


RESEARCH

Open Access



Registering prostate external beam radiotherapy with a boost from high-dose-rate brachytherapy: a comparative evaluation of deformable registration algorithms

Calyn R. Moulton^{1*} , Michael J. House¹, Victoria Lye², Colin I. Tang², Michele Krawiec², David J. Joseph^{2,3}, James W. Denham⁴ and Martin A. Ebert^{1,2}

Abstract

Background: Registering CTs for patients receiving external beam radiotherapy (EBRT) with a boost dose from high-dose-rate brachytherapy (HDR) can be challenging due to considerable image discrepancies (e.g. rectal fillings, HDR needles, HDR artefacts and HDR rectal packing materials). This study is the first to comparatively evaluate image processing and registration methods used to register the rectums in EBRT and HDR CTs of prostate cancer patients. The focus is on the rectum due to planned future analysis of rectal dose-volume response.

Methods: For 64 patients, the EBRT CT was retrospectively registered to the HDR CT with rigid registration and non-rigid registration methods in VelocityAI. Image processing was undertaken on the HDR CT and the rigidly-registered EBRT CT to reduce the impact of discriminating features on alternative non-rigid registration methods applied in the software suite for Deformable Image Registration and Adaptive Radiotherapy Research (DIRART) using the Horn-Schunck optical flow and Demons algorithms. The propagated EBRT-rectum structures were compared with the HDR structure using the Dice similarity coefficient (DSC), Hausdorff distance (HD) and average surface distance (ASD). The image similarity was compared using mutual information (MI) and root mean squared error (MSE). The displacement vector field was assessed via the Jacobian determinant (JAC). The post-registration alignments of rectums for 21 patients were visually assessed.

Results: The greatest improvement in the median DSC relative to the rigid registration result was 35 % for the Horn-Schunck algorithm with image processing. This algorithm also provided the best ASD results. The VelocityAI algorithms provided superior HD, MI, MSE and JAC results. The visual assessment indicated that the rigid plus deformable multi-pass method within VelocityAI resulted in the best rectum alignment.

Conclusions: The DSC, ASD and HD improved significantly relative to the rigid registration result if image processing was applied prior to DIRART non-rigid registrations, whereas VelocityAI without image processing provided significant improvements. Reliance on a single rectum structure-correspondence metric would have been misleading as the metrics were inconsistent with one another and visual assessments. It was important to calculate metrics for a restricted region covering the organ of interest. Overall, VelocityAI generated the best registrations for the rectum according to the visual assessment, HD, MI, MSE and JAC results.

Keywords: Contour propagation, Image similarity, Registration validation

*Correspondence: calyn.moulton@research.uwa.edu.au

¹School of Physics (M013), University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

Full list of author information is available at the end of the article

Introduction

Radiotherapy dose-volume parameters for specific organs have been associated with normal tissue toxicity [1]. However, the correlation between planned dose-volume parameters and observed toxicities is confounded by how well the planned dose reflects the dose delivered [2]. Hence, studies have focused on developing methods for accumulating dose from daily fractions [3] or combined treatments [4, 5].

Therapies with different fractionation can be adjusted for fractionation effects by converting to equieffective dose given in 2 Gy fractions ($EQD2_{\alpha/\beta}$) [4, 6]. However, the anatomy in CTs may not coincide due to motion and variations in reference coordinate systems. Consequently, a 'worst case' assumption that the same volumes will receive the high doses is not necessarily valid as it is possible that a volume planned to receive a specific dose from one component could receive a different dose after adjustments for motion [7]. A rigid registration is not sufficient as non-rigid registration, also called deformable image registration (DIR), is required due to deformations and shrinkage [5]. A total dose distribution could be obtained after DIR by performing voxel-by-voxel summation of the $EQD2_{\alpha/\beta}$ doses [4, 8]. Combining dose without applying DIR via post-planning the brachytherapy on the external beam radiotherapy (EBRT) planning CT has been explored [9] and is subject to whether post-planning the brachytherapy dose is adequate given anatomy changes.

The accuracies of DIR algorithms have been examined experimentally using deformed phantoms or image modification to include deformations [10, 11]. The reliability of DIR has been examined for each patient by checking the agreement between the manually-delineated structure for one CT and the DIR applied to the manually-delineated structure from the other CT [12, 13]. Clinical checks of the post-registration anatomical alignment can also be used [14, 15]. Additionally, metrics assessing the displacement vector field (DVF) and the similarity between one image and the deformed image have been proposed as tools for assessing the reliability of DIR [16]. One evaluation type may be more appropriate in certain situations [12, 16]. The deformed dose distribution can be used reliably when DIR is considered to be adequate [17].

Publications are lacking in the context of registering an EBRT pelvic CT to a high-dose-rate brachytherapy (HDR) pelvic CT. Image-intensity based DIR algorithms applied to such CTs are susceptible to errors when there are major image differences [18]. This application is problematic given that the time between the HDR and EBRT planning CTs can be months. The discrepancies between the CTs include varying amounts of bowel gas, rectal filling and general artefacts. Additionally, only the HDR CT contains the HDR needles, streak artefacts off the needles, low

CT number pixels around the needles and rectal packing materials.

This study examines the performance of image processing and non-rigid registration tasks available in commercial software and customizations to an open-source package when applied to register the rectums in prostate EBRT and HDR data. Specifically, how did they perform in terms of the Dice similarity coefficient [12], Hausdorff distance [12], average surface distance [12], root mean squared error [12], mutual information [19], Jacobian determinant [12] and visual assessment [14]? We focus on the rectum due to planned future analysis of rectal dose-volume response for combined EBRT/HDR prostate treatment.

