Kemppainen, R.; Vaara, T.; Joensuu, T.; Kiljunen, T.

Accuracy and precision of patient positioning for pelvic MR-only radiation therapy using digitally reconstructed radiographs

Published in:
Physics in Medicine and Biology

DOI:
10.1088/1361-6560/aaad21

Published: 02/03/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Kemppainen, R., Vaara, T., Joensuu, T., & Kiljunen, T. (2018). Accuracy and precision of patient positioning for pelvic MR-only radiation therapy using digitally reconstructed radiographs. *Physics in Medicine and Biology*, *63*(5), 1-10. Article 055009. https://doi.org/10.1088/1361-6560/aaad21

Accuracy and precision of patient positioning for pelvic MR-only radiation therapy using digitally reconstructed radiographs

R. Kemppainen[1,3], T. Vaara[1], T. Joensuu[2], T. Kiljunen[2]

1. Philips MR Therapy, Äyritie 4, FI-01510, Vantaa, Finland

2. Docrates Cancer Center, Saukonpaadenranta 2, FI-00180, Helsinki Finland.

3. Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Rakentajanaukio 2 C, FI-02150 Espoo, Finland

Address for correspondence: Reko Kemppainen, Philips MR Therapy, Äyritie 4 FI-01510 Vantaa, Finland

E-mail contact: reko.kemppainen@gmail.com

Keywords: Radiotherapy, MRI-based Radiotherapy, Image-guided radiotherapy, digitally reconstructed radioraphs

Abstract

Background and Purpose

Magnetic resonance imaging (MRI) has in recent years emerged as an imaging modality to drive precise contouring of targets and organs at risk in external beam radiation therapy. Moreover, recent advances in MRI enable treatment of cancer without computed tomography (CT) simulation. A commercially available MR-only solution, MRCAT, offers a single-modality approach that provides density information for dose calculation and generation of positioning reference images. We evaluated the accuracy of patient positioning based on MRCAT digitally reconstructed radiographs (DRRs) by comparing to standard CT based workflow.

Materials and Methods

Twenty consecutive prostate cancer patients being treated with external beam radiation therapy (EBRT) were included in the study. DRRs were generated for each patient based on the planning CT and MRCAT. The accuracy assessment was performed by manually registering the DRR images to planar kV setup images using bony landmarks. A Bayesian linear mixed effects model (LME) was used to separate systematic and random components (inter- and intra-observer variation) in the assessment. In addition, method agreement was assessed using a Bland–Altman analysis.

Results

The systematic difference between MRCAT and CT based patient positioning, averaged over the study population, were found to be (mean [95% CI]) -0.49 [-0.85 – -0.13] mm, 0.11 [-0.33 – +0.57] mm and -0.05 [-0.23 – +0.36] mm in vertical, longitudinal and lateral directions, respectively. The increases in total random uncertainty were estimated to be below 0.5 mm for all directions, when using MR-only workflow instead of CT.

Conclusions

The MRCAT pseudo-CT method provides clinically acceptable accuracy and precision for patient positioning for pelvic radiation therapy based on planar DRR images. Furthermore, due to the reduction of geometric uncertainty, compared to dual-modality workflow, the approach is likely to improve the total geometric accuracy of pelvic radiation therapy.

Introduction

In contemporary external beam radiation therapy (EBRT) of prostate cancer, computed tomography (CT) is the primary imaging modality providing anatomical and tissue density information. Magnetic resonance imaging (MRI) is widely used as a supplement to CT in EBRT of prostate cancer. The major advantages of MRI over CT are superior contrast-to-noise ration and better soft tissue differentiation resulting decreased contouring variability for prostate [1] and reduced organs at risk (OAR) dose [2]. Additional benefits include lack of ionizing radiation and versatility of existing MR imaging methods.

Major drawback of multi modal imaging in radiation therapy (RT) is the registration error introduced when images from two or more imaging modalities are registered [3]. Recent advances in usage of MR in RT promise to eliminate the registration error completely by using only MR images for EBRT planning of prostate cancer [4]–[7]. In the MR-only approach a so called pseudo-CT is obtained based on information available in the MR image. The pseudo-CT is then used to replace the CT image for both dose calculation and generation of digitally reconstructed radiographs (DRRs). Generation of DRRs based on MR images will be required if the patient alignment prior to treatment delivery must be verified by on-board planar X-ray imaging [8].

Numerous studies have been published recently that report adequate dose calculation accuracy with different MR-only strategies [9]. The focus in MR-only publications has been in dosimetric accuracy and, albeit equally important for total accuracy and treatment outcome, less attention has been paid to complete workflow and geometric accuracy of suggested methods. Prior work has demonstrated that manually delineated pelvic bones can provide sufficient accuracy for patient positioning based on DRR images using bulk assignment of HU values for bones in an MR-only workflow [10], [11]. However, feasibility of positioning methods for image guided radiation therapy (IGRT) may be very sensitive to differing number of density values used for modeling of bones, segmentation accuracy of bones and geometric accuracy of the chosen methods in general. Thus, the published results may not be generalizable to methods that use automated delineation of bones. Feasibility of such pseudo-CT method should be verified independently by realistic use of the images where a human observer registers the DRR images to kV-radiographs.

This study compares MRCAT and CT based DRRs in IGRT for positioning of patients for determining the treatment isocenter locations for daily treatments. The quantitative assessment of suitability of an MRCAT DRR for this purpose is done by simulating the standard clinical workflow of registering DRRs and planar radiographs manually. Planar radiographs (kV-images) taken before the treatment are registered with the MRCAT images and the registration results are compared with those obtained by using the CT-based DRRs. To authors knowledge, a thorough method comparison study has not been conducted to assess the sources of difference between CT and pseudo-CT based DRR patient positioning. For the first time, we analyze the impact to both systematic and random errors. Based on the impact to the required clinical target volume (CTV) margin, we derive an acceptance criterion for the use of pseudo-CT based DRRs.

Methods

Study design and image acquisition

The study cohort consists of 20 prostate cancer patients treated with external beam radiotherapy in the Docrates Cancer Center, Helsinki Finland. Patients were treated according to Docrates' clinical protocol for prostate cancer patients that is based on pelvic CT simulation and additional MR images that are registered with the planning CT for target and OAR contouring. In this study, secondary MR image set was taken that allowed retrospective comparison between CT and MR-based digitally reconstructed radiographs (DRRs). Thus, the treatment of the patients was not affected due to the study. The study protocol was approved by the Helsinki University Hospital Coordinating Ethics Committee.

