ARTICLE IN PRESS

The accuracy of prostate volume measurement from ultrasound images: A quasi-Monte Carlo simulation study using magnetic resonance imaging

David-Olivier D. Azulay [a],[*], Philip Murphy [a], Jim Graham [b]

[a] Precision Medicine, Pfizer, Sandwich CT13 9NJ, UK
[b] Centre for Imaging Science, Manchester Academic Health Sciences Centre, University of Manchester, M13 9PT, UK

ARTICLE INFO

ABSTRACT

Prostate volume is an important parameter to guide management of patients with benign prostatic hyperplasia (BPH) and to deliver clinical trial endpoints. Generally, simple 2D ultrasound (US) approaches are favoured despite the potential for greater accuracy afforded by magnetic resonance imaging (MRI) or complex US procedures. In this study, different approaches to estimate prostate size are evaluated with a simulation to select multiple organ cross-sections and diameters from 22 MRI-defined prostate shapes. A quasi-Monte Carlo (qMC) approach is used to simulate multiple probe positions and angles within prescribed limits resulting in a range of dimensions. The basic ellipsoid calculation which uses two scanning planes compares well to the MRI volume across the range of prostate shapes and sizes ($R = 0.992$). However, using an appropriate linear regression model, accurate volume estimates can be made using prostate diameters calculated from a single scanning plane.

1. Introduction

Estimation of a prostate volume is an integral component in the evaluation of patients with BPH. While measurement performance is important for routine clinical assessment, precision, reproducibility and practicality of measurement are crucial to enable derivation of clinical trial endpoints [6,15].

Volume estimation of the prostate is normally carried out using trans-rectal US images. The measurement can be made by a planimetric method where a stack of 2D slices is constructed by step movements of the ultrasound probe. Clinically it is more convenient to make NPUS measurements of prostate diameters in the latero-lateral (LL), anterio-posterior (AP) and cranio-caudal (CC) directions from 2D images using the assumption of an ellipsoidal shape. By displacing the probe the operator determines the orientation for the best estimate of diameters.

Bazinet et al. [3] have noted that volume measured in this way is inaccurate when compared with US (PUS) or magnetic resonance (MR) and conducted a reproducibility trial, finding that differences in volume estimates of up to 25% could be obtained in successive NPUS examinations of the same patient.

Direct 3D measurements of a volume by PUS or MR have the potential to deliver more accurate and reproducible results but are considerably more expensive, acquisition is time consuming and, without automated interpretation, analysis is highly labour-intensive [13,26,1]. US remains widely available to the urologist and therefore remains the most practical method of estimating prostate volume [22]. In this study we sought to investigate which estimation method can best be used with 2D NPUS to predict prostate volume.

Allen et al. [2] have investigated the use of active shape model (ASM) search for measuring the volumes of the complete gland and central gland using MR images. An ASM is a statistical model of shape, built from a training set of images which have been segmented (the important surfaces defined) manually [7,9]. For their study [2] collected twenty two fat-suppressed MR images of prostates from patients diagnosed with BPH attending the urology clinic at Salford Royal Hospital, UK.

The objective in this study is to simulate volume measurements by NPUS using the manual segmentation obtained from the ASM study in order to quantify the limitations on the accuracy and reproducibility of taking this approach. Rahmouni et al. [21] and Lee and Chung [13] have reported that measurements of volume using MR accurately represent the volumes of real specimens after prostatectomy. We therefore take the manually delineated borders of the prostate used by Allen et al. [2] to be realistic ground truth representing the prostate boundary.

* Corresponding author. Tel.: +44 1304 620 314.
E-mail addresses: david-olivier.azulay@pfizer.com,
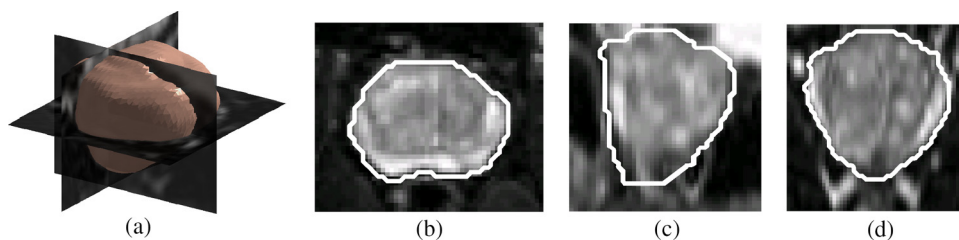david-olivier.azulay@wanadoo.fr (D.-O.D. Azulay).

Fig. 1. Example of a 3D rendered prostate: the approximately elliptical boundaries are taken to represent the projections that would be observed in NPUS images.

Fig. 1a shows a 3D rendered prostate volume from the manually annotated data. Axial, sagittal and coronal sections are indicated and the corresponding slices through the volume are shown in Fig. 1b–d, with the prostate contour superimposed.

While Allen et al. [2] addressed the segmentation of both the whole gland and the central gland, only the whole gland is considered here, as this is what may be measured using NPUS. Furthermore, internal structures are not necessarily comparable between MR and US images. We make the assumption that the US-defined borders are close approximations to the borders defined in MR images. Making realistic hypotheses about the range of possible orientations of the NPUS probe, we have implemented a qMC method to assess the distribution of the error in volume estimates produced by the different direct ellipsoidal formulae and the linear models extrapolated from these formulae. The errors have been successively quantified with standard linear regression goodness-of-fit parameters, correlation coefficients and Bland–Altman (BA) plots.

