Inter- and intra-observer variation of patient setup shifts derived using the 4D TPUS Clarity system for prostate radiotherapy

E P P Pang¹,², K Knight¹, M Baird¹ and J K L Tuan¹,³

¹ Faculty of Medicine, Nursing and Health Sciences, Department of Medical Imaging & Radiation Sciences, Monash University, Wellington Road, Clayton VIC 3800, Australia
² Division of Radiation Oncology, National Cancer Centre Singapore, 11 Hospital Drive, 169610 Singapore
³ Duke-NUS Graduate Medical School, Singapore, 8 College Road, 169857 Singapore

E-mail: eric.pang.p.p@nccs.com.sg

Keywords: prostate, US-derived setup shifts, inter-observer, intra-observer

Supplementary material for this article is available online

Abstract
This study assessed inter- and intra-observer variation in the derivation of setup positioning using four-dimensional (4D) transperineal ultrasound (TPUS). Seven radiation therapists with 4–16 months of TPUS clinical experience independently derived the setup position on a single fraction for 10 patients undergoing prostate radiotherapy. For each patient, the reference TPUS prostate contour was co-registered to a TPUS image acquired during treatment and inter-observer variation assessed. Intra-observer variation was assessed after two weeks. Inter-observer setup variation from the group median was ≤2 mm in 93.8% of cases (maximum deviation 5.4 mm). Spearman’s rho correlation test resulted in a moderate to high positive correlation for inter-observer variation. The median (range) Spearman’s rho correlation coefficients were 0.675 (0.439–0.939), 0.912 (0.669–1.000) and 0.964 (0.867–1.000) for the x, y and z directions respectively. The median (range) intra-observer variation was ≤2 mm in 93.3% (60%–100%) of cases (maximum deviation 4.9 mm). Wilcoxon matched-pairs signed ranks tests yielded no statistically significant difference for intra-observer TPUS-derived setup shifts (p-value range = 0.151–0.803). The magnitude of observer variation appeared to be influenced by training and/or the length of user experience. 4D TPUS is a promising non-invasive ultrasound-based image-guided radiation therapy solution for daily treatment setup with minimal inter- and intra-observer variation.

Nomenclature
RTs radiation therapists
TPUS transperineal ultrasound
PRV positioning reference volume

1. Introduction

Daily image-guided radiation therapy (IGRT) using either two-dimensional (2D) kilovoltage (kV) planar x-rays with gold seeds, or three-dimensional (3D) cone-beam computed tomography (CBCT), has traditionally been used to verify treatment positioning in prostate radiotherapy (Ye et al 2015). More recently, the use of IGRT strategies without additional imaging dose such as Calypso\textsuperscript{\textregistered} electromagnetic transponders (Varian Medical Systems, PA, USA) and the ultrasound (US)-based Clarity\textsuperscript{\textregistered} system (Elekta AB Stockholm, Sweden) have provided alternative strategies to the pre-treatment positioning and real-time tracking of the prostate during treatment (Kupelian et al 2007, Shelton et al 2011, Mayyas et al 2013, van der Meer et al 2013, O’Shea et al 2014, Ballhausen et al 2015, Ricardi et al 2015). Results have been promising in terms of the setup accuracy using the Calypso system (Willoughby et al 2006, Kupelian et al 2007). However, the Calypso system is an invasive procedure requiring the implantation of three electromagnetic transponders into the prostate as an internal surrogate for treatment setup and intra-fraction tracking (Willoughby et al 2006). Furthermore, although it was assumed that the electromagnetic transponders were fixed intra-prostastically, Tanyi et al
have reported that an average movement of <5 mm with 95% confidence takes place over the duration of the implantation. As well as there are some technical issues associated with this system that may influence the final outcome for the patient. For example, there is a lack of soft tissue information within the target region and limitations caused by the interference of the patient’s abdomen with the electromagnetic plate (Shah et al 2011).