Planning CT images were acquired using Siemens Sensation Open scanner with minimum of 40 cm FH coverage, 2 mm slices, 1 mm x 1 mm in-plane resolution, 120 kV tube voltage, mA modulation (Quality ref. mAs (CareDose) 190 mAs) and B31s reconstruction kernel. MR Images were acquired using Ingenia 1.5 T (n=8) or 3.0 T (n=12) scanners (software version 5.1.7 (n=11), 5.2 (2 cases) or 5.3 (n=7)) (Philips Medical Systems, B.V., Best, Netherlands). For all patients an axial 3D mDIXON sequence (T1-weighted, resolution 1.04 x 1.04 x 2.50 mm^3) covering full body contour in axial plane was taken and used for automatic generation of the MRCAT image. The scan protocol for MRCAT was fixed and specified by the manufacturer [12]. Full details of the sequence parameters are given in Appendix B. MR imaging time was below 200 seconds for all patients. Patients were immobilized similarly during the CT and MR simulation imaging. In the MR scan, patients were in supine position on a flat table and the anterior MR-coil was placed above the imaging volume using a coil holder provided by the manufacturer.

The kV positioning images were taken with an OBI system (On-Board Imager, Varian Medical Systems, Palo Alto, USA) integrated to a linear accelerator. Two orthogonal projections (AP or PA and LR or RL) were obtained for all patients (pixel-size: 0,388 x 0,388 mm^2, FOV: 30 x 40 cm^2).

MRCAT pseudo-CT generation

For pseudo-CT images, a commercially available MRCAT product, integrated with MR scanner software, was used to generate the images. In the MRCAT generating algorithm, CT-like density maps are generated fully automatically from an mDIXON image [12].

Bone structures in MRCAT are automatically segmented inside the body using the multiple contrasts provided by the mDIXON scan. Both the bone and outline segmentation employ a model-based segmentation approach trained on patient and volunteer mDIXON image datasets. The model contains information of average bone shape and how the shape varies in the training population. The model is adapted to an actual patient image using features (such as gray value edges) found within the image, while at the same time, a constraint for the shape of the segmented structure prevents the segmentation from being attracted to a wrong position [13].

Voxels inside the bone segmentation are assumed to contain either compact or spongy bone. An intensity threshold is used; lower intensities are considered to consist of compact bone, while voxels with higher intensity are assumed to contain spongy bone. The choice of the threshold value has been selected so that average bone density match with CT on population level.

Treatment planning

RT plans were created based on the simulation CTs according to the Docrates Cancer Center clinical practice. To study the DRRs generated from MRCAT, the structure sets and plans were copied from the original CT to rigidly registered MRCAT image and, thus, the isocenter position was the same, (within accuracy of the registration) in both plans. Pinnacle (Philips Healthcare, version 9.10) treatment planning system (TPS) was used for generating the digitally reconstructed radiographs (see Figure 1 for an example of CT and MRCAT DRRs used in the study).

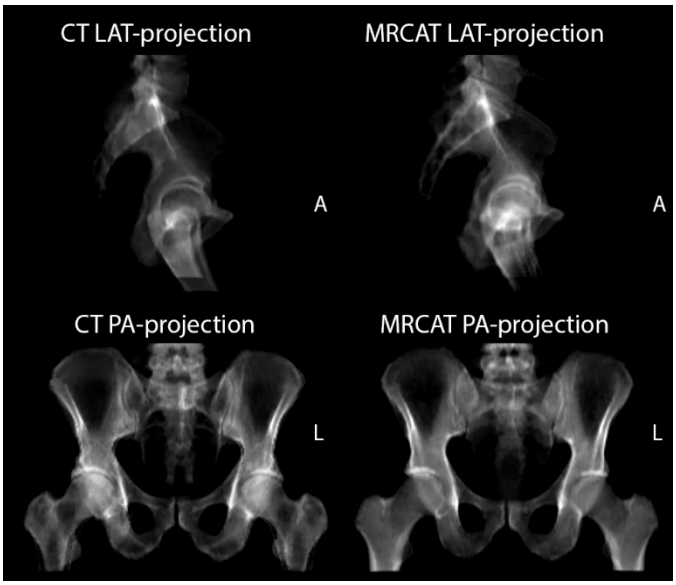

Figure 1: Example of CT (left column) and MRCAT DRR for anterior-posterior (bottom row) and left-right projections calculated in Pinnacle 9.10 treatment planning system (TPS). HU density table is modified to filter out soft tissue contribution.

Evaluation

Five observers (4 radiographers and one physicist) registered the CT- and MRCAT-based digitally reconstructed radiographs (DRRs) to daily localization radiographs (kV-setup images). Registration was fully manual as is the most common clinical workflow at the site. Each image pair was evaluated three times by the same observer to obtain repeated measurements enabling the assessment of measurement quality in terms of inter- and intra-observer variability. In total, 300 (5 observers, 20 patients, 3 repetition) registration were performed per method.

Dedicated software was used for manual registration that was implemented using a commercial software package (MATLAB® 8.4.0.150421 (R2014b), The MathWorks Inc., Natick, MA, 2014). The registration tool recorded the couch shifts for the longitudinal (SI), lateral (LR) and vertical (AP) directions based on the registration of the images by an expert observer. The images were shown to the observers in a random order without revealing whether the DRR originated from an MRCAT or a CT image. The registration software is available as precompiled Matlab program as Supplementary Data and the source code to academic users upon email request from the authors.

To obtain a reference for the manual registrations, a registration between CT and MRCAT must be defined. Therefore, the mDIXON source scan was rigidly registered to the planning CT image and this transformation is then used as the MRCAT to CT reference registration. The registration shall eliminate

minor orientation and translational difference between scans. See Figure 2 for an illustration of the registrations that were performed and evaluated in this study.

Bland Altman plots

Bland Altman plots for repeated measurements were used for visual assessment of method differences. Limits of agreement (LoA) were adjusted for repeated measurements as suggested by Bland and Altman [14]. Measurements from multiple observers were not separated in the calculation of LoAs.

