

Towards ultrasound-guided adaptive radiotherapy for cervical cancer: evaluation of Elekta's semi-automated uterine segmentation method on 3D ultrasound images

Sarah A. Mason,^{1, a)} Tuathan P. O'Shea,¹ Ingrid M. White,¹ Susan Lalondrelle,¹ Kate Downey,¹ Mariwan Baker,² Claus F. Behrens,² Jeffrey C. Bamber,¹ and Emma J. Harris^{1, b)}

¹⁾ *Joint Department of Physics at the Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Sutton and London, UK*

²⁾ *Department of Oncology, Herlev Hospital, University of Copenhagen, Herlev, Denmark*

(Dated: 12 May 2017)

Purpose: 3D ultrasound (US) images of the uterus may be used to adapt radiotherapy (RT) for cervical cancer patients based on changes in daily anatomy. This requires accurate on-line segmentation of the uterus. The aim of this work was to assess the accuracy of Elekta’s “Assisted Gyne Segmentation” (AGS) algorithm in semi-automatically segmenting the uterus on 3D transabdominal ultrasound images by comparison with manual contours.

Materials & methods: Nine patients receiving RT for cervical cancer were imaged with the 3D Clarity[®] transabdominal probe at RT planning, and 1 to 7 times during treatment. Image quality was rated from unusable (0) - excellent (3). Four experts segmented the uterus (defined as the uterine body and cervix) manually and using AGS on images with a ranking > 0 . Pairwise analysis between manual contours was evaluated to determine interobserver variability. The accuracy of the AGS method was assessed by measuring its agreement with manual contours via pairwise analysis.

Results: 35/44 images acquired (79.5%) received a ranking > 0 . For the manual contour variation, the median [interquartile range (IQR)] distance between centroids (DC) was 5.41 [5.0] mm, the Dice similarity coefficient (DSC) was 0.78 [0.11], the mean surface-to-surface distance (MSSD) was 3.20 [1.8] mm, and the uniform margin of 95% (UM95) was 4.04 [5.8] mm. There was no correlation between image quality and manual contour agreement. AGS failed to give a result in 19.3% of cases. For the remaining cases, the level of agreement between AGS contours and manual contours depended on image quality. There were no significant differences between the AGS segmentations and the manual segmentations on the images that received a quality rating of 3. However, the AGS algorithm had significantly worse agreement with manual contours on images with quality ratings of 1 and 2 compared with the corresponding interobserver manual variation. The overall median [IQR] DC, DSC, MSSD, and UM95 between AGS and manual contours was 5.48 [5.45] mm, 0.77 [0.14], 3.62 [2.7] mm, and 5.19 [8.1] mm, respectively.

Conclusions: The AGS tool was able to represent uterine shape of cervical cancer patients in agreement with manual contouring in cases where the image quality was excellent, but not in cases where image quality was degraded by common artifacts such as shadowing and signal attenuation. The AGS tool should be used with caution

for adaptive RT purposes, as it is not reliable in accurately segmenting the uterus on ‘good’ or ‘poor’ quality images. The interobserver agreement between manual contours of the uterus drawn on 3D US was consistent with results of similar studies performed on CT and MRI images.

^{a)}Electronic mail: sarah.mason@icr.ac.uk

^{b)}Electronic mail: emma.harris@icr.ac.uk

I. INTRODUCTION

Uterine motion reduces the accuracy of external beam radiotherapy (RT) for cervical
50 cancer,^{1,2} with positional changes ranging from 2 to 60 mm between treatments²⁻⁵. To
compensate for this positional uncertainty of the uterus, the planning target volume (PTV)
for the primary tumor site (i.e. excluding nodal disease) is commonly generated by expanding
the clinical target volume (CTV) by 6 - 40 mm⁶. This leads to increased dose to surrounding
normal tissues and incidence of adverse effects (such as both chronic and acute bladder,
55 gastrointestinal, and hematological toxicities) and in addition, may not be sufficient for
adequate uterus coverage in some cases^{2,7-11}.

At present, most verification schedules rely on either megavoltage portal imaging or cone
beam CT (CBCT) imaging of the bony anatomy. These images are commonly reviewed
immediately prior to radiation delivery, and are used to correct for random errors by shifting
60 the couch to align the patient's bony anatomy position during treatment with its position
during planning (i.e. position in the CT simulation [SIM] image)¹². However, a perfect
bone-match does not guarantee correspondence between the soft-tissues; residual uncertainty
regarding the shape and position of the uterus remains.^{1,2} One approach to correct for this
uncertainty uses fiducial markers as a surrogate for soft-tissue imaging. Markers can be
65 inserted into the uterus and imaged with x-ray based modalities, though this is invasive and
not always reliable as the fiducials can migrate^{6,13,14}.

The Clarity[®] ultrasound-guided RT (USGRT) system (Elekta Ltd., Stockholm, Sweden)
has been developed to provide soft-tissue imaging to improve the accuracy of RT for gyne-
cological cancer compared with bony anatomy-based image guidance. Briefly, the Clarity[®]
70 system may be used to acquire ultrasound images in the planning CT room (US-SIM) and
treatment room (US-Tx) frame of reference using an infrared-tracked transducer that is spa-
tially calibrated to the treatment co-ordinate system¹⁵. In the context of cervical cancer RT,
this technology allows the user to localize the uterus on US with respect to the isocenter
of the RT treatment room. This could enable: (i) soft-tissue-based couch shifts, and/or
75 (ii) adaptive RT, where the uterine shape at the time of treatment is explicitly taken into
account. Although soft-tissue based couch shifts resulting from USGRT may improve the
alignment of the uterine centroid with the treatment room isocenter, they do not address
the issue of healthy-tissue sparing because large margins to account for organ deformation

are still required. Adaptive RT is therefore an attractive alternative because the RT beam
80 aperture can be modified according to the shape and position of the target at the time of
RT delivery to ensure adequate target coverage whilst minimizing the organ at risk (OAR)
radiation exposure. Segmentation of the uterus could allow for automated selection of the
plan-of-the-day from a library of pre-defined treatment plans, or for on-line treatment re-
planning according to the patient’s anatomy at each treatment fraction.^{5,16,17}.

85 Manual contouring by an expert can be considered a gold standard for organ segmenta-
tion, though this is too time consuming to be a feasible option for on-line adaptive RT^{18,19}.
On-line segmentation must be achieved on a timescale of minutes so that the additional
time that the patient spends on the treatment couch during segmentation does not result in
patient discomfort and/or movement, a delay in the clinical workflow, or significant natural
90 changes in internal anatomy (such as bladder filling) that would displace the uterus from its
position when it was first imaged. For such applications a rapid method of capturing the
3D uterine outline at treatment time is greatly needed.

One method of localizing regions of interest (ROIs) at treatment is to incorporate *a priori*
knowledge of ROI shape and size, which can be obtained from US-SIM. The Clarity[®] sys-
95 tem implements this approach by requiring a user to manually shift a Reference Positional
Volume ([RPV] - the set of rigid manual ROI contours drawn on the US-SIM image) to best
match the apparent position of the ROI as visualized by US-Tx. This allows for estimation
of the ROI centroid position for soft-tissue based patient setup. However, in the context of
adaptive RT, this approach requires that the ROI undergo little or no deformation through-
100 out the course of treatment so that the RPV is still a valid representation of the patient’s
anatomy at the time of radiotherapy delivery. Since the large amount of deformation oc-
curring in the uterus violates this constraint, rigid registration-based techniques (including
Clarity’s[®] RPV method) for localizing the uterus at the time of treatment are not suitable
for adaptive radiotherapy, as shown in Figure 1²⁸.