In contrast, the four-dimensional (4D) transperineal ultrasound (TPUS) Clarity system, using an autoscan TPUS probe, is non-invasive and provides soft tissue visualisation of the adjacent normal tissues including the penile bulb, bladder and anterior rectal wall (Lachaine and Falco 2013). In addition, the use of high frequency (1–10 MHz) ultrasound imaging provides high spatial resolution which is useful for image guidance applications in radiotherapy (Xing et al 2006, O’Shea et al 2016). A recent review by O’Shea et al (O’Shea et al 2016) confirmed that medical ultrasound has a very useful role to play both as an imaging and positioning modality in the field of radiation therapy. Ultrasound may also potentially provide an alternative form of bladder volume assessment, replacing CBCT (Ung et al 2014). In spite of the potential that US can contribute to the pre-treatment setup of prostate patients, previous studies have highlighted inter- and intra-observer variation using US as concerning (Robinson et al 2012, Fiandra et al 2014).

In our institution routine practice for the first fraction requires a radiation oncologist to approve the co-registered verification images (i.e. 3DCBCT or 2D kV images) and setup shifts. All subsequent online IGRT image co-registrations and setup corrections are verified by two RTs. Hence, this study aims to quantify the degree of inter- and intra-observer variation in the derivation of the setup positioning error by RTs using the TPUS Clarity system. To our knowledge, this is the first study to report both inter- and intra-observer variation in the derivation of setup shifts using the 4D TPUS Clarity system.

2. Methods

This was a prospective quantitative study involving a total of seven RTs. Ethics approval was obtained from the SingHealth Centralised Institutional Review Board in November 2014 and informed consent was obtained from each subject. The study is registered on the National Institutes of Health clinical trial registry (ID: NCT02408497). The patients ($n = 10$) selected in this study were part of a larger cohort ($n = 60$) of participants in an on-going study evaluating the use of the non-invasive 4D TPUS Clarity system in real-time tracking of the target volume in prostate RapidArc® therapy. All patients (median age = 69.4 years (range: 58–80), Eastern Cooperative Oncology Group performance status = 0, staged between T1c-3b N0 M0, Gleason score = 3 + 4 to 4 + 5, and PSA range (nanogram per decilitre) = 5.8–83 ng dl$^{-1}$) underwent volumetric modulated radiation therapy and were prescribed a dose of 74 Gy in 37 fractions to the prostate gland (+/− seminal vesicles and pelvic lymph nodes) delivered over 7.5 weeks.

2.1. Image registration and assessment workflow

Patients were immobilised using the Clarity system autoscan probe kit (ASPK) with a knee rest and an autoscan TPUS probe (centre frequency of 4.5 MHz). Prior to the commencement of the study, five out of the seven RTs (RT1–RT5) completed a 5 d theoretical and technical training workshop on the operation of the TPUS Clarity system directly from an RT application specialist. This included hands-on practice and daily quality assurance of the TPUS Clarity system. The remaining two RTs (RT6 and RT7) joined in the study team later and received on-the-job training delivered by the senior RT in-charge of the treatment unit. Hence, the distribution of RT operator experience with the use of the TPUS Clarity system ranged from 4 to 16 months at the time of assessment in this study.

2.1.1. Reference TPUS image acquisition and defining the positioning reference volume (PRV)

In this study, a simulation TPUS was not available to be used as the reference image as autoscan TPUS was not integrated in the standard planning workflow due to limited technical resources. Instead the following modified workflow, specifically designed for institutions without simulation TPUS images and recommended to us by the manufacturer, was used. As a surrogate, the TPUS images acquired on the first day of treatment were used as the reference TPUS. In order to verify that the prostate position on the first day of treatment was accurately reproduced for the reference TPUS, the reference TPUS image was first registered with the simulation CT images to confirm the geometrical location of the prostate gland. Some anatomical changes (such as rectum and bladder volumes) between simulation CT and first day of treatment were expected and the following steps were carried out to ensure that the registration accurately reflected the prostate location during the simulation CT. The initial registration with the simulation CT images was automatic based on the isocentre coordinates associated to both the simulation CT and reference TPUS (i.e. the position where the TPUS was acquired on day 1 after isocentre shifts were applied in the treatment room). The day 1 CBCT shifts were then applied to the TPUS and simulation CT image registration. Finally, manual fine tuning of the TPUS images with reference to the bony and soft tissue anatomy (i.e. pubic bone, position of penile bulb, bladder and rectum) were performed. The quality of the reference TPUS was considered optimal when the prostate gland was easily recognisable and the
surrounding structures could be visualised (e.g. penile bulb, pubic bone, bladder wall and anterior rectal wall). The prostate boundaries were generally well visualised on the TPUS images and correlated well, though usually slightly smaller than, the simulation CT prostate contours. However, if substantial anatomical changes (e.g. rectal gas) were observed when compared to the simulation CT images the first fraction TPUS was not used as the reference TPUS and the process was repeated during the next fraction until an acceptable reference TPUS and registration was acquired. Once an acceptable reference TPUS image was acquired the prostate gland was contoured (hereafter to be referred to as the PRV) by an independent RT, and confirmed by a RO, on the reference TPUS to be used for daily registration against the treatment TPUS. The simulation CT contours (i.e. bladder, rectum and seminal vesicles) were ‘switched on’ for a final check on the PRV before approval for use.