2. Methods

The most common clinical approach to estimating volume using NPUS is to capture US images from two approximately orthogonal planes passing through the centre of the prostate. Rather than collecting US and MR images from volunteers, we simulate the US measurements using a qMC procedure that allows us to investigate the error distributions arising from varying the choices of the US planes used and the analytical methods applied to produce the volume estimates. The reference volume for the simulation procedure is a triangulated closed surface defining the 3D volume of the prostate derived from manual segmentation of MR images (see Fig. 1a). An evenly distributed selection of pairs of orthogonal cuts through this surface is simulated to generate perimeters that would have been obtained from the corresponding NPUS images. There is evidently an infinite number of planes cutting a 3D volume but the generation of a representative subset is required. Triplets of points distributed on the surface of a sphere are successively picked to define planes that belong to this subset.

2.1. Devising an evenly distributed set of orthogonal cuts

The general procedure is illustrated with the 2D example sketched in Fig. 2.

The blue line represents the outer boundary of the 2D slice of an artificial prostate, the dotted rectangle its bounding box. Cuts are simulated across this boundary by selecting random pairs of points on the circumference of a circle. The figure shows six points uniformly distributed on the circumference of a small circle near the centre of the prostate. A small circle is preferred, rather than a large one circumscribing the bounding box because the cuts should pass close to the centre of the prostate. There are several cuts that are geometrically equivalent in that they divide the

circle into segments whose areas are in the same ratio (e.g. AB, BC, CD, DE, EF and FA, or AC, AE, BD, BF, CE and DF). However, only some of those are suitably oriented. To start with the angle between the probe and the anterior-posterior axis should be small to reflect probe positioning. The selections are therefore limited by requiring that this angle should be less than 22.5°. While somewhat arbitrary, this tolerance probably represents an upper limit on a human observer's ability to estimate the orientation of the LL diameter. On Fig. 2, AD is an acceptable choice because the angle α with the AP axis is small, whereas CE is not acceptable because the angle β is too large. Then it is assumed that an NPUS operator should judge fairly accurately the position of the centre of the prostate which would be close to the middle zone of the bounding box. In determining next the LL diameter, it is also assumed that the two diameters should intersect close to the true centre.

The extension to 3D is less straightforward because there is no easy parameterisation that allows fine control over the likely position and orientation of the plane that is necessary to simulate the clinical situation. The obvious extension to Fig. 2 is to distribute points on the surface of a sphere. However, distributing points evenly on a sphere is a complex problem [23]. The solution we have adopted is due to Li et al. [14], which generates a "low discrepancy" sequence of points by using the terms of the sequence to approximate the sizes of surface on the sphere defined by neighbouring groups of points. In other words, each point on the surface is equidistant to its nearest neighbours. This guarantees the existence of a bounded error (low discrepancy with the uniform distribution).

A plane is fully determined from the selection of three distinct points. Because these points are drawn from an evenly spaced sequence (and not from a random sequence), the simulation is based on a qMC (and not a Monte Carlo) method [18]. Sloane et al. [24] have calculated a convenient set of distributions, which were downloaded and used for this study (see Fig. 3). Planes generated
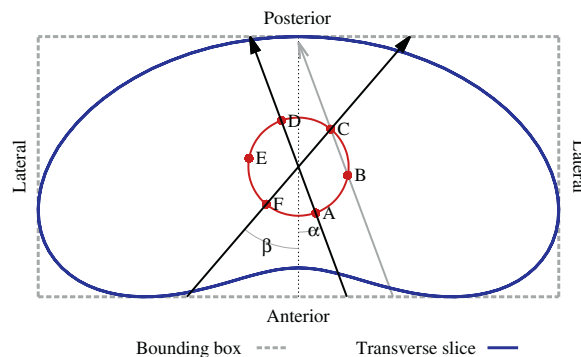

Fig. 2. Principle of the cut selection procedure on the schematic of a 2D slice: the α cut is accepted but not the β cut because it is too far from the AP axis. The BC cut would be equally accepted as its angle is α as well, even though it does not pass through the center of the circle.
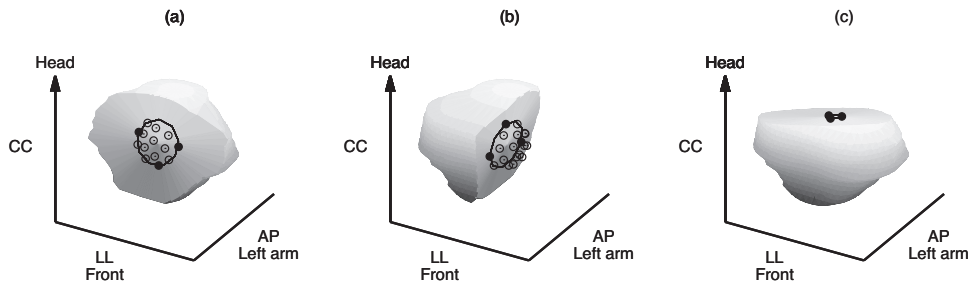
Fig. 3. Example of three possible cuts in 3D. Three points from the uniform distribution of 50 points on a sphere are chosen to define each cutting plane; (a) cut defining CC and LL on the coronal plane, (b) defining AP and CC on the sagittal plane, (c) defining AP and LL along the transverse plane. (c) is an example of an extreme cut that passes just through the central sphere trimming the top of the prostate horizontally.

