phys* 2005; **61**: 954–60.
77. van der Put RW, Raaymakers BW, Kerkof EM, van Vulpen M, Lagendijk JJW. A novel method for comparing 3D target volume delineations in radiotherapy. *Phys Med Biol* 2008; **53**: 2149–59.
78. Ackerly T, Andrews J, Ball D, Healy B, Williams I. Discrepancies in volume calculations between different radiotherapy treatment planning systems. *Australas Phys Eng Sci Med* 2003; **26**: 90–2.
79. Juhler-Nottrup T, Korreman SS, Pederson AN *et al*. Interfractional changes in tumour volume and position during entire radiotherapy courses for lung cancer with respiratory gating and image guidance. *Acta Oncol* 2008; **47**: 1406–13.
80. Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. *Med Phys* 2003; **30**: 979.
81. Smith WL, Lewis C, Bauman G *et al*. Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR. *Int J Radiat Oncol Biol Phys* 2007; **67**: 1238–47.
82. Remeijer P, Rasch C, Lebesque JV, van Herk M. A general methodology for three-dimensional analysis of variation in target volume delineation. *Med Phys* 1999; **26**: 931–40.
83. Christensen GE, Carlson B, Chao KS *et al*. Image-based dose planning of intracavitary brachytherapy: registration of serial-imaging studies using deformable anatomic templates. *Int J Radiat Oncol Biol Phys* 2001; **51**: 227–43.
84. Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. *Strahlenther Onkol* 2003; **179**: 21–30.
85. Lu WG, Olivera GH, Chen Q, Chen ML, Ruchala KJ. Automatic re-contouring in 4D radiotherapy. *Phys Med Biol* 2006; **51**: 1077–99.
86. Engels B, Soete G, Verellen D, Storme G. Conformal arc radiotherapy for prostate cancer: increased biochemical failure in patients with distended rectum on the planning computed tomogram despite image guidance by implanted markers. *Int J Radiat Oncol Biol Phys* 2009; **74**: 388–91.
87. Ebert MA, Haworth A, Kearvell R *et al*. Detailed review and analysis of complex radiotherapy clinical trial planning data: evaluation and initial experience with the SWAN software system. *Radiother Oncol* 2008; **86**: 200–10.