2.1.2. Treatment TPUS image acquisition
Treatment TPUS refers to the TPUS images acquired during daily treatment setup, for co-registration with the PRV contoured on day 1. From the second fraction onwards, each time a patient was positioned in the treatment room a TPUS-guided session was initiated once the patient’s treatment positioning was finalised using external skin marks (i.e. tattoos). Each session acquired a set of TPUS positioning images that were recorded purely for study purposes and transferred automatically to a standalone server at the Clarity automatic fusion and contouring (AFC) workstation. As this was a non-interventional study, after the collection of TPUS images, the Clarity guide review application was used using daily 3D CBCT prior to treatment delivery.

2.1.3. Measurement procedure-offline registration of PRV to determine setup shifts
The assessment of TPUS setup shifts was conducted offline using the Clarity guide review application (version 4.0.1.44) on the Clarity AFC workstation. At the AFC workstation, a TPUS Clarity-guided session was selected from each patient for the seven RTs to derive the setup position by registering the PRV to the treatment TPUS images. The session selected for each patient was randomly determined (using the randomisation function on an Excel spreadsheet) to eliminate selection bias. Each of the seven RTs independently derived the setup position using the TPUS Clarity system for one fraction on 10 patients who underwent radiotherapy to the prostate from April to December 2015. Each RT assessed the same fraction for the same patient and all the images were assessed offline. Once the RTs registered the PRV to the selected treatment TPUS images, the Clarity guide review application was used to derive the value of the setup shifts based on the geometrical centroid differences of the registered PRV on the treatment TPUS. The RTs were asked to assess the images independently one after the other and the images were reset once the calculations were completed. A total of 70 datasets were acquired, with each RT contributing 10 datasets. The TPUS-derived setup shifts calculated by each RT were recorded for inter-observer analysis. In order to assess intra-observer variation, a repeated measurement for each RT using the same treatment sessions of each patient was obtained two weeks later. The study design is shown in figure A1 in appendix.

2.2. Statistical analysis
Inter-observer variation of the TPUS-derived shifts was assessed among the RTs using the non-parametric Spearman’s rho correlation coefficient test. A non-parametric Wilcoxon matched-pairs signed ranks test was employed to assess the intra-observer variation. Non-parametric tests were used considering the small sample size in this study and the presence of outliers (Pallant 2016). Data collection and analysis was performed using PASW for windows, version 18.0 (SPSS Inc., Chicago, IL). All statistical tests were conducted under a two-tailed significance level, at a minimum 95% confidence interval.

3. Results
3.1. Inter-observer variation
The deviation of derived setup shifts from the group median (group median was chosen as a representation of the central tendency due to presence of outlier values among the observations) was ≤2 mm in all direction for all patients in 93.8% of the time, with a maximum deviation of 5.4 mm derived by RT7 (patient 4) (figure 1).