Patient data

This study used treatment plans for 64 prostate cancer patients who were treated with EBRT followed by a boost dose from Iridium-192 HDR via after-loading hollow metal needles at Sir Charles Gairdner Hospital in the period 2004–2008. Patient criteria and treatment methodology were as specified for the Trans-Tasman Radiation Oncology Group (TROG) 03.04 Randomized Androgen Deprivation and Radiotherapy (RADAR) trial [20, 21]. A planning CT was acquired at the start of each treatment component (e.g. Additional file 1: Figures A1 and A2). The number of slices (EBRT 32–77, HDR 32–59) and the voxel spacing (EBRT 0.809–0.977 mm, HDR 0.242–0.566 mm) for the CTs varied; however, there was a common slice thickness (3 mm) and dimension (512 by 512 pixels).

The external wall of the rectum was manually delineated by treating clinicians in the EBRT CTs using the Elekta Focal treatment planning software (Elekta AB, Stockholm, Sweden) and in the HDR CTs using the Brachyvision planning software (Varian Medical Systems, Palo Alto, US). Rectum outlines were reviewed (by author MK) for consistency between patients. The superior border of the rectum structures in the EBRT CTs were defined by the level that the rectum turns horizontally into the sigmoid colon and the inferior border defined on the most inferior axial image slice on which the ischial tuberosities were visible. Any further references to rectum 'structure' refer to the 3D manual outline of the external rectum wall while 'contour' refers to the 2D section of this outline on a particular image slice.

Ethics approval and consent to participate

The TROG 03.04 RADAR Trial is registered with the National Institutes of Health Clinical Trials Registry (number NCT00193856). This trial has approval from the Hunter New England Human Research Ethics Committee (Trial ID. 03/06/11/3.02), the Sir Charles Gairdner Group Human Research Ethics Committee (2003-050) and the

University of Western Australia Human Research Ethics Office (RA/4/1/5601). Patients participating in the trial signed consent forms.

Consent for publication

The signed patient consent forms for the trial informed patients that their medical information may be used to publish the results of the study. In accordance with the signed patient consent forms, this publication includes only anonymized information and does not include information identifying any patient.

Methods

Figure 1 illustrates the registration and evaluation process detailed in this section.

Rigid registration

A manual rigid registration (global translations and rotations) was performed in Velocity Advanced Imaging 2.8.1 (Varian Medical Systems, Palo Alto, US) to align the bony anatomy in the EBRT and HDR planning CTs. An automatic rigid registration was then performed to optimize the registration.

Copies of the HDR CT, the re-sampled rigidly-registered EBRT CT and the rectum structures from the HDR and rigidly-registered EBRT CTs were exported from Velocity Advanced Imaging (VelocityAI) in DICOM format for further image preprocessing in MATLAB™ R2010a (The MathWorks Inc., Massachusetts, US), CERR (version 4.1) [22] and DIRART (version 1.0a) [23]. At the time of export the rigidly-registered EBRT CTs

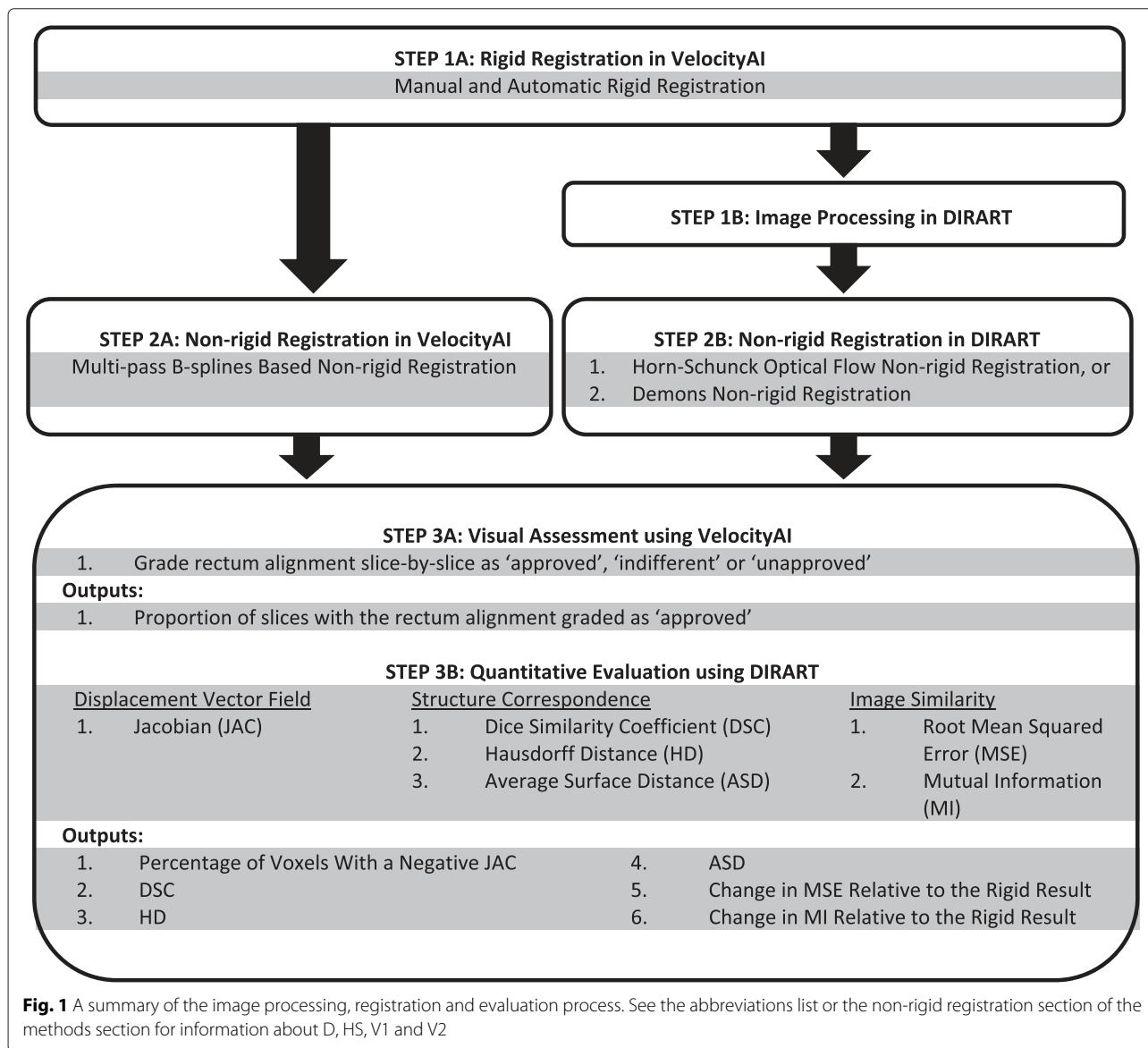


Fig. 1 A summary of the image processing, registration and evaluation process. See the abbreviations list or the non-rigid registration section of the methods section for information about D, HS, V1 and V2

were re-sampled to have the same voxel sizes and dimensions as the HDR CTs (see earlier section on patient data), which covered a smaller field-of-view.

