



Published in final edited form as:

Radiother Oncol. 2017 January ; 122(1): 66–71. doi:10.1016/j.radonc.2016.11.016.

Cardiac Atlas Development and Validation for Automatic Segmentation of Cardiac Substructures

Rongrong Zhou, MD, PhD^{a,d}, Zhongxing Liao, MD^a, Tinsu Pan, PhD^b, Sarah A Milgrom, MD^a, Chelsea C Pinnix, MD, PhD^a, Anhui Shi, MD^a, Linglong Tang, MD^a, Ju Yang, MD^a, Ying Liu, MD^a, Daniel Gomez, MD^a, Quynh-Nhu Nguyen, MD^a, Bouthaina S Dabaja, MD^a, Laurence Court, PhD^c, and Jinzhong Yang, PhD^c

^aDepartment of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas

^bDepartment of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas

^cDepartment of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas

^dDepartment of Radiation Oncology, Xiangya Hospital, Central South University, Changsha, Hunan, China

Abstract

Purpose—To develop and validate a set of atlases for auto-contouring cardiac substructures.

Methods—Eight radiation oncologists manually and independently delineated 15 cardiac substructures from noncontrast CT images of 6 patients by referring to their respective fused contrast CT images. Individual contours were fused together for each structure, edited by 2 physicians, and became atlases to delineate other 6 patients. The auto-delineated contours of the 6 additional patients became templates for manual contouring. These 12 patients with well-defined contours composed the final atlases for multi-atlas segmentation.

Results—The average time for manually contouring the 15 cardiac substructures was about 40 minutes. Inter-observer variability was small for the heart, the chambers, and the aorta compared with that for other structures that were not clearly distinguishable in CT images. The mean Dice similarity coefficient and mean surface distance of auto-segmented contours were within one standard deviation of expert contouring variability. Good agreement between auto-segmented and manual contours was observed for the heart, the chambers, and the great vessels. Independent validation on other 19 patients showed reasonable agreement for the heart chambers.

Corresponding Author: Jinzhong Yang, PhD, Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Unit 1420, Houston, TX 77030. Tel: (713) 792-2814; jyang4@mdanderson.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Conclusions—A set of cardiac atlases was created for auto-contouring from noncontrast CT images. The accuracy of auto-contouring for the heart, chambers, and great vessels was validated for potential clinical use.

Keywords

Cardiac atlas; automatic contouring; multi-atlas segmentation; radiotherapy

Introduction

Radiotherapy has improved the overall survival of cancer patients; however, radiation-induced cardiac side effects, which often manifest years after treatment, have been shown to offset this improvement [1–4]. Cardiac toxicity has been seen in patients treated for lymphoma, breast cancer, and lung cancer [5–9]. Cardiac exposure was shown a negative prognostic factor for patients with lung cancer after concurrent chemotherapy and radiotherapy. In the most recent report of a phase III trial, the volume of the heart received as low as 5 Gy negatively impacted on the overall survival [10]. Cardiac injury is not limited to the myocardium. Radiation may also damage the vascular endothelium and capillary vessels resulting in peripheral, coronary, and carotid artery disease [11]. Most previous reports used the whole heart as a single region of interest; however, the relationship between dose to cardiac substructures and subsequent toxicity is of great interest and has not been well defined, mainly due to the inconsistency and demanding of time in cardiac substructure delineation from computed tomography (CT) images [2].

Manual contouring of cardiac substructures from CT images requires clinical knowledge, time, and effort. Cardiac substructures are not clearly distinguishable in the noncontrast CT normally used for treatment planning. Although cardiac substructures are more visible on magnetic resonance images and contrast CT images, these image modalities are usually unavailable for treatment planning. In addition, cardiac and respiratory motions and anatomical variations also present a great challenge to defining cardiac substructures. Furthermore, manual contouring of structures with inadequate image contrast, like cardiac substructures, can be unreliable because of substantial inter-observer variability [12,13].

In this study, we developed a set of cardiac atlases with well-defined cardiac substructure contours for multi-atlas segmentation. We validated the auto-contouring using these cardiac atlases by comparing auto-segmentation with manual delineation and inter-observer variability. We demonstrated that accurate and consistent cardiac contours could be generated with this development.