Table 1
Linear regressions on the top 5% and 10% estimates. Every one of the 10 listed formulae generates a different set of 22 estimates that was used to predict the 22 fixed MR volumes. All the SE_b are statistically significant ($p^{\ddagger} < 0.001$) demonstrating the appropriateness of the linear fits. The likelihoods of the χ^2 observations were calculated as well to confirm the validity of the linear regressions ($Q'' > 0.1$, $Q' > 0.01$ and $Q^{\star} > 0.001$). The CI of the average $\overline{V_d}$ is printed in the last column. The first three linear regressions on the top 5% estimates are illustrated in Fig. 7.

| V_d (cm^3) | Top 5% | | | | | Top 10% | | | | |
	$MR = a + bV_d$	R	SE_b	$\overline{V_d}$	$CI(\overline{V_d})$	$MR = a + bV_d$	R	SE_b	$\overline{V_d}$	$CI(\overline{V_d})$
AP.CC.LL	$2.43 + 0.962V_d''$	0.992	$0.028^{\ddagger}$	44.2	1.14	$2.80 + 0.974V_d''$	0.990	$0.030^{\ddagger}$	43.3	1.23
AP . LL^2	$0.08 + 0.894V_d''$	0.983	$0.037^{\ddagger}$	50.2	1.63	$0.65 + 0.920V_d''$	0.987	$0.033^{\ddagger}$	48.2	1.40
CC . LL^2	$-5.16 + 0.818V_d''$	0.983	$0.034^{\ddagger}$	61.3	1.64	$-3.88 + 0.822V_d''$	0.981	$0.036^{\ddagger}$	59.4	1.71
LL . CC^2	$-6.38 + 0.819V_d''$	0.976	$0.041^{\ddagger}$	62.6	1.94	$-5.73 + 0.824V_d''$	0.975	$0.042^{\ddagger}$	61.5	1.96
LL . AP^2	$8.13 + 0.995V_{d'}$	0.969	$0.057^{\ddagger}$	37.0	2.21	$8.52 + 1.024V_{d'}$	0.968	$0.059^{\ddagger}$	35.6	2.23
CC . AP^2	$5.87 + 1.031V_{d'}$	0.965	$0.062^{\ddagger}$	37.9	2.32	$6.62 + 1.047V_{d'}$	0.964	$0.065^{\ddagger}$	36.6	2.38
AP . CC^2	$-1.37 + 0.900V_d^{\star}$	0.961	$0.058^{\ddagger}$	51.5	2.46	$-1.04 + 0.918V_{d'}$	0.965	$0.056^{\ddagger}$	50.1	2.35
AP^3	$11.36 + 1.090V_d^{\star}$	0.948	$0.082^{\ddagger}$	30.8	2.83	$12.30 + 1.114V_d^{\star}$	0.941	$0.090^{\ddagger}$	29.3	3.01
LL^3	$-7.34 + 0.708V_d^{\star}$	0.941	$0.057^{\ddagger}$	73.9	3.00	$-6.90 + 0.740V_d^{\star}$	0.956	$0.051^{\ddagger}$	70.1	2.61
CC^3	$-6.18 + 0.673V_d^{\star}$	0.887	$0.078^{\ddagger}$	75.9	4.11	$-5.83 + 0.686V_d^{\star}$	0.894	$0.077^{\ddagger}$	74.0	3.98

by close neighhbours would miss the center of the sphere and therefore the approximate center of the prostate. These extreme cases contribute to the observed variability in volume estimates.

There are three formulae for calculating volumes based on different ellipsoidal approximations, requiring up to three diameters (the $\pi/6$ term will be omitted when referring to a specific formula to simplify the notations):

- spherical $\pi/6 \times d_1^3$,
- spheroidal $\pi/6 \times d_1^2 \times d_2$ which falls into either the egg-shaped prolate ($d_1 < d_2$) or disk-shaped oblate ($d_1 > d_2$) categories,
- ellipsoidal $\pi/6 \times d_1 \times d_2 \times d_3$.

Each one of the three diameters d_1, d_2 and d_3 can successively be AP, CC and LL leading to ten possibilities (see Table 1).

The LL, AP and CC diameters in the transverse and sagittal planes should all intersect at their middle point and be perpendicular to each other, which means that the sagittal plane needs to be perpendicular to the transverse one while containing the AP diameter. Strict perpendicularity between planes, not diameters, is maintained during the qMC simulation to prevent the overestimation of the CC diameter. That is due to the latent inclination of the sagittal plane which artificially increases the CC diameter when manually measured [8].

2.2. Simulation algorithm

One potential approach could have been the exhaustive enumeration of perfectly orthogonal cutting diameters across the 3D volume. However, in a practical situation, the positions and angles of the diameters and planes are estimated visually and therefore of limited precision. Rahmouni et al. [21] and Kim and Kim [12]

have reported that the choice of the transducer orientation relative to the prostate is operator dependent. The transverse slice is determined first since it is less error prone when delimited by the human eye [17,11]. As a consequence, some flexibility has been introduced in the automated procedure: for instance the longest diameter is drawn at the expense of true orthogonality. We set the tolerances on perpendicular angles between diameters to 5° and on middle point localisations to a 5% precision of the lengths of respective diameters. The purpose is to compensate for both the visual imprecision of the operator and the potential shape deformation of the prostate due to the probe.

The steps implemented to achieve the qMC simulation are summarised in the following algorithm:

Algorithm 2.1.

1:     define points of a sphere such that each point is equidistant to its nearest neighbours
2:     **repeat**
3:       choose a triplet of points from this uniform distribution to define a transverse plane
4:       compute the 2D transverse slice
5:       search the longest AP diameter within the 22.5° tolerance angle of the real AP axis
6:       search the longest LL diameter which is orthogonal to the AP diameter within the 5° tolerance angle and intersects at their middle points within the 5% tolerance distance
7:       compute the 2D sagittal slice which is perpendicular to the transverse one and contains the AP segment
8:       search the longest CC diameter which is orthogonal to the AP diameter within the 5° tolerance angle and intersects at their middle points within the 5% tolerance distance
9:     **until** all triplets of points have been exhausted

