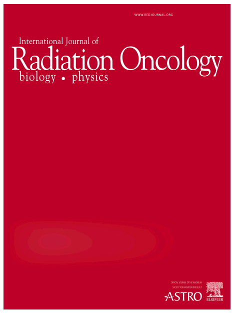
Accepted Manuscript



A comparison of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) for breast target volume delineation in prone and supine positions

Elise M. Pogson, BMedRadPhysAdvHons, PhD, Geoff P. Delaney, M.B.B.S. MD. PhD F.R.A.N.Z.C.R, Verity Ahern, FRANZCR, Miriam Boxer, BMBS BHSc(Hons) FRANZCR, Christine Chan, FRANZCR, Steven David, MBBS, FRANZCR, Marion Dimigen, MBChB, FRCR, FRANZCR, Jennifer A. Harvey, FRANZCR, Eng-Siew Koh, MBBS FRANZCR, Karen Lim, MBBS FRANZCR, George Papadatos, MBBS FRANZCR, Mei L. Yap, MBBS BSc (Med) FRANZCR, Vikneswary Batumalai, MHlthSc, Elizabeth Lazarus, FRANZCR, Kylie Dundas, BAppSc (MRT)(RT), MAppSc (MRS), Jesmin Shafiq, MBBS MCN MPH, Gary Liney, PhD, Catherine Moran, PhD, Peter Metcalfe, PhD, Lois Holloway, PhD

PII: S0360-3016(16)32996-0

DOI: 10.1016/j.ijrobp.2016.08.002

Reference: ROB 23749

To appear in: *International Journal of Radiation Oncology • Biology • Physics*

Received Date: 21 January 2016

Revised Date: 26 July 2016

Accepted Date: 1 August 2016

Please cite this article as: Pogson EM, Delaney GP, Ahern V, Boxer M, Chan C, David S, Dimigen M, Harvey JA, Koh E-S, Lim K, Papadatos G, Yap ML, Batumalai V, Lazarus E, Dundas K, Shafiq J, Liney G, Moran C, Metcalfe P, Holloway L, A comparison of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) for breast target volume delineation in prone and supine positions, *International Journal of Radiation Oncology • Biology • Physics* (2016), doi: 10.1016/j.ijrobp.2016.08.002.

A comparison of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) for breast target volume delineation in prone and supine positions

Elise M. Pogson[1,2,3] (BMedRadPhysAdvHons, PhD), Geoff P Delaney[2,3, 4,5] (M.B.B.S. MD. PhD F.R.A.N.Z.C.R), Verity Ahern[6] (FRANZCR), Miriam Boxer[2,4] (BMBS BHSc(Hons) FRANZCR), Christine Chan[7] (FRANZCR), Steven David[8] (MBBS, FRANZCR), Marion Dimigen[7] (MBChB, FRCR, FRANZCR), Jennifer A. Harvey[9,10] (FRANZCR), Eng-Siew Koh[2,3,4] (MBBS FRANZCR), Karen Lim[2] (MBBS FRANZCR), George Papadatos[2] (MBBS FRANZCR), Mei L. Yap[2,3,4,5] (MBBS BSc (Med) FRANZCR**),** Vikneswary Batumalai[2,3,4] (MHlthSc), Elizabeth Lazarus[7] (FRANZCR), Kylie Dundas[1,2,3] (BAppSc (MRT)(RT), MAppSc (MRS)), Jesmin Shafiq[3,4] (MBBS MCN MPH), Gary Liney[1,2,3] (PhD). Catherine Moran[11](PhD), Peter Metcalfe[1,2,3] (PhD), Lois Holloway[1,2,3,4,*] (PhD).

1. Centre for Medical Radiation Physics, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, NSW, 2522, Australia,
2. Liverpool and Macarthur Cancer Therapy Centres, Liverpool, NSW, Australia,
3. Ingham Institute for Applied Medical Research, Liverpool Hospital, Sydney, NSW, Australia,
4. South Western Sydney Clinical School, University of New South Wales, Sydney, NSW, Australia,
5. School of Medicine, University of Western Sydney, Sydney, NSW, Australia,
6. Crown Princess Mary Cancer Care Centre, Westmead Hospital, NSW, Australia,
7. Department of Radiology, Liverpool Hospital, NSW, Australia,
8. Peter MacCallum Cancer Institute, Melbourne, VIC, Australia,
9. School of Medicine, University of Queensland, QLD, Australia,
10. Princess Alexandra Hospital, QLD, Australia,
11. Department of Radiology, Stanford University, CA, America.

Running title: Inter-observer target delineation variation in Breast Radiotherapy

***Corresponding authors email: lois.holloway@sswahs.nsw.gov.au**
***Corresponding authors phone: +61 2 87389218**
*** Corresponding authors address: Ingham Institute for Applied Medical Research, 1 Campbell Street, Sydney, NSW, 2170, Australia****Conflict of interest statement:** The authors report a grant from the National Breast Cancer Foundation, during the conduct of the study. Dr. Ahern reports other from Commonwealth Serum Laboratories, outside the submitted work; Dr. Moran reports other from GE Healthcare, outside the submitted work.

Acknowledgements
This work was supported by a grant number APP1033237 from Cancer Australia and the National Breast Cancer Foundation. Thanks to Glen Osbourne, Drew Latty, and Chris Kelly for the importing/exporting of datasets into/from their centers treatment planning systems. Thanks to Jason Dowling for his assistance with MILXview.