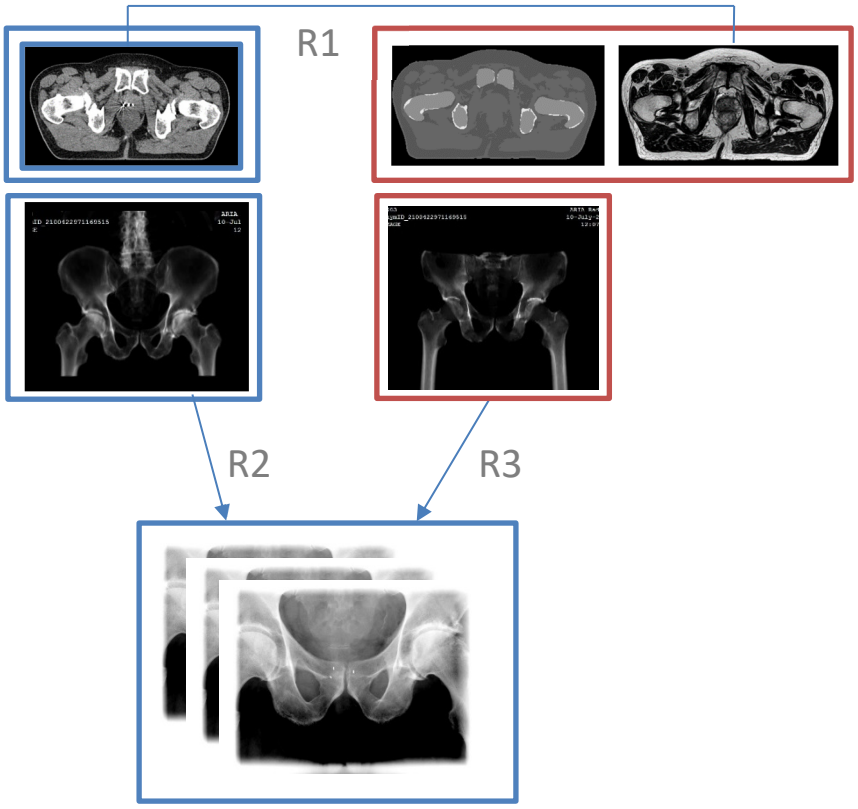

Figure 2: Illustration of registrations used in the study: MRCAT MR-source mDIXON images were rigidly registered to the simulation CT using bones making the reference registration (R1) for the comparison of manual 2D registrations between MRCAT DRRs (R3) and CT DRRs (R2) to the kV radiographs.

Linear Mixed Effects Model

The following mixed-effects model was used for evaluation of systematic difference and comparison of accuracy and reliability of the two methods:

$$y_{ijk} = \mu + a_i + ab_{ij} + \varepsilon_{ijk}$$

- $i \in \{1,2,\dots,20\}$ is the patient index
- $j \in \{1,2,\dots,5\}$ is the observer index
- $k \in \{1,2,\dots,3\}$ is the kth measurement made by an observer on a certain patient
- y_{ijk} is the observed couch shift for patient i, observer j and measurement k
- μ is the fixed effect population mean for the required couch-shifts
- $a_i \sim N(0, \sigma_a)$ is a patient level random effect that denotes the true deviation from μ for the ith patient

- $ab_{ij} \sim N(0, \sigma_{ab})$ is a random effect describing the bias of the jth reader when measuring the couch-shift for the ith patient. σ_{ab} is the inter-observer repeatability, attributed to the standard deviation of the bias terms amongst all observers
- $\varepsilon_{ijk} \sim N(0, \sigma_{\varepsilon})$ is a random error made by the readers when making their kth measurement of the couch-shift for the ith patient. Intra-observer repeatability is identified as the standard deviation parameter σ_{ε}.

Stan statistical computing language with R implementation was used to fit the LME model to the measured data and to estimate the parameters of interest [15]. Full details of the statistical methods used in this manuscript together with R and RStan source code are given in the Appendix. Similar LME model has been used for an IGRT method comparison study by Roy et al. [16].

Assessment of impact to total geometric deviation

Geometric accuracy of image guidance in radiation therapy directly affects required treatment margins and outcomes. Depending on whether a source of error is introduced in the planning phase or randomly during each fraction, the contribution to total uncertainty is either systematic or random [17]. Similarly, introduction of MRCAT reference image in RT can introduce both systematic and random error compared to CT-based workflow (see Table 1). Systematic and random errors were analysed separately since their contribution to total uncertainty is different [17].

Table 1: Grouping the method comparison disagreement to systematic and random components:

Difference component	*Explanation*	*Origin of difference*	*Contribution to MRCAT registration error*
$\Delta\mu$	Systematic difference in the population mean over all registrations of the two methods	Systematic difference between MRCAT and CT bone structures for the cohort of patients or systematic error in MR-to-CT reference registration	Systematic difference between MRCAT and CT bone structures results into **systematic** geometric deviation for the whole population in RT workflow.
Δa $(\sigma_{\Delta a})$	Difference in patient mean between CT and MRCAT based registration. SD of Δa is denoted as $\sigma_{\Delta a}$.	Difference between MRCAT and CT bone structures varying from patient to patient or random error in MR-to-CT reference registration	Random difference between MRCAT and CT bone structures results to **systematic** geometric error in RT workflow for a patient and random error for a population.
$\sigma^2_{ab.CT}$ and $\sigma^2_{ab.MRCAT}$	The inter-observer repeatability (variability) for CT and MRCAT DRR registrations	Difference in the bone representation of CT and MRCAT might cause changes to inter-observer variability reflected by σ_{ab}	Increased inter-observer variability would cause an increase of **random** error in RT workflow.
$\sigma^2_{\varepsilon.CT}$ and $\sigma^2_{\varepsilon.MRCAT}$	The repeatability (intra-observer variability) of registration in each method	Difference in the bone representation of CT and MRCAT might cause changes to inter-observer variability reflected by σ_{ε}	Increased intra-observer variability would cause an increase of **random** error in RT workflow.

To assess the differences in the systematic and random uncertainties between the two methods, the following derived metrics were used:

$$\Delta_{sys} = |\Delta\mu| + \sigma_{\Delta a}$$

$$\Delta\sigma_{rand} = \sqrt{\sigma_{ab.MRCAT}^2 + \sigma_{\varepsilon.MRCAT}^2} - \sqrt{\sigma_{ab.CT}^2 + \sigma_{\varepsilon.CT}^2},$$

where Δ_{sys} is a random variable describing the systematic uncertainty and difference in registrations. It consists of the difference in population mean $\Delta\mu$ and standard deviation of difference in patient registrations, $\Delta a_i = a_i^{CT} - a_i^{MR}$, denoted as $\sigma_{\Delta a}$. Furthermore, $\Delta\sigma_{rand}$ is a random variable describing the difference in random registration error between the two methods consisting of both inter- and intra-observer variability.