Image preprocessing

Prior to DIR in DIRART the image processing detailed below was applied as the image processing led to a considerably improved post-registration rectum alignment. In Additional file 1: Figures A3 and A4 provide examples of slices of the final HDR and rigidly-registered EBRT CTs after image processing. The image processing steps are explained in detail in section I of Additional file 1. They key components are:

1. The HDR needles, HDR rectum packing material and HDR rectum low CT number artefacts were replaced with the average CT number of neighboring tissue pixels.
2. A Gaussian smoothing and blurring process was applied to avoid features in the HDR image caused by the previous pixel adjustments.
3. Rectum painting [14] with a uniform high CT number (2500) was applied to the final HDR and rigidly-registered EBRT CTs.

Non-rigid registration (deformable image registration)

Image processing was not applied prior to DIR in VelocityAI as the post-registration alignment in VelocityAI was reasonable relative to registrations obtained in DIRART without image processing. The multi-pass DIRs in VelocityAI (version 2.8.1) were based on the B-spline algorithm with the Mattes mutual information metric [24]. Additionally, non-rigid registrations in VelocityAI were performed by applying a global scale registration immediately before DIR. The VelocityAI methods were rigid, rigid plus multi-pass DIR (V1) and rigid plus scale plus multi-pass DIR (V2).

The DIR in DIRART was applied to the EBRT rigidly-registered CT and the HDR CT after image processing as this led to a considerably-improved post-registration rectum alignment and made it more comparable with the VelocityAI alignments. The original Demons and original Horn-Schunck optical flow (HSOF) algorithms were used. These DIRs use the root of the mean of the squared-intensity differences as the image-similarity metric [23]. The default settings in DIRART were used [23, 24]. The image processing and DIRs applied in DIRART were rigid plus image-processing plus HSOF-DIR (HS) and rigid plus image-processing plus Demons-DIR (D).

Evaluation

Visual assessments

The anatomical alignment for 64 patients was initially inspected by the researcher running each registration (author CRM). The post-DIR anatomical alignments for

21 of the 64 patients were inspected by a combination of in-training (author VL) and experienced (author CIT) radiation oncologists. The alignment between the rectums in the HDR CT and the registered EBRT CT was graded slice-by-slice using the spyglass tool in VelocityAI. The grades were 'approved', 'indifferent' or 'unapproved'. The grading was based on whether the misalignment was clinically relevant and was similar to the situation where an observer has to decide if a contour is sufficiently inconsistent with anatomy to warrant re-contouring. The results were assessed by calculating the proportion of slices with grades of the 'approved' type.

Structure-correspondence metrics

The Dice similarity coefficient (DSC) was calculated as the volume of overlap of the two structures and normalized by the average volume of the structures. The DSC range is zero (no overlap) to one (perfect overlap) [12]. The Hausdorff distance (HD) was calculated as the maximum of the distances from a point on one 3D structure to the closest point on the other 3D structure [12]. The average surface distance (ASD) was calculated as the average of the distances from a point on one 3D structure to the closest point on the other 3D structure [12]. Due to considerable differences in the slice span of the rectum structures for the HDR and EBRT CTs, these metrics were calculated over slices where the HDR (fixed image) rectum structure existed.

Image-similarity metrics

Image similarity was examined via the percentage increase (decrease) in the image-similarity (dissimilarity) metric relative to that before the registration. The mutual information (MI) was used for similarity and the root of the mean squared error (MSE) for dissimilarity [12, 19]. Using these two metrics ensured assessment with at least one image-similarity metric that was different to the metric used in the DIR algorithm to optimize the registration.

Displacement-vector-field metric

Physically unachievable organ deformations are indicated by negative Jacobian determinants (JAC) of the DVF [12]. Consequently, the physically-unachievable characteristics of the DVF can be summarized via the percentage of voxels with a negative JAC.

Statistical analysis

Paired percentage differences between the absolute DSC/ASD/HD results for different registration comparisons were tested for significance via exact Wilcoxon signed-rank tests against a zero median. The percentage JAC metric and the proportion of approved rectum-alignments for different registration comparisons were expressed in absolute difference and subject to the same

test. Quantile-quantile plots showed that differences were not normally distributed. The tests were performed in R (version 2.15.2) [25] using the Coin package [26] and the Pratt method for zeros [27]. *P*-values were considered significant if less than 0.05.

Results

Visual assessments

The major misalignments after DIR were observed around the pubic symphysis, ischium near the inferior extent of the obturator foramen, superior ramus of pubis near the obturator canal, coccyx, medial aspect of the acetabulum and anterior side of the rectum (see Additional file 1: Figure A5 for labeling of anatomy).

The medians of pairwise differences in the proportions of slices with the alignment of the rectum approved for various DIR comparisons of the V1, V2, D and HS methods are provided in Table 1. According to the median differences in rectum approval-proportions between registrations, the most useful to least useful alignments came from the V1, V2 and D/HS methods respectively. The median approval-proportions for the V1, V2, D and HS methods were 0.626, 0.574, 0.385 and 0.385 respectively.

