Radiotherapy and Oncology 131 (2019) 215–220

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients

Robert Kaderka, Erin F. Gillespie, Robert C. Mundt, Alex K. Bryant, Camila B. Sanudo-Thomas, Anna L. Harrison, Emilie L. Wouters, Vitali Moiseenko, Kevin L. Moore, Todd F. Atwood, James D. Murphy *

Department of Radiation Medicine and Applied Sciences, University of California San Diego, United States

ARTICLE INFO

Article history:
Received 30 January 2018
Received in revised form 20 June 2018
Accepted 16 July 2018
Available online 11 August 2018

Keywords:
Atlas based auto-segmentation
Breast cancer
Heart contouring

ABSTRACT

Background and purpose: Auto-segmentation represents an efficient tool to segment organs on CT imaging. Primarily used in clinical setting, auto-segmentation plays an increasing role in research, particularly when analyzing thousands of images in the "big data" era. In this study we evaluate the accuracy of cardiac dosimetric endpoints derived from atlas based auto-segmentation compared to gold standard manual segmentation.

Material and methods: Heart and cardiac substructures were manually delineated on 54 breast cancer patients. Twenty-seven patients were used to build the auto-segmentation atlas, the other 27 to validate performance. We evaluated accuracy of the auto-segmented contours with standard geometric indices and assessed dosimetric endpoints.

Results: Auto-segmented contours overlapped geometrically with manual contours of the heart and chambers with Dice-similarity coefficients of 0.93 ± 0.02 (mean ± standard deviation) and 0.79 ± 0.07 respectively. Similarly, there was a strong link between dosimetric parameters derived from auto-segmented and manual contours (R^2 = 0.955–1.000). On the other hand, the left anterior descending artery had little geometric overlap (Dice-similarity coefficient 0.09 ± 0.07), though acceptable representation of dosimetric parameters (R^2 = 0.646–0.992).

Conclusions: The atlas based auto-segmentation approach delineates heart structures with sufficient accuracy for research purposes. Our results indicate that quality of auto-segmented contours cannot be determined by geometric values only.

Auto-segmentation in radiation oncology has the capacity to reduce variation, improve accuracy, and increase efficiency compared to manual segmentation. The digital age in radiation oncology has produced vast quantities of imaging data which has helped to transform radiation oncology research. In particular, studies evaluating the link between radiation dose and toxicity could potentially involve hundreds or thousands of images – making auto-segmentation critical to allow for accurate, consistent, and reproducible contouring of normal organs.

Auto-segmentation routines, which have permeated radiation oncology for several years, have been evaluated with a focus on clinical application and accuracy [1–4]. The accuracy of auto-segmentation techniques in these studies was evaluated by the geometric overlap between auto-segmented contours compared to some gold standard manual set of contours. Multiple geometric indices exist though more commonly used metrics include the Dice-similarity coefficient, average and max Hausdorff's distance, or concordance index. Geometric indices represent useful metrics when assessing accuracy of auto-contouring in the clinical setting, though in the research setting when retrospectively evaluating images of treated patients the *dosimetric data* derived from auto-segmented contours represents a much more meaningful endpoint that has not been thoroughly investigated yet. Classic radiation epidemiology studies depend on the accuracy of these dosimetric endpoints (such as mean radiation dose) [5–7], though the question of whether using auto-segmentation in this context can produce accurate dosimetric endpoints remains unclear.

The purpose of this study was to define the utility of using atlas based auto-segmentation (ABAS) to delineate cardiac substructures among breast cancer patients treated with radiation. We focus on the question of whether auto-segmentation can produce accurate dosimetric endpoints compared to gold standard manual contours in a real-world retrospective cohort of breast cancer patients. Lastly, we tested whether a Pearson-correlation exists

* Corresponding author at: Department of Radiation Medicine and Applied Sciences, University of California, San Diego, 3960 Health Sciences Dr., MC0865, La Jolla, CA 92093-0865, United States.
E-mail address: j2murphy@ucsd.edu (J.D. Murphy).

https://doi.org/10.1016/j.radonc.2018.07.013
0167-8140/

between geometric indices and the accuracy of dosimetric endpoints for auto-segmentation compared to manual contours.