A comparison of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) for breast target volume delineation in prone and supine positions

Running title: Inter-observer target delineation variation in Breast Radiotherapy

Abstract

Purpose/Objective(s): MRI provides no ionizing radiation dose (allowing inter and intra fraction imaging), improved soft tissue contrast and potentially improved conformity in delineation than CT. This study aimed to determine if T2 weighted magnetic resonance imaging improves seroma cavity (SC) and Whole Breast (WB) inter-observer conformity for radiotherapy purposes compared with the gold standard of CT, both in the prone and supine positions.

Methods and Materials: Eleven observers (two Radiologists and nine Radiation Oncologists) delineated SC and WB Clinical Target Volumes (CTVs) on T2-weighted MRI and CT supine and prone scans (4 scans per patient) for 33 patient datasets. Individual observer's volumes were compared using the Dice Similarity Coefficient (DSC), Volume overlap Index (VOI), Centre of Mass (COM) shift and Hausdorff Distances (HD). An average Cavity Visualization Score (CVS) was also determined.

Results: Imaging modality did not affect inter-observer variation for WB CTVs. Prone WB CTVs were larger in volume and more conformal than Supine CTVs (on both MRI and CT). SC volumes were larger on CT than MRI. SC volumes proved to be comparable in inter-observer conformity in both modalities (VOI of 0.57±0.03 for CT supine, 0.52±0.04 for MR supine, 0.56±0.03 for CT prone and 0.55±0.04 for MR prone) however after registering modalities together the inter-modality variation (DSC of 0.41±0.05 for supine and 0.38±0.04 for prone) was larger than the inter-observer variability for SC despite the location typically remaining constant.

Conclusions: MRI inter-observer variation was comparable to CT for the WB CTV and SC delineation, in both prone and supine positions. Whilst the CVS and inter-observer concordance

was not significantly higher for MRI than CT, the SCs were smaller on MRI, potentially due to clearer SC definition, especially on T2-weighted MR images.

Introduction

Approximately 80% of patients with breast cancer should receive radiotherapy [1]. However uncertainties in defining the Whole Breast (WB) Clinical Target Volume (CTV) exist and the optimal method to define the seroma cavity (SC) remains unclear [2]. Targeting the SC is crucial because, if missed, local control rates may be adversely affected [3,4]. Contouring variability of the SC has important implications for accelerated partial breast irradiation [5].

Soft tissues such as the SC walls and postoperative complexes (POCxs) are better visualized on MRI and 3D ultrasound (US) compared to CT (which only images the SC fluid [6]). Small observer studies have shown conflicting results, with the SCs proven to be both smaller (with less inter-observer variation from 3 observers, [7]) and larger (one observer, [8]) on MRI than CT. Clips provide another option [9], however these are not always consistent with the SC walls [10], can migrate and are larger than CT and US defined volumes [11].

Inter-observer variation quantification and minimization for the WB is necessary, as techniques improve and set-up uncertainties are reduced (such as Deep Inspiration Breath Hold (DIBH) [12]). Studies with a limited number of observers comparing WB CTV for MRI and CT have shown no difference in delineation variations [13],[14]. One study showed larger MRI volumes than CT [13], the other showed CT to be larger than volumes optimised on registered MRI/CT (initially delineated on CT) [14].

Breast MRIs for diagnostic purposes are undertaken in a prone position with MRI coils designed to achieve optimum signal-to-noise ratio (SNR) in this position. CT is the current standard of care for delineation of radiotherapy breast volumes, however MRI is the preferred modality for diagnostic purposes [15]. Despite earlier work demonstrating challenges with supine MRI [15] recent investigations have demonstrated feasibility [16]. MRI is non-ionizing, unlike CT, and could enable additional imaging throughout the patients treatment in future, to improve accuracy without exposing the patient to further ionising radiation. This study has undertaken a larger inter-observer variation study for whole breast CTV and SC radiotherapy delineations for CT and MRI in both prone and supine positions.

Materials and Methods

Cohort of patients

Following ethics approval 33 breast cancer patients receiving adjuvant breast radiotherapy were recruited from July 2012 to November 2014. Patients were above 18 years, had undergone breast conservation surgery, and been recommended adjuvant breast radiotherapy. Patients were excluded if they had a contraindication for MRI. Patient characteristics are displayed in Table 1.

Imaging

Patients underwent scans in two positions; 1) supine with a vacbag on a flat wing board (MTWB09 Wingboard, CIVCO Medical Solutions, Orange City, IA) with arms raised above their head, and 2) prone position, with the MRI coil housing as a set-up device to ensure consistent set-up between CT and MRI. CT scans were undertaken on a Philips CT scanner with 2 mm slice thickness and 1.024 px/mm resolution, wires were placed around the breast tissue for supine CT for clinical purposes. MRIs were undertaken on a 70 cm bore MAGNETOM Skyra 3 T (Siemens Medical Systems, Erlangen, Germany) scanner with a 3 T non-contrast T2 weighted turbo spin echo with 2 mm slices and no slice gap, no fat suppression, acquisition scan time of

2.9-7.4 minutes (this does not include patient set-up time), TE of 76-88 ms, TR of 2430-12360 ms, flip angle of 120-180^{o} and pixel bandwidth of 220-450 Hz/pixel. The prone MRI was taken using a dedicated 16 channel breast coil (Sentinelle Breast MRI System, Hologic Inc, Bedford, USA) with 1.318 px/mm resolution. The supine MRI was acquired with 18 channel surface coils held close to but not touching the breast tissue (to avoid deformation of the breast) utilizing a foam bridge, with 1.707 px/mm resolution. MRIs were completed within 14 days of the CT scans for all patients except 1 (17 days) and was performed on the same day for 26 patients.