Methods and Materials

Patient data

This study was approved by our institutional review board. First, 6 patients with malignancies located in the thorax were retrospectively identified as the first group of patients for this study. The selection criteria was that they all had both contrast and noncontrast diagnostic chest CT scans acquired in the same imaging session and the

noncontrast CT images could be reconstructed the same display field of view similar to those used for thoracic radiation treatment planning. These 6 patients were used as the development dataset to generate gold standard cardiac atlas. As a regular procedure, the images were acquired while the patients held their breath and had their arms above their heads. A pre-contrast CT image was acquired first, and with the patient still on the table, a post-contrast CT image was acquired about 3 minutes later. For post-contrast CT image acquisition, systolic- and diastolic-phase images were scanned and averaged. Next, 25 patients with non-small cell lung cancer (NSCLC) treated with definitive chemoradiation were retrospectively identified for this study. They were selected randomly without special restrictions on gender, heart anatomy, or body anatomy. Patients with significant lung necrosis or lung collapse were excluded though. Treatment simulation was performed with the patients' arms above their heads in the supine position and immobilization with custom Vac-Lok cradles (Civco Medical Solutions, Kalona, IA). All 25 patients underwent four-dimensional CT (4DCT) for treatment planning, and the averaged 4DCT image was used for planning CT, on which the cardiac contours were delineated. All CT images in two groups had an in-slice resolution of 1.0 mm and slice spacing of 2.5 mm. The CT image quality is normal in our current clinical practice.

Cardiac atlas development

Combined previous studies [14,15] and our own experience, we found that a total of 12 atlases can be sufficient for obtaining robust multi-atlas segmentation for thoracic CT images. Because of the limitation of available contrast CT images, here we developed 12 cardiac atlases in two phases. First, contours were drawn manually on the first group of 6 patients with both contrast and noncontrast CT images. Because cardiac substructures are almost indistinguishable in noncontrast CT images, contrast CT is considered gold standard for delineation of the heart substructures and can help reduce ambiguity in manual contouring. Second, the set of atlases was expanded to include a second group of 6 additional patients. Overall atlas development process was described in Figure 1.

Phase I: manual contouring—The diagnostic images of first group of 6 patients were imported into the Pinnacle treatment planning system (Philips Medical Systems, Fitchburg, WI) for manual contouring. The contrast CT image was fused with the noncontrast CT image using rigid-registration in Pinnacle. Manual contouring was performed by 8 radiation oncologists: 2 specialists in thoracic cancer radiotherapy, 2 specialists in lymphoma radiotherapy, and 4 radiation oncologists trained outside the United States with various levels of experience. Before manual delineation, the 8 radiation oncologists reviewed the RTOG (Radiation Therapy Oncology Group) 1106 organ-at-risk contouring guideline [16] and a published cardiac atlas consensus contouring guideline [17] as a group. Each of the 8 oncologists manually and independently delineated 15 cardiac structures on noncontrast CT images by referring to the fused contrast CT image and Netter's Atlas of Human Anatomy [18] and by following the aforementioned contouring guidelines. The 15 delineated structures were the whole heart, the 4 heart chambers (left atrium, right atrium, left ventricle, and right ventricle), 4 coronary arteries (left main coronary artery, left anterior descending artery, left circumflex artery, and right coronary artery), and 6 great vessels (superior vena cava, inferior vena cava, pulmonary artery, pulmonary vein, ascending aorta with the aortic

arch, and descending aorta). The contours of each structure delineated by the 8 radiation oncologists were fused to one using the simultaneous truth and performance level estimation (STAPLE) algorithm for each patient [19]. The fused contours were further reviewed and edited by 2 radiation oncologists to ensure consistency across patients.

Phase II: atlas expansion—The noncontrast CT images of the first group of 6 patients and the modified contours were used in conjunction with the in-house Multi-Atlas Contouring Service (MACS) software (detailed in the next section) to automatically delineate the cardiac structures of 6 patients that were randomly selected from the 25 NSCLC patients. The auto-segmented contours for the second group of 6 patients were used as templates for manual contouring, and 2 radiation oncologists jointly contoured these 6 patients on the basis of the templates. Thus, a total of 12 patients, including the noncontrast CT images with well-defined contours, served as the final atlases for multi-atlas segmentation.

Multi-atlas segmentation

The in-house MACS software was used to perform multi-atlas segmentation. MACS has a user interface in the Pinnacle treatment planning system. For a new image to be segmented, users can submit the request in Pinnacle. MACS processes segmentation on a central server. First, each atlas was separately registered to the new image using dual-force Demons deformable registration [20]. The resultant deformation vector fields that characterized the individual deformable registration were used to deform the contours in each atlas to obtain individual segmentations [21]. Finally, we used the STAPLE algorithm with a built-in tissue appearance model [22] to combine the individual segmentations, generating a fusion contour that approximated the true segmentation. The contour fusion process allowed us to minimize variations among segmentations obtained from different atlases and random errors in deformable image registration. The MACS server sends the final fusion contour back to Pinnacle to complete the auto-segmentation. The auto-segmentation in MACS is independent to treatment planning systems.