105 An alternative to manual contouring is to use a segmentation algorithm to automatically
or semi-automatically (i.e. where user-interaction is required) contour the uterus in 3D in
place of an expert. To our knowledge, Elekta is the first to develop an automated solution
for segmenting the uterus on 3D transabdominal US images via the “Assisted Gyne Seg-
mentation” (AGS) tool²⁰. However, similarly to the RPV method, the AGS tool is currently
110 only used to guide soft-tissue based couch shifts according to the apparent centroid position

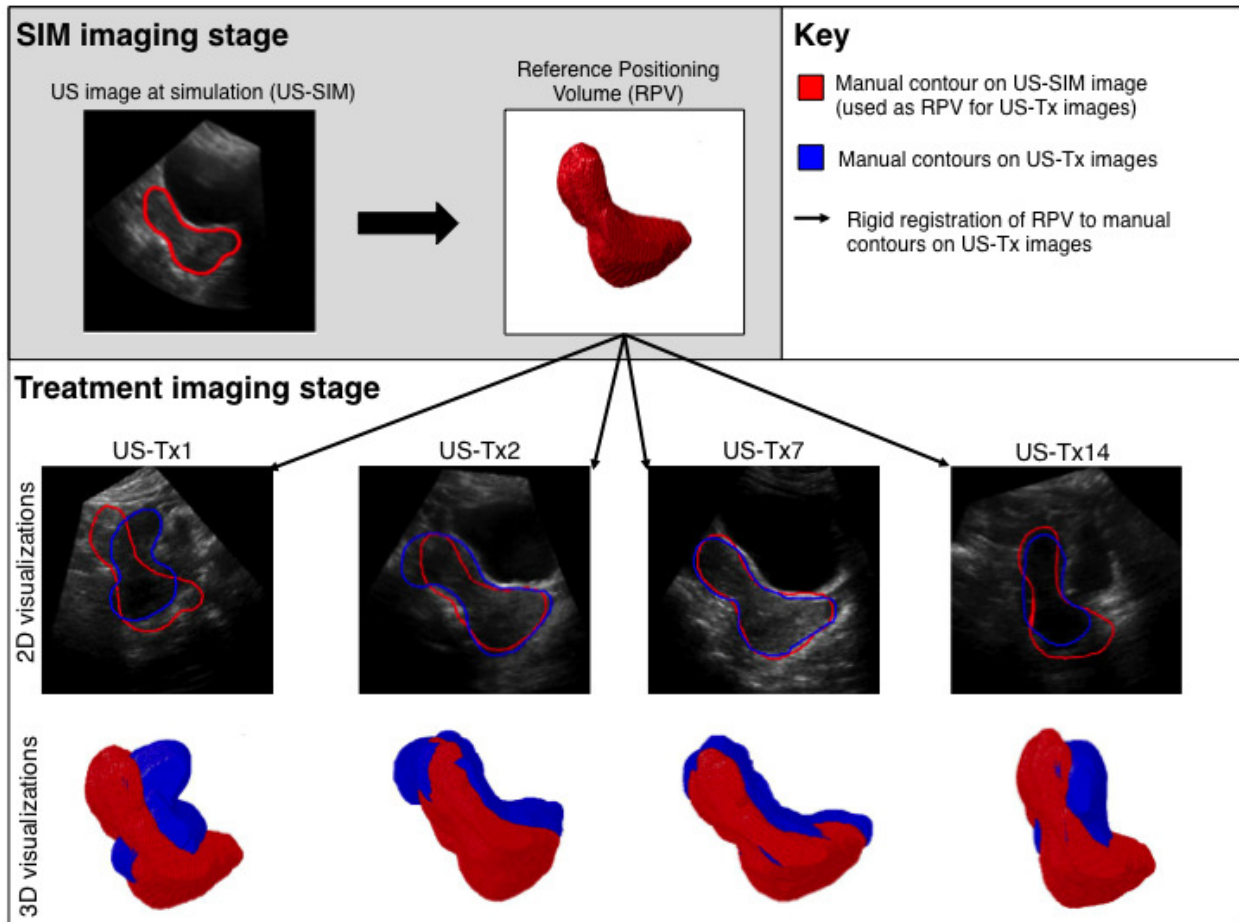


FIG. 1. Example of rigid registration ROI localization technique, where a Reference Positional Volume (RPV) from SIM is used to localize the uterus as visualized by US at treatment (US-Tx). Note that the RPV is often a poor representation of the uterus at the time of treatment (particularly at US-Tx1 and US-Tx14) due to the large amount of uterine deformation.

at treatment.

There may be considerable patient benefit in adaptive RT from employing a method that can automatically, and hence rapidly, segment the 3D uterine shape on 3D US images. However, neither the AGS tool nor any other method for automatically segmenting the
 115 uterus has yet been assessed for its accuracy and hence potential for application in adaptive RT. In this work, the following research questions were addressed:

1. What is the accuracy of the AGS tool in segmenting the uterus on 3D transabdominal US images? This was quantified by pairwise comparison with corresponding manual contours, which led to the secondary research question:

- 120 2. What is the interobserver variability in contouring the uterus on 3D transabdominal US images? This variability was used as a reference for the ideal accuracy of a semi-automated segmentation method.
3. What is the effect of image quality on both (1) AGS tool accuracy and (2) interobserver contour variation.

125 All analyses were performed on 3D transabdominal US images acquired from nine cervical cancer patients.

II. MATERIALS & METHODS

A. Data Acquisition

Nine patients receiving radiotherapy for cervical cancer were included in this study: six
130 from Herlev Hospital, Copenhagen, Denmark (23 US images acquired) and three from the Royal Marsden NHS Foundation Trust, London, UK (21 US images acquired). Ethical approval for these studies was obtained from the ‘De Videnskabsetiske Komiteer’ and the ‘NHS Research Ethics Committees (reference: 15/LO/1438)’ respectively. Median patient age was 49.5 years (range 36 - 65 years), median body mass index (BMI) was 27.6 (range 21.5
135 - 40.7), and median FIGO cervical cancer stage was IIB (range IIB - IIIB). The six patients from Herlev were instructed not to pass urine approximately one hour prior to RT treatment. The three patients from the Royal Marsden Hospital were asked to drink 200mL of liquid and to refrain from passing urine in the hour prior to treatment. After being positioned on the couch, 3D transabdominal US images of the uterus were acquired for each patient at 2 to
140 8 times (once at US-SIM and 1-7 times at US-Tx) during the course of treatment. All scans were acquired with the Clarity[®] USGRT system (Clarity[®] Model 310C00, Elekta, Montreal, Canada), using a 3D mechanically swept convex 5 MHz transducer (m4DC7-3/40), with the pressure between the US transducer and the patient’s skin as low as possible to minimize soft-tissue displacement.

145 B. Segmentation

Manual Segmentation: Four experts (two clinical oncologists (IMW and SL), one radiologist (KD), and one researcher trained by an oncologist (SAM)) manually contoured the uterus in the sagittal plane on a RayStation 5.0 workstation (RaySearch Laboratories, Stockholm, Sweden) for all US-SIM and US-Tx images analyzed. In this study, the ‘uterus’
150 is referred to as a single structure containing both the uterine body and cervix.

AGS segmentation: The core of the AGS tool is a discrete dynamic contouring (DDC) algorithm, which is a gradient-based segmentation technique commonly used in prostate segmentation applications²². Elekta have adapted the methods employed by Ladak et al.¹⁸, Hu et al.²³, and Ghanei et al.²⁴, such that the algorithm semi-automatically segmented the
155 uterus on US. The same four experts who performed the manual uterine segmentations used the AGS tool to segment the uterus on all US image volumes. This required an initialization step where four hint points were placed on uterine features (the uterine fundus, both isthmus points, and base of the cervix) on a central sagittal slice (Figure 2).