The T2 sequence was selected based on a previous study on a separate cohort of volunteers [17]. The selected T2 sequence is also supported in the literature for visualizing SCs [18].

Image-processing

The wires on the CT Supine image set were removed by post processing in MILXview (Australian e-Health Research Centre (AEHRC), Australia). Images were imported into Treatment Planning Systems (TPSs) at the four hospitals of the observers, including Pinnacle^{3} versions 9.0 and 9.8 (Philips, Netherlands) at two sites and Eclipse Planning System 11.0.64 (Varian Medical Systems, Palo Alto USA) at two sites.

Delineation method

A study-specific delineation protocol was utilized by all observers (9 radiation oncologists and 2 radiologists) for the WB CTV and SC (supplementary Material 1). This protocol was edited after a single complete patient dataset had been contoured by all participants and the anonymized volumes reviewed together.

Contouring was undertaken in batches of 8-20 datasets with images from only a single-modality (CT or MRI) for each position (prone or supine) included in each batch to avoid recall of individual patient data. The order in which the 4 image datasets were circulated for each patient was varied. Observers were blinded to others contours. Patient notes or other information on SC location were not available to the clinicians, to ensure that the study tested the utility of the images alone. If no SC was able to be visualized then no SC was contoured.

Cavity Visualization Score (CVS)

CVSs adopted from [19-21] for each image were established by two radiation oncologists (at least 4 months post the last delineation) and one radiation therapist (RT), utilizing the images alone, at the same window levels as for contouring.

Analysis

Review of the SCs showed that some observers had contoured axillary Lymph Node (LN) cavity rather than the SC, although the protocol specified contouring of the SC. Thus for the generation of the SC gold standard volumes only 17 patients and 8 observer SCs were utilized to rule out the effects of unidentified SC or contours of the axillary Lymph Node (LN) SC. The combination of 17 patients and 8 observers were selected to retain large numbers of both observers and patients, with SC contours on all 4 datasets. Observers excluded were not less experienced however were more frequent in contouring the axillary Lymph Node (LN) SC or not visualizing a SC. Gold standard WB CTV and SC volumes were generated, to aid visual comparison of modalities, using the Simultaneous Truth And Performance Level Estimation (STAPLE) algorithm [22] for each dataset. All observers were weighted equallyin the STAPLE generation. The STAPLE, generalised kappa statistic (κ) [23,24] and volume were calculated utilizing CERR (a computational environment for radiotherapy research) software [25] in the MATLAB R2012a (Mathworks Incorporated, Natick USA) platform, with a subset of metrics verified in MILXview to agree to within 2 significant figures, similarly to other reported values [26]. Generalized kappa (κ) statistics determine inter-observer delineation agreement which occurs by chance alone. Generalised kappa (κ) values (where +1 represents perfect agreement, 0 represents no agreement above chance and −1 represents complete disagreement) were generated for each of the patients WB and SC volumes.. Pairwise analysis, such as the Volume Overlap Index (VOI) which is the Dissimilarity Index (DSC) performed pairwise [27,28], was completed in MILXview. The average of all three observers' CVSs were compared to contour conformity.

MRI datasets were rigidly registered to the corresponding CT by a Radiation Therapist (RT) in Pinnacle using the 'normalized mutual information algorithm' within a region of interest which included the breast of interest, sternum, chest-wall muscles and ribs. Difficult cases were verified by a separate experienced RT. MRI volumes were transferred onto the CT dataset for comparison of individual observer volumes between MRI and CT using DSC, COM shift, and Hausdorff Distance (HD) (in MILXview). These were then averaged over observers and patients. The impact of MRI geometric systematic distortion on the patient MRIs was assessed in a related study and found to be minimal [29,30].

Results

Whole Breast CTV

Both CT and MRI supine WB CTVs were smaller in volume than the corresponding prone volumes ($p<0.001$), with MRI supine volumes smaller than all other datasets ($p<0.001$) (Fig 1a).

High inter-observer agreement was seen for WB CTV for all metrics across observers (Fig. 1b) and across datasets (Table 2). Prone datasets had slightly higher inter-observer conformity than supine for both modalities ($p<0.001$ for VOI ($DSC_{pairwise}$), and κ). The average CT supine HDs are lower than all other datasets CTVs ($p<0.001$).

Average concordance for registered datasets was comparable to the inter-observer concordance (Fig. 1(b) and Table 3).

SC volumes

There was obvious disagreement between SC volumes particularly where some observers delineated a volume and others did not.. In total 120 SC volumes out of the potential 1452 (11 observers on 4 modalities for 33 patients) were not delineated. Of these 26/120 were on CT supine, 33/120 on MRI supine, 32/120 on CT prone and 29/120 on MRI prone. There was no discernible correlation between the omitted SC volumes, in terms of modality, nor position. Only 4/33 patients had their SC volume delineated by all observers, and only 3/11 observers contoured SC volumes for all patients. For the majority of patients where the SC volumes were not delineated by at least 1 observer on at least 1 of the 4 datasets (CT supine, CT prone, MRI supine and MRI prone), delineations were drawn on some datasets but not others. The SC volume was compared across datasets in Table 2.

