

Physics Contribution

# 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

Mingyao Zhu, PhD,\* Karl Bzdusek, PhD,<sup>†</sup> Carsten Brink, PhD,<sup>‡,§</sup>  
Jesper Grau Eriksen, MD, PhD,<sup>||</sup> Olfred Hansen, MD,<sup>‡,||</sup> Helle Anita Jensen, MD,<sup>||</sup>  
Hiram A. Gay, MD,\* Wade Thorstad, MD,\* Joachim Widder, MD, PhD,<sup>¶</sup>  
Charlotte L. Brouwer, MSc,<sup>¶</sup> Roel J.H.M. Steenbakkers, MD, PhD,<sup>¶</sup>  
Hubertus A.M. Vanhauten, MD,<sup>¶</sup> Jeffrey Q. Cao, MD,<sup>#</sup> Gail McBrayne, RTT,<sup>#</sup>  
Salil H. Patel, MD,\*\* Donald M. Cannon, MD,<sup>††</sup>  
Nicholas Hardcastle, BSc (Hons), PhD,<sup>‡‡</sup> Wolfgang A. Tomé, PhD, FAAP,<sup>§§</sup>  
Matthias Guckenberger, MD,<sup>||||</sup> and Parag J. Parikh, MD\*

\*Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri; <sup>†</sup>Philips Healthcare, Fitchburg, Wisconsin; <sup>‡</sup>Institute of Clinical Research, University of Southern Denmark, Odense, Denmark; <sup>§</sup>Laboratory of Radiation Physics, Odense University Hospital, Odense, Denmark; <sup>||</sup>Department of Oncology, Odense University Hospital, Odense, Denmark; <sup>¶</sup>Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; <sup>#</sup>London Regional Cancer Centre, Ontario, Canada; \*\*Department of Radiation Oncology, Rhode Island Hospital, Providence, Rhode Island; <sup>††</sup>Department of Human Oncology, University of Wisconsin—Madison; <sup>‡‡</sup>Department of Physical Science, Peter MacCallum Cancer Centre, Melbourne, Australia; <sup>§§</sup>Montefiore Medical Center and Institute of Onco-Physics, Albert Einstein College of Medicine, Bronx, New York; and <sup>||||</sup>University of Würzburg, Department of Radiation Oncology, Würzburg, Germany

Received Apr 8, 2013, and in revised form Jul 11, 2013. Accepted for publication Aug 3, 2013.

Reprint requests to: Parag J. Parikh, MD, Department of Radiation Oncology, Washington University School of Medicine, 660 South Euclid Ave, Campus Box 8224, St. Louis, MO 63110. Tel: (314) 362-9703; E-mail: [pparikh@radonc.wustl.edu](mailto:pparikh@radonc.wustl.edu)

Conflict of interest: Karl Bzdusek, PhD, is employed by Philips Healthcare. Hubertus A.M. Vanhauten, MD, participated in an advertising review for Ferring and received lecture payment from Astellas. Nicholas Hardcastle, BSc (Hons), PhD, was employed under a Philips Radiation Oncology Systems grant awarded to Professor Wolfgang Tomé at the University of Wisconsin—Madison. Wolfgang A. Tomé, PhD, FAAP, has received research grants from the NCI, Philips Radiation Oncology Systems, and Accuray; payment for patents (planned, pending, or issued) from WI Alumni Research Foundation; royalties as sole author of

a monograph in mathematical physics titled *Path Integrals on Group Manifolds*; and serves on Scientific Advisory Board for ViewRay, Inc. Parag J. Parikh, MD, has received grants from ViewRay, Inc, and Philips Healthcare; equipment and technical support from ViewRay, Inc; grants/grants pending from Varian Medical, Philips Healthcare; and payment for lectures including service on speakers bureaus from Varian Medical.

Supplementary material for this article can be found at [www.redjournal.org](http://www.redjournal.org).

*Acknowledgment*—We thank Nesrin Dogan, Carol Johnson, Barb Millman, Wendy Schoen, and Nikolaj Jensen for their contributions to the review of the autocontours. We also thank Dr. Todd DeWees, Washington University School of Medicine in St Louis, for his helpful discussions on the data analysis.

## Summary

Smart Probabilistic Image Contouring Engine (SPICE), an atlas- and model-based autosegmentation tool, was quantitatively evaluated and clinically validated using computed tomography images in the head and neck, thorax, liver, and male pelvis regions from 125 patients at 7 institutions. Mean autosegmentation times were 3.1 to 11.1 minutes/patient. The quantitative evaluation showed good agreement between the autocontours and clinically drawn ones; expert review favored autosegmentation over manual contouring in 38 of the 40 structures validated.

**Purpose:** Clinical validation and quantitative evaluation of computed tomography (CT) image autosegmentation using Smart Probabilistic Image Contouring Engine (SPICE).

**Methods and Materials:** CT images of 125 treated patients (32 head and neck [HN], 40 thorax, 23 liver, and 30 prostate) in 7 independent institutions were autosegmented using SPICE and computational times were recorded. The number of structures autocontoured were 25 for the HN, 7 for the thorax, 3 for the liver, and 6 for the male pelvis regions. Using the clinical contours as reference, autocontours of 22 selected structures were quantitatively evaluated using Dice Similarity Coefficient (DSC) and Mean Slice-wise Hausdorff Distance (MSHD). All 40 autocontours were evaluated by a radiation oncologist from the institution that treated the patients.