Evaluation metrics

Let R and T represent two binary volumes and ∂R and ∂T represent the surface of these two volumes, respectively. The surface was generated from the binary volume using the marching cube algorithm from Insight Toolkit software [23] and represented by a triangular mesh with vertices resampled to be approximately 3 within 1 mm³ space. We calculated the Dice similarity coefficient (DSC) for volumes R and T as:

$$DSC(R, T) = \frac{2|\partial R \cap \partial T|}{|\partial R| + |\partial T|}.$$

The DSC has a value between 0 and 1, with 1 indicating perfect agreement and 0 indicating no overlap. We also calculated the symmetric mean surface distance (MSD) between the two surfaces ∂R and ∂T [24,25]:

$$MSD(\partial R, \partial T) = \frac{\sum_{r \in \partial R} \min_{t \in \partial T} d(r, t) + \sum_{t \in \partial T} \min_{r \in \partial R} d(t, r)}{|\partial R| + |\partial T|}.$$

where $d(\cdot, \cdot)$ stands for the Euclidian distance of two points and $|\cdot|$ denotes the number of points on a surface.

Atlas validation

Inter-observer variability—Inter-observer variability was evaluated by comparing individual expert contours with the STAPLE fused contours for the first group of 6 patients. For each structure in each patient, the 8 expert contours were compared with the fused contours individually using DSC and MSD metrics. For each structure, the mean and standard deviation of 48 DSC and 48 MSD values were calculated to measure inter-observer variability. At the same time, we performed leave-one-out auto-segmentation using the fused contours without any modification for the 6 patients and evaluated the auto-segmented contours against the fused contours using DSC and MSD metrics. The mean and standard deviation of DSC and MSD were calculated for the 6 tests. Auto-segmentation was compared with inter-observer variability to evaluate the performance of auto-contouring using the atlases.

Independent validation—The remaining 19 of the 25 NSCLC patients were used for independent validation. Another radiation oncologist, who did not participate in the manual contouring for the atlases, independently delineated the 4 heart chambers on the averaged 4DCT image without any reference. She also followed the published cardiac contouring guideline [17]. In parallel, we used the 12 atlases for multi-atlas segmentation to automatically delineate the 4 chambers for the 19 patients. The auto-segmented contours were compared with the manual contours using DSC and MSD to evaluate the agreement. In addition, the volume of auto-segmented contours was compared to that of the manual contours and the difference between them was tested for statistical significance using the Wilcoxon signed ranks test in SPSS version 23.0 (SPSS, Chicago, IL, USA), which allows for non-normally distributed volumes without the assumption of normality.

Results

Cardiac atlases

The final cardiac atlases were composed of 12 patients: the first group of 6 patients with diagnostic CT scans and the second group of 6 NSCLC patients. The first group included the noncontrast CT image and the edited fused contours, while the second group included the averaged 4DCT image and the edited auto-segmented contours. The atlases were used together with MACS to automatically delineate cardiac substructures in the Pinnacle treatment planning system. Figure 2 shows one atlas with contours overlaid on the contrast and noncontrast CT images. The average time for manually contouring the 15 cardiac substructures was about 40 minutes (range: 35 to 50 minutes), with the contrast CT available for reference. The average time for the MACS to automatically delineate the 15 cardiac

substructures was about 10 minutes per patient, with the MACS running on a Microsoft Windows PC with an eight-core CPU (2 Intel Xeon X5472 quad-core processors at 3GHz) and 8GB of memory.

Inter-observer variability

Figure 3 shows a comparison of inter-observer variability in manual contouring with auto-segmentation for the first group of 6 patients. Inter-observer variability was smaller for the heart, the 4 chambers, and the aorta than for other structures that were not clearly distinguishable on the CT image, such as the pulmonary vein and coronary arteries. The mean DSCs for individual expert contours relative to the fused contours were less than 0.5 for coronary arteries and the pulmonary vein, and the mean MSDs were greater than 4.0 mm, indicating a very large inter-observer variability. The largest MSD for deviation of expert contours from the fused contours was 5.4 cm for the pulmonary vein (Supplementary Tables 1 and 2). This disagreement is mostly due to the complex shape of these structures and their low contrast on the CT images. The interface of pulmonary vein to heart chambers has the most disagreement among experts. The mean DSCs and mean MSDs of auto-segmented contours were within one standard deviation of the mean DSCs and mean MSDs of expert contouring variability except for the right coronary artery (Supplementary Table 3). These findings demonstrated that auto-segmentation of most cardiac structures was at least comparable to manual delineation. Supplementary figure A1 shows examples of inter-observer variability for the heart and the 4 chambers. Contouring variability can be clearly seen for different structures. For example, the most variable heart contours were in the superior region.