To assess that a MR-only solution will improve the total geometric accuracy of RT workflow, the increase in registration uncertainty must be weighed against the reduction of the registration uncertainty present in a dual modality workflow. In RT, uncertainties can be linked directly to the required PTV margins and assessment of method agreement should reflect the impact to total uncertainty of compared workflows.

As proposed by van Herk et. al. [16,17], the sufficient clinical target volume (CTV) to planning target volume (PTV) margin can be expressed as:

$$m_{PTV} = 2.5\Sigma + 0.7\sigma$$

where Σ (here Δ_{sys}) is the systematic and σ ($\Delta\sigma_{rand}$ in this study) is the random spatial uncertainty. We use this weighting of the two error components in the assessment of significance of the difference between CT and MRCAT registrations. Consequently, for MRCAT based workflow to be as good as or better than CT based workflow, the following condition must be met for all three directions:

$$\Pr(m_{PTV}^{MRCAT} - m_{PTV}^{CT} < 0) \geq 0.95$$

That is, the probability that the required PTV margin in MR-only workflow is smaller than in the dual modality workflow is larger than 95 %.

An estimate for registration error of 2 mm from literature was used is our assessment [18]. In addition, we calculate an estimate of registration error for which there is 95% probability that required margin in MR-only workflow is smaller.

Results

Bland Altman plots

The mean difference between the methods [±95% limits of agreement (LoA)] were -0.5 [-3.2 – +2.3] mm, +0.1[-2.4 – +2.6] mm and +0.1[-1.6 – +1.7] for vertical, longitudinal and lateral directions, respectively (see Figure 3). The repeatability coefficients were (CT vs. MRCAT) 2.1 mm vs. 2.6 mm, 1.4 mm vs. 2.1 mm and 1.2 mm vs. 1.4 mm for vertical, longitudinal and lateral directions.

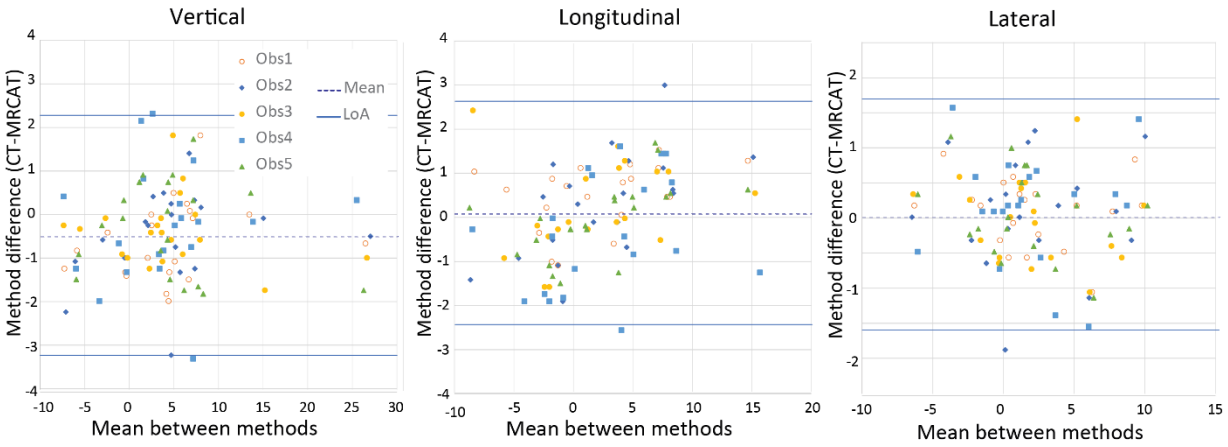

Figure 3: Bland Altman plot for vertical, longitudinal and lateral direction with limits of agreement (LoA) and mean difference. Obs1-5 show the registration differences between methods per observer.

Systematic differences

No difference in population mean $\Delta\mu$ between CT and MRCAT based positioning were observed in lateral or longitudinal directions (see Figure 4, $P(\Delta\mu < 0)$=0.34). However, for vertical direction a statistically significant offset of (mean [95% CI]) -0.49 [-0.85 – -0.13] mm was detected ($P(\Delta\mu < 0)$>0.99). The standard deviation of patient offsets $\sigma_{\Delta a}$ was 0.29 [0.21 – 0.36] mm, 0.50 [0.40 – 0.63] mm and 0.63 [0.48 – 0.76] mm lateral, longitudinal and vertical directions.

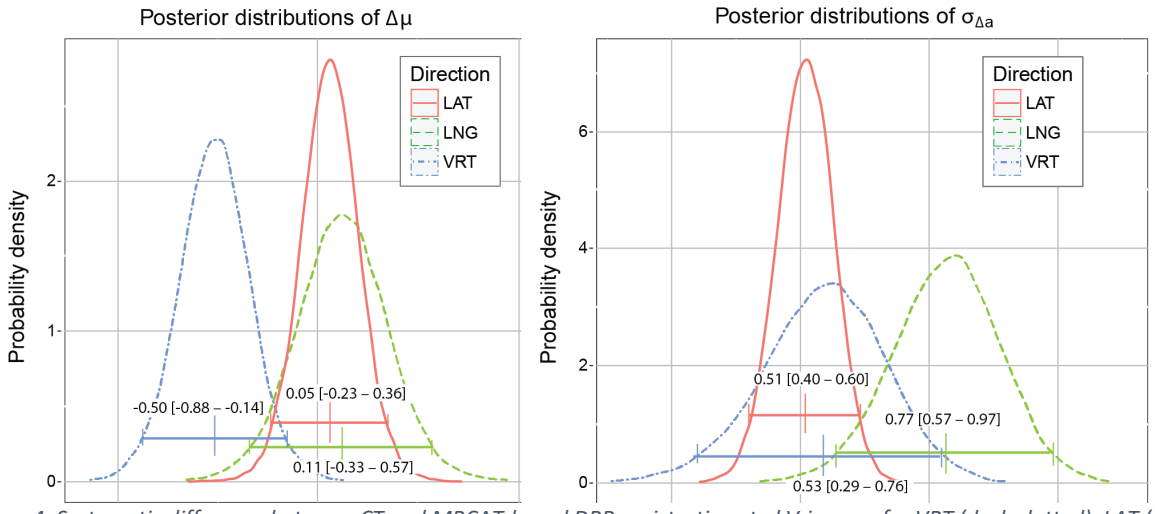

Figure 4: Systematic difference between CT and MRCAT-based DRR registrations to kV-images for VRT (dash-dotted), LAT (solid) and LNG (dashed) directions with mean (vertical bar) and 95% credible interval (horizontal bar). Left: difference of population means $\Delta\mu$. Right: standard deviation of patient offset $\sigma_{\Delta a}$.