**Results:** The mean computational times to autosegment all the structures using SPICE were 3.1 to 11.1 minutes per patient. For the HN region, the mean DSC was  $>0.70$  for all evaluated structures, and the MSHD ranged from 3.2 to 10.0 mm. For the thorax region, the mean DSC was 0.95 for the lungs and 0.90 for the heart, and the MSHD ranged from 2.8 to 12.8 mm. For the liver region, the mean DSC was  $>0.92$  for all structures, and the MSHD ranged from 5.2 to 15.9 mm. For the male pelvis region, the mean DSC was  $>0.76$  for all structures, and the MSHD ranged from 4.8 to 10.5 mm. Out of the 40 autocontoured structures reviews by experts, 25 were scored useful as autocontoured or with minor edits for at least 90% of the patients and 33 were scored useful autocontoured or with minor edits for at least 80% of the patients.

**Conclusions:** Compared with manual contouring, autosegmentation using SPICE for the HN, thorax, liver, and male pelvis regions is efficient and shows significant promise for clinical utility. © 2013 Elsevier Inc.

## Introduction

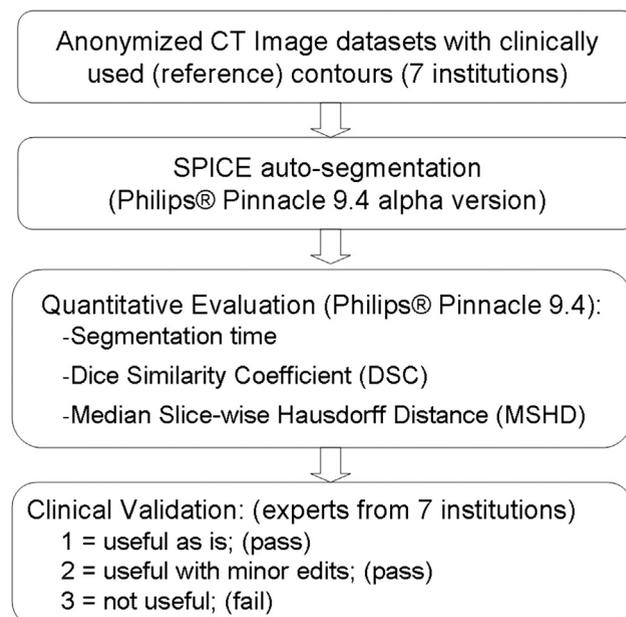
Modern radiation therapy allows for highly conformal dose distributions especially using techniques such as intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT). These methods offer the ability to deliver high therapeutic dose to the targets while significantly reducing the dose to nearby organs at risk (OARs), which may lead to a decrease in radiation related toxicity. An essential step in enabling these methods is to segment or delineate the organs at risk and target structures accurately. This procedure can be tedious, time-consuming, and subjective, and it requires a high level of concentration. A recent study documents that an average of 2.7 hours is required to fully contour a single head and neck (HN) patient for IMRT or VMAT (1). Organ delineation also suffers from inter- and intraobserver variability that may even exceed planning and setup errors and is considered the biggest uncertainty in radiation therapy planning (2).

Automatic segmentation aims to remove most of the manual work by providing initial structures that require minimal modifications for all patient images in a consistent manner. There are a number of semiautomatic and automatic approaches that are generally atlas-based, model-based, or a hybrid. Many approaches to automated segmentation have been published (3–6). Although technically these approaches are fully automated, the results are not always ideal, especially for the tissue types that have poor contrast relative to neighboring structures.

Commercial systems with autosegmentation have been available for the past several years, and there had been some reports on the evaluation of the performance of such software (7–11). Most of the evaluations were done for a single anatomical region, and sometimes within a single institution. La Macchia et al evaluated 3 commercial software (ABAS, MIM, and Velocity) for 15 patients treated for head-and-neck, prostate, and lung cancers and showed

that the automatic contouring can significantly shorten the adaptive radiation therapy workflow.

This article aims to evaluate a new commercial package using multiple data sets and users. This paper investigates the clinical effectiveness of an alpha version of this package by validating the automatic segmented structures using both common quantitative



**Fig. 1.** Flow chart of Smart Probabilistic Image Contouring Engine (SPICE) segmentation, quantitative evaluation, and clinical validation. CT = computed tomography; DSC = Dice Similarity Coefficient; MSHD = Mean Slice-wise Hausdorff Distance.

**Table 1** List of autocontoured structures and the expert review score ratios

Structure	Expert scoring results			Structure	Expert scoring results		
	Score 1 rate	Score 2 rate	Score 3 rate		Score 1 rate	Score 2 rate	Score 3 rate
<b>Head and Neck</b>							
<b>Brainstem</b>	21.2	72.7	6.1	Thyroid	9.4	50	40.6
<b>Spinal cord</b>	84.8	15.2	0	L optical nerve	12.5	53.1	34.4
<b>Brain</b>	68.2	31.8	0	R optical nerve	12.5	56.3	31.3
<b>L eye</b>	31.2	68.8	0	L lens	40.6	43.8	15.6
<b>R eye</b>	34.4	65.6	0	R lens	40.6	43.8	15.6
<b>Mandible</b>	51.5	48.5	0	Pharyngeal	28	72	0
<b>L parotid</b>	15.2	81.8	3	Glottis	10	55	35
<b>R parotid</b>	12.1	81.8	6.1	L cochlea	39.4	45.5	15.2
<b>L sub</b>	33.3	57.6	9.1	R cochlea	27.3	51.5	15.2
<b>R sub</b>	16	80	4	Oral cavity	21.2	78.8	0
L neck	0	0	100	Sublingual	20	70	10
R neck	0	4	96	Soft palate	12	80	8
Optical chiasm	22.6	41.9	35.5				
<b>Thorax</b>							
<b>L lung</b>	31.1	67.7	2.2	Trachea	77.8	22.2	0
<b>R lung</b>	37.8	62.2	0	Bronchi	60	37.8	2.2
<b>Spinal canal</b>	84.4	11.1	4.4				
<b>Spinal cord</b>	60	35.6	4.4				
<b>Heart</b>	55.6	40	4.4				
<b>Liver</b>							
<b>Liver</b>	39.1	60.9	0	L kidney			
<b>R kidney</b>	60	30	10				
<b>Male pelvis</b>							
<b>Bladder</b>	30.8	57.7	11.5	Seminal vesicle	38.5	42.3	19.2
<b>L femur head</b>	76.9	23.1	0				
<b>R femur head</b>	73.1	26.9	0				
<b>Rectum</b>	15.4	69.2	15.4				
<b>Prostate</b>	11.5	69.2	19.3				

Abbreviations: L = left; R = right; sub = submandibular glands.