The inter-observer SC volume agreement was lower than the WB CTV agreement. The SC volumes were of moderate inter-observer agreement (κ between 0.41-0.6 [23]) for all datasets. There was no statistical difference between any datasets for κ or the COM shift. The average HDs were slightly larger for CT than MRI in prone position ($p<0.05$).

Variation between modalities and across observers was small (Fig 1b); however variation between patients was large (Fig 2). The registration conformity (DSC) for MRI to CT SC volumes (Table 3) was lower than inter-observer conformity. The majority of MRI SC volumes were smaller than the corresponding patients CT SC volume. The relatively low registered DSC was due to these volume differences between MRI and CT SC volumes. One patient had a DSC of zero due to different SC locations, patient number 27 (Fig. 2). There are small average COM shifts between modality contours (for supine=1.13±0.19 cm and prone=1.16±0.16 cm). These registration COM shifts are slightly larger than the inter-observer COM shift within datasets (0.74±0.11 cm for CT supine, 0.81±0.16 for MRI supine, 0.78±0.12 for CT prone and 0.72±0.16 cm for MRI prone), as shown in Fig. 2(b).

The 8 volumes on CT and MRI and the registered gold standard STAPLE volumes are shown for differing CVSs in supplementary material 2. The CVS was highly correlated with conformity measures, as shown in Supplementary Material 3.

Discussion

WB CTV inter-observer variability was consistently smaller on the prone datasets than the supine ($p<0.001$ for both MRI and CT metrics: VOI and κ). It has been well established that the main contouring discrepancies are at the medial and lateral edges of the WB CTV in supine imaging [31,32], however in prone datasets, geometrically these areas are reduced. Both intra-observer delineation variations [33] and differences due to planning system and data transfer [26] are much smaller than inter-observer variation and thus expected to have minimal impact on these WB results.

Importantly no clinically significant differences were seen in inter-observer or inter-modality variability between the prone and supine MRI datasets, demonstrating that supine MRI for the purpose of radiotherapy contouring can be considered equivalent to prone MRI, despite the differences in coils and positioning. This is different to the diagnostic setting where supine breast MRI is not standard.

Inter-observer variability when delineating the SC was not highly concordant e.g. CT supine average VOI of 0.57. There are many factors that may influence inter-observer seroma delineation. For some patients, there may be several cycles occurring between lumpectomy and imaging for radiotherapy, resulting in the SC volume no longer being able to be visualized [18]. In this study any correlation between the lengths of time post-surgery or the presence of chemotherapy, and the CVS was not able to be differentiated, due to multiple factors obscuring any correlation. These factors include differences across individuals including size of tumor, size of cavity, surgeon, and post-op bleeding or infection. Registration uncertainties also affect CT-MRI registered results.

The ~1.1 cm COM shift (and low DSC of ~0.4) between CT and MRI modalities may be of concern clinically for APBI and boost treatments. Adding this large uncertainty into the treatment margins may adversely affect normal tissue, furthermore if you are only treating the cavity for APBI the risk of geographic miss increases. However, this would not be the case for intraoperative radiation therapy (IORT) including APBI and boost treatments.

The closeness of the various inter-observer variability scores suggests MRI does not offer any advantages over CT in determining the WB CTV. However the impact of differences in SCs (which may impact on whole breast treatment planning as well as partial breast approaches) between CT and MRI warrant further investigation. MRI SC volumes (average of 22 cm^3) were smaller than CT SC volumes (average of 28 cm^3) in both positions. This result was similar to another study of 15 prone patients where this volume difference indicated superior delineation on MRI [34]. MRIs were performed without a contrast injection because SC volumes are well-visualized on non-contrast enhanced T2-weighted images and post-surgery, ideally no tumor remains to take up contrast to aid localization.

Limitations of this study include; some observers not being as experienced with MRI as they are with CT, the relatively small cohort size, especially the smaller subset included in the SC analysis, the fact that surgical clips were not mandatory nor excluded giving a small subset of 8/17 patients with surgical clips. The clips were titanium and are evident on both CT, shown clearly as a bright white marker and MRI as a black artifact. In order to test the use of the imaging modalities ability to determine seroma location and volume in isolation, no clinical history or localisation information (such as ultrasound location of tumour) were provided. This is therefore an artificial situation compared to clinical practice where this information (as well as surgical clips) are routinely provided. Consequently the identification of the SC volume is image-based and different to the true surgical volume utilised clinically, similarly to other studies that volumed the postoperative complex [6]. The interobserver concordance (VOI) for the patients with (≥4 clips) and without clips denoting SC boundaries was in agreement for all datasets except CT supine, where average conformity for those without clips was higher. This may be because clips are not always consistent with the SC walls [10], can migrate significantly over days/weeks postoperatively [35] and are larger than CT and US defined volumes [11]. Registered CT/MRI delineation has not been evaluated here, as the primary goal was to evaluate MRI delineation alone, to assess any benefit or equivalence towards MRI only planning.

Conclusion

MRI inter-observer variation was considered for 33 patient datasets and 11 observers and found to be comparable to CT for the WB CTV in both prone and supine positions. For the SC, the dataset was reduced due to inconsistency in contoured volumes. Whilst the SC inter-observer concordance was not significantly higher for MRI than CT, the SCs were smaller on MRI, potentially due to clearer SC definition on non-contrast enhanced T2 weighted MRI.

Concordance between individual MRI-CT registered observer volumes was lower than inter-observer concordance for the SC volumes and comparable for the WB. Further investigations, with increased patient numbers, into the impact of the differences in SCs from CT to MRI are required.