Independent validation

For the 19 independent-validation patients, the DSCs between the manual and auto-segmented contours (mean \pm standard deviation) for the left atrium, right atrium, left ventricle, and right ventricle were 0.77 ± 0.07 , 0.72 ± 0.14 , 0.79 ± 0.13 , and 0.70 ± 0.15 , respectively; MSDs were 3.8 ± 1.4 , 4.9 ± 3.3 , 4.5 ± 3.2 , and 5.4 ± 3.7 mm, respectively. Supplementary figure A2 shows some comparisons between the manual and auto-segmented contours. Reasonable agreement was observed, if considering the inter-observer variability because the manual contours were drawn by an independent radiation oncologist without reference to contrast CT images. Detailed results are listed in Supplementary Table 4. The volume of auto-segmented contours (V2) were compared with that of manual contours (V1) in the Supplementary Table 5. The average volume ratio (V2/V1) was 0.87, 0.97, 1.06, and 1.27 for the left atrium, right atrium, left ventricle, and right ventricle, respectively. The volume of the right atrium and the left ventricle did not show statistically significant difference between the auto-segmented and manual contours ($P=0.469$ and $P=0.334$, respectively); however, it showed statistically significant difference for the left atrium and the right ventricle ($P=0.003$ and $P=0.000$, respectively).

Discussion

In this study we developed a set of cardiac atlases to be used with multi-atlas segmentation for auto-contouring cardiac substructures. We showed that manual contouring has high inter-

observer variability and validated auto-contouring using the atlases by comparing auto-segmentation with manual delineation and inter-observer variability. We have demonstrated that accurate and consistent contours can be automatically delineated for cardiac substructures except for coronary arteries.

Our results have clinically significant implications. First, the cardiac atlases can aid treatment planning in radiotherapy and potentially save time for clinicians in cardiac delineation. Our results have shown that auto-segmentation of most cardiac substructures was at least comparable to manual delineation. Accurate and consistent contours can also improve heart dose sparing in treatment planning and thereby improve treatment outcomes. Second, accurate and consistent cardiac substructure contouring can facilitate volume dose analysis, allowing assessment of radiation doses to specific cardiac structures for toxicity analysis to achieve cardiac risk control [2]. Recent efforts have shown that with the advances of radiation therapy planning avoiding heart structures is possible and desirable, especially for young patients with a high cure rate and long-term survival like lymphoma patients [8]. A standard and consistent heart contour is also important to reporting and comparison of the heart dose volume histogram and heart toxicities [10].

Validation of auto-segmentation is frequently subject to the impact of inter-observer variability when manual contours are used as the ground truth [26–28]. The relatively low agreement shown in Supplementary Table 4 mostly owes to the large inter-observer variability, particularly when the manual contours were drawn without referring to corresponding contrast CT. In this study, we performed an inter-observer variability analysis, used consensus contours as the ground truth to evaluate auto-segmented contours, and compared auto-segmentation with inter-observer variability to evaluate auto-segmentation. We thereby essentially minimized the impact of human factors in manual contouring and evaluated auto-segmentation for practical clinical use as an alternative to traditional manual contouring.

Our study had some limitations. The large inter-observer variability indicates that the manual contours were far from perfect. The major cause of the uncertainty in manual contouring may be cardiac motion. The contrast CT used in this study had both systolic- and diastolic-phase images, but the noncontrast CT did not. In general, cardiac-phase images are not available for treatment planning. Due to cardiac motion, fusion of contrast and noncontrast CT images may not be perfect. This imperfect fusion may have a considerable effect on contouring of small structures such as coronary arteries because a small deviation in fusion may cause a large discrepancy in contouring. On the other hand, even with contrast CT, the boundaries of some structures, such as coronary arteries, were still not clearly discernible. Contouring those structures still relied a lot on individual judgment, thereby resulting in large inter-observer variability in manual contouring. Accurate anatomical identification of coronary arteries therefore remains challenging and will be a major barrier for auto-contouring.

Image resolution may limit the capability of auto-contouring the small structures as well. Coronary arteries may not be discernible in older planning CT scans with a large voxel size. Auto-contouring is also subject to the effects of inter-patient variation. A tumor close to the

heart can significantly change the shape and appearance of a cardiac structure in CT images. Our auto-segmentation relies on accurate matching of the anatomy between atlas images and the new image for segmentation. If a large difference exists, the matching is not accurate. We indeed found that when a large tumor was near the heart, the accuracy of auto-contouring was low.

The choice of metrics can also affect evaluation results. DSC-based evaluations are known to be favorable for large volumes. It is not surprising to see larger DSCs for the heart, chambers, or great vessels than for coronary arteries. Therefore, distance-based metrics should also be used. However, distance-based metrics are subject to the effect of voxel size. Images with higher resolution (or smaller voxel size) tend to produce better distance-based evaluation for auto-segmentation. In our study, all images had slice spacing of 2.5 mm, which may limit distance-based assessment. By considering the limitation of image resolution, our auto-segmentation approach is comparable to state-of-the-art cardiac segmentation methods [29–31], which were applied to contrast CT images with sub-millimeter resolution for diagnosis of cardiovascular diseases. On the other hand, this showed that the atlases we developed are useful for auto-contouring the cardiac substructures.