Score 1, 2, and 3 correspond to useful as autocontoured, useful with minor edits, and not useful, respectively. The Dice Similarity Coefficient and Median Slice-wise Hausdorff Distance were calculated for selected structures (in bold) using clinical contours as references and the values are listed separately in [Table E2](#) and [E3](#) in the online supplement.

metrics (Dice Similarity Coefficient and Mean Slice-wise Hausdorff Distance) and qualitative expert scoring.

## Methods and Materials

### Autosegmentation software

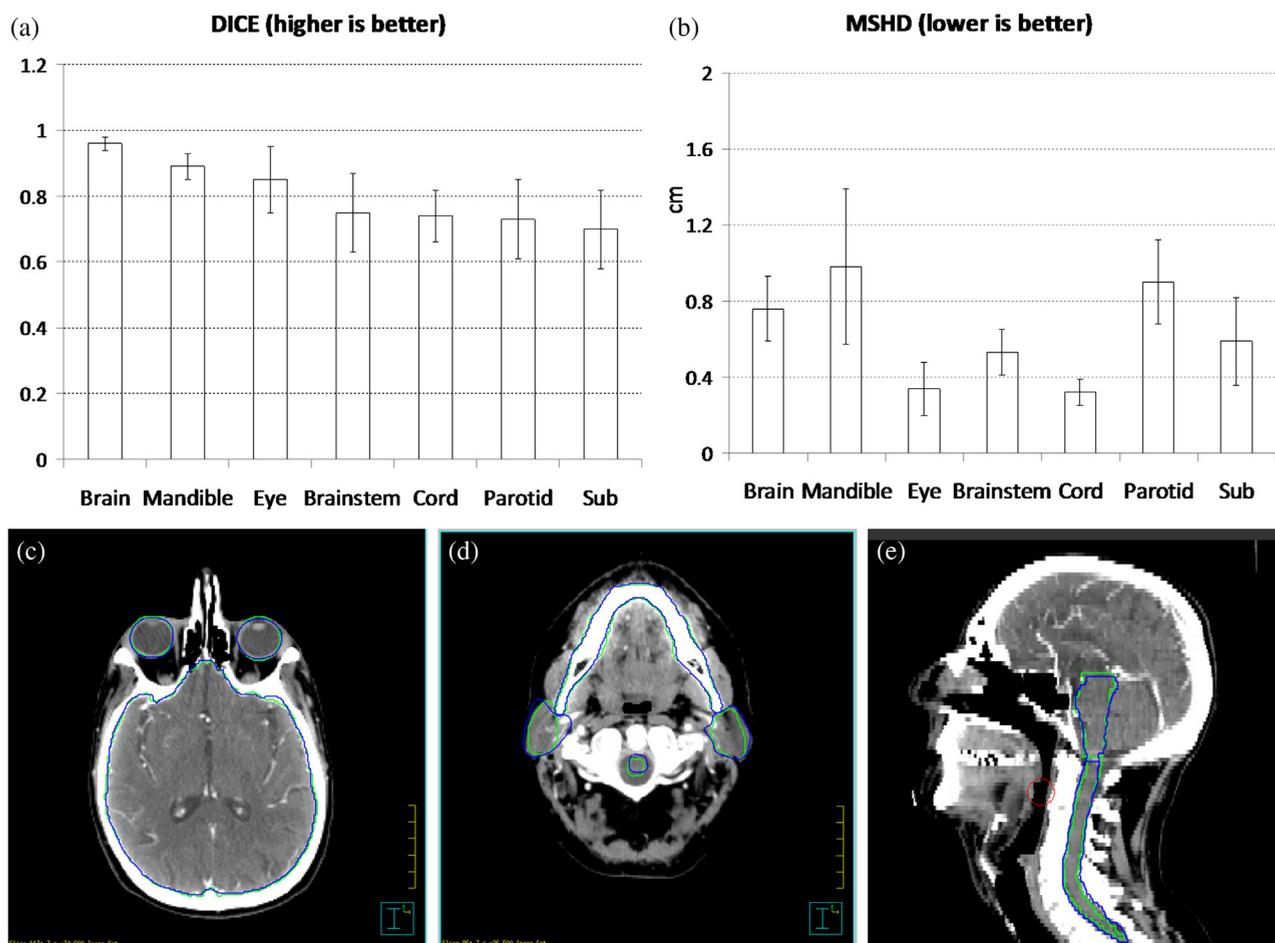
Recently, several fully automated, feature driven, model-based methods were developed to automatically segment the normal structures and target volumes in HN, thorax, liver, and male pelvis CT images (12, 13). These methods generally include multiple steps consisting of an initial registration, dense deformable registration, and refinement. A commercial treatment planning system (Pinnacle alpha, version 9.4, Philips, Madison, WI) with a new code package named Smart Probabilistic Image Contouring Engine (SPICE) was used for this study. SPICE integrates both an atlas- and model-based segmentation system into a joint segmentation system (the detailed steps are described in the [supplementary content](#)). Multiple variations of autocontours were provided for selected structures (eg, the parotids, submandibular glands, heart, and rectum) to accommodate the differences in the requirements of the 7 participating institutions.