Spearman’s rho produced a moderate to high positive correlation in all directions (supplementary tables 1–3 are available online at stacks.iop.org/BPEx/3/025014/mmedia demonstrate shifts in the x, y and z directions respectively). The median (range) Spearman’s rho correlation coefficient is 0.675 (0.439–0.939), 0.912 (0.669–1.000) and 0.964 (0.8674–1.000) for the corresponding x, y and z directional setup shifts. Figures 2–4 illustrate the observed TPUS-derived setup variation of the RTs for each patient for each directional vector. On the boxplots, circles (o) denote mild outliers and asterik (*) denotes extreme outlier; the associated number represents the RT involved. Mild outliers are defined as values that exceeded 1.5 times of the interquartile range. Extreme outliers are defined as values that exceeded 3 times of the interquartile range.

3.2. Intra-observer variation
The median (range) intra-observer variation of derived setup shifts from the individual RTs were ≤2 mm in all direction for all patients in 93.3% (60%–100%) of the time, with a maximum deviation of 4.9 mm derived by RT7 (patient 4) (figure 5).
The results of the Wilcoxon matched-pairs signed ranks test yielded no statistically significant difference for the TPUS-derived setup shifts within each of the individual observers. The \( p\)-value ranged from 0.110 to 0.779 indicating that there was no significant difference observed between the paired TPUS-derived setup shifts (supplementary table 4). The median (range) Spearman rho correlation coefficient of the paired readings the individual RTs is 0.962 (0.726–0.987) (supplementary table 5).

### 3.3 Effect of training and experience on observer variation

To further analyse the outliers observed in figures 2–4, the effect of training and experience on inter-observer variation was considered. The correlation coefficient

---

**Figure 1.** Inter-observer variation of derived setup shifts from group median for all directions.

**Figure 2.** Inter-observer variation of derived setup shifts in the left/right (x-plane) direction. Circles (o) denote mild outliers and the associated number represents the RT involved.
was strongest (>0.9) for the first 5 RTs (RT1–5) who had 9–16 months experience ($p < 0.001$), compared to 0.782–0.889 for the two RTs (RT6 and RT7) who had less experience and completed different training ($p < 0.001$) (supplementary table 6). The effect of training and experience on intra-observer variation was also assessed. Whilst RT1–RT5 demonstrated repeated measures of ≤2 mm in 96.7% of cases (range 93.3%–100%), RT6 and RT7 were within 2 mm on 60% and 80% of cases respectively.

4. Discussion

The accuracy of the TPUS Clarity system has previous been reported by Abramowitz et al (2012) as being less than 1.3 mm. In 2013, Lachaine and Falco also reported accuracy within 1 mm in all directions (Lachaine and Falco 2013). Whilst our study was not able to replicate the same level of accuracy at 1 mm (variation ≤1 mm = 74.8%), the TPUS-derived setup shifts observed between all seven RTs were well
correlated and demonstrate a clinically acceptable level of consistency ($\leq 2$ mm). The highest correlation coefficient among the RTs were observed in the Superior/Inferior (SI: z-plane) followed by the Anterior/Posterior (AP: y-plane) direction. From this it could be inferred that the sagittal image plane of the TPUS images produced superior image quality for visualisation, interpretation and co-registration where the cross-sectional soft tissue information of prostate, penile bulb, bladder and anterior rectal were sufficiently defined. This is consistent with earlier findings by Fiandra et al. (2014) who reported that prostate gland is better visualised in the AP direction on US and less well defined in left/right direction.

Although the range of the TPUS-derived setup shift variation observed among the RTs (figures 2–4) was relatively small, and clinically acceptable ($\leq 2$ mm in 93.8% of cases), it is noted that RT6 and RT7 produced several outliers when compared to the remaining RTs. As such, further analysis showed that they produced larger inter and intra-observer variations. Whilst any statistical analysis on such small sample groups needs to be interpreted with caution, it does indicate that the increased variation observed could be attributed to several factors. One factor is the length of the user experience as both RT6 and RT7 had received a shorter training experience (i.e. 4 months) on the TPUS Clarity system compared to the remaining RTs (range 9–16 months). In 2014, Fiandra et al. reported statistically significant inter-observer variation between two groups of ROs with varying experience (>1 or <1 year) when localising the prostate (Fiandra et al. 2014). The importance of the user experience has also been previously reported by Ricardi et al. (2015) who highlighted an initial learning curve effect on accuracy as a potential issue when implementing US-based IGRT. The learning curve effect for TPUS would affect the accurate interpretation of the anatomical features, probe placement and technical use of the Clarity acquisition software.