A distribution of $n = 50$ points on the sphere was chosen, describing $C_3^{50} = 19,600$ distinct planes, out of which $m = 4808$ are properly oriented. Each accepted pair of planes provides estimates of the LL,
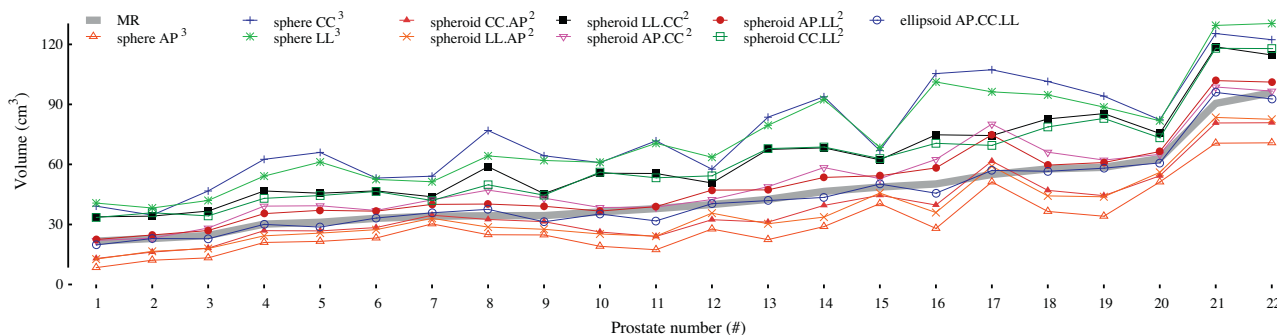
Fig. 4. Averages of the top 5% estimates for the ten formulae and each prostate, sorted by MR volume.

CC and AP diameters from which ten estimated volumes are computed: one with each of the formulae and every combination of parameters. These are the estimates to be compared with the true volume obtained from the manual segmentation of the MR images. In NPUS imaging an operator would normally seek to maximise the diameters. Kim and Kim [12] have shown that experienced operators produce larger estimates than beginner and trained ones. So all the main statistics in Section 3 are derived from the top 5% (240 largest) volume estimates to reflect how a user would be likely to decide on the maximum measured prostate size.

The figure of 5% is selected to provide sufficient statistics to construct a linear model, constrained to produce the largest estimates. The threshold is rather arbitrary and we include figures in Table 1 representing models calculated using the largest 10% for comparison.

2.3. Error prediction for the linear models

Each formula for calculating the volume from measured diameters corresponding to each pair of planes provides average and variance values for the volume of each of the 22 images. This forms a training set that is used to define a linear model of the form $y = a + bx$. The predictor, x, corresponds to one of the spherical, spheroidal or ellipsoidal estimates and the outcome, y, is the true MR volume [5].

Leave-one-out cross-validation analyses were completed in order to quantify the sensitivity of these linear models [20]. For every formula, 22 distinct training sets are created including 21 averages, by leaving one average out from the original set and a new linear regression calculated for each of them.

2.4. Assessing the agreement of the estimates with the MR volumes

Bland and Altman [4] described a statistical procedure that here can be used to evaluate the relative concordance between assumed ground truth from MR and simulated NPUS imaging as a function of prostate size. The BA plot shows how the difference between measurements of the same quantity is related to the mean of the measurement. If the differences follow a normal distribution with average ave and standard deviation SD then 95% of their values should be within the limits $ave \pm 1.96 \times SD$, also called limits of agreement (LoA).

3. Results

Fig. 4 indicates that application of different ellipsoidal approximations from the top 5% estimates tend to systematically overestimate or underestimate the true prostate volume. This error can be corrected by training a regression model that allows measured

values to predict the true volume. Here we use a linear model of the form $y = a + bx$ for simplicity. A slope lower or higher than 1 will correct for overestimates and underestimates respectively [27,11].

Table 1 shows the linear models derived from the training data corresponding to each of the ellipsoidal calculations along with their correlation coefficients R, standard error of the slope (SE_b), mean value of the estimated volume and 95% confidence interval around the mean.

The statistical significance of the linear regression parameters can be estimated by the probability (Q) of the chi-square distribution [19]. Values of Q greater than 0.1 represent believable fits. The linear model is adequate to extrapolate the true volume of the prostate, especially when the $AP.CC.LL$, $AP.LL^2$, $CC.LL^2$ or $LL.CC^2$ formula is used.

The size of the 95% confidence interval (CI) of the average as a proportion of the average itself is about 3% for the first four ranked estimates. The arbitrary top 5% criterion reflects the experience of the operator searching for the maximum diameters. A less experienced operator is more likely to miss these maxima; this is quantified by the extension to 10%, that is to say the addition of 5% lower volumes to the previous calculations. Using the top 10% clearly results in slightly lower volume estimates (around 2 cm^3), however the model parameters and standard errors are not affected greatly by the acceptance threshold.

Fig. 5 shows the variation in the correlation coefficient that arises when predicting the true (MR) volume for each of the training set using a model constructed in a leave-one-out manner from the remaining data. Values for the three-diameter ellipsoidal calculation and the two estimates using only two diameters with the highest correlation are shown. Fig. 6 shows the Bland–Altman plots for the same three volume estimation formulae. The grey bands indicate the 95% confidence limits around the limits of agreement values.
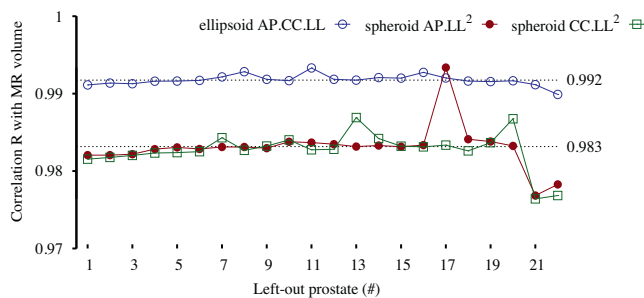