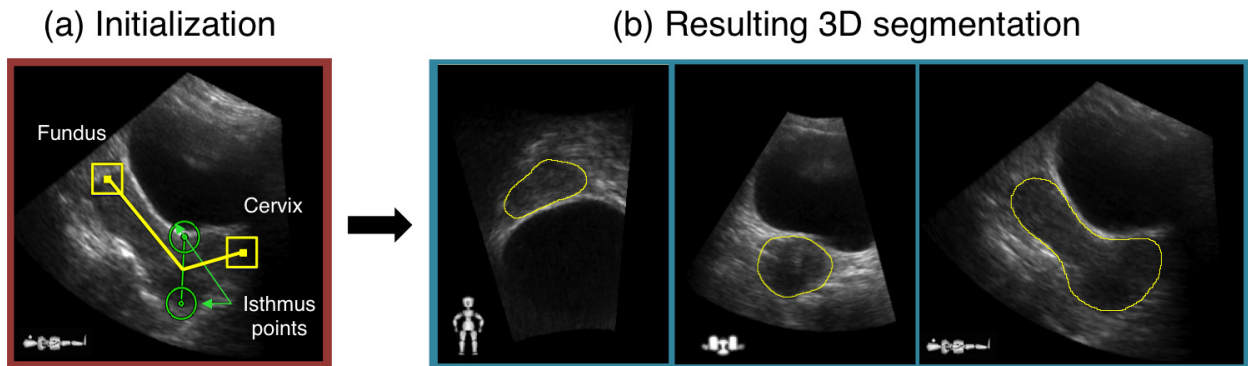


FIG. 2. (a) Snapshot of AGS tool user interface and (b) resulting segmentation, where the placement of 4 anatomical hint points is required to initialize AGS segmentation. The hint points are the uterine fundus and base of the cervix (yellow squares), and the anterior and posterior isthmus points (green circles). The resulting segmentation is shown in yellow on three slices from the coronal, transversal, and sagittal planes, respectively

TABLE I. Scale used to rate image quality of transabdominal US uterine scans

Rating	Criteria
0 - Unusable	Impossible to identify any structures in the US image
1 - Poor	Uterine boundaries blurred in the majority of planes and slices. Imaging artefacts severe and/or prevalent
2 - Good	Uterine boundaries may be partially obscured or difficult to discern due to moderate imaging artifacts, but still visible in the majority of slices and anatomical planes
3 - Excellent	Clearly defined uterine boundaries in all three anatomical planes. Subtle or no imaging artifacts present

C. Data Analysis

160 1. *Image Quality rating*

Each 3D US image was rated twice on a 4-point scale according to the criteria listed in Table I by one observer (SAM), with at least 10 days in between ratings of the same image. Any image receiving a rating of 0 at least once was excluded from further analysis. The final rating for the remaining images was calculated as the mean rating for each image, rounded
165 to the nearest integer.

2. *Contour Agreement*

Interobserver manual contouring variation was assessed by measuring the pairwise agreement between the four manual contours drawn on each US image; i.e each observer contour
170 was compared with the other 3 observers' contours giving 12 pairwise comparisons per image. The accuracy of the AGS tool was quantified by measuring its agreement with manual contours via pairwise analysis; blue i.e. each AGS contour was compared with each manual contour, giving 16 pairwise comparisons per image. In some instances, the AGS algorithm did not produce a contour at all; these cases were referred to as failed segmentation attempts,
175 and were excluded from the quantitative analyses. The AGS segmentation attempts that failed were reported as a percentage of all AGS segmentations attempted. In all cases,

‘contour agreement’ was assessed using the following 4 metrics, where **A** and **B** represent hypothetical 3D contours:

1. The Euclidian distance between the centroids (DC) of **A** and **B**. The centroid of the uterus (a point identified by its x, y, and z coordinates in the treatment room frame of reference) is currently used in the Clarity[®] workflow to suggest soft-tissue-based couch shifts; discrepancies between **A** and **B** were considered to be setup errors in the patient position. A perfect DC was defined as 0 mm.
- 180 2. The 3D Dice similarity coefficient (DSC), defined as $(2|\mathbf{A} \cap \mathbf{B}|)/(|\mathbf{A}| + |\mathbf{B}|)$, where a DSC of 0 and a DSC of 1 indicate zero and perfect overlap respectively. Good agreement (across a range of anatomical sites and imaging modalities) was considered to be $\sim > 0.75^{25-27}$.
- 190 3. The mean surface-to-surface distance (MSSD) was defined as the mean of the Euclidean distances between every vertex on the surface of **A** and its nearest neighboring vertex on the surface of **B**. Like the DSC, the MSSD is a measure of segmentation accuracy, though it is more sensitive to strong local deviations in shape. A perfect MSSD was defined as 0 mm, and good contour agreement (across a range of anatomical sites and imaging modalities) was considered to have an MSSD of ~ 3 mm or less²⁸⁻³¹.
- 195 4. The Uniform Margin of 95% (UM95)²⁸ was defined as the margin required (in mm) to uniformly expand **A** to create **A'**, such that at least 95% of the volume of **B** was included in the volume of **A'**. The UM95 was used to indicate the contribution of localization accuracy to the overall treatment margin required in RT.

3. *Statistical analyses*

Interobserver manual contour agreement: A Wilcoxon rank sum test with Bonferroni correction was used to test for differences in DC, DSC, MSSD and UM95 between manual contours in each image quality rating group (1,2, and 3) to see whether agreement between observers increased with improving image quality.

AGS segmentation accuracy: A Wilcoxon rank sum test was used to test for differences in DC, DSC, MSSD and UM95 between AGS and manual contours for all images, and

205 when the images were grouped according to image quality (ratings 1, 2, and 3). The interobserver manual contour agreement was used as a benchmark to gauge the performance of semi-automatic segmentation methods; ideally, the agreement between an algorithmically-derived contour and a manually-derived contour should be the same as the variation in agreement between manual contours. To investigate whether better image quality improved 210 AGS segmentation performance, a Wilcoxon rank sum test with Bonferroni correction was used to test for differences in DC, DSC, MSSD, and UM95 within each group (image quality ratings of 1, 2, and 3).

III. RESULTS

Image Quality Rating: 35 of the 44 US images acquired had an image quality rating of 215 1 or higher, and were included in subsequent quantitative analyses: 6/35, 18/35, and 11/35 US images received ratings of 1, 2, and 3, respectively.

Interobserver manual contour agreement: The median [interquartile range (IQR)] DC, DSC, MSSD, and UM95 results for the interobserver manual contouring variation are given in Table II, and Figure 3. The overall medians [IQR] for the DC, DSC, MSSD and UM95 220 were 5.41 [5.0] mm, 0.78 [0.11], 3.20 [1.8] mm, and 4.04 [5.8] mm respectively. Images with a quality rating of 2 had a significantly lower ($p < 0.05$) DC, DSC, and MSSD than images with a quality ratings of 1 or 3 in every metric but UM95 (Table II). There was no statistical difference between images with a ranking of 1 and 3 in any of the agreement metrics considered.

225 *AGS contours acquired:* Out of 140 attempts at using the AGS tool to segment the uterus (35 US images * 4 observers), 113 AGS contours were successfully obtained (80.7%), whereas the algorithm failed to return a result in 27 cases (19.3%). The 27 cases with no result were excluded from the quantitative analysis.