Geometric evaluation measured by the DSC or MSD serves directly for the purpose to validate the accuracy of auto-segmentation. On the other hand, dosimetric evaluation by comparing the dose volume difference between the manual and auto-segmented contours may be able to evaluate the effect of contour disagreement to the treatment plan, but it is not able to validate the accuracy of auto-segmentation. A large geometric discrepancy without any absorbed dose in the region may show no difference in dosimetric evaluation. In general, dosimetric evaluation depends on the treatment location and is treatment specific, while auto-segmentation is not. Dosimetric evaluation normally needs a large patient cohort to reach a definitive conclusion. Evaluation of the dosimetric impact of auto-segmentation will be our future study.

To conclude, it is possible to automatically delineate the heart, the heart chambers, and the great vessels from noncontrast CT images for radiation oncology applications, especially on clinical trials to evaluate the radiation-induced cardiac toxicities. We have validated the accuracy and consistency of auto-contouring for potential clinical use. However, accurate identification of coronary arteries in CT images is difficult and their auto-contouring needs further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Arthur Gelmis from the Department of Scientific Publication for reviewing the manuscript. This research was supported in part by the National Institutes of Health through MD Anderson's Cancer Center Support Grant CA016672.

References

1. Albin A, Pennesi G, Donatelli F, Cammarota R, De Flora S, Noonan DM. Cardiotoxicity of anticancer drugs: the need for cardio-oncology and cardio-oncological prevention. *Journal of the National Cancer Institute*. 2010; 102:14–25. [PubMed: 20007921]
2. Gagliardi G, Constine LS, Moiseenko V, et al. Radiation dose-volume effects in the heart. *Int J Radiat Oncol Biol Phys*. 2010; 76:S77–85. [PubMed: 20171522]
3. Lancellotti P, Nkomo VT, Badano LP, et al. Expert consensus for multi-modality imaging evaluation of cardiovascular complications of radiotherapy in adults: a report from the European Association of Cardiovascular Imaging and the American Society of Echocardiography. *J Am Soc Echocardiogr*. 2013; 26:1013–1032. [PubMed: 23998694]
4. Mulrooney DA, Yeazel MW, Kawashima T, et al. Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: retrospective analysis of the Childhood Cancer Survivor Study cohort. *BMJ*. 2009; 339:b4606. [PubMed: 19996459]
5. Demirci S, Nam J, Hubbs JL, Nguyen T, Marks LB. Radiation-induced cardiac toxicity after therapy for breast cancer: interaction between treatment era and follow-up duration. *Int J Radiat Oncol Biol Phys*. 2009; 73:980–987. [PubMed: 19251085]
6. Darby SC, Cutter DJ, Boerma M, et al. Radiation-related heart disease: current knowledge and future prospects. *Int J Radiat Oncol Biol Phys*. 2010; 76:656–665. [PubMed: 20159360]
7. Sardaro A, Petruzzelli MF, D'Errico MP, Grimaldi L, Pili G, Portaluri M. Radiation-induced cardiac damage in early left breast cancer patients: risk factors, biological mechanisms, radiobiology, and dosimetric constraints. *Radiother Oncol*. 2012; 103:133–142. [PubMed: 22391054]
8. Maraldo MV, Brodin NP, Vogelius IR, et al. Risk of developing cardiovascular disease after involved node radiotherapy versus mantle field for Hodgkin lymphoma. *Int J Radiat Oncol Biol Phys*. 2012; 83:1232–1237. [PubMed: 22270170]
9. van Eggermond AM, Schaapveld M, Lugtenburg PJ, et al. Risk of multiple primary malignancies following treatment of Hodgkin lymphoma. *Blood*. 2014; 124:319–327. quiz 466. [PubMed: 24740811]
10. Bradley JD, Paulus R, Komaki R, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. *Lancet Oncol*. 2015; 16:187–199. [PubMed: 25601342]
11. Yusuf SW, Sami S, Daher IN. Radiation-induced heart disease: a clinical update. *Cardiology research and practice*. 2011; 2011:317659. [PubMed: 21403872]
12. Collier DC, Burnett SS, Amin M, et al. Assessment of consistency in contouring of normal-tissue anatomic structures. *Journal of Applied Clinical Medical Physics*. 2003; 4:17–24. [PubMed: 12540815]
13. Yang J, Woodward WA, Reed VK, et al. Statistical modeling approach to quantitative analysis of interobserver variability in breast contouring. *Int J Radiat Oncol Biol Phys*. 2014; 89:214–221. [PubMed: 24613812]
14. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*. 2009; 46:726–738. [PubMed: 19245840]
15. Yang J, Amini A, Williamson R, et al. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. *Practical Radiation Oncology*. 2013; 3:e139–e147. [PubMed: 24674411]
16. RTOG. Atlases for Organs at Risk (OARs) in Thoracic Radiation Therapy. 2015 vol.
17. Feng M, Moran JM, Koelling T, et al. Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. *International journal of radiation oncology, biology, physics*. 2011; 79:10–18.
18. Netter, FH. Atlas of human anatomy. Elsevier Health Sciences; 2010.
19. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*. 2004; 23:903–921. [PubMed: 15250643]