Fig. 5. Leave-one-out correlation coefficient R from the linear regressions on the top 5% estimates. The horizontal dotted lines display the corresponding values of the complete training set shown in Table 1.
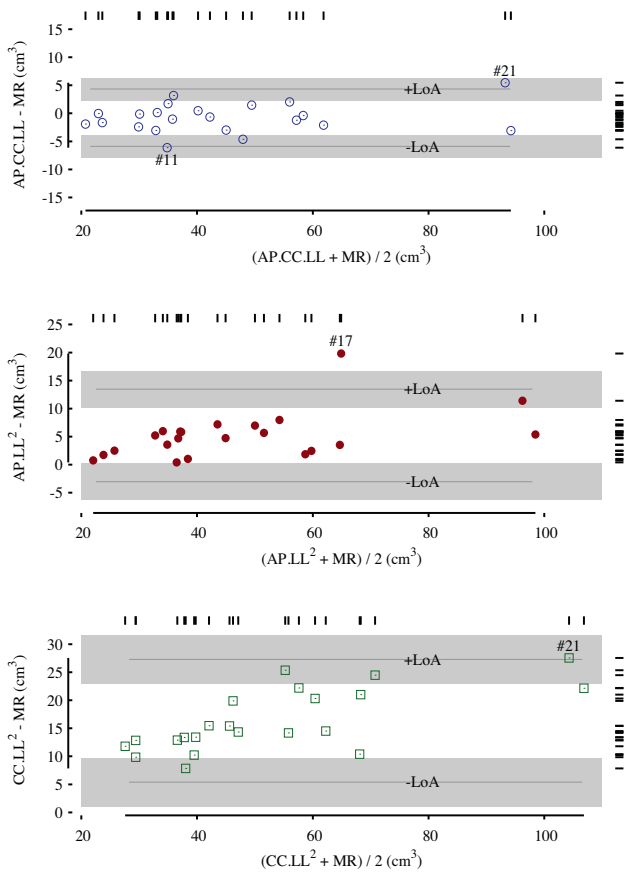
Fig. 6. BA plots for the three direct formulae. Estimates are obtained from the top 5% largest volumes and scales are preserved across the plots to facilitate the visual comparisons.

The means μ and standard deviations σ of three of the formulae and their respective diameters are listed in Table 2 along with the measured ground-truth volume (MR).

Table 2 also shows the distribution of the 4808 simulated ellipsoidal volumes for each prostate as a histogram. The vertical bar in the histogram marks the reference volume, that is to say the 0% deviation from the MR volume.

4. Discussion

A good linear model depends on the reproducibility of the deviation, rather than its amplitude. For instance, in Table 2, the $CC.LL^2$ direct estimate systematically overestimates the MR by 39% on

average; however its derived linear model is acceptable as an estimate according to the goodness-of-fit parameters R and SE_b in Table 1.

[4] explain why correlation coefficients can be misleading to quantify the quality of the agreement between two measurements and suggest more appropriate calculations to perform better assessments. They complement the standard linear regression plot presented in Fig. 7 by the examination of the estimate errors, i.e. the differences between the real and approximated values (Fig. 6). The CI of the LoA are delimited by the gray bands whose widths are both equal to $1.96 \times CI(\pm LoA)$.

The $CC.LL^2$ formula exhibits wide gray bands delimiting the 95% LoA confidence intervals. Those are inadequate for an approximation method even if the correlation coefficient R for the $CC.LL^2$ model is appealingly high in Table 1. Conversely, the spheroidal $AP.LL^2$ estimate is satisfactory according to both its correlation coefficient and its BA plot notwithstanding the fact that its performance is undermined by the extreme outlier #17 (see also Fig. 5). The high correlation of the $AP.LL^2$ model approximation indicates that most of the prostates in the sample can be reasonably approximated as oblate spheroids with circular cross-section in the coronal plane (the LL and CC diameters are often of similar length in Table 2). Examination of the shape of prostate #17 indicates that its shape approximates a prolate spheroid with a circular cross-section in the transverse plane. Fig. 5 reveals that the correlation coefficient of the $AP.LL^2$ spheroidal model is better than those of the ellipsoidal model when this prostate does not contribute to the model. The correlation coefficient is slightly higher than the ellipsoidal model in that case.

Fig. 4 demonstrates that the direct ellipsoidal calculation is consistently close to the truth when restricted to the top 5% estimates. Nevertheless there are significant variations as indicated by the horizontal range of all the histograms displayed in Table 2: estimates can vary from −50% to +20% which is in agreement with the observation of a high variability in reproducibility studies by [3]. The histograms also reveal that the majority of the qMC ellipsoidal volumes underestimate the true value. Because the prostate is not a perfectly ellipsoidal object [21], the probability of coming across an orientation which captures the three largest diameters in a given pair of orthogonal planes at the same time is quite low. This explains why underestimation is expected and justifies the use of the largest volumes only in the analysis [16,10,22].

The volume based on the ellipsoidal choice benefits from narrower confidence intervals, but requires the acquisition of both the transverse and sagittal planes. The $AP.LL^2$ linear regression appears to be of good accuracy but is obtained from the transverse plane only; [10], [25] and [28] raised the same practical consideration in terms of clinical feasibility. As determined by Fig. 4, the spherical formula provides a lower limit on the volume estimate if the
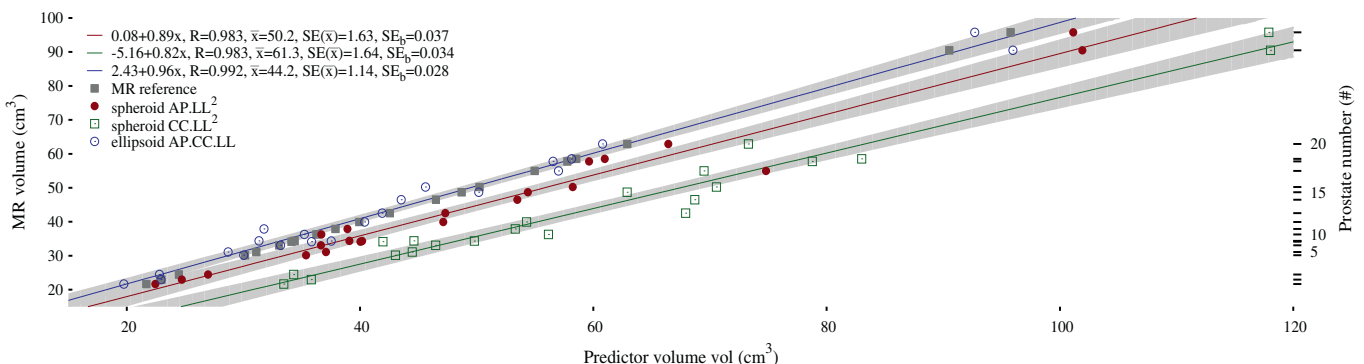