AGS segmentation accuracy: The median [IQR] DC, DSC, MSSD, and UM95 results 230 for the AGS segmentation accuracy are given in Table II. The AGS segmentations had a significantly better accuracy (i.e agreement with manual contours) on images with a rating of 3 than images rated 1 or 2. However, there was no difference in segmentation performance between Rating 1 images and Rating 2 images. The AGS algorithm agreed with manual contours on images that received a rating of 3, as there was no significant difference between

TABLE II. *Col 3*: Agreement (median [IQR]) between manual contours from different observers and *Col 4*: accuracy of the AGS tool, measured by pairwise analysis with manual contours. Symbols indicate statistical differences ($p < 0.05$) between image ratings within a particular group: γ - statistically different to Rating 1, δ - statistically different to Rating 2, and ϕ - statistically different to Rating 3. Note that the AGS tool accuracy was significantly better in Rating 3 images than Rating 2 images (all cases) and rating 1 images (DSC, MSSD, UM95). (*Abbreviations*: DC = distance between centroids, DSC = Dice similarity coefficient, MSSD = mean surface-to-surface distance and UM95 = uniform margin of 95%.)

	Image Quality	Inter-observer variability (manual contours)	AGS tool accuracy
DC (mm)	All Images	5.41 [5.0]	5.48 [5.45]
	Rating 1	4.33 [2.9]	5.46 [2.8]
	Rating 2	6.03 [6.7] γ, ϕ	6.29 [7.0]
	Rating 3	5.26 [4.3]	4.64 [5.12] δ
DSC	All Images	0.78 [0.11]	0.77 [0.14]
	Rating 1	0.79 [0.08]	0.74 [0.11]
	Rating 2	0.76 [0.13] γ, ϕ	0.71 [0.16]
	Rating 3	0.81 [0.09]	0.81 [0.09] γ, δ
MSSD (mm)	All Images	3.20 [1.8]	3.62 [2.7]
	Rating 1	3.13 [0.08]	4.47 [2.1]
	Rating 2	3.33 [1.0] ϕ	4.38 [2.8]
	Rating 3	3.06 [1.5]	2.58 [1.4] γ, δ
UM95 (mm)	All Images	4.40 [5.8]	5.19 [8.1]
	Rating 1	4.33 [4.0]	8.08 [8.5]
	Rating 2	4.62 [6.9]	6.35 [8.7]
	Rating 3	4.01 [5.8]	2.89 [5.2] γ, δ

235 them in all metrics considered (Figure 3). However, the AGS algorithm was less accurate in segmenting the uterus on Rating 1 images according to all metrics considered, and also less accurate on Rating 2 images according to DSC, MSSD, and UM95. Overall, the AGS

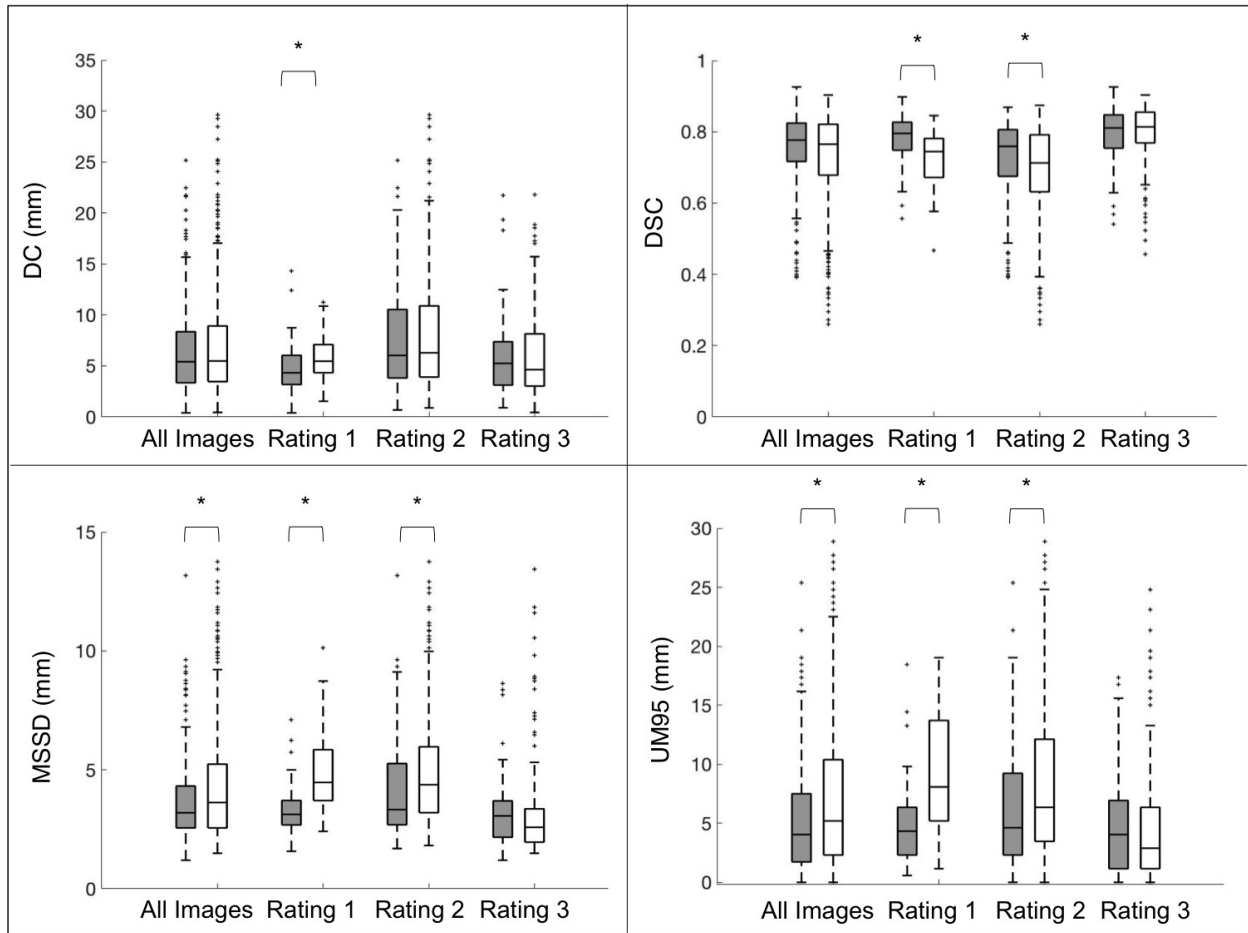


FIG. 3. Boxplot showing interobserver variability between manual contours (shaded boxes) and the accuracy of the AGS algorithm as measured by agreement with manual contours (white boxes). The asterisks denote statistical differences between manual and AGS segmentations ($p < 0.05$). Note that there were no significant differences between the AGS and manual segmentations in images with a quality rating of 3 (excellent) on any metric considered. Also note the that the AGS segmentations were significantly different from manual contours on rating 1 (poor) quality images for every metric considered. *Abbreviations:* DC = distance between centroids, DSC = Dice similarity coefficient, MSSD = mean surface-to-surface distance, and UM95 = uniform margin of 95%

algorithm was statistically equivalent to manual contouring in terms of DC and DSC, but not in terms of MSSD and UM95.