Methods and materials

Patients

This study consisted of 54 breast cancer patients who received radiotherapy at the University of California, San Diego between 2014 and 2016. We sought to evaluate a diverse real-life group of patients, and our study cohort included those treated with radiation after breast conservation surgery, as well as those that received post-mastectomy radiation, including a subset with breast tissue expanders. We included patients with left- and right-sided breast cancer as well as those in different setup positions (supine and prone). For left-sided breast cancer patients, standard cardiac avoidance techniques included deep inspiration breath hold or treatment in prone position.

Patients were treated to the whole breast or chest wall with either hypofractionated radiation (40.05 Gy in 15 fractions) or standard fractionation (50.4 Gy in 28 fractions or 50 Gy in 25 fractions) using 3D conformal radiotherapy. Most patients received boost radiation via *en face* electrons, or mini-tangents, or a combination of both for an additional 10–14 Gy in 5–7 fractions. For simplicity this study focused on the whole breast or chest wall radiation dose, and did not include dose from regional irradiation.

Target delineation

A single expert contoured the whole heart, left anterior descending artery (LAD) as well as left and right atria and ventricles in all patients according to the guidelines published by Feng et al. [8]. Most notably, the whole heart was contoured superiorly to the last slice before the right pulmonary artery appears, and inferiorly to the apex of the heart including pericardial sac.

Atlas based auto-segmentation

CT simulation included a slice thickness of 2.5 mm. We used MIM 6.8.2 (MIMvista Corp., Cleveland, OH, USA) for the atlas based auto-segmentation, described in detail in [2]. For patients imaged in supine position, a multi-atlas approach with three matches was employed for the whole heart and its chambers. The matching was followed by intensity-based free-form deformable image registration and the finalization of contours was performed with STAPLE [9]. For the left anterior descending artery and for patients imaged in prone position, a single-atlas approach was employed, where the best rigid match was deformed with the image registration described above. Examples of the auto-segmentation are illustrated in Fig. 1. The auto-contouring time took between one to two minutes per patient.

We divided our patient cohort into two groups: the first group ($n = 27$) was used to create the atlas based auto-segmentation model, and the second group ($n = 27$) was used to validate the accuracy of the auto-segmented contours. To improve the performance of the auto-segmentation, we ensured a diverse population of patients in the atlas group with 18 left-sided breast cancer patients, 9 right-sided breast cancer patients, 23 patients imaged in supine position, 4 in prone position and 3 patients with breast expanders.

Analysis

In the validation dataset we measured geometrical indices commonly used in contouring studies such as Dice-similarity coefficient, average and max Hausdorff's distance [2,10] between auto-segmented contours and gold-standard manual contours. Dice-similarity coefficients are a measure of contour overlap scored from 0 to 1, with 1 indicating a perfect overlap [11]. Average and max Hausdorff's distances are a measure of the distance between two contours, where a perfect overlap would achieve a distance of 0 cm [10,12].

We generated dose–volume histograms (representative patient shown in Fig. 2) and extracted a variety of dosimetric parameters for each patient from the auto-segmented contours and the gold-standard contours. Examples of these dosimetric endpoints include D_{mean} (mean dose), V1Gy (the volume of a structure receiving 1 Gy or more) and D5% (the minimum dose to the hottest 5% of a structure). The full list of parameters investigated in this study is given in Table 1. Correlation between dosimetric endpoints derived from auto-segmented and manual contours was assessed with R^2 of the linear regression.

Lastly, we tested whether geometric indices correlated with the dose differences between manual and auto-segmented contours, where the dose difference was defined as $\Delta D_{mean} = |(D_{mean,GS} - D_{mean,ABAS})/(D_{mean,GS})|$, with $D_{mean,GS}$ equal to the mean gold standard dose, and $D_{mean,ABAS}$ equal to the mean ABAS dose. The correlation between geometric indices and dose difference was evaluated with a Pearson correlation coefficient r.