Fig. 7. Linear regressions on the averages of the top 5% volumes for three formulae (ellipsoidal and two spheroidal models). An overestimate is found on the right side of the MR reference value, an underestimate on the left side. The CI are delimited by the grey bands centered around their line equation.

Table 2

Means (μ) and standard deviations (σ) for the volume estimates for three of the formulae: ellipsoid ($AP.CC.LL$) and spheroids ($AP.LL^2$ and $CC.LL^2$) together with the values of the diameters giving rise to these volumes for each image of the 22 subjects. These are derived from the 5% highest volumes in the qMC simulation. The histograms show the distribution of all the ellipsoidal volumes (grey) as a percentage deviation from the MR volume (indicated by the vertical dotted line), while the bins shown in red represent the 5% largest estimates used to derive the statistics shown. The average deviation of each estimate from the true volume is denoted Δ. Overall estimates of μ and σ appear at the foot of the table.

#		MR	AP	CC	LL		AP.CC.LL	Δ	AP	LL	$AP.LL^2$	Δ	CC	LL	$CC.LL^2$	Δ
1	μ	21.7	2.4	4.1	3.9		19.8	−8.9%	2.4	4.3	22.4	+3.6%	3.7	4.2	33.5	+54.4%
	σ		0.1	0.2	0.2		0.6	2.8%	0.0	0.1	0.5	2.3%	0.3	0.1	1.0	4.6%
2	μ	23.0	2.7	3.9	4.2		22.9	−0.2%	2.7	4.2	24.7	+7.6%	3.9	4.2	35.8	+56.0%
	σ		0.1	0.1	0.1		0.4	1.7%	0.1	0.1	0.8	3.3%	0.1	0.1	1.1	4.8%
3	μ	24.5	2.8	4.0	3.9		22.8	−6.8%	2.8	4.3	27.0	+10.2%	3.6	4.3	34.3	+40.2%
	σ		0.1	0.3	0.3		0.3	1.3%	0.1	0.1	1.4	5.5%	0.2	0.1	1.1	4.4%
4	μ	30.1	3.2	4.6	3.9		30.0	−0.4%	3.1	4.7	35.4	+17.3%	3.7	4.7	43.0	+42.7%
	σ		0.1	0.4	0.4		0.5	1.6%	0.1	0.1	1.9	6.4%	0.1	0.1	2.2	7.3%
5	μ	31.1	3.3	4.4	3.9		28.7	−7.7%	3.1	4.8	37.1	+19.2%	3.6	4.9	44.5	+43.0%
	σ		0.1	0.6	0.6		0.8	2.5%	0.2	0.2	2.9	9.5%	0.1	0.2	2.4	7.7%
6	μ	33.1	3.4	4.3	4.3		33.2	+0.4%	3.3	4.6	36.6	+10.8%	4.2	4.6	46.5	+40.6%
	σ		0.1	0.2	0.3		0.4	1.3%	0.1	0.1	0.9	2.7%	0.1	0.1	1.5	4.6%
7	μ	34.1	3.8	4.4	4.2		35.8	+5.0%	3.7	4.6	40.0	+17.3%	4.0	4.5	42.0	+23.0%
	σ		0.1	0.2	0.2		0.4	1.1%	0.2	0.1	1.5	4.3%	0.3	0.2	1.3	3.7%
8	μ	34.3	3.5	5.0	4.1		37.5	+9.3 %	3.2	4.9	40.2	+17.0 %	4.2	4.8	49.8	+45.0 %
	σ		0.2	0.3	0.3		1.3	3.7%	0.2	0.2	3.2	9.4%	0.5	0.4	4.0	11.7%
9	μ	34.4	3.6	4.5	3.8		31.3	−8.9%	3.2	4.8	39.1	+13.7%	3.6	4.9	44.6	+29.7%
	σ		0.2	0.5	0.5		1.3	3.7%	0.2	0.2	2.0	5.9%	0.1	0.1	1.6	4.6%
10	μ	36.3	3.2	4.7	4.5		35.2	−2.9%	3.0	4.8	36.7	+1.1 %	4.5	4.9	56.1	+54.8 %
	σ		0.1	0.2	0.2		0.7	2.0%	0.2	0.1	1.4	3.8%	0.1	0.1	1.5	4.2%
11	μ	37.9	3.1	4.6	4.4		31.8	−16.2 %	2.9	5.1	38.9	+2.7 %	4.0	5.0	53.3	+40.7 %
	σ		0.2	0.5	0.4		1.1	2.8%	0.2	0.2	1.7	4.5%	0.2	0.2	2.1	5.6%
12	μ	39.9	3.7	4.3	4.9		40.4	+1.2%	3.7	4.9	47.1	+18.0%	4.3	4.9	54.3	+35.9%
	σ		0.1	0.1	0.1		0.5	1.4%	0.1	0.1	1.5	3.8%	0.1	0.1	0.7	1.8%
13	μ	42.6	3.4	5.0	4.8		41.9	−1.6%	3.2	5.3	47.3	+11.2 %	4.7	5.3	67.9	+59.6 %
	σ		0.1	0.3	0.3		0.9	2.1%	0.2	0.1	2.2	5.1%	0.2	0.1	1.3	3.2%
14	μ	46.5	3.7	5.3	4.2		43.5	−6.4 %	3.3	5.6	53.5	+15.0 %	4.2	5.6	68.7	+47.7 %
	σ		0.2	0.4	0.3		1.1	2.4%	0.1	0.1	2.2	4.7%	0.1	0.1	2.3	5.0%
15	μ	48.7	4.1	4.8	4.9		50.2	+3.0 %	4.1	5.0	54.4	+11.7 %	4.7	5.1	62.9	+29.1 %
	σ		0.1	0.2	0.2		0.5	1.0%	0.1	0.1	1.5	3.1%	0.1	0.1	1.5	3.1%
16	μ	50.2	3.6	5.4	4.5		45.6	−9.2 %	3.4	5.7	58.2	+15.9 %	4.1	5.8	70.5	+40.4 %
	σ		0.2	0.7	0.6		1.8	3.6%	0.2	0.2	3.8	7.5%	0.1	0.1	3.0	6.0%