Image quality rating: Low bladder volume and high BMI can increase the attenuation of US and reduce image quality^{32,33}. Not only does a full bladder help with tissue sparing in RT treatment for cervical cancer, but it has the added benefit of providing an acoustic “window” to the uterus, as urine has a low US attenuation coefficient compared with surrounding tissues. Patients with a high BMI are likely to have a greater amount of adipose tissue through which the US must travel, which may be important because fat has a relatively low speed of sound and its presence can cause image aberrations due to acoustic refraction, wave aberration, reverberations, steering errors, focusing errors and spatially dependent image scale mis-calibration. These factors may explain why eight of the nine unusable images (i.e. received an image rating of ‘0’) were acquired from patients who did not follow a stringent drinking protocol (the Herlev cohort), and why four of the nine unusable images were obtained from the same patient, who had the highest BMI (36.5) of the patients included in this study. Additionally, care was taken to apply low pressure to the abdomen when acquiring the US images to avoid internal soft-tissue displacement; though this is crucial for RT applications, this comes at the cost of poorer image quality as contact between the transducer and the skin surface is decreased^{34,35}. A larger study is needed to investigate methods of overcoming these challenges associated with implementing US guidance in adaptive RT to reduce the risk of obtaining an unusable image. One potential solution could be to ensure an adequate level of bladder filling at the time of treatment by enforcing a stringent drinking protocol, or by finding ways to compensate for variables such as poor hydration over the previous twenty four hours prior to treatment or reduced bladder capacity often occurring during treatment. Another solution could be establishing inclusion/exclusion criteria to identify good candidates for transabdominal US scanning. However, it should be noted that even without such measures in place, approximately 80% of the US images acquired in this study were used to successfully identify the position and shape of the uterus at the time of RT treatment.

Interobserver manual contour agreement: The DC, DSC, and MSSD values reported here (medians of 5.4 mm, 0.78, and 3.20 mm respectively) are consistent with those reported in similar studies, though a direct comparison was not possible due to differences in: imaging modalities used, the disease status of the cohort investigated, the anatomical site contoured,

and the number of observers. Baker et al. reported a median DC of 6.0 mm between contours of two observers in manually delineating the uterus on 3D US on a healthy volunteer cohort³⁶. In the literature, reported values of the DSC between manual contours drawn on CT and MRI images for a variety of anatomical sites ranged from $\sim 0.7 - 0.98$ ^{25,26,37,38} with $\sim 0.7-$
275 0.8 generally considered acceptable²⁵⁻²⁷. The MSSD between manual contours drawn on US, CT, and MRI images reported in the literature for a variety of anatomical sites ranged from $\sim 1 - 5$ mm^{26,31,39,40}. The fact that the UM95 required to overcome interobserver contouring variability in this study (median [IQR] of 4.04 [5.8] mm) was much smaller than the interfractional uterine motion commonly observed, (which can be as much as 60 mm)
280 supports the idea that USGRT could reduce the size of the margins needed to compensate for organ motion, even in the presence of contouring uncertainties⁵.

As shown in Figure 4, common areas of disagreement between manual contours observed in this study arose from determining the left-right extent of the uterus, and distinguishing the base of the cervix from the top of the vagina. This may be attributed to problems
285 associated with contouring in the sagittal plane. The agreement between manual contours did not correlate with improving image quality, despite the uterine boundary becoming sharper in higher quality images. This may be due to the experts' abilities to infer the boundary of the uterus in places where it was obscured using prior knowledge of uterine shape and/or relative orientation of other anatomical landmarks in the US field of view. Even
290 in the presence of these sources of disagreement, the manual contour agreement reported here is comparable with previous contouring variability studies, indicating that the uterus can be visualized with 3D transabdominal US at the time of RT treatment. Furthermore, USGRT could be dosimetrically beneficial to cervical cancer patients as the component of the margin needed to compensate for contouring variability (represented by the UM95) is
295 still much smaller than the margin that is needed to compensate for uterine motion without any form of soft-tissue guidance.

AGS tool performance: When applied to images acquired from cervical cancer patients at RT treatment, the AGS tool failed to return a result in nearly 20% of segmentation attempts, which is unacceptable for use in adaptive RT considering that an ideal segmentation method
300 should produce a result in 100% of segmentation attempts. This occurred in cases where the image quality rating was 2 or lower, indicating that a clearly defined boundary in all three anatomical planes is required to ensure that the AGS tool functions. Potential solutions

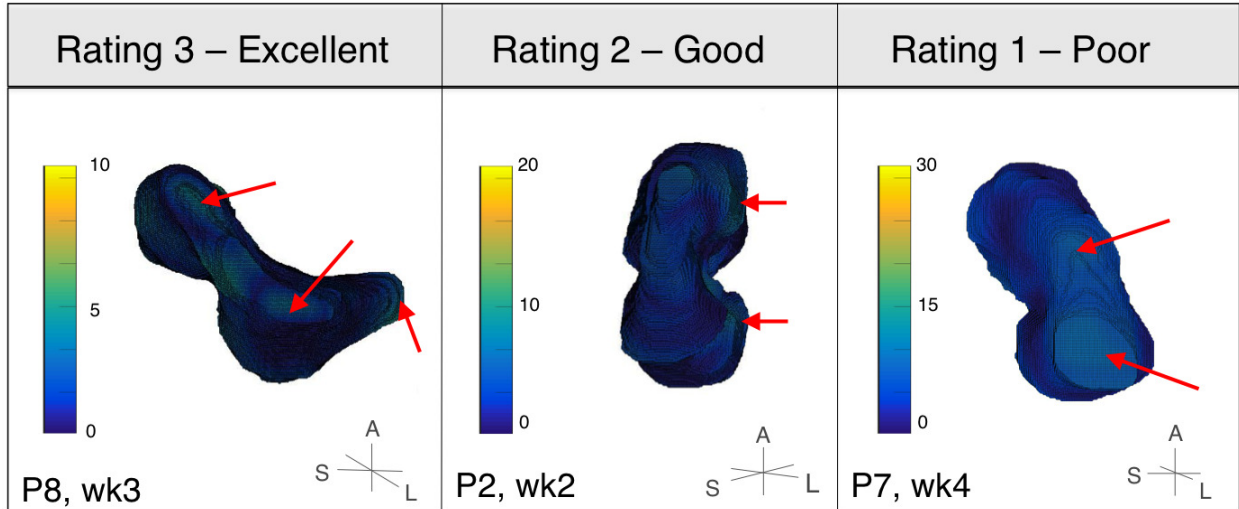


FIG. 4. Heatmaps showing interobserver manual contour variability displayed on the uterine isosurface. Dark blue indicates 0mm deviation, and yellow indicates > 10 mm deviation. The scale of the heatmap is different for each image. Note that the largest deviations tend to be in the left-right uterine edges and the base of the cervix (red arrows). The orientation of the uterus is given by axes in bottom right corner (A = anterior, S = superior, and L = left) Labels in bottom left corner of each image indicate the patient number (P2, P7, or P8) and time point where image was obtained (week = wk). For corresponding US image, see row 1 of Figure 5.

for improving the image quality such that the probability of AGS returning a result is increased may include introducing a selection criteria at baseline to identify patients who have characteristics conducive to obtaining excellent US images (e.g. low BMI), or applying
 305 US image processing/acquisition techniques such as speckle reduction or image compounding to improve the contrast to noise ratio between the uterus and background tissues^{42–44}.