References

1. Delaney G, Barton M Jacob S. Estimation of an optimal radiotherapy utilization rate for breast carcinoma. *Cancer* 2003;98:1977-1986.
2. Petersen RP, Truong PT, Kader HA, et al. Target volume delineation for partial breast radiotherapy planning: Clinical characteristics associated with low interobserver concordance. *Int. J. Radiat. Oncol. Biol. Phys.* 2007;69:41-48.
3. Veronesi U, Marubini E, Mariani L, et al. Radiotherapy after breast-conserving surgery in small breast carcinoma: Long-term results of a randomized trial. *Annals. of oncol. ESMO* 2001;12:997-1003.
4. Whipp E, Beresford M, Sawyer E, et al. True local recurrence rate in the conserved breast after magnetic resonance imaging–targeted radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* 2010;76:984-990.
5. Kosztyla R, Olson R, Carolan H, et al. Evaluation of dosimetric consequences of seroma contour variability in accelerated partial breast irradiation using a constructed representative seroma contour. *Int. J. Radiat. Oncol. Biol. Phys.* 2012;84:527-532.
6. Whipp EC Halliwell M. Magnetic resonance imaging appearances in the postoperative breast: The clinical target volume–tumor and its relationship to the chest wall. *Int. J. Radiat. Oncol. Biol. Phys.* 2008;72:49-57.
7. Jolicoeur M, Racine M-L, Trop I, et al. Localization of the surgical bed using supine magnetic resonance and computed tomography scan fusion for planification of breast interstitial brachytherapy. *Radiother. & Oncol.* 2011;100:480-484.
8. Sabine B, Giovanna D, Peter P, et al. Open low-field magnetic resonance (mr) versus ct scanner (ct) imaging in breast radiotherapy treatment planning. *Int. J. Radiat. Oncol. Biol. Phys.* 2005;63, Supplement 1:S232-S233.
9. Kirova YM, Castro Pena P, Hijal T, et al. Improving the definition of tumor bed boost with the use of surgical clips and image registration in breast cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.* 2010;78:1352-1355.
10. Yang Z, Chen J, Hu W, et al. Target definition: Planning the breast boost: How accurately do surgical clips represent the ct seroma? *Radiother. & Oncol.* 2010;97:530-534.
11. Rabinovitch R, Finlayson C, Pan Z, et al. Radiographic evaluation of surgical clips is better than ultrasound for defining the lumpectomy cavity in breast boost treatment planning: A prospective clinical study. *Int. J. Radiat. Oncol. Biol. Phys.* 2000;47:313-317.
12. Latty D, Stuart KE, Wang W, et al. Review of deep inspiration breath-hold techniques for the treatment of breast cancer. *J. of Med. Radiat. Sci.* 2015;62:74-81.
13. Giezen M, Kouwenhoven E, Mast ME, et al. Magnetic resonance imaging-versus computed tomography-based target volume delineation of the glandular breast tissue (clinical target volume breast) in breast-conserving therapy: An exploratory study. *Int. J. Radiat. Oncol. Biol. Phys.* 2011;81:804-811.
14. Mast M, Coerkamp E, Heijenbrok M, et al. Target volume delineation in breast conserving radiotherapy: Are co-registered ct and mr images of added value? *Radiat. Oncol.* 2014;9:65.
15. Ahn K-H, Hargreaves BA, Alley MT, et al. Physics contribution: Mri guidance for accelerated partial breast irradiation in prone position: Imaging protocol design and evaluation. *Int. J. Radiat. Oncol. Biol. Phys.* 2009;75:285-293.
16. Siegler P, Holloway CM, Causer P, et al. Supine breast mri. *JMRI* 2011;34:1212-1217.

17. L. Holloway, V. Batumalai, C. Moran, et al. Comparison of prone and supine mri sequences for the purpose of radiotherapy treatment planning for breast cancer. *AAPM 55th Annual Meeting and Exhibition.* Indianapolis, Indiana. 2013
18. Jacobson G, Zamba G, Betts V, et al. Image-based treatment planning of the post-lumpectomy breast utilizing ct and 3tmri. *Int. J. of Breast Cancer* 2011:1-5.
19. Kirby AM, Yarnold JR, Evans PM, et al. Tumor bed delineation for partial breast and breast boost radiotherapy planned in the prone position: What does mri add to x-ray ct localization of titanium clips placed in the excision cavity wall? *Int. J. Radiat. Oncol. Biol. Phys.* 2009;74:1276-1282.
20. Giezen M, Kouwenhoven E, Scholten AN, et al. Mri- versus ct-based volume delineation of lumpectomy cavity in supine position in breast-conserving therapy: An exploratory study. *Int. J. Radiat. Oncol. Biol. Phys.* 2012;82:1332-1340.
21. Smitt MC, Birdwell RL Goffinet DR. Breast electron boost planning: Comparison of ct and us. *Radiol.* 2001;219:203-206.
22. Commowick O Warfield SK. Estimation of inferential uncertainty in assessing expert segmentation performance from staple. *IEEE transactions on medical imaging* 2010;29:771-780.
23. Landis JR Koch GG. The measurement of observer agreement for categorical data. *Biom.* 1977:159-174.
24. Lim K, Small W, Jr., Portelance L, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2011;79:348-355.
25. Deasy JO, Blanco AI Clark VH. Cerr: A computational environment for radiotherapy research. *Med. Phys.* 2003;30:979-985.
26. Pogson EM, Begg J, Jameson MG, et al. A phantom assessment of achievable contouring concordance across multiple treatment planning systems. *Radiat. Oncol.* 2015.
27. 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.
28. Louie AV, Rodrigues G, Olsthoorn J, et al. Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4d-ct era. *Radioth. & Oncol.* 2010;95:166-171.
29. Walker A, Liney G, Holloway L, et al. Ep-1282l: An assessment of mri distortion for the purpose of radiotherapy treatment planning. *Radioth. & Oncol.* 2013;106:S484.
30. Walker A, Metcalfe P, Liney G, et al. Mri geometric distortion: Impact on tangential whole-breast imrt. *J. of Appl. Clin. Med. Phys. (in press)* 2016;14.
31. Batumalai V, Koh ES, Delaney GP, et al. Interobserver variability in clinical target volume delineation in tangential breast irradiation: A comparison between radiation oncologists and radiation therapists. *Clin. Oncol.* 2011;23:108-113.
32. Pogson E, Bell L, Batumalai V, et al. A comparison of coordinate systems for use in determining a radiotherapy delineation margin for whole breast. *J. of Phys.: Conference Series. IOP Publishing.* 2014;489:012057.
33. Hurkmans CW, Borger JH, Pieters BR, et al. Variability in target volume delineation on ct scans of the breast. *Int. J. Radiat. Oncol. Biol. Phys.* 2001;50:1366-1372.
34. Huang W, Currey A, Chen X, et al. A comparison of lumpectomy cavity delineations between use of magnetic resonance imaging and computed tomography acquired with patient in prone position for radiation therapy planning of breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2016;94:832-840.