The registration package providing the best rectum registration according to the other metrics detailed in the following sections was consistent irrespective of whether the metrics were calculated for the 64 patients or the subsample used for the visual assessments (see Additional file 2 for the results when metrics are calculated for the subsample). Consequently, the results for the metrics when they were calculated across the full analyzed data set were compared with the visual assessment results.

Structure-correspondence metrics

Figure 2 shows the DSCs after the HS, D, V1 and V2 methods for the 64 patients. Additionally, the median and interquartile range for the rigid registration DSCs were 0.641 and 0.142. The medians of the percentage differences between the DSC results for most comparisons of

the rigid, V1, V2, D and HS registrations were significantly different from zero given the Wilcoxon test *Z*-values and *p*-values. The significant differences for the HS, D, V1 and V2 registration comparisons are indicated in Fig. 2. The HS method achieved the best DSC results in terms of percentage differences with the other methods (Fig. 2).

Figure 3a and b show the ASD and HD results after the HS, D, V1 and V2 registration methods for the 64 patients. The significant differences for the HS, D, V1 and V2 registration comparisons via Wilcoxon signed-rank tests on pairwise percentage differences are indicated in Fig. 3a and b. The ASDs for the HS method were significantly smaller (smaller average shape discrepancy) than those for the D, V1 and V2 methods (Fig. 3a). However, the HDs for the V1 and V2 methods were significantly smaller (smaller extreme shape discrepancy) than those for HS and D methods (Fig. 3b).

All non-rigid registration methods led to a significant percentage improvement of the DSC, ASD and HD from the rigid registration result (see Additional file 1: Table A1 for statistical results).

Image-similarity metrics

Figure 4 summarizes the image similarity results by ranking the V1, V2, D and HS methods according to the MI and MSE values for the 64 patients (alternatively, see Additional file 1: Figure A6 for the values). The registrations with insignificant pairwise differences in metric values according to Wilcoxon signed-rank tests were assigned the same ranking in Fig. 4. Alternatively, see Additional file 1: Figure A7, Tables A2 and A4 for the statistical results.

Considering similarity over the entire images, the HS method led to the best change (greatest percentage reduction) in the median MSE relative to the rigid registration value (Fig. 4), whereas the HS/V1/V2 methods inseparably led to the best change (greatest percentage increase) in the median MI for similarity over the entire images (Fig. 4). However, the V1 and V2 methods inseparably provided

Table 1 Registration comparisons via pairwise differences in the proportions of slices with the rectum alignment approved. See the abbreviations list or the non-rigid registration section of the methods section for information about D, HS, V1 and V2

Registration comparison	<i>N</i>	Rectum approval-proportion median difference	<i>Z</i> -value	<i>P</i> -value
V2 versus V1	21	-0.032	-3.44	0.0002
HS versus V1	21	-0.169	-3.84	<0.0001
HS versus V2	21	-0.124	-3.22	0.0006
D versus V1	21	-0.241	-4.02	<0.0001
D versus V2	21	-0.189	-4.00	<0.0001
D versus HS	21	0	-	-

N is the number of patients to whom the registrations were applied. The pairwise difference is calculated as the first mentioned registration subtract the second mentioned registration. The *Z*-value and *p*-value are from the Wilcoxon signed-rank test of the pairwise differences against a median of zero (this test is not appropriate when the median difference is zero). A significant negative difference indicates the second mentioned registration is superior

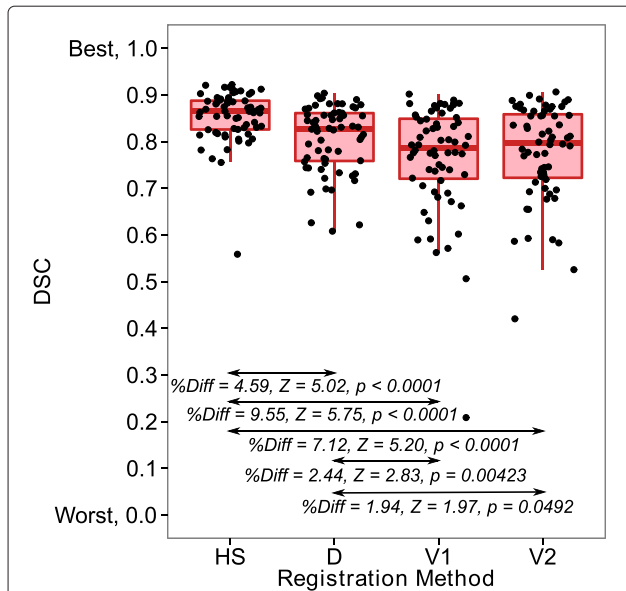


Fig. 2 The Dice similarity coefficient (DSC) results for registrations applied to 64 patients. The figure includes the median (thick horizontal line), interquartile ranges (large boxes), maximums/minimums without outliers (vertical lines from large boxes) and raw data points (filled circles). The median pairwise percentage difference (%Diff) between the indicated registrations is provided alongside the Z-values (Z) and p-values (p) from exact Wilcoxon signed-rank tests of a median of zero for the pairwise percentage difference. The difference was calculated as the registration on the left subtract the registration on the right and this difference was expressed as a percentage of the registration on the right. A significant positive percentage difference in DSC indicates that the registration on the left is superior. See the abbreviations list or the non-rigid registration section of the methods section for information about D, HS, V1 and V2

the best changes in the median MSE and median MI when considering similarity within the 3D bounding box enclosing both the HDR CT and rigidly-registered EBRT CT rectum structures (Fig. 4). For the DIRART methods, the MI decreased (deteriorated) relative to the rigid registration result and the MSE increased (deteriorated) relative to the rigid result when considering similarity in the 3D bounding box (Additional file 1: Figure A6).