### Patients and image data sets

One hundred twenty-five patient CT image data sets from 7 institutions were analyzed in this publication. There were 32, 40, 23, and 30 patients for the HN, thorax, liver, and male pelvis areas, respectively. Each institution participated independently and provided randomly selected patient data depending on the availability and expertise. [Table E1](#) shows the distribution of the patient number, scanner model and parameter, and scanning protocols for the thoracic studies. The patients were all clinically planned. The clinical structures were drawn based on each particular institution's guidelines and used as references. Because the structures were drawn before this study, no manual contouring time was recorded and not all patients had every structure of interest contoured. None of the patient images reviewed were used for training the algorithms.

### Autosegmentation and evaluation workflow

[Figure 1](#) outlines the overall workflow of this study. First, CT image data sets with original (clinically used) contours were collected and anonymized at each institution. Then the image datasets were sent to Philips, the SPICE autosegmentations were



**Fig. 2.** (a-b) Mean and standard deviation of Dice Similarity Coefficient (DSC) and Mean Slice-wise Hausdorff Distance (MSHD) of autocontours versus reference contours for head and neck area. (c-e) An example of the reference contours (green) and the autocontours (blue) for a head and neck patient image data set. Sub = submandibular.

performed remotely by M.Z. on a Pinnacle 9.4 alpha Enterprise server as an integrated functionality. The autosegmentation functionality is integrated in the Pinnacle treatment planning software and fully automated. After completing the CT data loading, the user only needs to select the atlas area (ie, HN, thorax, liver, or male pelvis), then Pinnacle starts the autosegmentation. There is no need for the user to specify the landmarks. The computational times were recorded as the machine running time from the selection of atlas area to the completion of autocontouring all structures for the selected anatomic area.

The SPICE algorithm provided autosegmentations for a series of targets and normal structures for each anatomic area, as listed in Table 1. These included 25 structures in the HN, 7 in the thorax, 3 in the liver, and 6 in the male pelvis areas, respectively. Following the completion of the autosegmentation, a script in Pinnacle calculates the Dice Similarity Coefficient (DSC) and Median Slice-wise Hausdorff Distance (MSHD) of the autocontours using the clinically used contours as a reference (discussed subsequently). Finally, the image data sets with autocontours were sent back to each institution for clinical validation review.

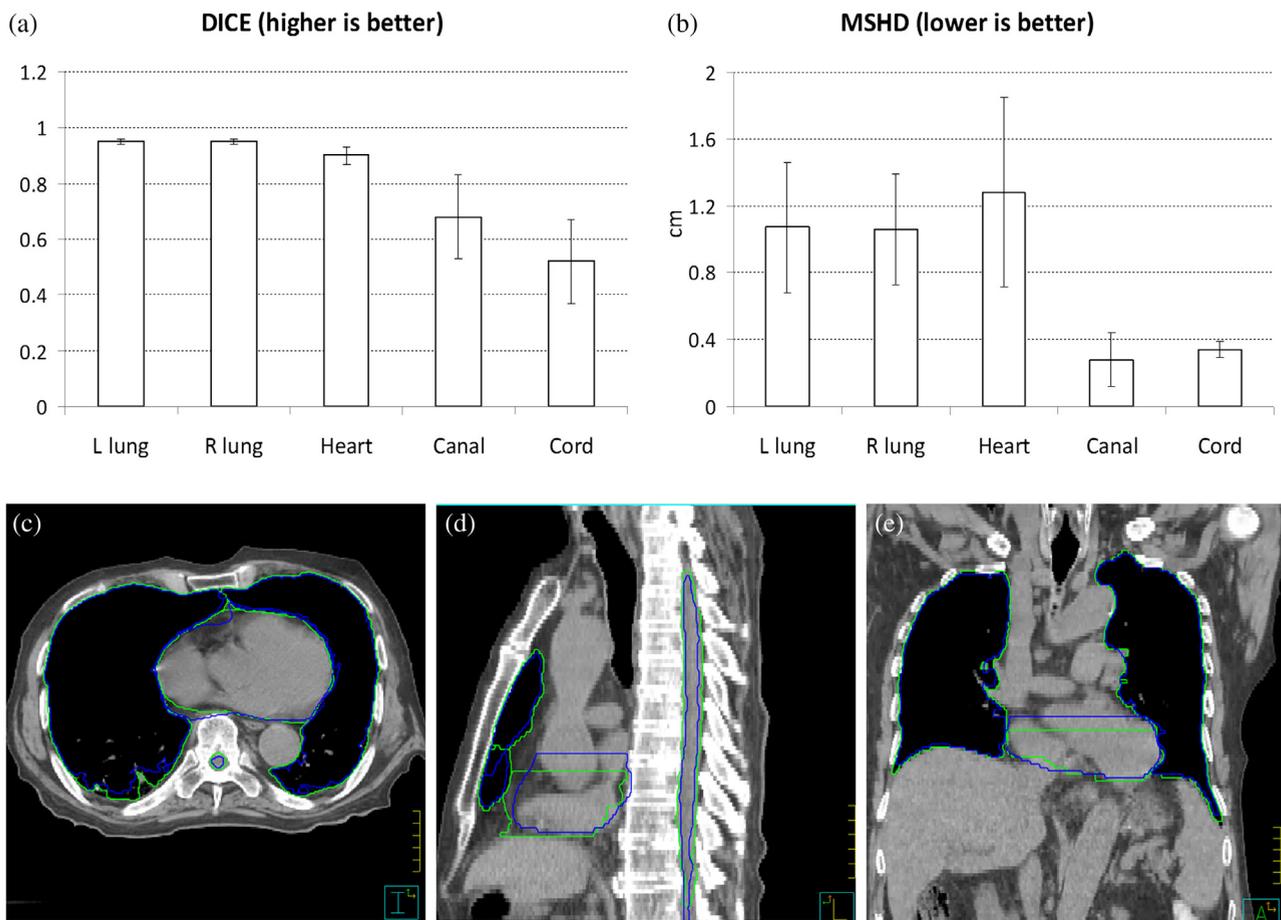
## Quantitative evaluation

Twenty-five selected structures (bold text in Table 1) were quantitatively evaluated based mainly on the availability of reference