20. Wang H, Dong L, Lii MF, et al. Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy. *International journal of radiation oncology, biology, physics*. 2005; 61:725–735.
21. Yang, J., Zhang, Y., Zhang, L., Dong, L. Automatic Segmentation of Parotids from CT Scans Using Multiple Atlases. In: vanGinneken, B.Murphy, K.Heimann, T.Pekar, V., Deng, X., editors. *Medical Image Analysis for the Clinic: A Grand Challenge*. 2010. p. 323-330.
22. Yang J, Garden A, Zhang Y, Zhang L, Court L, Dong L. WE-E-213CD-09: Multi-Atlas Fusion Using a Tissue Appearance Model. *Med Phys*. 2012; 39:3961.
23. Ibanez, L., Schroeder, W., Ng, L., Cates, J. *The ITK Software Guide*. Clifton Park, NY: Kitware Inc.; 2005.
24. 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; 73:944–951. [PubMed: 19215827]
25. Babalola KO, Patenaude B, Aljabar P, et al. An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage*. 2009; 47:1435–1447. [PubMed: 19463960]
26. 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. [PubMed: 20231069]
27. Teguh DN, Levendag PC, Voet PW, 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. [PubMed: 20932664]
28. Yang J, Wei C, Zhang L, Zhang Y, Blum RS, Dong L. A statistical modeling approach for evaluating auto-segmentation methods for image-guided radiotherapy. *Comput Med Imaging Graph*. 2012; 36:492–500. [PubMed: 22673541]
29. Ecabert O, Peters J, Walker MJ, et al. Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. *Med Image Anal*. 2011; 15:863–876. [PubMed: 21737337]
30. Kirisli HA, Schaap M, Klein S, et al. Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: a large-scale, multicenter, and multivendor study. *Med Phys*. 2010; 37:6279–6291. [PubMed: 21302784]
31. Zhuang X, Bai W, Song J, et al. Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection. *Med Phys*. 2015; 42:3822–3833. [PubMed: 26133584]