In the 80% of cases where a result was returned, the values of DC, DSC, and MSSD between AGS and manual contours were dependent on image quality. The agreement between
 310 the AGS algorithm and manual contours was statistically equivalent to the interobserver agreement between manual contours for images with a rating of 3; this indicates that the AGS algorithm can accurately segment the uterus on US images containing virtually no imaging artifacts/imperfections. This is shown in column 1 of Figure 5, where the AGS (red) segmentations agree well with the manual (green) segmentations in on the US images with distinct, continuous uterine boundaries. Note that in these cases, the patients all had
 315

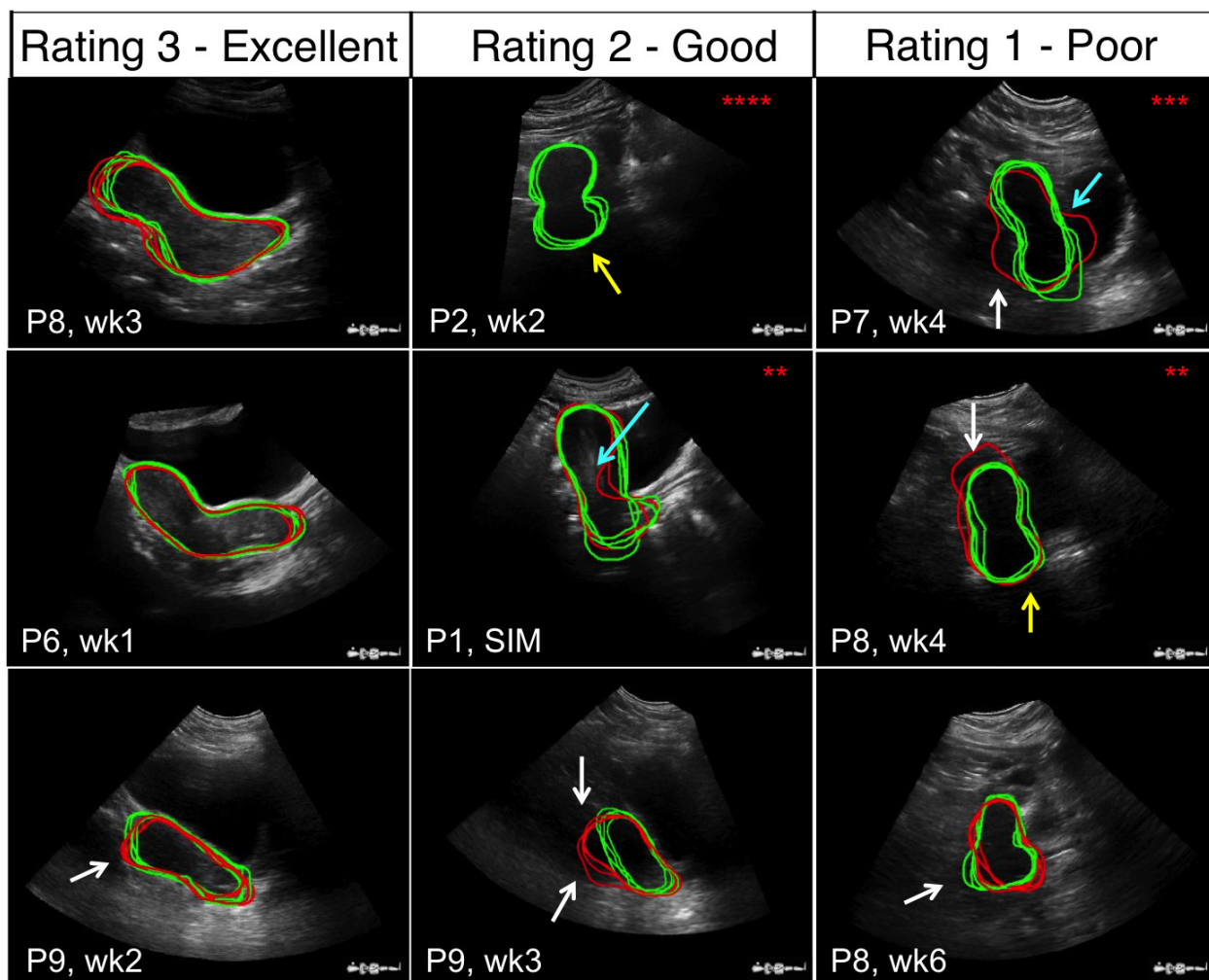


FIG. 5. Examples of manual (green) and AGS (red) contours superimposed over a central slice of corresponding 3D US images. Images are grouped by column according to image quality rating. Labels in bottom left corner indicate the patient number (P1-P9) and time point where image was obtained (week = wk and SIM = ultrasound acquired at CT simulation). Arrows indicate artifacts in the US image that have led to errors in the AGS algorithm. White arrows = shadowing, yellow arrows = signal attenuation, and cyan arrows = misinterpretation of other anatomical boundaries as the uterine boundary (e.g. the endometrium in P1 SIM and the bladder in P7 wk4). Red asterisks indicate number of times where the AGS algorithm failed to give a result for the corresponding image.

full bladders extending across the length of the uterus. However, the majority of the US images acquired in this study had some form of image artifact partially obscuring the true

uterine boundary (image quality ratings 1 and 2). In these cases, the AGS algorithm performance was significantly poorer than its manual segmentation counterpart on all metrics considered (with the exception of the DC on rating 2 images), which may be attributed to the fact gradient-based algorithms are susceptible to errors due to the speckle, shadowing, and signal variation with ultrasound beam angle commonly present in US images taken of cervical cancer patients during RT treatment¹⁸, as shown in columns 2 and 3 of Figure 5. In these examples, the image artifacts either caused the AGS contour to deviate from the true uterine boundary (as defined by the manual segmentations), or confounded the US image to the extent that the resulting shape of the uterus defined by the AGS tool was either corrupted, or unobtainable, despite good agreement between the corresponding manual contours. Furthermore, the statistical analyses performed to check for differences in AGS algorithm accuracy between image rating groups showed that AGS segmentations on images with a rating of 3 were significantly better than AGS segmentations on images with ratings 1 or 2.

When comparing the overall performance of the AGS algorithm with the interobserver manual contours, there were significant differences in MSSD and UM95, but no significant differences in DC or DSC. Note that (1) DC doesn't take shape into account and (2) the DSC is only sensitive to changes in shape if that shape is accompanied by changes in the volume of overlap; for example, thin extrusions of the contour produced by the AGS algorithm in the presence of shadowing or speckle had little effect on the DSC, (3) the MSSD is a direct measure of contour surfaces, and therefore much more sensitive to local deviations in shape, and (4) the UM95 represents the volume expansion needed to account for contouring errors. Taking this into account, the statistical results were interpreted to mean that even though the AGS tool may be sufficient in terms of centroid position and volume, it's overall shape was often incorrect. This is of great concern when considering adaptive RT, which aims to modify the beam aperture such that it conforms to the boundary of the target. Furthermore, this difference in shape manifested itself in an increase in the UM95, suggesting that AGS segmentation errors would likely have a dosimetric effect.

Future Work: This work highlights that there remains a need for a segmentation technique that is capable of conforming to the uterine boundaries at the time of treatment to accurately represent the position and shape of the RT target. Although the AGS tool is capable of achieving this in US images with excellent image quality, it is inaccurate and

350 unreliable in images where the uterine boundary is blurred or partially obstructed. To overcome some of the pitfalls of the AGS tool, a new algorithm is being developed that is less dependent on image gradient to semi-automatically segment the uterus; one potential solution includes incorporating shape models into a gradient-based segmentation framework to overcome errors associated with US shadowing^{29,45}. Additional work will investigate meth-
355 ods of improving US image quality, image processing techniques to further distinguish the uterus from surrounding tissues, quantitative methods of directly comparing other imaging modalities (such as MRI, CT, and CBCT) with US in the ability to accurately represent the uterus, and dosimetric studies assessing the relationship between uterine segmentation accuracy and target coverage and OAR sparing⁴¹.