Displacement-vector-field metrics

To determine orderings, Wilcoxon signed-rank tests of the pairwise differences in the percentage of voxels with a negative JAC between the HS, D, V1 and V2 methods were performed for the 64 patients. When the V2, V1, HS and D methods were compared for the DVF across the whole image, the ordering of methods according to increasing percentage of voxels with a negative JAC was D, V2 and V1/HS. However, the medians of the percentages of voxels with a negative JAC were zero for the VelocityAI methods when calculations were restricted to the region contained by the volume of the rigidly-registered EBRT rectum structure. For this region, the ordering of registrations in terms of increasing percentages of voxels

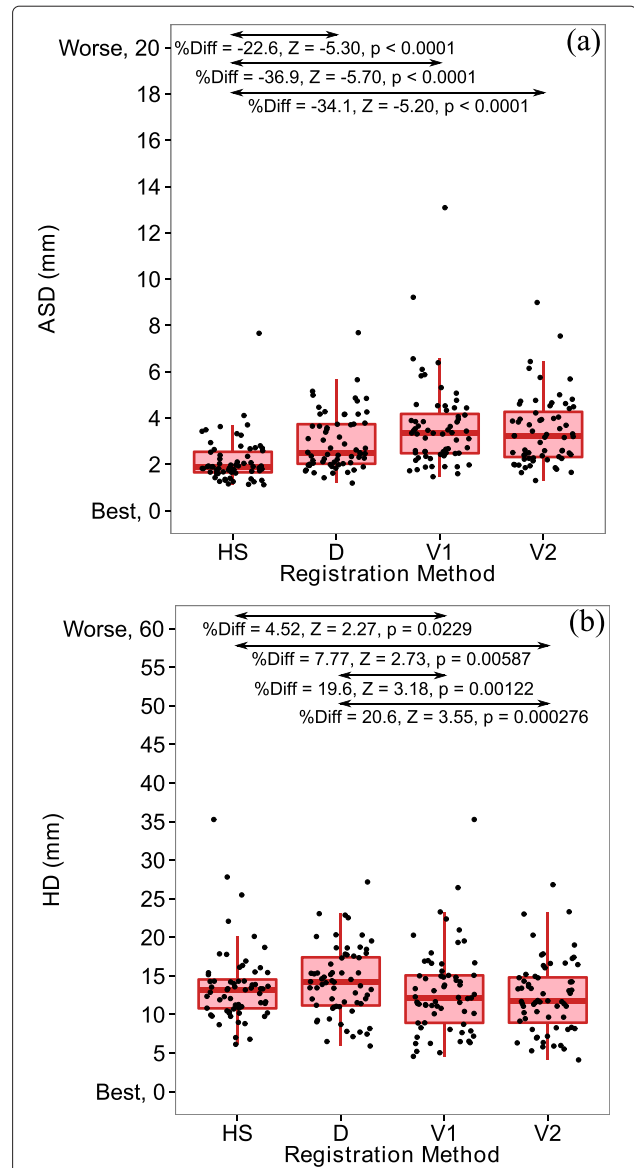


Fig. 3 a Average surface distance (ASD) and **b** Hausdorff (HD) results for registrations applied to 64 patients. The figures include the median (thick horizontal line), interquartile ranges (large boxes), maximums/minimums without outliers (vertical lines from large boxes) and raw data points (filled circles). The median pairwise percentage difference (%Diff) between the indicated registrations is provided alongside the Z-values (Z) and p-values (p) from exact Wilcoxon signed-rank tests of a median of zero for the pairwise percentage difference. The difference was calculated as the registration on the left subtract the registration on the right and this difference was expressed as a percentage of the registration on the right. A significant negative percentage difference in ASD indicates that the registration on the left is superior. A significant positive percentage difference in HD indicates that the registration on the right is superior. See the abbreviations list or the non-rigid registration section of the methods section for information about D, HS, V1 and V2

with a negative JAC was V1/V2, D and HS. Alternatively, Additional file 1: Figure A6 provides values with the test results detailed in Figure A7, Tables A2 and A3.

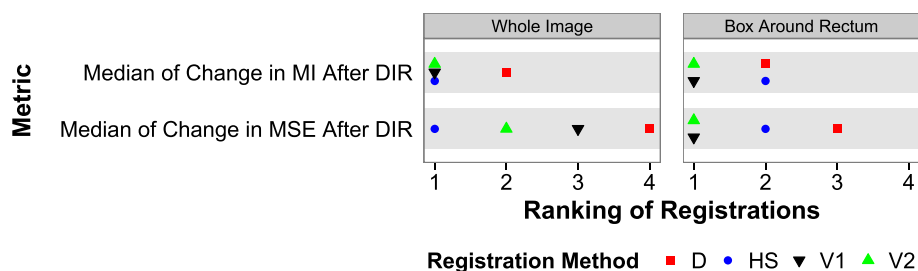


Fig. 4 Ranking of registration methods according to image-similarity results for registrations applied to 64 patients. The medians of the percentage changes in the mean square error (MSE) and mutual information (MI) were calculated via $100 \times \frac{\text{after}}{\text{before}} - 100$. Increasing ranking indicates less image similarity and more image dissimilarity. Some registrations share the same ranking due to insignificant ($p > 0.05$) paired differences for metric values (alternatively, see Additional file 1: Figures A6 and A7 for metric values and the results of the statistical significance tests). The MSE and MI metrics were calculated for two regions of interest, which were the entirety of the images and the bounding box enclosing both the HDR CT and rigidly-registered EBRT CT rectum structures. See the abbreviations list or the non-rigid registration section of the methods section for information about D, HS, V1 and V2

Discussion

Visual assessments were important

The V1 method was superior to the V2, HS and D methods in terms of significant differences in the proportions of slices with the rectum-alignment approved according to the visual assessments. Additionally, the VelocityAI methods (V1 and V2) resulted in superior rectum alignment approval-proportions compared to the DIRART methods. This was inconsistent with the structure-correspondence metric results, where the HS and D methods achieved better DSCs with worse HDs. The inconsistency between the results of metrics and visual assessments has been identified before [28]. Additionally, in this case it supports the current practice that a sole structure-correspondence metric cannot be used for the remaining registrations of the larger dataset as a filtering measure in lieu of a slice-by-slice visual assessment by expert observers.