contours. The structures that only have reference contours from a limited number of patients were not included for quantitative evaluation. However, autocontours of all 40 structures were clinically validated. To provide consistency with the Medical Image Computing and Computer Assisted Interventions (MICCAI) autosegmentation challenge publications (14, 15), the DSC and MSHD were computed using the same method as described in Pekar et al (15). In brief, the DSC represents the ratio of volume overlap between 2 contours but does not provide positional information. A higher DSC indicates better similarity of the 2 structures; a DSC of 1 means the 2 contours are identical and 0 means no overlap at all. The MSHD, on the other hand, is a measure of the distance disagreement between the 2 contours within slices. The smaller the MSHD, the closer the surfaces of the 2 contours, and an MSHD of 0 means the 2 contours are identical. These 2 metrics were commonly used in contouring studies (2, 9-11).

## Clinical validation

All autocontours were evaluated by experts from the original institution. Because of the large amount of patients and autocontours, the review was usually done by a single observer (radiation oncologist) for each case. In addition, radiation therapists and dosimetrists from Institution 1 also provided reviews for normal structures. The



**Fig. 3.** (a-b) Mean and standard deviation of Dice Similarity Coefficient (DSC) and Mean Slice-wise Hausdorff Distance (MSHD) of autocontours versus reference contours for thorax area. (c-e) An example of the reference contours (green) and the autocontours (blue) for a thoracic image data set.

reviewers were guided to turn off the original contour display, and score all autocontours using 3 levels: useful as autocontoured (= 1), useful with minor edits (=2), and not useful (=3). The definition of minor edits was that editing the SPICE autocontours was preferred over manual contouring, and the reviewers made the decisions based on their experiences and personal preferences.

## Results

### Computational time

The mean ( $\pm$  standard deviation [SD]) time to autosegment all structures was 11.1 ( $\pm$ 1.9) minutes per patient for the HN area, 3.1 ( $\pm$ 0.3) minutes per patient for the thorax area, 6.6 ( $\pm$ 0.2) minutes per patient for the liver area, and 3.7 ( $\pm$ 0.7) minutes per patient for the male pelvis area. The results are also shown in Table E2 and Figure E3 in the online supplement. Because this was a retrospective study, the times for creating the clinical contours were not available.

### HN area

Figure 2a-2b show the DSC and MSHD of the autocontoured structures in the HN area of all 32 patients. The DSC mean ( $\pm$ SD)

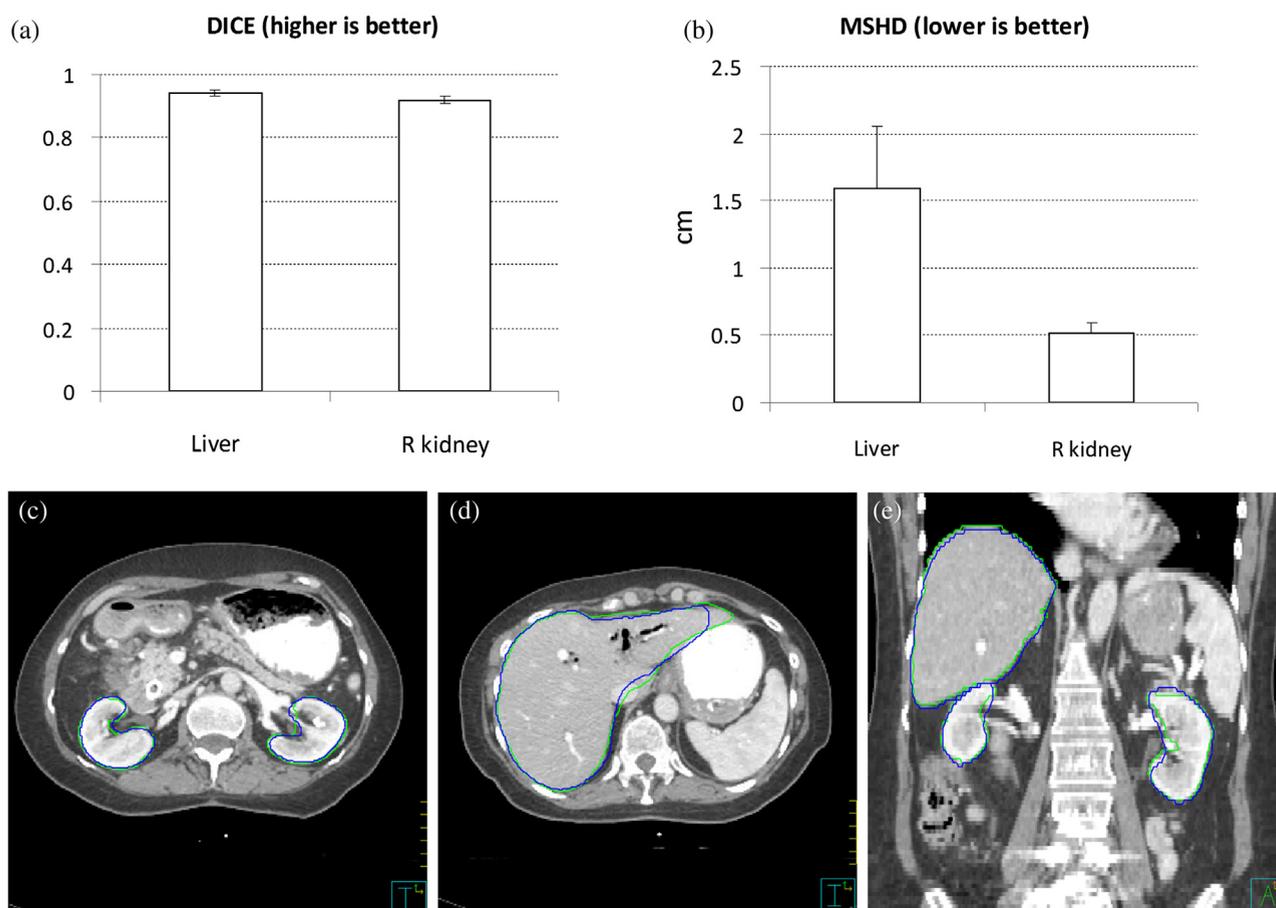
were 0.96 ( $\pm$ 0.02) for brain, 0.89 ( $\pm$ 0.04) for mandible, 0.85 ( $\pm$ 0.08) for eyes, 0.75 ( $\pm$ 0.12) for brainstem, 0.74 ( $\pm$ 0.08) for spinal cord, 0.70 ( $\pm$ 0.12) for submandibular nodes, and 0.72 ( $\pm$ 0.12) for parotids. Figure 2c-2e presents 3 examples of the autocontours (in blue) and the reference contours (in green) of 1 HN case. Table 1 lists the expert review score ratios of the autocontours of the selected structures.