Random setup errors

Inter-observer variation was not significantly different between CT and MRCAT based DRR to kV-image registrations in longitudinal ($P(\sigma_{ab,MRCAT} > \sigma_{ab,CT})$ =0.58) or lateral ($P(\sigma_{ab,MRCAT} > \sigma_{ab,CT})$ =0.52) directions. However, the largest difference was observed in vertical direction being statistically significant

($P(\sigma_{ab,MRCAT} > \sigma_{ab,CT}) > 0.99$) whereas smaller deviation was measured in longitudinal direction (see Figure 5). In lateral direction, inter-observer was smallest and almost identical between the two methods.

A statistically significant difference was observed in intra-observer variability between CT and MRCAT based DRR to kV-image registrations in vertical, longitudinal and lateral directions (see Figure 5, $P(\sigma_{\varepsilon,MRCAT} > \sigma_{\varepsilon,CT}) > 0.99$). The largest difference between methods was observed in longitudinal direction and smallest in lateral direction.

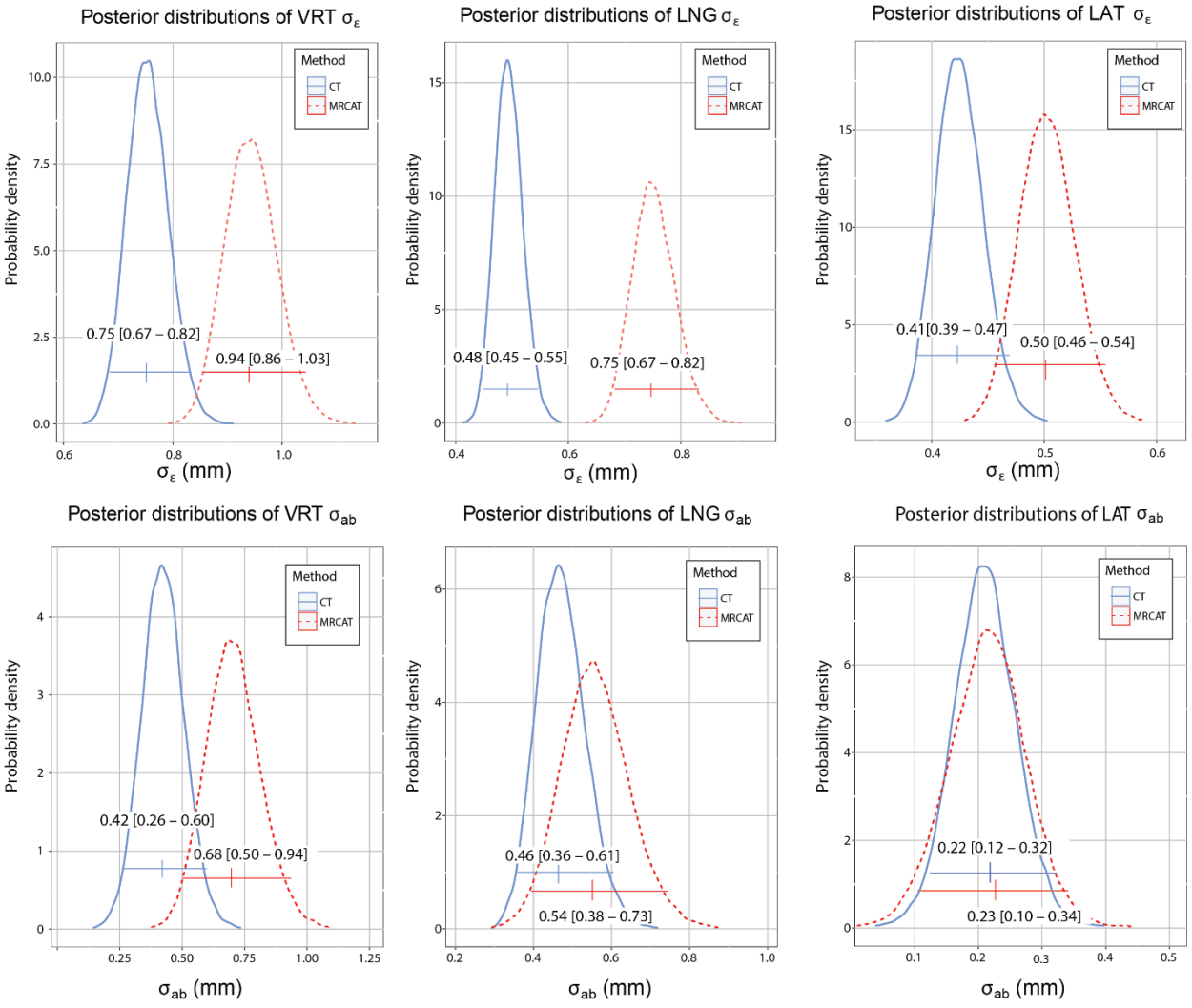

Figure 5: Posterior distribution of intra- (to row) and inter-observer (bottom row) variability in vertical (left), longitudinal (middle) and lateral (right) directions for CT (solid) and MRCAT (dashed) DRR to kV-radiograph registrations. Horizontal bars indicate 95% credible intervals [95% CI] and vertical bars show the maximum a posteriori (MAP) point of the distribution.