Results

We found a range of geometric overlap that depended on cardiac substructure (Table 1). The whole heart had the highest degree of geometric overlap as demonstrated by the highest average Dice-similarity coefficient (0.93 ± 0.02), and the lowest average Hausdorff's distance (0.18 ± 0.06 cm). On the other hand the left-anterior descending artery had the lowest degree of geometric overlap with the lowest Dice-similarity coefficient (0.09 ± 0.07).

When considering the dosimetric parameters generated with auto-segmentation, we found a high degree of agreement between the auto-segmented contours and the manually generated contours for most of the cardiac substructures with most of the dosimetric endpoints (Table 1, and Fig. 4). The R^2 values were uniformly above 0.955 for the whole heart and the individual chambers for all dosimetric endpoints studied. The auto-segmented left anterior descending artery had a slightly lower degree of agreement, with a R^2 value of 0.65 (with V10Gy). Fig. 3 gives a qualitative representation of a scenario commonly observed in our cohort, where the left anterior descending artery had low geometric overlap, yet the doses received by auto-segmented and the manual contours were very similar.

Geometrical measures did not predict the accuracy of dosimetric parameters determined by auto-segmentation. Table 2 shows the correlation coefficient r of geometrical and dosimetric indices for the whole heart, with r ranging from -0.49 to 0.53. Among all organs and indices, we found no $|r| > 0.53$.

Discussion

This study highlights an atlas based auto-segmentation approach to contour the heart and cardiac substructures. These auto-segmentation methods will prove vital in next generation large scale radiation epidemiology research, and other applications that require segmentation of normal anatomy where manual contouring is not feasible due to time constraints coupled to the massive number of available images. With cardiac toxicity specifically, the landmark case–control study by Darby et al. [5] found that the mean heart dose correlated with the risk of ischemic heart disease. An important limitation of the Darby study was that the investigators lacked imaging data from their patients, i.e. the heart doses