The visual assessment results can be confounded by intra-observer and inter-observer assessment variations [28]; however, the impact of these variations was reduced by conducting the analysis via paired registration differences and the same observer assessing the four registrations per patient in a consecutive manner.

Deformable image registration improved the rigid registration results

DIR was useful as, for example, the DSC, ASD and HD results were improved by applying DIR methods after rigid registration. The improvement in the median DSC was 35 % for the HS algorithm with image processing as compared to rigid registration. This compares well with the 31 % improvement in the mean DSC obtained by a study using the same algorithm with similar image processing tasks in the context of registering daily megavoltage CT images to treatment planning kilo-voltage CT images [29]. The results and comparisons are confounded by inter/intra-observer variations in contouring [30].

The choice of metrics and the way they were calculated were important

The results for the structure-correspondence metrics indicate that the selection of structure-correspondence metrics should be made carefully. The HS method was superior to the D, V1 and V2 methods in terms of a better structure-volume match (DSC) and less overall shape discrepancy (ASD). The V1 and V2 methods were superior to the HS and D methods in terms of extreme shape discrepancy (HD). The inconsistency of these metrics contrasts with another study where they were useful for evaluations [12]. In this case, the most extreme shape discrepancy (HD) is important from a dosimetric perspective as the anterior side of the registered rectum-structure could deviate from the fixed structure by extending over the brachytherapy high-dose area. Consequently, the correlation between the most extreme shape discrepancy and the high-dose parameters after registration may be useful when checking the validity of deformed dose.

It is important to calculate metrics over a restricted region that covers the area of concern or the organ at risk rather than the whole image when assessing whether the registration is acceptable in that area or for the organ at risk. The reason is that the registration is optimized over a region of interest and the performance can vary locally. For example, the V1/V2 methods provided optimal rectum results in terms of MI and MSE calculated in the region defined by the volume of the rigidly-registered EBRT rectum structure, whereas the HS method provided the best MSE result when calculated over the whole image. Additionally, unlike the DIRART algorithms the VelocityAI algorithms led to improvements in the image-similarity metrics calculated across the rectum relative to the rigid registration result. The choice of metrics can be important as elsewhere the MSE was found to not be useful for evaluation [12].

The B-splines based registration resulted in the best registration of the rectum

The results for the rectum were sufficiently different to distinguish the best-performing VelocityAI registration from the best-performing DIRART registration. Relative to the DIRART algorithms, the VelocityAI algorithms did achieve better image similarity and visual alignment over the region contained by the volume of the rectum structure. Additionally, the VelocityAI algorithms appeared to do so with less physically-unrealistic displacements (smaller percentage of displacements with negative JACs) and less extreme shape discrepancy between the fixed and propagated rectum structures (smaller HD). As there was no image processing prior to the VelocityAI algorithms, the VelocityAI algorithms achieved these superior results whilst exposed to rectum discrepancies. As such, this study demonstrates the VelocityAI DIRs (B-splines based) appeared to result in the best rectum alignment and achieve DVFs with the least physically-unrealistic displacements.

This evaluation is based on the algorithms in the form they were released. Also, the user cannot change the registration parameters in VelocityAI. If the parameters in both packages were adjustable it would be a useful and difficult task to find optimal performance [24].

The comparative evaluations of rectum registrations from different registration systems is important for adequately accumulating dose for combined EBRT/HDR prostate cancer treatment and correlating it with observed gastrointestinal toxicities. The assessment of impact of image registration on dose-outcomes correlation will provide additional validation of the alternative approaches, and this is the subject of ongoing investigation.

Recommendations and future considerations

- Registrations may benefit from images immediately prior to the HDR insertion of needles as this may allow changes over the preceding months to be separated from changes due to HDR needles and treatment positioning.
- Given the image discrepancies, it would be useful to evaluate registrations including a recently-developed penalty term minimizing the volume of missing information [18], methods that exclude the rectum discrepancies [31, 32] or changes to DIRART to use other image-similarity metrics (e.g. mutual information).
- Evaluation of registrations customized for the urethra, bladder, prostate and seminal vesicles would be useful as they require work on considerable image issues (e.g. HDR needles in the prostate and the urethra catheter balloon in the bladder).
- Registration evaluation for patients can be difficult and involve a variety of methods as there is no direct

measure of registration error due to no known ground truth. Information obtained from other evaluation methods such as landmarks, phantoms and deformed dose uncertainty [24, 33–36] would be useful if applied to HDR CTs given the image contents.

Conclusion

This study demonstrated that structure correspondence, image similarity and visual assessments are useful for assessing registrations applied to EBRT and HDR CTs of prostate cancer patients. We found that using non-rigid registrations in VelocityAI or image processing plus non-rigid registrations in DIRART improved the alignment of the rectum according to visual assessment and various metrics. It would have been misleading to use a structure-correspondence metric as a sole indicator of rectum alignment given that such metrics were inconsistent with other metrics and visual assessments. It is recommended that image-similarity and displacement-vector-field metrics be calculated for a restricted region covering the organ of interest instead of using global values. Applying the DIR methods in VelocityAI provided the most optimal registration result for the rectum as assessed by the greatest rectum alignment approval-proportion, the least extreme shape discrepancy between rectum structures and the most optimal rectum image similarity. We encourage the development of registrations for the prostate and urethra in EBRT and HDR CTs as doses to the prostate and urethra are key clinical concerns in the RADAR trial.