Effect on total uncertainty and on required PTV margin

The estimate of the parameter describing the systematic difference between MRCAT and CT based DRR-to-kV registrations, Δ_{sys}, was (Maximum a Posteriori, MAP [95% CI]) 1.04[0.65 – 1.50] mm, 0.93 [0.65 – 1.44] mm, and 0.58 [0.44 – 0.91] mm in vertical, longitudinal and lateral directions (see Figure 6). The

increases in total random uncertainty were estimated to be below 0.5 mm for all directions being (MAP [95% CI]) 0.31 [0.15 – 0.48] mm, 0.24 [0.12 – 0.40] mm and 0.06 [-0.00 – 0.15] mm for vertical, longitudinal and lateral directions.

When registration errors were larger than 1.7 mm, 1.5 mm, and 1.1 mm in vertical, longitudinal and lateral directions, statistically significantly smaller CTV to PTV margin was needed in MR-only workflow.

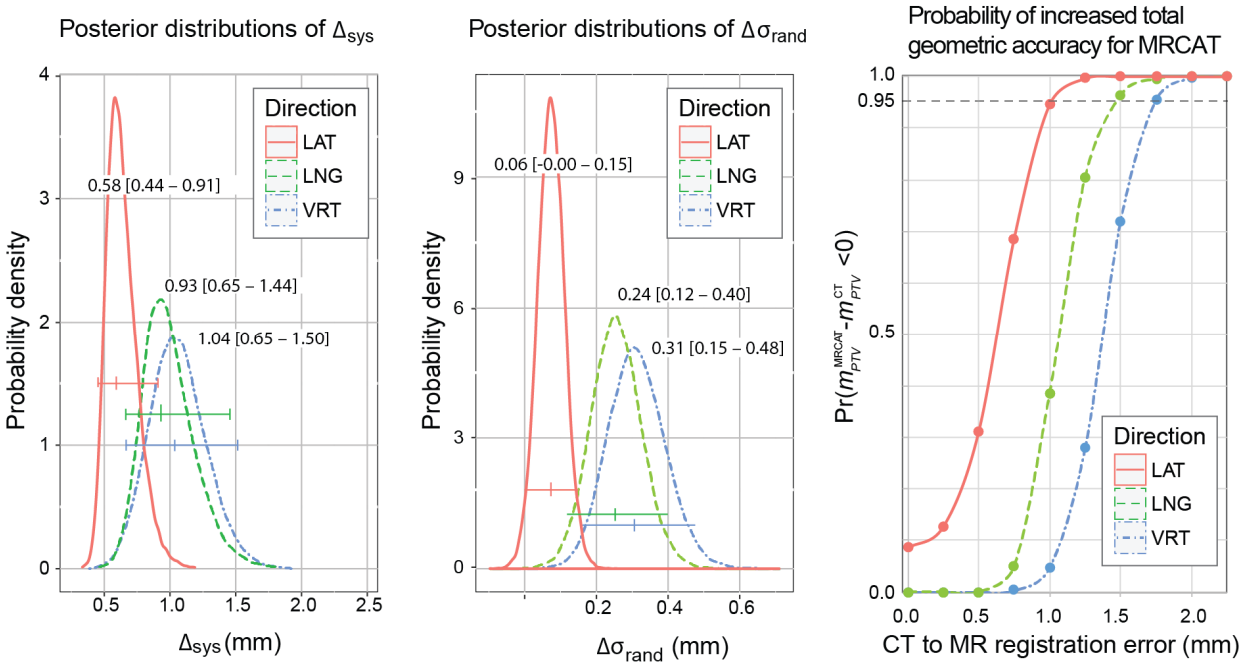

Figure 6: Effect on total systematic (left) and random (middle) geometric uncertainty and probability of reduced required CTV to PTV margin for MRCAT compared to CT workflow (right) for VRT (dash-dotted), LAT (solid) and LNG (dashed) directions with MAP (vertical bar) with 95% credible interval (horizontal bar). On right, above points where probability is larger than 95%, required margin for MRCAT is statistically significantly lower. Circular marks indicate the points where probability was evaluated.

Discussion

New pseudo-CT methods are increasingly introduced in the field [9]. It is essential to note that assessment of the impact of using such methods to total geometric accuracy is as important as dosimetric accuracy. Currently, assessments of dosimetric accuracy outnumbers assessments of impact to other areas of uncertainty in RT. Thus, studies and methods assessing overall impact of pseudo-CTs beyond dose accuracy are needed.

In this study, we measured both accuracy and precision of MRCAT DRR to planar kV-image registration for IGRT of pelvic EBRT. According to the results, average positioning differences were -0.5 mm, 0.1 mm and -0.1 mm in vertical, longitudinal and lateral directions, respectively. These figures are similar to what has been reported for DRR to kV registrations for pseudo-CTs based on manually contoured bone segmentations in a similar measurement setup [11]. Namely, the authors reported average positioning differences (+/- SD) of -0.3 +/- 1.0 mm, 0.2 +/- 0.9 mm and 0.1 +/- 0.5 mm for vertical, longitudinal, and

lateral, directions. Prior MRCAT DRR evaluation by Tyagi studied the quality of DRRs using a non-clinical workflow of automatically registering CT and MRCAT DRRs with an algorithm. The average match for all the patients was 0.3 ± 0.4 mm, 0.03 ± 0.6 and 0.5 ± 0.8 mm in lateral, vertical and longitudinal direction respectively. Unfortunately, their results can't be compared to our manual DRR to kV-image registration study since the figures by Tyagi et al. don't contain the uncertainty resulting from DRR to kV-image registration nor observer variation and, consequently, is not estimate of clinical performance of the method.

Alternative methods for assessment of DRR feasibility have also been reported. Unlike in our study, Kim et al. [7] used a bounding box width difference for assessment of geometric accuracy of their pseudo-CT method giving 0.4 +/- 1.7 mm and -0.6 +/- 1.0 mm AP and RL DRRs, respectively. Furthermore, Chen et al. used a set of eight measurement points defined in both CT and manually segmented pseudo-CT to assess the feasibility of calculated DRRs for IMRT treatment of prostate carcinoma. The maximum difference for a point location between pseudo-CT and CT was 1.3 +/- 1.6 mm when averaged over the study population. However, either Kim or Chen et al. do not report performance metrics measured based on image registrations. In addition, several authors [7], [11] have reported image similarity scores between pseudo-CT and planning CT DRR. Unfortunately, the similarity metrics have a disadvantage of not measuring positioning accuracy or precision achievable with the investigated method. Furthermore, when using similarity scores, it's difficult to assess the impact to total geometric accuracy and acceptance criteria may not be defined.