### Thorax area

Figure 3a-3b show the DSC and MSHD of the autocontoured structures in the thorax area of 30 patients. The DSC mean ( $\pm$ SD) were 0.95 ( $\pm$ 0.01) for the left lung, 0.95 ( $\pm$ 0.00) for the right lung, 0.90 ( $\pm$ 0.03) for the heart, 0.68 ( $\pm$ 0.15) for the spinal canal, and 0.52 ( $\pm$ 0.15) for the spinal cord. Figure 3c-3e presents 3 examples of the autocontours (in blue) and the reference contours (in green) of 1 thorax case. Table 1 lists the expert review results of the autocontours of selected structures, and the score “3” (not useful) ratios are <5% for all structures.

### Liver area

Figure 4a-4b show the DSC and MSHD of the autocontoured structures in the liver area of all 23 patients. The DSC mean



**Fig. 4.** (a-b) Mean and standard deviation of Dice Similarity Coefficient (DSC) and Mean Slice-wise Hausdorff Distance (MSHD) of autocontours versus reference contours for liver area. (c-e) An example of the reference contours (green) and the autocontours (blue) for a liver patient image data set.

( $\pm$ SD) were 0.94 ( $\pm$ 0.01) for the liver, 0.92 ( $\pm$ 0.01) for the right kidney, and 0.85 ( $\pm$ 0.15) for the left kidney. Figure 4c-4e shows 3 examples of the autocontours (in blue) and the reference contours (in green) of 1 liver case. Table 1 lists the expert review score ratios: there is no score 3 for the liver and the right kidney scored 3 for 10% of the patients, although not enough left kidney clinical contours were available for evaluation.

### Male pelvis area

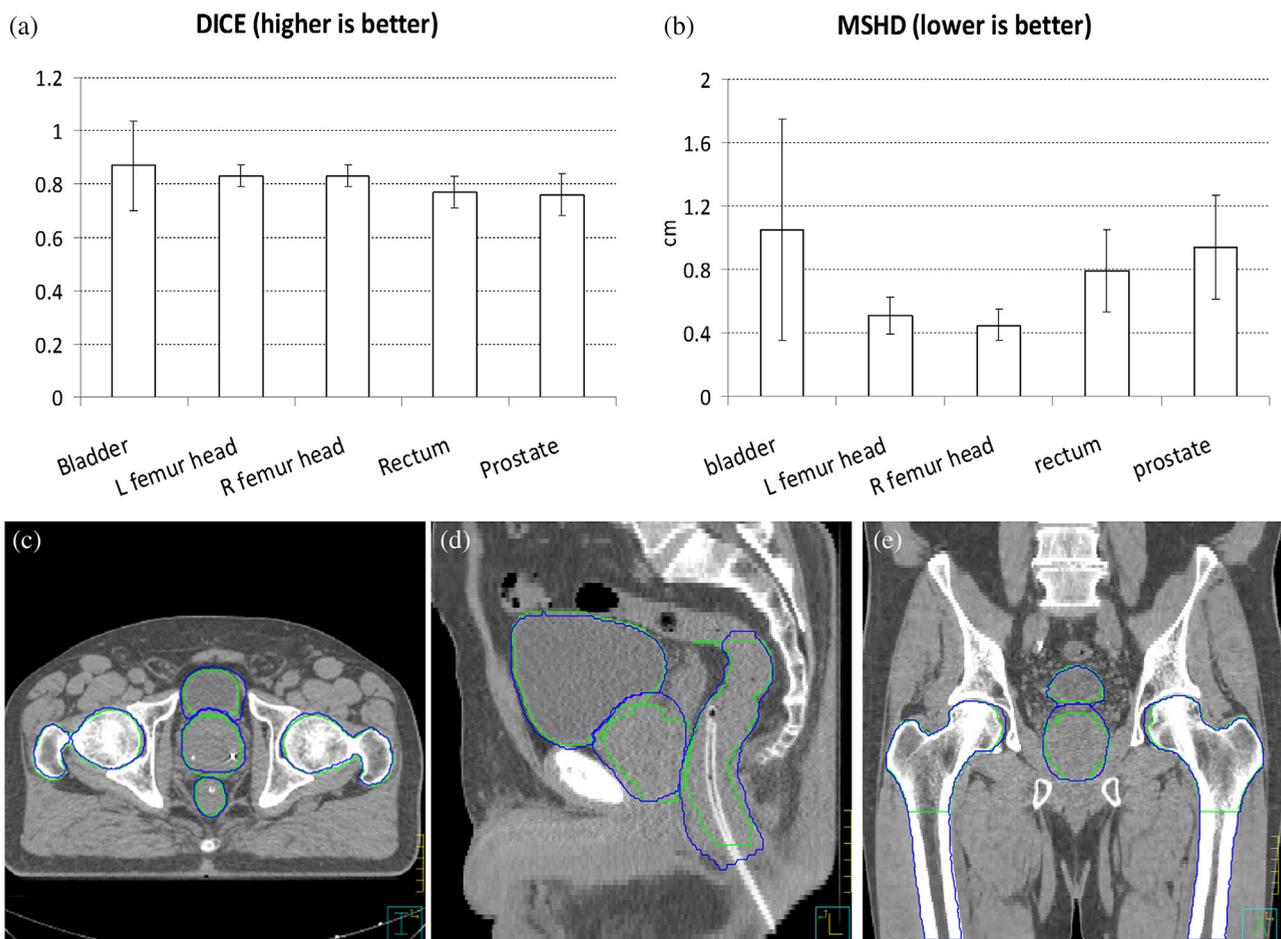
Figure 5a-5b show the DSC and MSHD of the autocontoured structures in the male pelvis area of all 30 patients. The DSC mean ( $\pm$ SD) were 0.87 ( $\pm$ 0.17) for the bladder, 0.83 ( $\pm$ 0.04) for the left and right femoral heads, 0.77 ( $\pm$ 0.06) for the rectum, and 0.76 ( $\pm$ 0.08) for the prostate. Figure 5c-5e presents 3 examples of the autocontours (in blue) and the reference contours (in green) of 1 male pelvis case. Table 1 lists the expert review results of the autocontours of selected structures. No femoral heads were scored 3, and the ratio of score 3 were <20% for the remaining 3 structures.