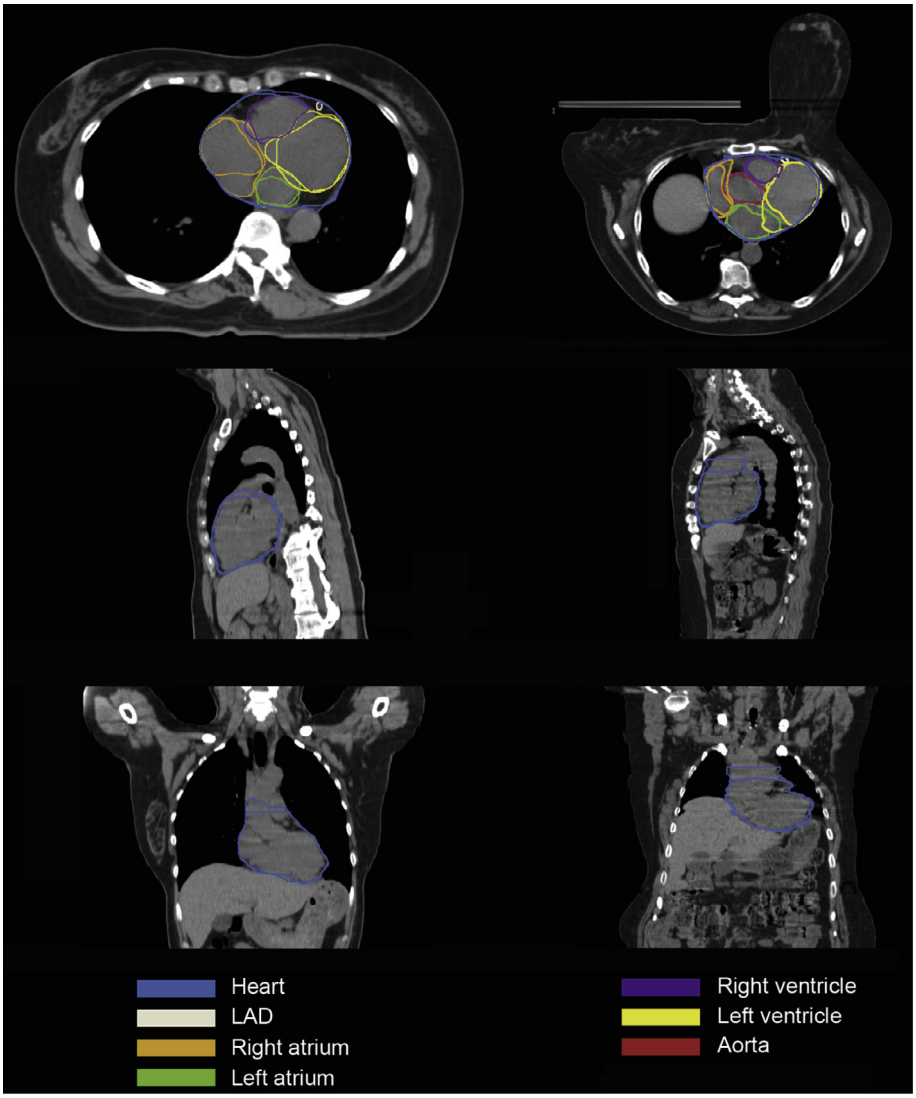

Fig. 1. CT images showing the overlay of manual and atlas based auto-segmented contours. Left side shows a patient imaged in supine position with an average Dice-similarity coefficient for the whole heart (0.93). Right side shows the patient with the lowest whole heart Dice-similarity coefficient (0.86). The right patient was imaged in prone position, which was rare in the atlas. LAD denotes the left anterior descending artery.

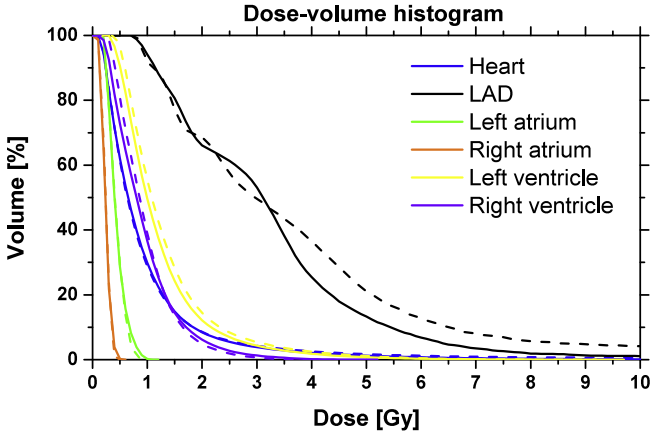

Fig. 2. Dose–volume histogram obtained with gold standard contours (solid lines) and auto-segmented contours (dashed lines) for the left patient in Fig. 1. LAD denotes the left anterior descending artery.

were estimated through different techniques [13,14], which introduces the possibility of biased results [15]. The next generation of radiation epidemiology research will have access to patient images and radiation dose-distributions. Processing these imaging data will require efficient validated auto-segmentation algorithms such as the one presented in this project.

Other studies have investigated atlas based auto-segmentation approaches with cardiac contouring [2–4,16,17], though to our knowledge this work represents the first investigation of the impact of an automated contouring model on estimates of radiation dose to the heart and cardiac substructures in breast cancer patients. This study demonstrates that auto-contouring could be used to extract dosimetric information from large numbers of CT images where manual contouring is not feasible. We found that conventional geometric parameters such as Dice-similarity coefficient – which is heavily influenced by the volume of a structure – may not consistently reflect whether an auto-segmented contour accurately represents actual radiation dose. Naturally, in situations with a perfect Dice-similarity coefficient one would expect to have perfect accuracy with dosimetric endpoints. However, the left

Table 1
Geometric and dosimetric comparison of atlas based auto-segmentation and manually drawn contours. Dosimetric endpoints denote the R^2 in the linear regression of metrics determined by manual and auto-segmented contours.