Additional files

Additional file 1: Online supplementary material providing additional method details, patient images, metric values and statistical analysis. (PDF 0.46 kb)

Additional file 2: Online supplementary material providing metric results for the 21 patient subsample. (PDF 796 kb)

Abbreviations

%Diff: median pairwise percentage difference; ASD: average surface distance; CERR: computational environment for radiotherapy research; D: rigid plus image processing plus Demons-DIR in DIRART; DIR: deformable image registration; DIRART: deformable image registration and adaptive radiotherapy research; DSC: dice similarity coefficient; DVF: displacement vector field; EBRT: external beam radiotherapy; EQD2: equieffective dose given in 2 Gy fractions; HD: Hausdorff distance; HDR: high-dose-rate brachytherapy; HS: rigid plus image processing plus HSOF-DIR in DIRART; HSOF: Horn-Schunck optical flow registration algorithm; JAC: Jacobian determinant; MI: mutual information; MSE: root of the mean squared error; p : p -value from Wilcoxon signed-rank test; RADAR: randomized androgen deprivation and radiotherapy; TROG: trans-tasman radiation oncology group; V1: rigid plus multi-pass DIR in VelocityAI; V2: rigid plus scale plus multi-pass DIR in VelocityAI; VelocityAI: velocity advanced imaging; Z: z-value from Wilcoxon signed-rank test.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CRM, MJH, CIT and MAE have made substantial contributions to design of analysis. CRM, VL, CIT, MK, DJJ, JWD and MAE have made substantial

contributions to acquire the data. CRM, MJH, VL, CIT and MAE substantially contributed to analysis and interpretation of data. CRM has been involved in drafting the manuscript. MJH, VL, CIT, DJJ, JWD and MAE revised it critically for important and correct content. JWD and DJJ were involved in the design and coordination of the RADAR trial. All authors read and approved the version to be published.

Acknowledgements

We acknowledge funding from the National Health and Medical Research Council (300705, 455521, 1006447), the University of Western Australia, an Australian Postgraduate Award, an Ana Africh Scholarship, the Hunter Medical Research Institute, the Health Research Council (New Zealand), Abbott Laboratories and Novartis Pharmaceuticals. We thank Annette Haworth and radiation oncology staff at Sir Charles Gairdner Hospital for their contributions.

Author details

¹School of Physics (M013), University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia. ²Radiation Oncology, Sir Charles Gairdner Hospital, Hospital Avenue, Nedlands, Western Australia 6009, Australia. ³School of Surgery, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia. ⁴School of Medicine and Population Health, University of Newcastle, University Drive, Callaghan, New South Wales 2308, Australia.