Measured difference between registrations in CT and MRCAT based workflow include both true difference and measurement noise. The measurement noise stems mainly from two sources. First, registrations between CT and kV-radiographs are affected by uncertainties stemming from the ∼1 mm-pixel sizes and the ∼2 mm-slice thickness in the reference images contributing to errors in calculated DRRs [20]. Second, the residual errors from planning CT to mDixon source image registrations are contributing to the systematic offsets for each patient. In this study patient positioning uncertainties with MRI-based reference images were analyzed by assuming that the registrations between MRCAT and CT can be regarded as an error free reference for correct patient position. Since such baseline assignment is not truly error free and CT based registration is affected by uncertainties, it is likely that the actual registration uncertainty with MRI- based DRR is smaller than reported in this study. Particularly, when considering the significance of the observed small systematic differences one needs to acknowledge that due to chosen study design these differences can result from MRCAT properties or from the registration uncertainties in the MR to CT reference registration (R1). To our knowledge, there is no simple direct method to measure the performance of MRCAT DRRs for patient position verification purposes using human observers. Thus, our method was to compare to the "gold standard" CT-based workflow.

Unlike the systematic differences, inter- and intra-observer variability are not affected by the uncertainties of the reference registration. Thus, they can be used for measuring the random positioning error independently for both MRCAT and CT-based DRRs. More importantly, in any method comparison study, simultaneous estimation of repeatability, through repeated measurements of identical conditions, and agreement are necessary in order to analyze the origin of discrepancies between methods [21]. No prior studies assess intra-observer variation in image guided patient positioning workflow for prostate cancer patients.

Based on obtained registrations, the inter-observer variability was not significantly different in longitudinal or lateral directions but was slightly increased compared to CT in vertical direction. The increase might be due to degraded visibility of the anterior outline of the pubic bone in the MRCAT DRRs. Intra-observer variability was significantly larger for MRCAT DRRs than CT DRRs in vertical and longitudinal directions. However, the total increase in random registration error was below 0.5 mm for all directions and can be considered insignificant in the context of total geometric uncertainties for pelvic RT.

Based on the results, given that there is an uncertainty associated with CT-to-MR registration larger than 1.7 mm, 1.5 mm, and 1.1 mm in vertical, longitudinal and lateral direction, it's likely that use of MR-only has a positive effect to the total geometric accuracy. In the literature uncertainties of 2 mm have been reported for prostate RT [18]. Thus, we conclude that in terms of total geometric accuracy benefits of using MRCAT likely outweigh the small increase in MRCAT DRR-based positioning uncertainty. However, as registration errors were not quantitated in this cohort of patients, further research is needed to confirm our findings.

An important patient positioning workflow for EBRT of prostate cancer is based on implantable fiducials [22]. The workflow is also feasible with MRCAT [19] since most of the fiducials are accurately localized in MR images [22]–[24] and their contours on DRR images can be used for target position verification and the clinical workflow is feasible [19], [23]. However, automatic and robust detection of fiducial markers from MR-images remains a challenge and is a subject of active research [25]–[28].

The study population consisted of only prostate cancer patients. For other indications in the pelvis, including also female patients, we have demonstrated the feasibility of MRCAT generation, geometric accuracy and dose accuracy, in an earlier study [29]. Thus, the results obtained in this study are considered applicable for pelvic anatomy in general. However, further studies are needed to verify the findings and assess robustness of MRCAT for female population.

Currently, contraindications for the use of MRCAT are metal in the imaging volume, such as a metal prosthesis in the hip region, bone anomalies or bone disease in the pelvic area and body diameter in the pelvic area exceeding 50 cm in AP direction. The risk of using contraindicated MRCAT images for RT is mitigated by implementing a safety check in the MRCAT algorithm that prevents the reconstruction of the pseudo-CT in case of atypical bony anatomy or presence of hip implants, for example. Use of MRCAT DRRs for contraindicated patients is beyond the scope of this article and requires further studies.

Use of volumetric positioning image (kV CBCT) is becoming more common due to improved visualization of the daily variation in the anatomy [30]. For prostate, MRCAT could be used also with CBCT-based position verification utilizing implantable fiducial's for registration to daily CBCT [19]. In general, however, soft-tissue based registration between daily CBCT and MRCAT for position verification needs to be investigated in subsequent studies since MRCAT, with only two soft-tissue density values, might not be robustly registered to a CBCT image.

This study advances research and supports future clinical implementation of MRI-only IGRT by presenting a Bayesian linear mixed effects model (LME) method comparison approach for assessing accuracy and precision of patient positioning. Our analysis method is available, uses widely used open source software and could be easily utilized for assessment of other anatomies or methods. Particularly, we used the LME to compare MRCAT pseudo-CT method to a CT-based workflow and demonstrated the feasibility of the method for kV planar imaging IGRT for pelvic anatomy.

Conclusions

The MRCAT pseudo-CT method provides clinically acceptable accuracy and precision for patient positioning for pelvic radiation therapy based on planar DRR images. Use of MRCAT method is associated with a small increase in precision and small systematic difference compared to CT. However, when the slight increase in uncertainty is compared to the uncertainty in CT-to-MR registration in dual modality workflow, MRCAT can be considered more accurate in terms of total accuracy when registration errors are larger than 1.7 mm, 1.5 mm, and 1.1 mm in vertical, longitudinal and lateral directions.

Acknowledgments

The authors thank associate professor Aki Vehtari for comments on Stan model and Bayesian statistics, and Jukka-Pekka Kauppi for suggestions for improving the manuscript text.

The authors would also like to show gratitude to Docrates Cancer Center radiographers for performing the manual registrations required for the study. Especially we would like to thank Aili Aaltonen for her effort on recruitment of the patients.

Disclosure statement

Authors Reko Kemppainen and Teuvo Vaara declare that they were employed by Philips MR Therapy, Finland by the time of conducting the research.

Bibliography

[1] M. Debois *et al.*, "The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer.," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 45, no. 4, pp. 857–865, 1999.

[2] C. Rasch, I. Barillot, and P. Remeijer, "Definition of the prostate in CT and MRI: a multi-observer study," *Int. J. …*, vol. 43, no. 1, pp. 57–66, 1999.

[3] T. Nyholm, M. Nyberg, M. G. Karlsson, and M. Karlsson, "Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments," *Radiat. Oncol.*, vol. 4, no. 1, p. 54, Jan. 2009.