A second factor could be the human effect of the training programs received by the RTs. RT1-5 completed training with the same applications expert. RT6 and RT7 were subsequently trained using the same content, but through in-house training. It is possible that not only the length of the user experience affects competency. Perhaps, the training environment, teaching techniques and experience of the trainer may also influence the training outcome. Salminen et al. (2011) highlighted the importance of devising strategies to develop trainers who can in turn produce adequately trained staff. Additionally, Odle and Martino (Odle 2007) have highlighted the importance of standardisation of training methods, exploration of having a ‘super user’ as a trainer, and the need to address potential learning barriers of trainees as paramount for the training to be effective. The causes of this increased observer variation should be a focus of further research, but our results suggest the above factors need to be considered by clinical centres implementing new technology.

However, despite observer variation being greater for the RTs with less experience and different training, the results still demonstrated that TPUS was able to be used to derive set up shifts with clinically acceptable variation ($\leq 2$ mm in 93.8% and $\leq 3$ mm in 95% of cases). These low levels of variation may have been because the design of the setup device (mounted TPUS probe on the ASPK) and the availability of live guidance of probe position (based on the optical markers on the probe) during TPUS scanning at the Clarity acquisition cart were optimal. These features reduce the uncertainties in probe pressure and allow the user to reproduce the scan angle during daily setup. Other factors that may have contributed to high user consistency may be due to the improved TPUS image quality acquired via the perineum due to the shorter scan path length (Lachaine and Falco 2013). This is in contrast to traditional transabdomen US scanning.
which tends to be more susceptible to poor image quality due to multiple factors including high user dependency due to handheld probe and scanning technique, variations in probe pressure, bladder filling, and patient characteristics such as obesity (Artignan et al 2004, Xing et al 2006, Boda-Heggemann et al 2009, Baker and Behrens 2015).

The minimal intra-observer variation observed in this study (93.3% <2 mm) also reassures our institution that uncompromised consistency in the reproduction of the TPUS-derived setup shifts is achievable.

The results of this study are only applicable for intact prostate cases. Further training and an assessment of competency needs to completed prior to TPUS being used clinically for other indications, such as post-prostatectomy cases. Further study is required to validate the TPUS-derived setup shifts against the standard 3D CBCT-derived shifts, and this will be completed as part of our larger study protocol.

5. Conclusion

Inter-observer variation in derived setup shifts was minimal, with 93.8% of the measured differences ≤2 mm. The five RTs with longer user experience and manufacturer delivered training managed to achieve high intra-observer consistency of ≤2 mm in 96.7% of cases (range 93.3%–100%). However, a larger variability in the repeated measurements was observed for the two RTs who had less user experience and in house training. Our results indicate that despite observer variation potentially being affected by human factors, such as training quality and length of user experience, the 4D TPUS Clarity system is a promising non-invasive US-based IGRT solution for daily treatment setup with minimal inter- and intra-observer variation.

Acknowledgments

The authors would like to express our gratitude towards Elekta Pte Ltd for providing the 4D TPUS Clarity equipment for evaluation in this study. We would like to thank Elekta Pte Ltd for the logistics involved in installation and training. We would like to thank the radiation therapists for participating in the assessment. Finally, we would like to thank the radiation physicists (Ms Wendy Chow and Mr Zubin Master) for providing the assistance with the commissioning of the equipment.

Conflict of interest statement

The authors have no conflict of interest related to this study.

Appendix

Figure A1. Illustration of study design.
References


Baker M and Behrens C F 2015 Prostate displacement during transabdominal ultrasound-guided radiotherapy assessed by real-time four-dimensional transperineal monitoring Acta Oncol. 54 1308–14


Ricardi U et al. 2015 Three-dimensional ultrasound-based image-guided hypofractionated radiotherapy for intermediate-risk prostate cancer: results of a consecutive case series Cancer Invest. 33 23–8


van der Meer S et al. 2013 Critical assessment of intramodality 3D ultrasound imaging for prostate IGRT compared to fiducial markers Med. Phys. 40 071707