Agreement of manual and auto-segmented contours						
	Whole heart	LAD	Left atrium	Right atrium	Left ventricle	Right ventricle
Geometric indices						
Dice-similarity coefficient	0.93 ± 0.02	0.09 ± 0.07	0.76 ± 0.06	0.76 ± 0.06	0.85 ± 0.04	0.79 ± 0.06
Average Hausdorff's Distance (cm)	0.18 ± 0.06	0.73 ± 0.61	0.26 ± 0.09	0.26 ± 0.09	0.26 ± 0.07	0.26 ± 0.08
Dosimetric endpoints (R^2)						
D_{mean}(Gy)	0.996	0.947	0.996	0.997	0.994	0.988
D5%(Gy)	0.998	0.846	0.986	0.994	0.995	0.998
D95%(Gy)	0.999	0.646	0.995	0.992	0.991	0.975
V1Gy(%)	0.999	0.992	0.959	0.994	0.995	0.988
V10Gy(%)	0.988	0.698	–	–	0.955	1.000

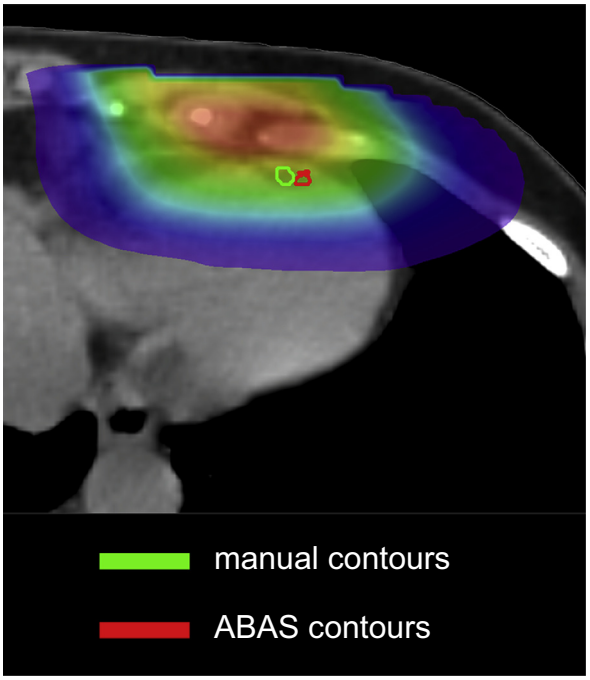

Fig. 3. Example of discordance between geometric overlap and dosimetric values in the left anterior descending artery. The image shows a patient with a low Dice-similarity coefficient for the left anterior descending artery, though the dosimetric impact was negligible, because of the proximity and low dose gradient.

anterior descending artery in this study provides a prime example of a structure with a very low Dice-similarity coefficient (likely due to its small size and lack of intravenous contrast), but still a high degree of dosimetric concordance between auto-segmented contours and manual contours. On the other hand, one could hypothesize a scenario where dosimetric concordance is low for a critical structure, despite a high Dice-similarity coefficient. For researchers seeking to use auto-segmentation algorithms in epidemiology studies, we recommend using a prudent validation approach involving both geometric measures and the dosimetric endpoints considered for a particular epidemiological question.

This study focused on an atlas based auto-segmentation approach in the context of research, though this method could have clinical implications as well. Current guidelines by the National Comprehensive Cancer Network recommend delineating the heart and employing cardiac avoidance techniques such as respiratory control in breast cancer radiation planning [18]. A UK Consensus Statement further specifies using mean heart dose to guide cardiac sparing efforts [19]. A recent retrospective review of US patients showed a decrease in the mean heart dose in routine practice compared to prior studies [20]. However, increasing evidence points toward the importance of individual cardiac substructures. Research has implicated the left anterior descending artery [21,22], as well as the left ventricle [7] as important components of the heart associated with radiation heart disease. Auto-segmentation could help to standardize contouring for these structures. A study by Feng et al. found that radiation oncologists without explicit contouring guidelines have a high degree of variability in contouring the heart with a concordance index of 0.79 ± 0.11 [8]. The variability is reduced when using guidelines with a concordance index of 0.89 ± 0.03. Our auto-segmentation approach yielded a comparable concordance index of 0.87 ± 0.03 compared to a single expert's contours, indicating the potential of using this method for contouring the whole heart in a clinical environment. For the left anterior descending artery, the dosimetric concordance achieved by auto-segmentation is deemed acceptable for epidemiology research, but auto-contouring routines for clinical purposes will require further development or improved image quality by adding contrast agents during CT image acquisition. Efficient standardization of contouring in the clinical setting remains a key focus to ensure delivery of high quality radiation – auto-segmentation could help move closer to this goal.