[4] J. Korhonen, M. Kapanen, J. Keyriläinen, T. Seppälä, and M. Tenhunen, "A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer," *Med. Phys.*, vol. 41, no. 1, p. 11704, 2014.

[5] C. Siversson *et al.*, "Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm," *Med. Phys.*, vol. 42, no. 10, pp. 6090–6097, 2015.

[6] J. A. Dowling *et al.*, "An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 83, no. 1, pp. e5-11, May 2012.

[7] J. Kim, C. Glide-Hurst, and A. Doemer, "Implementation of a Novel Algorithm For Generating Synthetic CT Images From Magnetic Resonance Imaging Data Sets for Prostate Cancer Radiation Therapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 91, no. 1, pp. 39–47, 2015.

[8] M. Karlsson, M. G. Karlsson, T. Nyholm, C. Amies, and B. Zackrisson, "Dedicated magnetic resonance imaging in the radiotherapy clinic.," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 74, no. 2, pp. 644–51, Jun. 2009.

[9] J. M. Edmund and T. Nyholm, "A review of substitute CT generation for MRI-only radiation therapy," *Radiat. Oncol.*, vol. 12, no. 1, p. 28, 2017.

[10] L. Chen *et al.*, "Magnetic Resonance-Based Treatment Planning for Prostate Intensity-Modulated Radiotherapy: Creation of Digitally Reconstructed Radiographs," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, no. 3, pp. 903–911, 2007.

[11] J. Korhonen *et al.*, "Feasibility of MRI-based reference images for image-guided radiotherapy of the pelvis with either cone-beam computed tomography or planar localization images," *Acta Oncol. (Madr).*, vol. 54, no. 6, pp. 889–895, Sep. 2015.

[12] M. Köhler, T. Vaara, M. Van Grootel, R. Hoogeveen, R. Kemppainen, and S. Renisch, "White paper: MR-only simulation for radiotherapy planning," *Philips*, pp. 1–16, 2015.

[13] O. Ecabert *et al.*, "Automatic model-based segmentation of the heart in CT images," *IEEE Trans. Med. Imaging*, vol. 27, no. 9, pp. 1189–1202, 2008.

[14] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods Med. Res.*, vol. 8, no. 2, pp. 161–179, 1999.

[15] B. Carpenter *et al.*, "Stan: A probabilistic programming language," *J. Stat. Softw.*, vol. 76, no. 1,

pp. 1–32, 2017.

[16] A. Roy, C. D. Fuller, D. I. Rosenthal, and C. R. Thomas, “Comparison of measurement methods with a mixed effects procedure accounting for replicated evaluations (COM3PARE): method comparison algorithm implementation for head and neck IGRT positional verification.,” *BMC Med. Imaging*, vol. 15, no. 1, p. 35, 2015.

[17] M. Van Herk, “Errors and Margins in Radiotherapy,” *Semin. Radiat. Oncol.*, vol. 14, no. 1, pp. 52–64, 2004.

[18] P. L. Roberson, P. W. McLaughlin, V. Narayana, S. Troyer, G. V. Hixson, and M. L. Kessler, “Use and uncertainties of mutual information for computed tomography/magnetic resonance (CT/MR) registration post permanent implant of the prostate,” *Med. Phys.*, vol. 32, no. 2, pp. 473–482, 2005.

[19] N. Tyagi *et al.*, “Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis,” *Phys. Med. Biol.*, vol. 62, pp. 2961–2975, 2017.

[20] C. W. Hurkmans, P. Remeijer, J. V Lebesque, and B. J. Mijnheer, “Set-up verification using portal imaging; review of current clinical practice.,” *Radiother. Oncol.*, vol. 58, no. 2, pp. 105–20, 2001.

[21] P. S. Myles and J. Cui, “I. Using the Bland Altman method to measure agreement with repeated measures,” *Br. J. Anaesth.*, vol. 99, no. 3, pp. 309–311, 2007.

[22] C. Parker and A. Damyanovich, “Magnetic resonance imaging in the radiation treatment planning of localized prostate cancer using intra-prostatic fiducial markers for computed tomography co-registration,” *Radiother. Oncol.*, vol. 66, pp. 217–224, 2003.

[23] M. Kapanen, J. Collan, A. Beule, T. Seppälä, K. Saarilahti, and M. Tenhunen, “Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate.,” *Magn. Reson. Med.*, vol. 70, no. 1, pp. 127–35, Jul. 2013.

[24] J. H. Jonsson, A. Garpebring, M. G. Karlsson, and T. Nyholm, “Internal fiducial markers and susceptibility effects in MRI-simulation and measurement of spatial accuracy.,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 82, no. 5, pp. 1612–8, Apr. 2012.

[25] C. Gustafsson, J. Korhonen, E. Persson, A. Gunnlaugsson, T. Nyholm, and L. E. Olsson, “Registration free automatic identification of gold fiducial markers in MRI target delineation images for prostate radiotherapy:,” *Med. Phys.*, vol. 44, no. 11, pp. 5563–5574, 2017.

[26] C. Dinis Fernandes *et al.*, “Prostate fiducial marker detection with the use of multi-parametric magnetic resonance imaging,” *Phys. Imaging Radiat. Oncol.*, vol. 1, pp. 14–20, 2017.

[27] S. Ghose *et al.*, “MRI-alone radiation therapy planning for prostate cancer: Automatic fiducial marker detection,” *Med. Phys.*, vol. 43, no. 5, pp. 2218–2228, 2016.

[28] M. Maspero *et al.*, “Evaluation of an automatic MR-based gold fiducial marker localisation method for MR-only prostate radiotherapy,” *Phys. Med. Biol.*, vol. 62, no. 20, pp. 7981–8002, Oct. 2017.

[29] R. Kemppainen, S. Suilamo, T. Tuokkola, P. Lindholm, M. H. Deppe, and J. Keyriläinen, “Magnetic resonance-only simulation and dose calculation in external beam radiation therapy: a feasibility study for pelvic cancers,” *Acta Oncol. (Madr).*, vol. 56, no. 6, pp. 792–798, 2017.

[30] D. J. Moseley *et al.*, “Comparison of localization performance with implanted fiducial markers and cone-beam computed tomography for on-line image-guided radiotherapy of the prostate,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 67, no. 3, pp. 942–953, 2007.