Table 2 (*Continued*)

#		MR	AP	CC	LL		AP.CC.LL	Δ	AP	LL	AP.LL²	Δ	CC	LL	CC.LL²	Δ
17	μ	55.0	4.5	5.3	4.6		57.0	+3.7 %	4.4	5.7	74.8	+36.1 %	4.1	5.7	69.5	+26.4 %
	σ		0.1	0.7	0.7		0.9	1.6%	0.1	0.2	5.5	9.9%	0.1	0.2	4.7	8.5%
18	μ	57.8	4.0	5.4	5.0		56.6	−2.1 %	3.7	5.5	59.6	+3.2 %	4.8	5.6	78.8	+36.3 %
	σ		0.2	0.4	0.3		2.8	4.9%	0.3	0.2	2.4	4.2%	0.2	0.2	3.2	5.6%
19	μ	58.5	3.9	5.4	5.3		58.1	−0.7 %	3.8	5.5	61.0	+4.2 %	5.4	5.4	83.0	+41.8 %
	σ		0.1	0.1	0.2		0.9	1.5%	0.1	0.1	1.4	2.3%	0.2	0.1	1.5	2.6%
20	μ	62.9	4.4	5.2	5.2		60.8	−3.4 %	4.4	5.4	66.4	+5.6 %	5.1	5.2	73.3	+16.5 %
	σ		0.2	0.2	0.1		1.8	2.9%	0.2	0.1	2.5	3.9%	0.3	0.2	2.3	3.7%
21	μ	90.5	5.0	6.0	6.1		96.0	+6.0 %	5.0	6.2	101.9	+12.6 %	6.0	6.1	118.0	+30.4 %
	σ		0.1	0.1	0.1		1.3	1.5%	0.1	0.1	1.0	1.1%	0.1	0.0	0.7	0.8%
22	μ	95.8	5.0	5.8	6.1		92.7	−3.2 %	5.0	6.2	101.1	+5.6 %	5.8	6.3	117.9	+23.1 %
	σ		0.2	0.2	0.2		1.5	1.6%	0.2	0.1	3.0	3.1%	0.2	0.1	2.5	2.7%
	$\overline{\mu}$	44.9	3.6	4.8	4.6		44.2	−2.3 %	3.5	5.1	50.2	+11.8 %	4.4	5.1	61.3	+39.0 %
	$\overline{\sigma}$		0.1	0.3	0.3		1.0	2.2 %	0.1	0.1	2.0	4.8 %	0.2	0.1	2.0	4.8%

Table 3

Mapping between measured values of first three non-planimetric approximations of Table 1 and the estimates of true (MR) volume V_e using the regression models of Fig. 7. The left column provides some example values of the measured direct volume V_d. The remaining columns show the corresponding estimates V_e of the true (MR) volume derived from each model, together with their CI.

V_d	$V_d = AP.CC.LL$ $V_e = 2.43 + 0.962V_d$ $V_e \pm CI(V_e)cm^3$	$V_d = AP.LL^2$ $V_e = 0.08 + 0.894V_d$ $V_e \pm CI(V_e)cm^3$	$V_d = CC.LL^2$ $V_e = -5.16 + 0.818V_d$ $V_e \pm CI(V_e)cm^3$
20	21.68 ± 1.81	17.97 ± 2.85	11.19 ± 3.39
30	31.31 ± 1.41	26.91 ± 2.26	19.37 ± 2.79
40	40.93 ± 1.17	35.86 ± 1.81	27.54 ± 2.25
50	50.56 ± 1.19	44.80 ± 1.63	35.72 ± 1.83
60	60.18 ± 1.46	53.75 ± 1.79	43.90 ± 1.65
70	69.81 ± 1.88	62.69 ± 2.24	52.07 ± 1.76
80	79.43 ± 2.37	71.64 ± 2.83	60.25 ± 2.12
90	89.06 ± 2.90	80.58 ± 3.49	68.43 ± 2.64
100	98.68 ± 3.44	89.53 ± 4.20	76.61 ± 3.23

AP diameter is used and an upper limit if the *LL* diameter is used. Generally speaking the AP diameter is the shortest and the LL diameter is the longest, with the result that the $AP.LL^2$ oblate spheroid provides a suitable estimate of volume.

The regression lines specified in Fig. 7 allow us to estimate the true volume of the prostate given specific measurements of the ellipsoidal ($AP.CC.LL$), oblate ($AP.LL^2$) and prolate ($CC.LL^2$) approximations. Table 3 lists the estimated values V_e corresponding to examples of each of these direct measurements V_d together with associated confidence limits.

For instance, a direct volume of 50 cm³ using the $AP.LL^2$ approximation will result in a 95% chance that the true volume is 44.8 cm³.

Using this table as a guide, measured values of each of these three approximations can be translated into an estimated true volume.