360 V. CONCLUSIONS

The good agreement between manual contours when compared with results from other imaging modalities such as CT and MRI supports the use of transabdominal US to visualize the uterus prior to RT treatment for cervical cancer patients. The AGS tool was able to accurately determine the uterine shape of cervical cancer patients as well as manual
365 contouring in cases where the image quality was excellent, but not in cases where image quality was degraded by common artifacts such as shadowing and signal attenuation. The AGS tool should be used with caution for adaptive RT purposes, as it is not reliable in accurately segmenting the uterus on ‘good’ or ‘poor’ quality images. However, there may be potential to improve the performance of the AGS algorithm if the US image quality
370 is improved. The unreliable performance of the AGS tool highlights a continuing need for a rapid method of segmenting the uterus at treatment to obtain both uterine position and shape; this is a critical step in implementing US-guided adaptive RT for patients with cervical cancer.

DISCLOSURE OF CONFLICTS OF INTEREST

375 The data collected at the Herlev Hospital, University of Copenhagen site was part of a 3-year PhD research project, which was granted by Elekta Inc.

ACKNOWLEDGMENTS

The authors would like to thank Helen McNair and Naomi Hopkins for their help in coordinating the patient trial at the Royal Marsden Hospital site. Research at The Institute of Cancer Research is also supported by Cancer Research UK under Programme Grant C33389/A19727. We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and The Institute of Cancer Research. We would also like to thank Gullander L, Pedersen S K, Zarp T, Jacobsson S, Pazhang S, and Lynnerup VK for helping to conduct the study at the Herlev Hospital, University of Copenhagen site. We also acknowledge Elekta’s support for the this study.

REFERENCES

- ¹M. Baker, J. A. Jensen, and C. F. Behrens, “Determining inter-fractional motion of the uterus using 3D ultrasound imaging during radiotherapy for cervical cancer,” *Proc. of SPIE* **9040**, 90400Y–90400Y–10 (2014).
- ²R. Jadon, C. a. Pembroke, C. L. Hanna, N. Palaniappan, M. Evans, a. E. Cleves, and J. Staffurth, “A systematic review of organ motion and image-guided strategies in external beam radiotherapy for cervical cancer.” *Clinical oncology (Royal College of Radiologists (Great Britain))* **26**, 185–96 (2014).
- ³P. Chan, R. Dinniwell, M. a. Haider, Y.-B. Cho, D. Jaffray, G. Lockwood, W. Levin, L. Manchul, A. Fyles, and M. Milosevic, “Inter- and intrafractional tumor and organ movement in patients with cervical cancer undergoing radiotherapy: a cinematic-MRI point-of-interest study.” *International journal of radiation oncology, biology, physics* **70**, 1507–15 (2008).
- ⁴C. Collen, B. Engels, M. Duchateau, K. Tournel, M. De Ridder, S. Bral, D. Verellen, and G. Storme, “Volumetric imaging by megavoltage computed tomography for assessment of internal organ motion during radiotherapy for cervical cancer.” *International journal of radiation oncology, biology, physics* **77**, 1590–5 (2010).
- ⁵M. L. Bondar, M. S. Hoogeman, J. W. Mens, S. Quint, R. Ahmad, G. Dhawtal, and B. J. Heijmen, “Individualized nonadaptive and online-adaptive intensity-modulated radiotherapy treatment strategies for cervical cancer patients based on pretreatment acquired

variable bladder filling computed tomography scans,” *International Journal of Radiation Oncology Biology Physics* **83**, 1617–1623 (2012).

⁶K. Lim, W. Small, L. Portelance, C. Creutzberg, I. M. Jürgenliemk-Schulz, A. Mundt, L. K. Mell, N. Mayr, A. Viswanathan, A. Jhingran, B. Erickson, J. De los Santos, D. Gaffney, C. Yashar, S. Beriwal, A. Wolfson, A. Taylor, W. Bosch, I. El Naqa, and A. Fyles, “Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer.” *International journal of radiation oncology, biology, physics* **79**, 348–55 (2011).

⁷L. K. Mell, J. D. Kochanski, J. C. Roeske, J. J. Haslam, N. Mehta, S. D. Yamada, J. A. Hurteau, Y. C. Collins, E. Lengyel, and A. J. Mundt, “Dosimetric predictors of acute hematologic toxicity in cervical cancer patients treated with concurrent cisplatin and intensity-modulated pelvic radiotherapy,” *International Journal of Radiation Oncology Biology Physics* **66**, 1356–1365 (2006).

⁸A. J. Mundt, L. K. Mell, and J. C. Roeske, “Preliminary analysis of chronic gastrointestinal toxicity in gynecology patients treated with intensity-modulated whole pelvic radiation therapy,” *International Journal of Radiation Oncology*Biolog*Physics* **56**, 1354–1360 (2003).

⁹C. Brixey, J. Roeske, A. Lujan, and S. Lumada, “Impact of intensity modulated radiotherapy on haemotologic toxicity in women with gynecologic malignancies,” *International Journal Radiation Oncology Biology Physics* **54**, 1388 (2002).

¹⁰L. T. Tan, S. Rusell, and L. Burgess, “Acute toxicity of chemo-radiotherapy for cervical cancer: The Addenbrooke’s experience,” *Clinical Oncology* **16**, 255–260 (2004).

¹¹J. C. Roeske, A. Lujan, J. Rotmensch, S. E. Waggoner, D. Yamada, and A. J. Mundt, “Intensity-modulated whole pelvic radiation therapy in patients with gynecologic malignancies,” *International Journal of Radiation Oncology Biology Physics* **48**, 1613–1621 (2000).

¹²D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, “Flat-panel cone-beam computed tomography for image-guided radiation therapy,” *International Journal of Radiation Oncology Biology Physics* **53**, 1337–1349 (2002).

¹³D. Fontanarosa, S. van der Meer, J. Bamber, E. Harris, T. O’Shea, and F. Verhaegen, “Review of ultrasound image guidance in external beam radiotherapy: I. Treatment planning and inter-fraction motion management,” *Physics in Medicine and Biology* **60**, R77–R114

(2015).

- ¹⁴A. Taylor and M. E. B. Powell, “An assessment of interfractional uterine and cervical
440 motion: Implications for radiotherapy target volume definition in gynaecological cancer,”
Radiotherapy and Oncology **88**, 250–257 (2008).
- ¹⁵M. Lachaine and T. Falco, “Intrafractional prostate motion management with
the Clarity Autoscan system,” Medical Physics International **1**, 72–80 (2013),
arXiv:arXiv:1011.1669v3.
- ¹⁶R. Mohan, X. Zhang, H. Wang, Y. Kang, X. Wang, H. Liu, K. K. Ang, D. Kuban, and
445 L. Dong, “Use of deformed intensity distributions for on-line modification of image-guided
IMRT to account for interfractional anatomic changes,” International Journal of Radiation
Oncology Biology Physics **61**, 1258–1266 (2005).
- ¹⁷L. Tuomikoski, J. Collan, J. Keyrilainen, H. Visapaa, K. Saarilahti, and M. Tenhunen,
450 “Adaptive radiotherapy in muscle invasive urinary bladder cancer - An effective method
to reduce the irradiated bowel volume,” Radiotherapy and Oncology **99**, 61–66 (2011).
- ¹⁸H. M. Ladak, F. Mao, Y. Wang, D. B. Downey, D. a. Steinman, and A. Fenster, “Prostate
boundary segmentation from 2D ultrasound images,” Medical Physics **27**, 1777–1788
(2000).
- ¹⁹W. Qiu, J. Yuan, J. Kishimoto, J. McLeod, Y. Chen, S. de Ribaupierre, and A. Fenster,
455 “User-Guided Segmentation of Preterm Neonate Ventricular System from 3-D Ultrasound
Images Using Convex Optimization,” Ultrasound in Medicine and Biology **41**, 542–556
(2015).
- ²⁰Elekta Ltd., “Software 4.0 User Guide,” **0400** (2014).
- ²¹P. Besl and N. McKay, “A Method for Registration of 3-D Shapes,” IEEE Transactions
460 on Pattern Analysis and Machine Intelligence **14**, 239–256 (1992).
- ²²S. Lobregt and M. A. Viergever, “A Discrete Dynamic Contour Model,” IEE Transactions
on Medical Imaging **14**, 12 – 24 (1995).
- ²³N. Hu, D. B. Downey, A. Fenster, and H. M. Ladak, “Prostate surface segmentation
465 from 3D ultrasound images,” Proceedings IEEE International Symposium on Biomedical
Imaging **30**, 1648–1659 (2002).
- ²⁴A. Ghanei, H. Soltanian-Zadeh, A. Ratkewicz, and F. F. Yin, “A three-dimensional de-
formable model for segmentation of human prostate from ultrasound images,” Medical
Physics **28**, 2147–2153 (2001).