This study has several limitations to consider. Our algorithm focused on breast cancer patients, and did not include other cancer subtypes that involve radiation to the heart. Given the differences in radiation dose distribution to the heart in other diseases (lung cancer, esophageal cancer, or lymphoma) one might find that the accuracy of cardiac auto-segmentation from a dosimetric perspective varies by disease site. Further studies in other cancer subtypes would be necessary before extrapolating this to other sites.

Another limitation relates to the accuracy of gold standard contours on the planning CT. Scans for breast cancer patients usually do not include intravenous contrast, have variable amounts of respiratory motion and naturally include cardiac motion. These technical factors could influence the reproducibility or accuracy of our gold-standard manual contours, and could also impact the auto-segmentation registration process. Despite these potential sources of variation we found that our auto-segmentation algorithm performed relatively well in our patient cohort. Furthermore, this study included gold standard contours delineated by a single expert. Given the time required for contouring the heart and substructures employing multiple experts was not feasible in our study. This study represents a proof of principle, and a gold standard atlas for research or clinical purposes would include contours drawn by multiple experts.

In summary, we present a method to automatically contour the heart and cardiac substructures in breast cancer patients which represents an efficient tool to estimate radiation dose and dosimetric endpoints. With sufficient validation this atlas-based approach

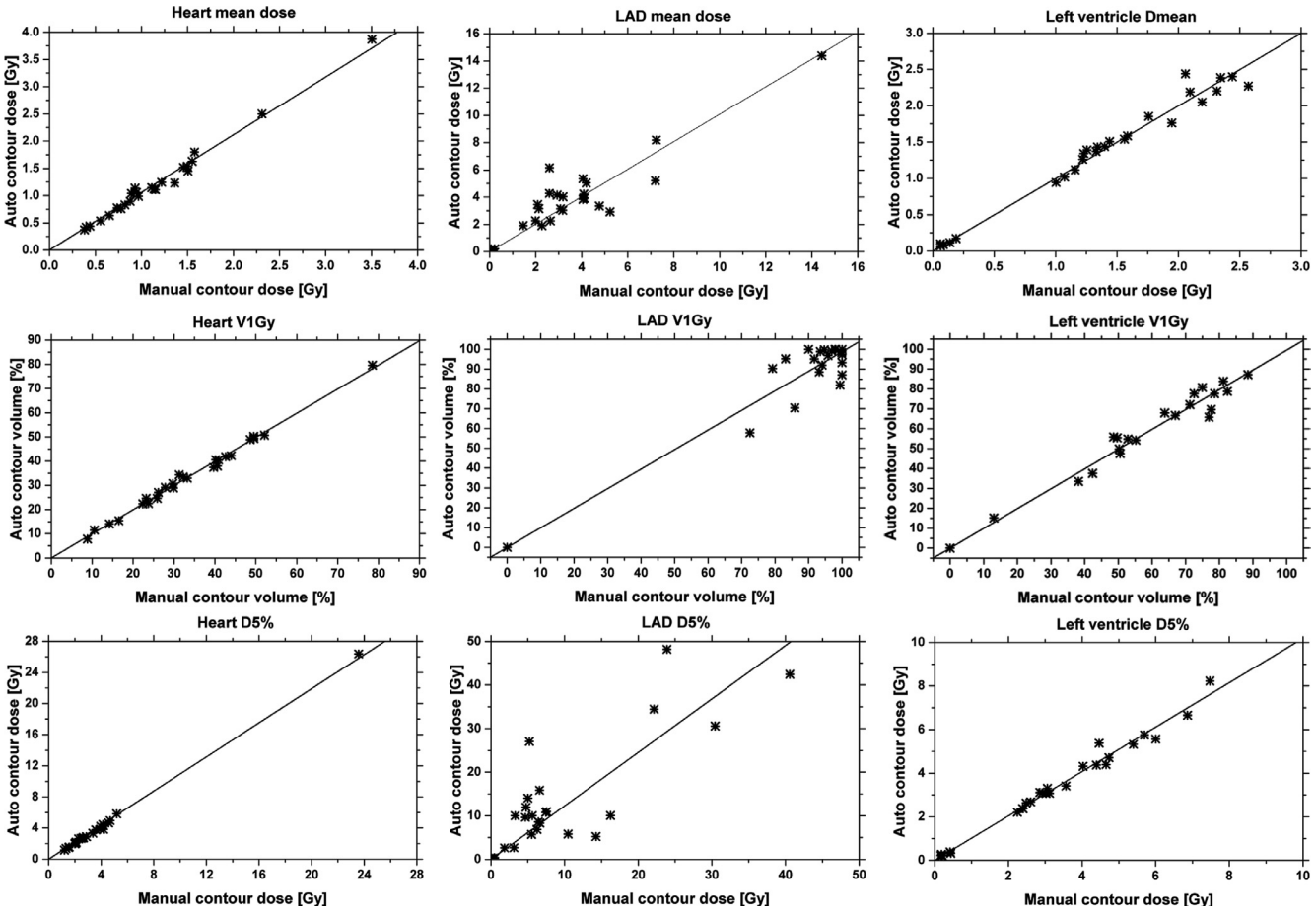

Fig. 4. Comparison of dose parameters for manually entered and atlas based auto-segmentation contours. Plots show the $D_{mean,}$ (top), V1Gy (middle) and D5% (bottom) of the heart, left anterior descending artery, and left ventricle for all patients. Solid line represents the linear regression.

Table 2
Pearson's correlation *r* for geometric measures and normalized dose differences between gold standard and auto-segmented contours. Dose parameters denote the absolute value of the difference between gold standard and auto-segmented contours normalized to gold standard dose values.

Pearson's correlation coefficient *r* of geometric and dosimetric indices			
	Dice-similarity coefficient	Average Hausdorff's distance	Max Hausdorff's Distance
Dice-similarity coefficient	NA	−0.959	−0.201
Average Hausdorff's distance	−0.959	NA	0.275
Max Hausdorff's Distance	−0.201	0.275	NA
ΔD_{mean}	−0.080	0.129	0.069
ΔD5%	−0.025	0.056	0.111
ΔD95%	−0.494	0.525	0.182
ΔV1Gy	−0.394	0.356	−0.123

has the capacity for use in radiation epidemiology research to automatically estimate radiation dose to large numbers of patients.

Conflict of interest statement

The authors have no relevant conflicts of interest to disclose.