35. Hurkmans C, Admiraal M, van der Sangen M, et al. Significance of breast boost volume changes during radiotherapy in relation to current clinical interobserver variations. *Radiother. & Oncol.* 2009;90:60-65.

Figure Legends

Fig. 1. Average WB (33 patients with 11 observers) and SC (17 patients with 8 observers) inter-observer comparison for: a) Volume b) $DSC_{pairwise}$ (VOI) within datasets and average DSC for CT-MRI registered datasets.

Fig. 2. The average SC inter-observer and registration variation from eight observers for; a) $DSC_{pairwise}$ (inter-observer) and DSC(registration), and b) the COM $shift_{pairwise}$ (cm), across the 17 patients.

Table Legends

Table 1. Patient characteristics with 4 image-sets MR-supine (MRs), CT-supine (CTs), MR-prone (MRp) and CT-prone (CTp).

Table 2. Average metrics for 4 independent datasets for WB (33 patients and 11 observers) and SC volumes (17 patients and 8 observers) with 95% Confidence Intervals.

Table 3. Average for both WB CTV and SC volumes of each observers contour on CT compared to their contour on MRI for various metrics with 95% Confidence Intervals.

Table 1. Patient characteristics with 4 image-sets MR-supine (MRs), CT-supine (CTs), MR-prone (MRp) and CT-prone (CTp).

		No. of patients- WB	*No. of patients- SC*
Total No. of patients		33	17
Laterality	Left	15	5
	Right	18	12
Topography	UOQ	13	5
	UIQ	9	6
	LIQ	1	0
	LOQ	2	2
	Central Portion	2	2
	Unspecified	3	1
	Intraductal (CIS)	2	1
	Overlapping	1	0
Age in years	Median (Range)	59 (38-72)	62 (41-72)
T-stage	T1b	8	6
	T1c	17	6
	T2	6	4
	Tis (DCIS)	2	1
Axillary lymph node status/N-stage	N0	26	15
	N1	3	0
	N1a	3	1
	N3a	1	1