Received: 15 October 2015 Accepted: 7 December 2015

Published online: 14 December 2015

References

- Michalski JM, Gay H, Jackson A, Tucker SL, Deasy JO. Radiation dose-volume effects in radiation-induced rectal injury. *Int J Radiat Oncol Biol Phys.* 2010;76(Suppl 3):123–9. doi:10.1016/j.ijrobp.2009.03.078.
- Thor M, Apte A, Deasy JO, Karlsdóttir Å, Moiseenko V, Liu M, et al. Dose/volume-response relations for rectal morbidity using planned and simulated motion-inclusive dose distributions. *Radiother Oncol.* 2013;109(3):388–93. doi:10.1016/j.radonc.2013.10.021.
- Godley A, Ahunbay E, Peng C, Li XA. Accumulating daily-varied dose distributions of prostate radiation therapy with soft-tissue-based KV CT guidance. *J Appl Clin Med Phys.* 2012;13(3):98–107. doi:10.1120/jacmp.v13i3.3859.
- Vásquez Osorio EM, Hoogeman MS, Teguh DN, Al-Mamgani A, Kolkman-Deurloo IKK, Bondar L, et al. Three-dimensional dose addition of external beam radiotherapy and brachytherapy for oropharyngeal patients using non-rigid registration. *Int J Radiat Oncol Biol Phys.* 2011;80(4):1268–1277. doi:10.1016/j.ijrobp.2010.10.006.
- Van de Kamer JB, De Leeuw AAC, Moerland MA, Jürgenliemk-Schulz IM. Determining DVH parameters for combined external beam and brachytherapy treatment: 3D biological dose adding for patients with cervical cancer. *Radiother Oncol.* 2010;94(2):248–53. doi:10.1016/j.radonc.2009.12.011.
- Bentzen SM, Dörr W, Gahbauer R, Howell RW, Joiner MC, Jones B, et al. Bioeffect modeling and equieffective dose concepts in radiation oncology - terminology, quantities and units. *Radiother Oncol.* 2012;105(2):266–8. doi:10.1016/j.radonc.2012.10.006.
- Andersen ES, Noe KØ, Sørensen TS, Nielsen SK, Fokdal L, Paludan M, et al. Simple DVH parameter addition as compared to deformable registration for bladder dose accumulation in cervix cancer brachytherapy. *Radiother Oncol.* 2013;107(1):52–7. doi:10.1016/j.radonc.2013.01.013.
- Abe T, Tamaki T, Makino S, Ebara T, Hirai R, Miyaoura K, et al. Assessing cumulative dose distributions in combined radiotherapy for cervical cancer using deformable image registration with pre-imaging preparations. *Radiat Oncol.* 2014;9:293. doi:10.1186/s13014-014-0293-4.
- Kikuchi K, Nakamura R, Tanji S, Yamaguchi S, Kakuhara H, Yabuuchi T, et al. Three-dimensional summation of rectal doses in brachytherapy combined with external beam radiotherapy for prostate cancer. *Radiother Oncol.* 2013;107(2):159–64. doi:10.1016/j.radonc.2013.03.003.
- Nie K, Chuang C, Kirby N, Braunstein S, Pouliot J. Site-specific deformable imaging registration algorithm selection using patient-based simulated deformations. *Med Phys.* 2013;40(4):041911. doi:10.1118/1.4793723.
- Yeo UJ, Supple JR, Taylor ML, Smith R, Kron T, Franich RD. Performance of 12 DIR algorithms in low-contrast regions for mass and density conserving deformation. *Med Phys.* 2013;40(10):101701. doi:10.1118/1.4819945.
- Varadhan R, Karangelis G, Krishnan K, Hui S. A framework for deformable image registration validation in radiotherapy clinical applications. *J Appl Clin Med Phys.* 2013;14(1):4066. doi:10.1120/jacmp.v14i1.4066.
- Hardcastle N, Tomé WA, Cannon DM, Brouwer CL, Wittendorp PW, Dogan N, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. *Radiat Oncol.* 2012;7:90. doi:10.1186/1748-717X-7-90.
- Thörnqvist S, Petersen JBB, Høyer M, Bentzen LN, Muren LP. Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration. *Acta Oncol.* 2010;49(7):1023–1032. doi:10.3109/0284186X.2010.503662.
- Ramadaan IS, Peick K, Hamilton DA, Evans J, Iupati D, Nicholson A, et al. Validation of Varian's SmartAdapt[®] deformable image registration algorithm for clinical application. *Radiat Oncol.* 2015;10:73. doi:10.1186/s13014-015-0372-1.
- Hub M, Karger CP. Estimation of the uncertainty of elastic image registration with the demons algorithm. *Phys Med Biol.* 2013;58(9):3023–3036. doi:10.1088/0031-9155/58/9/3023.
- Bender ET, Hardcastle N, Tomé WA. On the dosimetric effect and reduction of inverse consistency and transitivity errors in deformable image registration for dose accumulation. *Med Phys.* 2012;39(1):272–80. doi:10.1118/1.3666948.
- Berendsen FF, Kotte ANTJ, de Leeuw AAC, Jürgenliemk-Schulz IM, Vieregger MA, Pluim JPW. Registration of structurally dissimilar images in MRI-based brachytherapy. *Phys Med Biol.* 2014;59(15):4033–4045. doi:10.1088/0031-9155/59/4/4033.
- Oliveira FPM, Tavares JMRS. Medical image registration: a review. *Comput Methods Biomech Biomed Engin.* 2014;17(2):73–93. doi:10.1080/10255842.2012.670855.
- Denham JW, Steigler A, Joseph D, Lamb DS, Spry NA, Duchesne G, et al. Radiation dose escalation or longer androgen suppression for locally advanced prostate cancer? data from the TROG 03.04 RADAR trial. *Radiother Oncol.* 2015;115(3):301–7. doi:10.1016/j.radonc.2015.05.016.
- Denham JW, Joseph D, Lamb DS, Spry NA, Duchesne G, Matthews J, et al. Short-term androgen suppression and radiotherapy versus intermediate-term androgen suppression and radiotherapy, with or without zoledronic acid, in men with locally advanced prostate cancer (TROG 03.04 RADAR): an open-label, randomised, phase 3 factorial trial. *Lancet Oncol.* 2014;15(10):1076–1089. doi:10.1016/S1470-2045(14)70328-6.
- Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. *Med Phys.* 2003;30(5):979–85. doi:10.1118/1.1568978.
- Yang D, Brame S, El Naqa I, Aditya A, Wu Y, Goddu SM, et al. Technical note: DIRART - a software suite for deformable image registration and adaptive radiotherapy research. *Med Phys.* 2011;38(1):67–77. doi:10.1118/1.3521468.
- Kadoya N, Fujita Y, Katsuta Y, Dobashi S, Takeda K, Kishi K, et al. Evaluation of various deformable image registration algorithms for thoracic images. *J Radiat Res.* 2014;55(1):175–82. doi:10.1093/jrr/rrt093.
- R Development Core Team. R: A Language and Environment for Statistical Computing. 2015. <http://www.R-project.org>, Accessed 14 Oct 2015.
- Hothorn T, Hornik K, van de Wiel MA, Zeileis A. A lego system for conditional inference. *Am Stat.* 2006;60(3):257–63. doi:10.1198/000313006X118430.
- Pratt JW. Remarks on zeros and ties in the Wilcoxon signed rank procedures. *J Am Stat Assoc.* 1959;54(287):655–67. doi:10.1080/01621459.1959.10501526.
- Thor M, Petersen JBB, Bentzen L, Høyer M, Muren LP. Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer. *Acta Oncol.* 2011;50(6):918–25. doi:10.3109/0284186X.2011.577806.
- Yang D, Chaudhari SR, Goddu SM, Pratt D, Khullar D, Deasy JO, et al. Deformable registration of abdominal kilovoltage treatment planning CT and tomotherapy daily megavoltage CT for treatment adaptation. *Med Phys.* 2009;36(2):329–38. doi:10.1118/1.3049594.
- Geraghty JP, Grogan G, Ebert MA. Automatic segmentation of male pelvic anatomy on computed tomography images: a comparison with

- multiple observers in the context of a multicentre clinical trial. *Radiat Oncol.* 2013;8:106. doi:10.1186/1748-717X-8-106.
31. Gu X, Dong B, Wang J, Yordy J, Mell L, Jia X, et al. A contour-guided deformable image registration algorithm for adaptive radiotherapy. *Phys Med Biol.* 2013;58(6):1889–1901. doi:10.1088/0031-9155/58/6/1889.
 32. Xie Y, Chao M, Lee P, Xing L. Feature-based rectal contour propagation from planning CT to cone beam CT. *Med Phys.* 2008;35(10):4450–459. doi:10.1118/1.2975230.
 33. Stanley N, Glide-Hurst C, Kim J, Adams J, Li S, Wen N, et al. Using patient-specific phantoms to evaluate deformable image registration algorithms for adaptive radiation therapy. *J Appl Clin Med Phys.* 2013;14(6):4363. doi:10.1120/jacmp.v14i6.4363.
 34. Tilly D, Tilly N, Ahnesjö A. Dose mapping sensitivity to deformable registration uncertainties in fractionated radiotherapy - applied to prostate proton treatments. *BMC Med Phys.* 2013;13(1):2. doi:10.1186/1756-6649-13-2.
 35. Yeo UJ, Taylor ML, Supple JR, Smith RL, Dunn L, Kron T, et al. Is it sensible to deform dose? 3D experimental validation of dose-warping. *Med Phys.* 2012;39(8):5065–072. doi:10.1118/1.4736534.
 36. Murphy MJ, Salguero FJ, Siebers JV, Staub D, Vaman C. A method to estimate the effect of deformable image registration uncertainties on daily dose mapping. *Med Phys.* 2012;39(2):573–80. doi:10.1118/1.3673772.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