- 470 ²⁵V. Carillo, C. Cozzarini, L. Perna, M. Calandra, S. Gianolini, T. Rancati, A. E. Spinelli, V. Vavassori, S. Villa, R. Valdagni, and C. Fiorino, “Contouring variability of the penile bulb on CT images: Quantitative assessment using a generalized concordance index,” *International Journal of Radiation Oncology Biology Physics* **84**, 841–846 (2012).
- 475 ²⁶S. K. Vinod, K. Lim, L. Bell, J. Veera, L. Ohanessian, E. Juresic, N. Borok, P. Chan, R. Chee, V. Do, G. Govindarajulu, S. Sridharan, C. Johnson, D. Moses, S. Van Dyk, and L. Holloway, “High-risk CTV delineation for cervix brachytherapy: Application of GEC-ESTRO guidelines in Australia and New Zealand,” *Journal of Medical Imaging and Radiation Oncology* (2016), 10.1111/1754-9485.12509.
- 480 ²⁷D. Thomson, C. Boylan, T. Liptrot, A. Aitkenhead, L. Lee, B. Yap, A. Sykes, C. Rowbottom, and N. Slevin, “Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk.” *Radiation oncology (London, England)* **9**, 173 (2014).
- ²⁸M. L. Bondar, M. Hoogeman, W. Schillemans, and B. Heijmen, “Intra-patient semi-automated segmentation of the cervix-uterus in CT-images for adaptive radiotherapy of cervical cancer.” *Physics in medicine and biology* **58**, 5317–32 (2013).
- 485 ²⁹D. Freedman, R. J. Radke, T. Zhang, Y. Jeong, D. M. Lovelock, and G. T. Y. Chen, “Model-based segmentation of medical imagery by matching distributions,” *IEEE Transactions on Medical Imaging* **24**, 281–292 (2005).
- ³⁰T. Zhang, Y. Chi, E. Meldolesi, and D. Yan, “Automatic Delineation of On-Line Head-And-Neck Computed Tomography Images: Toward On-Line Adaptive Radiotherapy,” *International Journal of Radiation Oncology Biology Physics* **68**, 522–530 (2007).
- 490 ³¹C. Garnier, J. J. Bellanger, K. Wu, H. Shu, N. Costet, R. Mathieu, R. De Crevoisier, and J. L. Coatrieux, “Prostate segmentation in HIFU therapy,” *IEEE Transactions on Medical Imaging* **30**, 792–803 (2011).
- ³²M. G. Jackson, J. Ludmir, and T. J. Bader, “The Accuracy of Digital Examination and
495 Ultrasound in the Evaluation of Cervical Length,” (1992).
- ³³D. Paladini, “Sonography in obese and overweight pregnant women: clinical, medicolegal and technical issues.” *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* **33**, 720–9 (2009).
- 500 ³⁴X. Artignan, M. H. P. Smitsmans, J. V. Lebesque, D. A. Jaffray, M. Van Her, and H. Bartelink, “Online ultrasound image guidance for radiotherapy of prostate cancer: Impact of image acquisition on prostate displacement,” *International Journal of Radiation*

Oncology Biology Physics **59**, 595–601 (2004).

- ³⁵M. Fargier-Voiron, B. Presles, P. Pommier, S. Rit, A. Munoz, H. Liebgott, D. Sarrut, and M. C. Biston, “Impact of probe pressure variability on prostate localization for ultrasound-based image-guided radiotherapy,” *Radiotherapy and Oncology* **111**, 132–137 (2014).
- ³⁶M. Baker, D. T. Cooper, and C. F. Behrens, “Evaluation of uterine ultrasound imaging in cervical radiotherapy ; a comparison of autoscan and conventional probe,” (2016).
- ³⁷B. E. Nelms, W. A. Tomé, G. Robinson, and J. Wheeler, “Variations in the contouring of organs at risk: Test case from a patient with oropharyngeal cancer,” *International Journal of Radiation Oncology Biology Physics* **82**, 368–378 (2012).
- ³⁸E. L. Lorenzen, C. W. Taylor, M. Maraldo, M. H. Nielsen, B. V. Offersen, M. R. Andersen, D. O’Dwyer, L. Larsen, S. Duxbury, B. Jhitta, S. C. Darby, M. Ewertz, and C. Brink, “Inter-observer variation in delineation of the heart and left anterior descending coronary artery in radiotherapy for breast cancer: A multi-centre study from Denmark and the UK,” *Radiotherapy and Oncology* **108**, 254–258 (2013).
- ³⁹V. K. Reed, W. A. Woodward, L. Zhang, E. A. Strom, G. H. Perkins, W. Tereffe, J. L. Oh, T. K. Yu, I. Bedrosian, G. J. Whitman, T. A. Buchholz, and L. Dong, “Automatic Segmentation of Whole Breast Using Atlas Approach and Deformable Image Registration,” *International Journal of Radiation Oncology Biology Physics* **73**, 1493–1500 (2009).
- ⁴⁰X. A. Li, A. Tai, D. W. Arthur, T. A. Buchholz, S. Macdonald, L. B. Marks, J. M. Moran, L. J. Pierce, R. Rabinovitch, A. Taghian, F. Vicini, W. Woodward, and J. R. White, “Variability of Target and Normal Structure Delineation for Breast Cancer Radiotherapy: An RTOG Multi-Institutional and Multiobserver Study,” *International Journal of Radiation Oncology Biology Physics* **73**, 944–951 (2009).
- ⁴¹C. F. Behrens, T. B. Andreasen, H. Lindberg, S. K. Buhl, A. Vestergaard, U. V. Elstrøm, and D. Sjöström, “Quantitative image quality evaluation of pelvic computed tomography-based imaging systems: a novel concept in radiotherapy.” *Acta oncologica (Stockholm, Sweden)* **52**, 1579–82 (2013).
- ⁴²Y. Yu, J. Molloy, and S. Acton, “Three-dimensional speckle reducing anisotropic diffusion,” *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003* **2**, 1260–1270 (2003).
- ⁴³K. Z. Abd-Elmoniem, A. B. M. Youssef, and Y. M. Kadah, “Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion,”

IEEE Transactions on Biomedical Engineering **49**, 997–1014 (2002).

⁵³⁵ ⁴⁴S. Ji, D. W. Roberts, A. Hartov, and K. D. Paulsen, “NIH Public Access,” *Med Image Comut Comput Asist Interv* **12**, 795–802 (2009).

⁴⁵T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 681–685 (2001).