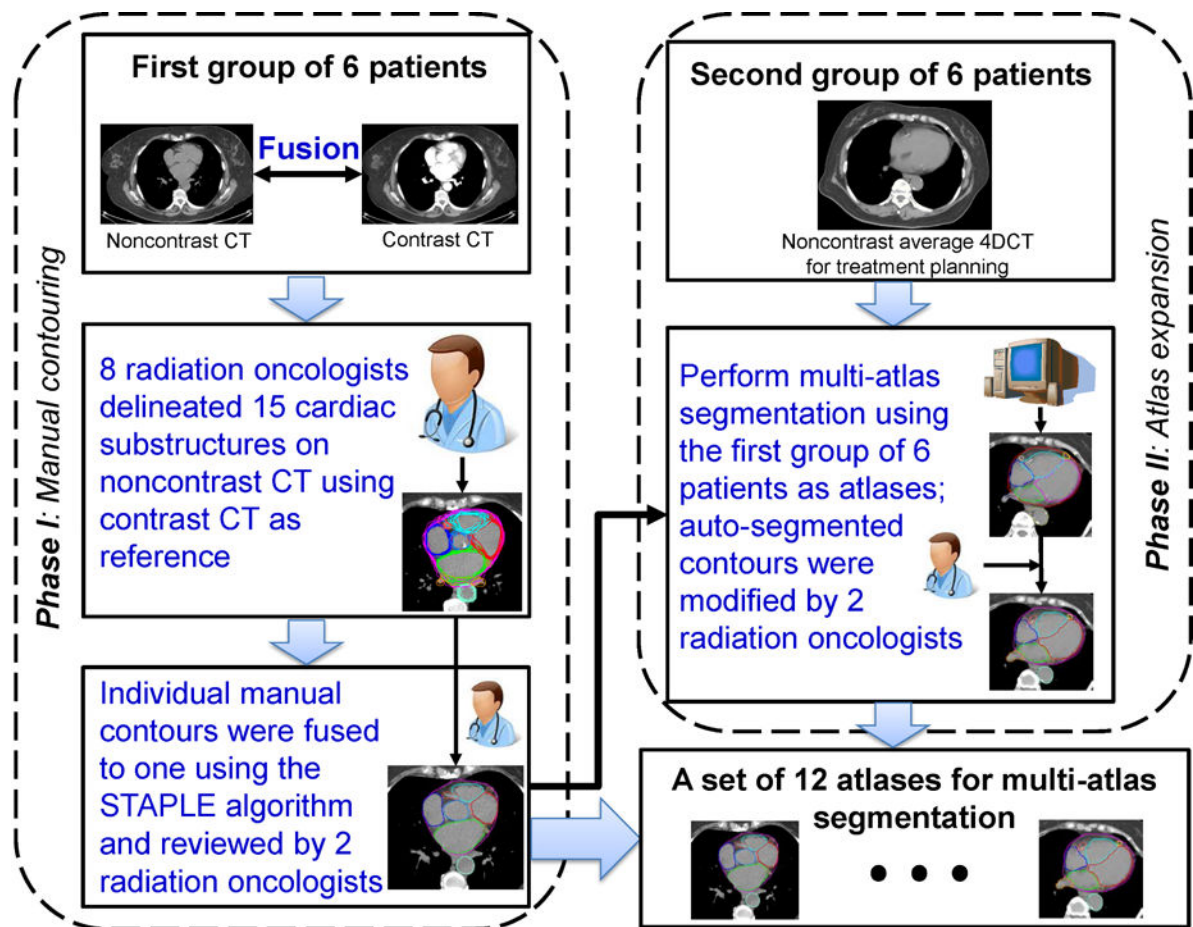


Figure 1.

The overall framework of the atlas development. Please refer to the text for details.

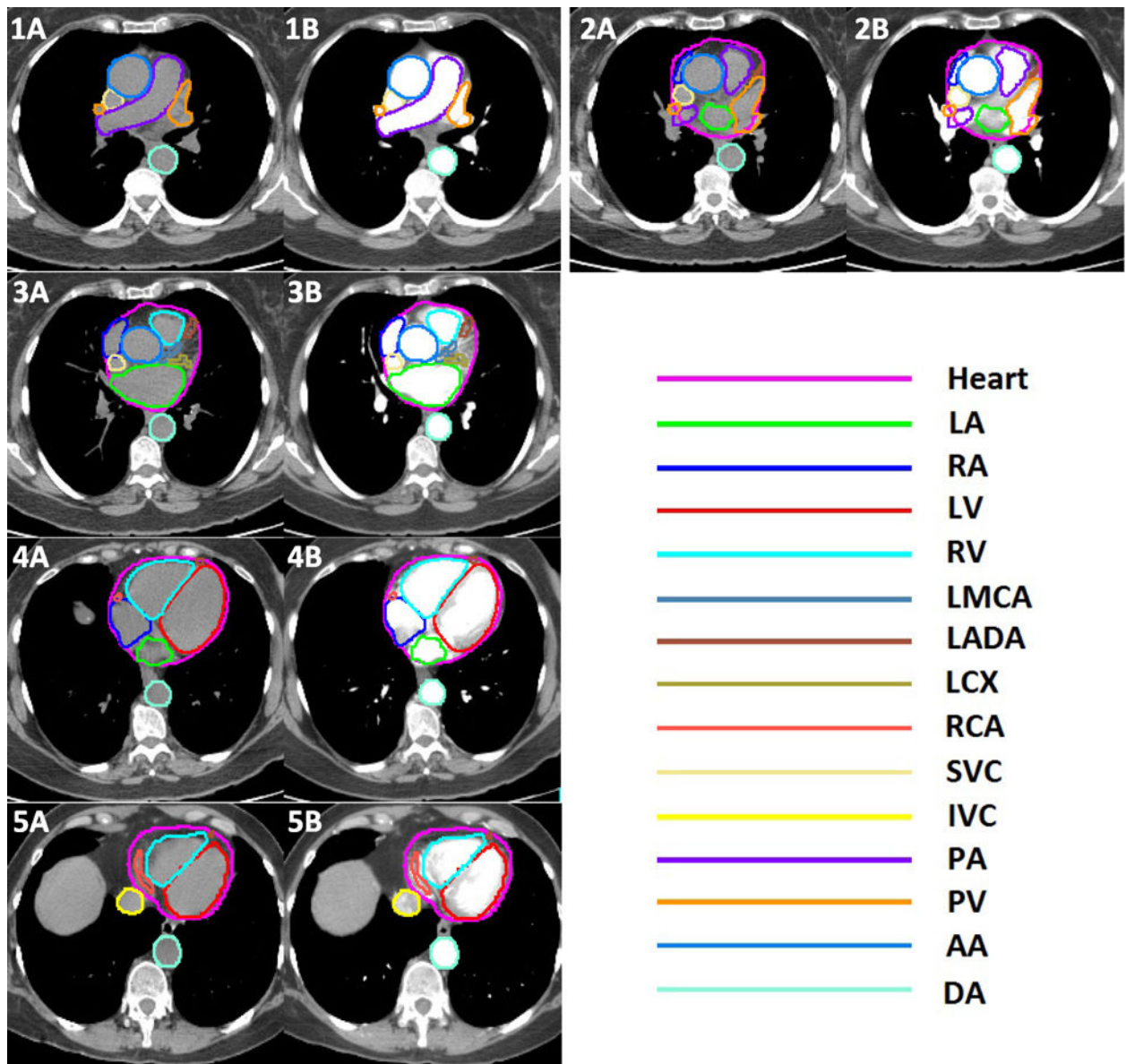


Figure 2.