The detailed results of the DSC and MSHD of the data from all institutions and individual institutions are also listed in Tables E2 and E3 and plotted in Figure E4-E7 in the online supplement.

### Discussion

Target and normal structure manual contouring is not only time consuming but also contributes to uncertainty in modern radiation therapy treatment planning. Inter- and intraobserver variations of target and normal structure contouring have been reported for many anatomic areas (2, 16-19). Recent developments on the methods and algorithms for automated segmentation have improved both the efficiency and accuracy of organ delineation (10, 11). To date, the evaluation and clinical validation of commercial autosegmentation software tools have been limited to single anatomic areas (8-11) and often to patient images from a single institution (7, 9-11).

In this study, we assessed the autosegmentation performance of SPICE for 4 anatomic regions on CT images from 7 institutions (3 in Europe and 4 in North America). To our knowledge, this is the first multi-institutional and multianatomic area evaluation and validation of an autosegmentation tool. As described in the supplementary content and Figures E1 and E2, 2 pipelines were developed and included in the SPICE software package for the autosegmentation of the HN, thorax, liver, and male pelvis areas. Because both pipelines used in SPICE require a large amount of organ specific training and configuration, this process is not well suited for user customization. During this study, we noticed that for the normal structures, despite the recent publication of the



**Fig. 5.** (a-b) Mean and standard deviation of Dice Similarity Coefficient (DSC) and Mean Slice-wise Hausdorff Distance (MSHD) of autocontours versus reference contours for male pelvic area. (c-e) An example of the reference contours (green) and the auto-contours (blue) for a prostate patient image data set.

Radiation Therapy Oncology Group atlases and delineation guidelines (20-22), individual institutions still have their own contouring preferences and substantial differences may exist. Therefore, multiple variations of autocontours were provided for a subset of the structures (parotids, submandibular glands, heart, rectum, and femoral heads). In practice, each institution or physician could commission the software to output the contouring version that best matches their specific requirements. For this reason, we used the contour version with the highest DSC, lowest MSHD, and lowest expert review scores for our evaluation and validation in the present study.

The autosegmentation time by SPICE is 3 to 11 minutes, depending on the anatomic area and total number of structures delineated. This is in the same order of magnitude as reported by La Macchia et al (7) for 3 other software. For the organs that were commonly evaluated, Table E4 compares the DSC of SPICE (this work) to the values reported in reference (7). Again, the performance of SPICE is similar to ABAS, MIM, and Velocity, with some organ/structure slightly better and others slightly worse.

Because this is a retrospective study, the manual contouring times were not recorded. In general, manual contouring is a tedious and time-consuming process. Teguh et al reported an average contouring time of 180 minutes for HN CT images (11), Hermoye et al documented an average of 25 minutes for liver

manual contouring (23), and the study by Huyskens et al showed a mean time of 25 minutes to manually contour a prostate patient (8). In addition, manual contouring times for the HN, lung, and prostate of 163 minutes, 80 minutes, and 78 minutes were reported by La Macchia et al (7). Compared with these reported results, the computational time of SPICE autosegmentation is substantially lower and a background process, representing a great improvement in contouring efficiency.

There are shortcomings in our study. First of all, there is no “true structure” to compare with for the DSC and MSHD calculations. Because of the large amount of data, it was impractical to have multiple experts manually contour all patient data sets to generate the probabilistic estimate of the “true” contours. Instead, we decided to use the original contour that had actually been drawn for treatment planning purpose as the reference. Although this is not ideal, having input from 3 to 5 independent institutions for each anatomic area may have reduced this bias. A similar argument applies to the clinical validation of the autocontours because reviews of the autosegmentations were conducted by experts from 3 to 5 institutions.

The other limitation of this study is that we did not edit autocontours that needed minor correction and consequently could not evaluate the similarity between the edited autocontours and the reference contours, again because of the large amount of patient

data sets involved in this study. In the evaluation of the atlas-based auto segmentation on HN CT images, Teguh et al pointed out that editing autocontours took about one-third of the time as manually contouring the same structures (11). Time savings by auto-contouring software of about 1 hour for HN, 40 minutes for prostate, and 20 minutes for lung areas were recently reported by La Macchia et al (7). We feel it is reasonable to believe that the SPICE auto segmentation tool is more efficient than manual contouring, although editing of some autocontours may be needed.