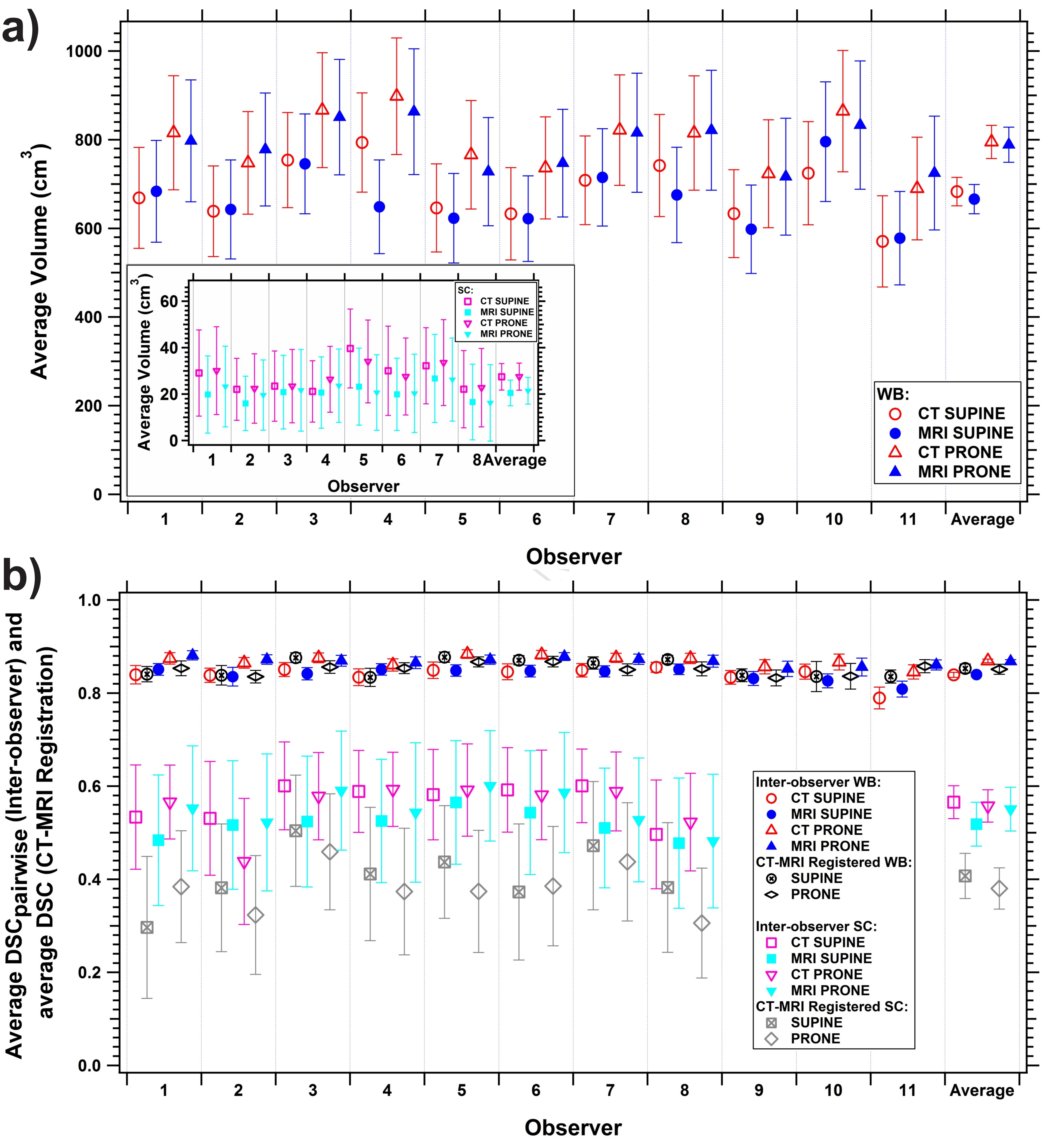
	NX	1	0
Receive neoadjuvant chemotherapy	Yes	13	7
	No	20	10
Stage	IA	21	11
	IIA	10	4
	IB	1	0
	IIIC	1	1
	0 (Tis (DCIS))	2	1
No. of patients that had surgical clips	4-6 clips	14	8
	1-2 surgical staples	6	6
	None	13	3
Time between surgery and Computed Tomography	< 1 month	2	2
	1-2 months	14	5
	2-3 months	4	2
	3-4 months	5	3
	4-5 months	6	5
	6-7 months	2	0
Cavity Visualisation Score category (averaged over all 4 datasets and 3 observers)	5(CTs/MRs/CTp/MRp)	2/5/2/5	2/5/2/4
	4(CTs/MRs/CTp/MRp)	7/5/5/6	4/3/4/4
	3(CTs/MRs/CTp/MRp)	14/7/16/10	8/3/7/5
	2(CTs/MRs/CTp/MRp)	9/15/10/12	3/6/4/4
	1(CTs/MRs/CTp/MRp)	1/1/0/0	0/0/0/0

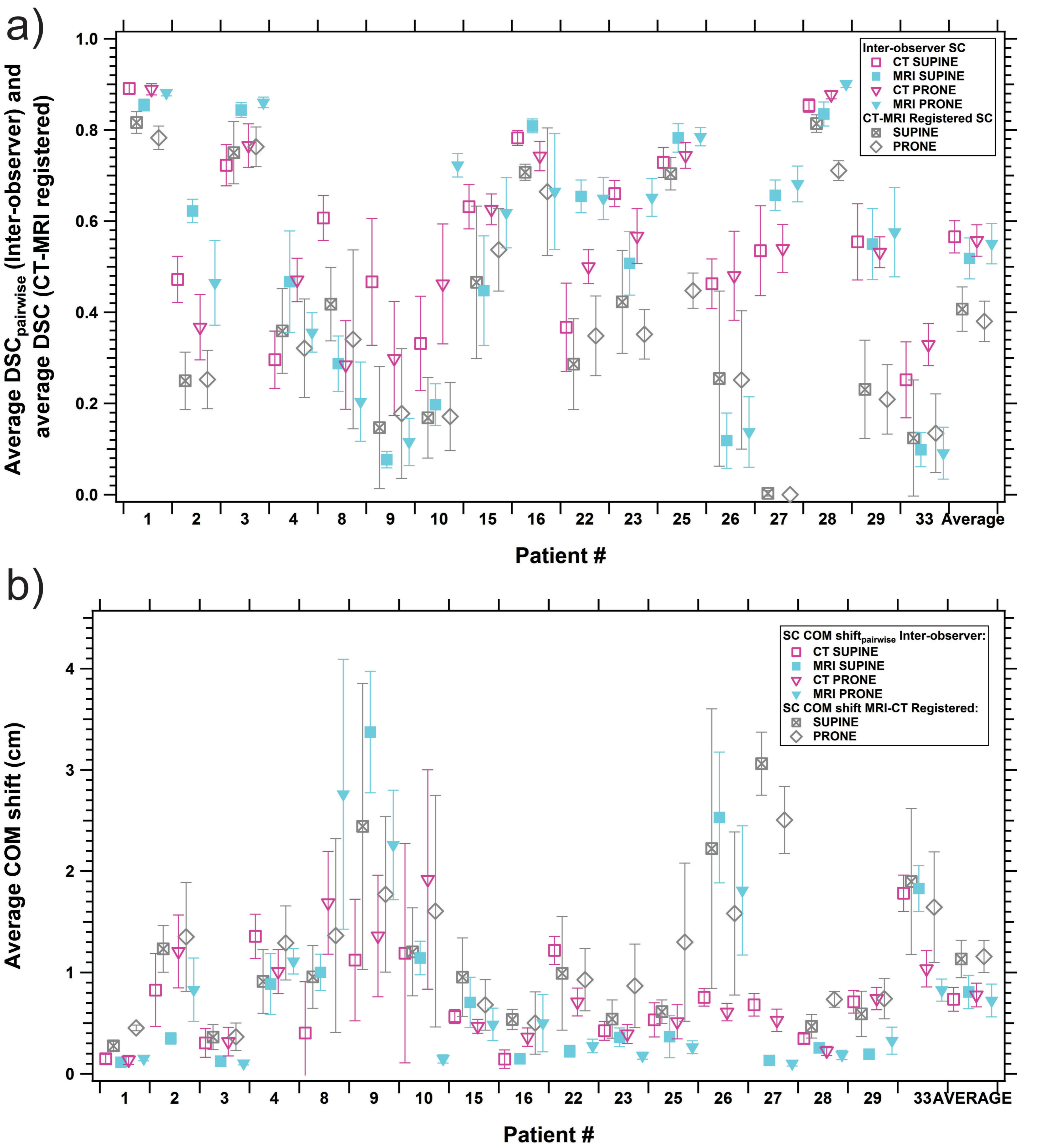
Table 2. Average gold standard and pairwise metrics for 4 independent datasets for WB (33 patients and 11 observers) and SC volumes (17 patients and 8 observers) with 95% Confidence Intervals.