Atlas contours overlaid on noncontrast (panels A) and contrast (panels B) CT images from one patient. LA = left atrium; RA = right atrium; LV = left ventricle; RV = right ventricle; LMCA = left main coronary artery; LADA = left anterior descending artery; LCX = left circumflex artery; RCA = right coronary artery; SVC = superior vena cava; IVC = inferior vena cava; PA = pulmonary artery; PV = pulmonary vein; AA = ascending aorta and aortic arch; DA = descending aorta.

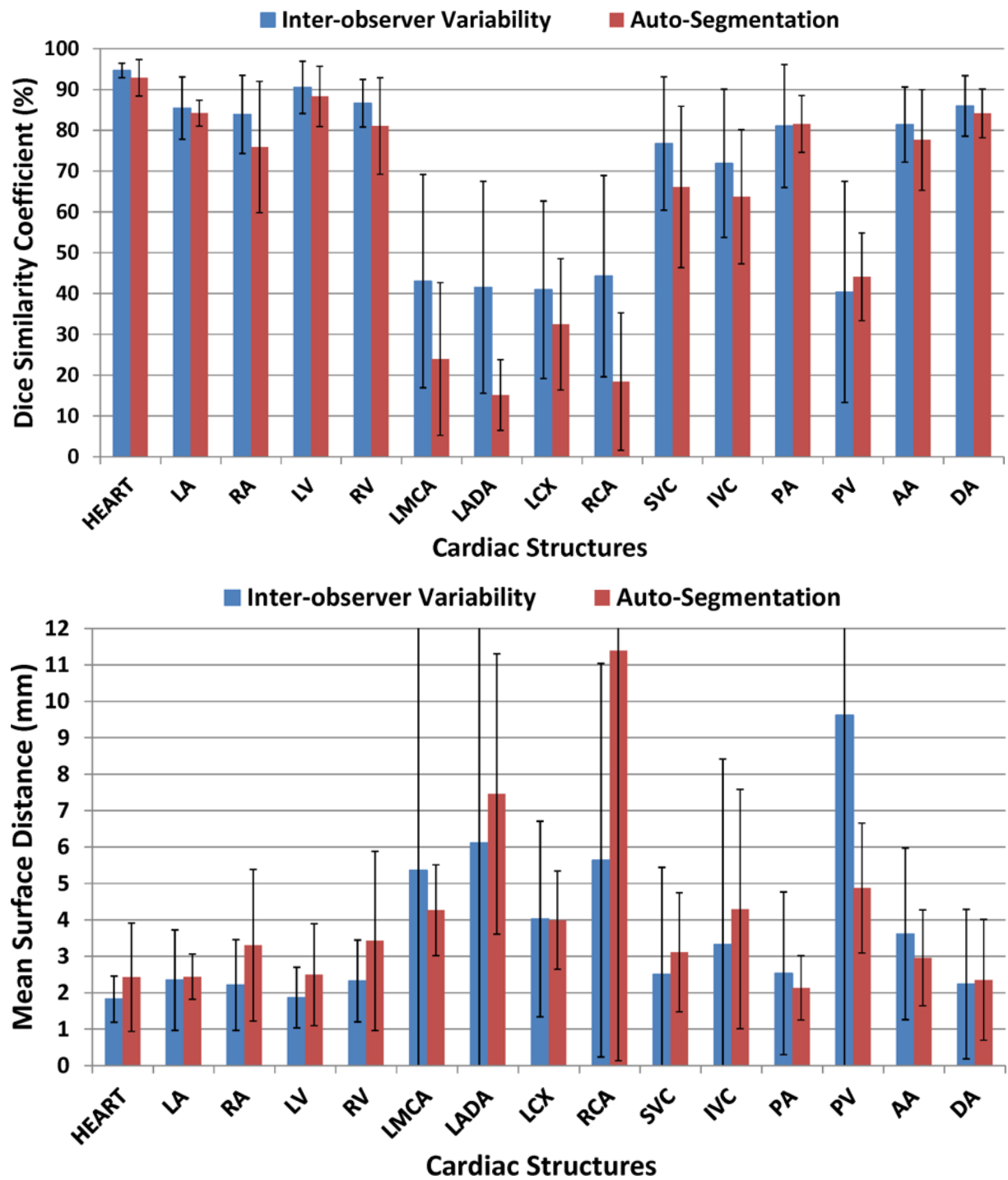


Figure 3.

Comparison of inter-observer variability with auto-segmentation using (a) Dice similarity coefficient (DSC) and (b) mean surface distance (MSD) metrics. The error bars indicate one standard deviation. For abbreviations, see the legend to Figure 2.