References

[1] Van Dijk-Peters FBJ, Sijtsema NM, Kierkels RGJ, Vliegenthart R, Langendijk JA, Maduro JH, et al. OC-0259: Validation of a multi-atlas based auto-segmentation of the heart in breast cancer patients. Radiother Oncol 2015;115:S132–3.
[2] Ciardo D, Gerardi MA, Vigorito S, Morra A, Dell'acqua V, Diaz FJ, et al. Atlas-based segmentation in breast cancer radiotherapy: evaluation of specific and generic-purpose atlases. Breast (Edinburgh, Scotland) 2016;32:44–52.
[3] Eldesoky AR, Yates ES, Nyeng TB, Thomsen MS, Nielsen HM, Poortmans P, et al. Internal and external validation of an ESTRO delineation guideline – dependent automated segmentation tool for loco-regional radiation therapy of early breast cancer. Radiother Oncol 2016;121:424–30.
[4] Zhou R, Liao Z, Pan T, Milgrom SA, Pinnix CC, Shi A, et al. Cardiac atlas development and validation for automatic segmentation of cardiac substructures. Radiother Oncol 2017;122:66–71.
[5] Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Bronnum D, et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. New Engl J Med 2013;368:987–98.
[6] van Nimwegen FA, Schaapveld M, Cutter DJ, Janus CP, Krol AD, Hauptmann M, et al. Radiation dose-response relationship for risk of coronary heart disease in survivors of Hodgkin lymphoma. J Clin Oncol 2016;34:235–43.
[7] van den Bogaard VA, Ta BD, van der Schaaf A, Bouma AB, Middag AM, Bantema-Joppe EJ, et al. Validation and modification of a prediction model for acute cardiac events in patients with breast cancer treated with radiotherapy based on three-dimensional dose distributions to cardiac substructures. J Clin Oncol 2017. Jco2016698480.
[8] Feng M, Moran JM, Koelling T, Chughtai A, Chan JL, Freedman L, et al. Development and validation of a heart atlas to study cardiac exposure to

radiation following treatment for breast cancer. Int J Radiat Oncol Biol Phys 2011;79:10–8.
[9] 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;23:903–21.
[10] Dong-Gyu S, Oh-Kyu K, Rae-Hong P. Object matching algorithms using robust Hausdorff distance measures. IEEE Trans Image Process 1999;8:425–9.
[11] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302.
[12] Sim DG, Kwon OK, Park RH. Object matching algorithms using robust Hausdorff distance measures. IEEE Trans Image Process 1999;8:425–9.
[13] Taylor CW, Bronnum D, Darby SC, Gagliardi G, Hall P, Jensen MB, et al. Cardiac dose estimates from Danish and Swedish breast cancer radiotherapy during 1977–2001. Radiother Oncol 2011;100:176–83.
[14] Taylor CW, Nisbet A, McGale P, Goldman U, Darby SC, Hall P, et al. Cardiac doses from Swedish breast cancer radiotherapy since the 1950s. Radiother Oncol 2009;90:127–35.
[15] Kaderka R, Gillespie EF, Mundt R, Bryant AK, Cornell M, Atwood T, et al. Potential pitfalls when using surrogate radiation dose measures in place of real dosimetric data in radiation epidemiology research, Int J Radiat Oncol Biol Phys, 99: S217.
[16] Eldesoky AR, Francolini G, Thomsen MS, Yates ES, Nyeng TB, Kirkove C, et al. Dosimetric assessment of an Atlas based automated segmentation for loco-regional radiation therapy of early breast cancer in the Skagen Trial 1: a multi-institutional study. Clin Trans Radiat Oncol 2017;2:36–40.
[17] Mitchell RA, Wai P, Colgan R, Kirby AM, Donovan EM. Improving the efficiency of breast radiotherapy treatment planning using a semi-automated approach. J Appl Clin Med Phys 2017;18:18–24.
[18] Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, et al. Guidelines insights: breast cancer, Version 1.2017. J Nat Compr Cancer Network JNCCN 2017;15:433–51.
[19] Postoperative radiotherapy for breast cancer: UK consensus statements.
[20] Hong JC, Rahimy E, Gross CP, Shafman T, Hu X, Yu JB, et al. Radiation dose and cardiac risk in breast cancer treatment: An analysis of modern radiation therapy including community settings. Pract Radiat Oncol 2017.
[21] Correa CR, Litt HI, Hwang WT, Ferrari VA, Solin LJ, Harris EE. Coronary artery findings after left-sided compared with right-sided radiation treatment for early-stage breast cancer. J Clin Oncol 2007;25:3031–7.
[22] Nilsson G, Holmberg L, Garmo H, Duvernoy O, Sjögren I, Lagerqvist B, et al. Distribution of coronary artery stenosis after radiation for breast cancer. J Clin Oncol 2012;30:380–6.