As discussed in a review by Jameson et al, there is no single metric that can fully describe the similarity between 2 contours (2). In our quantitative evaluation, we used the DSC and the MSHD, 2 commonly used metrics that can provide us with complementary information. All the structures evaluated in this study had a mean DSC >0.7 except for the spinal cord and spinal canal in the thorax region. Despite the relatively low DSC, the MSHD for these structures were among the lowest of all structures evaluated, and 95.6% of these autocontours were useful as is or with minor edits. The low DSC could be caused by the elongated shape of the spinal cord and spinal canal, and even small mismatches in overlap can represent a relatively large percentage of the volume. Regarding the MSHD, the largest value observed (15.9 mm) was for the liver, while it had a DSC of 0.94 and all scored useful as autocontoured or with minor edits. The high MSHD could be attributed to the large volume of the liver. The high useful rate could indicate that the relative accuracy for this structure is not as critical as for other structures or that the necessary edits appeared relatively easy to perform.

As demonstrated by our evaluation results, there exists a good agreement between the SPICE autosegmented contours and the reference contours for a majority of structures. Furthermore, the expert reviews indicated that the autocontours (as is or with edits) were generally preferred over manual contouring, with the exception of the neck nodes and optic chiasm. Substantial improvements on these 2 structures in the product were made after this study; unfortunately, the evaluation results were not available in time for this review. The relatively short computation time of SPICE autocontouring is clearly an advantage over manual contouring. For these reasons, we feel that SPICE is an efficient and promising tool for clinical use.

## Conclusions

We performed SPICE autosegmentation for the HN, thorax, liver, and male pelvis regions, quantitatively evaluated the similarity between the autocontours and manual contours, and validated the autocontours by expert review. We demonstrated that using SPICE for autocontouring was efficient and promising for clinical utility.

## References

- Harari PM, Song S, Tomé WA. Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2010;77:950-958.
- Jameson MG, Holloway LC, Vial PJ, et al. A review of methods of analysis in contouring studies for radiation oncology. *J Med Imaging Radiat Oncol* 2010;54:401-410.
- Heimann T, van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging* 2009;28:1251-1265.
- Costa M, Delngette H, Novellas S, et al. Automatic segmentation of bladder and prostate using coupled 3d deformable models. *Med Image Comput Comput Assist Interv* 2007;10:252-260.
- Dowling JA, Fripp J, Chandra S, et al. Fast automatic multi-atlas segmentation of the prostate from 3D MR images. In: *MICCAI 2011 Workshop on Prostate Cancer Imaging: Image Analysis and Image-Guided Interventions* 2011;6963:10-21.
- Li W, Liao S, Feng Q, et al. Learning image context for segmentation of prostate in CT-guided radiotherapy. *Med Image Comput Assist Interv* 2011;14:570-578.
- 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* 2012;7:160.
- Huyskens DP, Maingon P, Vanuytsel L, et al. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. *Radiother and Oncol* 2009;90:337-345.
- Hwee J, Louie AV, Gaede S, et al. Technology assessment of automated atlas based segmentation in prostate bed contouring. *Radiat Oncol* 2011;6:110.
- Stapleford LJ, Lawson JD, Perkins C, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2010;77:959-966.
- Teguh DN, Levendag PC, Voet PWJ, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys* 2011;81:950-957.
- Qazi AA, Pekar V, Kim J, et al. Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach. *Medical Physics* 2011;38:6160-6170.
- Bzdusek KA, D B, Peters J, et al. Smart probabilistic image contouring engine (SPICE). *Philips Healthcare* 2012;June. (Available at: <http://www.google.com/#q=Smart+probabilistic+image+contouring>).
- Pekar V, Allaire S, Kim J, et al. Head and neck auto-segmentation challenge. *MIDAS J Head and Neck Auto-Segmentation Challenge* 2009. Available at: <http://www.midasjournal.org/download/viewpdf/703/2>. Accessed on April 6, 2013.
- Pekar V, Allaire S, Qazi AA, et al. Head and neck auto-segmentation challenge: Segmentation of the parotid glands Medical image analysis for the clinic: A grand challenge. Seattle, WA: CreatSpace; 2010:273-280.
- Steenbakkens RJHM, Duppen JC, Fitton I, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: A "big brother" evaluation. *Radiother Oncol* 2005;77:182-190.
- Cooper JS, Mukherji SK, Toledano AY, et al. An evaluation of the variability of tumor-shape definition derived by experienced observers from CT images of supraglottic carcinomas (ACRIN Protocol 6658). *Int J Radiat Oncol Biol Phys* 2007;67:972-975.
- Brouwer CL, Steenbakkens RJ, van den Heuvel E, et al. 3D variation in delineation of head and neck organs at risk. *Radiat Oncol* 2012;7:32.
- Fotina I, Lütgendorf-Caucig C, Stock M, et al. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. *Strahlenther Onkol* 2012;188:160-167.
- Gay HA, Barthold HJ, O'Meara E, et al. Pelvic normal tissue contouring guidelines for radiation therapy: A radiation therapy oncology group consensus panel atlas. *Int J Radiat Oncol Biol Phys* 2012;83:e353-e362.
- Kong F-M, Quint L, Machtay M, et al. Atlas for organs at risk (OARS) in thoracic radiation therapy. Available at: <http://www.rtog.org/CoreLab/ContouringAtlases/LungAtlas.aspx>. Accessed on April 6, 2013.
- van de Water TA, Bijl HP, Westerlaan HE, et al. Delineation guidelines for organs at risk involved in radiation-induced salivary dysfunction and xerostomia. *Radiother Oncol* 2009;93:545-552.
- Hermoye L, Laamari-Azjal I, Cao Z, et al. Liver segmentation in living liver transplant donors: Comparison of semiautomatic and manual methods. *Radiology* 2005;234:171-178.