	CT SUPINE	MRI SUPINE	CT PRONE	MRI PRONE
WB				
Volume (cm^3)	675±32	658±33	785±38	778±40
VOI (DSC pairwise)	0.839±0.005	0.840±0.004	0.867±0.004	0.868±0.004
Kappa	0.81±0.01	0.81±0.01	0.84±0.01	0.84±0.01
HD pairwise (cm)	2.78±0.06	3.07±0.07	3.16±0.08	3.09±0.08
AV COM shift pairwise (cm)	0.53±0.02	0.67±0.03	0.54±0.02	0.54±0.02
SC				
Volume (cm^3)	27±6	20±6	28±6	21±6
VOI (DSC pairwise)	0.57±0.03	0.52±0.04	0.56±0.03	0.55±0.04
Kappa	0.53±0.09	0.47±0.12	0.52±0.09	0.49±0.13
HD pairwise (cm)	1.8±0.1	1.6±0.2	1.9±0.3	1.6±0.5
AV COM shift pairwise (cm)	0.74±0.11	0.81±0.16	0.78±0.12	0.72±0.16

Table 3. Average for both WB CTV and SC volumes of each observers contour on CT compared to their contour on MRI for various metrics with 95% Confidence Intervals.

	Supine Registration (MRI to CT)	Prone Registration (MRI to CT)
WB		
DSC	0.853±0.005	0.851±0.005
HD (cm)	2.7±0.2	3.0±0.1
COM shift (cm)	0.87±0.15	0.88±0.16
SC		
DSC	0.41±0.05	0.38±0.04
HD (cm)	2.1±0.2	2.2±0.2
COM shift (cm)	1.13±0.19	1.16±0.16


a)
Average Volume (cm3)
0
200
400
600
800
1000
Observer
1
2
3
4
5
6
7
8
9
10
11
Average
WB:
CT SUPINE
MRI SUPINE
CT PRONE
MRI PRONE
SC:
CT SUPINE
MRI SUPINE
CT PRONE
MRI PRONE
0
20
40
60
8Average
b)
Average DSCpairwise (Inter-observer) and average DSC (CT-MRI Registration)
0.0
0.2
0.4
0.6
0.8
1.0
Inter-observer WB:
CT-MRI Registered WB:
SUPINE
PRONE
Inter-observer SC:
CT-MRI Registered SC:



a)
Average $DSC_{pairwise}$ (Inter-observer) and average DSC (CT-MRI registered)
Inter-observer SC
CT SUPINE
MRI SUPINE
CT PRONE
MRI PRONE
CT-MRI Registered SC
SUPINE
PRONE
0.0
0.2
0.4
0.6
0.8
1.0
1 2 3 4 8 9 10 15 16 22 23 25 26 27 28 29 33 Average
Patient #
b)
Average COM shift (cm)
SC COM $shift_{pairwise}$ Inter-observer:
CT SUPINE
MRI SUPINE
CT PRONE
MRI PRONE
SC COM shift MRI-CT Registered:
SUPINE
PRONE
0
1
2
3
4
1 2 3 4 8 9 10 15 16 22 23 25 26 27 28 29 33AVERAGE
Patient #

Summary

Inter-observer variation and volume differences in seroma cavity (SC) and Whole Breast (WB) Clinical Target Volumes (CTVs) from MRI and CT images, in prone and supine positions, were compared. Prone WB CTVs were larger and exhibited less inter-observer variation than supine in both imaging modalities. CT SC volumes were significantly larger than MRI in both positions. No clinically significant differences in inter-observer variation were demonstrated between MRI and CT for WB or SC volumes.