The results of the leave-one-out exercise in Fig. 5 shows how sensitive the linear models are to large prostates. [16], [13] and [10] already noticed that correlation coefficients were different if small or large prostates were categorised. The ellipsoidal model estimate is clearly more stable across all the leave-one-out tests.

5. Conclusion

Using ground truth from manually segmented MR images of the prostates the precision of direct volume estimates given by several approximation formulae was evaluated. This has been achieved with simulations of NPUS measurements covering a range of reasonable combinations of capture planes and estimates of the prostate dimension in each of these.

Despite the relatively small number of subjects, our observations are consistent with previously published results, such as the high variability of replicates [3], the sensitivity of the direct volume estimations to large prostates [16,13,22] and the good accuracy and practicality of the $AP.LL^2$ estimates [10].

We have measured the parameters of a regression formula that generates an estimate of prostate volume, using an ellipsoidal approximation, that is more accurate than simply using the direct ellipsoidal calculation. This estimate requires the measurement of prostate diameters in two orthogonal planes. Using a single plane, a volume estimate of good accuracy can be also obtained from the regression model for the spheroidal model based on $AP.LL^2$. The slope of this model is less than unity, indicating that

an estimate based only on the direct spheroidal formula would tend to overestimate the volumes of large prostates.

We have demonstrated the use of a simulation study for interrogating sources of variability in estimating prostate volumes. This variability arises from factors such as image planes not intersecting at the centre or not being accurately perpendicular, operator proficiency in finding the largest diameters, oddly shaped prostates, etc.. This has resulted in an understanding of the sources of variability, and hence has enabled us to identify the most appropriate measurement strategy.

The approach could also be extended to optimise volume estimation in other organs such as the liver by studying the 3D shape.

References

[1] Aarnink R, De La Rosette J, Debruyne F, Wijkstra H. Reproducibility of prostate volume measurements from transrectal ultrasonography by an automated and a manual technique. Br J Urol 1996;78:219–23.

[2] Allen PD, Williamson D, Graham J, Hutchinson C. Differential segmentation of the prostate in MR images using combined 3D shape modelling and voxel classification. In: Kovacevic J, Meijering E, editors. IEEE international symposium on biomedical imaging. Arlington; 2006 April. p. 410–3.

[3] Bazinet M, Karakiewicz PI, Aprikian AG, Trudel C, Péloquin F, Dessureault J, et al. Reassessment of nonplanimetric transrectal ultrasound prostate volume estimates. Urology 1996 June;47:857–62.

[4] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986:307–10.

[5] Bland M. An introduction to medical statistics. Oxford: Oxford University Press; 2000.

[6] Chen ME, Troncoso P, Johnston D, Tang K, Babaian RJ. Prostate cancer detection: relationship to prostate size. Urology 1999;53:764–8.

[7] Cootes T, Taylor C, Cooper D, Graham J. Active shape models-their training and application. Comput Vis Image Understanding 1995;61:28–59.

[8] Dähnert WF. Determination of prostate volume with transrectal US for cancer screening. Radiology 1992:625–6.

[9] Davies R, Twining C, Cootes T, Waterton J, Taylor C. A minimum description length approach to statistical shape modeling. IEEE Trans Med Imaging 2002;21:525–37.

[10] Eri L, Thomassen H, Brennhovd B, Håheim L. Accuracy and repeatability of prostate volume measurements by transrectal ultrasound. Prostate Cancer Prostatic Dis 2002;5:273–8.

[11] Jeong CW, Park HK, Hong SK, Byun S-S, Lee HJ, Lee SE. Comparison of prostate volume measured by transrectal ultrasonography and MRI with the actual prostate volume measured after radical prostatectomy. Urol Int 2008;81:179–85.

[12] Kim SH, Kim SH. Correlation between the various methods of estimating prostate volume: transabdominal, transrectal, and three-dimensional US. Kor J Radiol 2008;9:134–9.

[13] Lee JS, Chung BH. Transrectal ultrasound versus magnetic resonance imaging in the estimation of prostate volume as compared with radical prostatectomy specimens. Urol Int 2007;78:323–7.

[14] Li X, Wang W, Martin RR, Bowyer A. Using low-discrepancy sequences and the Crofton formula to compute surface areas of geometric models. Computer-aided Des 2003;35:771–82.

[15] Loeb S, Han M, Roehl KA, Antenor JAV, Catalona WJ. Accuracy of prostate weight estimation by digital rectal examinantion versus transrectal ultrasonography. J Urol 2005;173:63–5.

[16] Matthews GJ, Motta J, Fracchia JA. The accuracy of transrectal ultrasound prostate volume estimation: clinical correlations. J Clin Ultrasound 1996;24:501–5.

[17] Nathan M, Seenivasagam K, Mei Q, Wickham J, Miller R. Transrectal ultrasonography: why are estimates of prostate volume and dimension so innacurate? Br J Urol 1996;77:401–7.

[18] Niederreiter H. Quasi-Monte Carlo methods and pseudo-random numbers. Bull Am Math Soc 1978;84:957–1041.

[19] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes: the art of scientific computing. 3rd ed. Cambridge Press; 2007. p. 780–5. Ch. Modeling of data.

[20] Quenouille MH. Note on bias in estimation. Biometrika 1956;43:353–60.

[21] Rahmouni A, Yang A, Tempany CMC, Frenkel T, Epstein J, Walsh, et al. Accuracy of in-vivo assessment of prostatic volume by MRI and transrectal ultrasonography. J Comput Assist Tomogr 1992;16(6):935–40.

[22] Rodriguez Jr E, Skarecky D, Narula N, Ahlering TE. Prostate volume estimation using the ellipsoid formula consistently underestimates actual gland size. J Urol 2008;179:501–3.

[23] Saff EB, Kuijlaars ABJ. Distributing many points on a sphere. Math Intell 1997;19:5–11.

[24] Sloane N, Hardin R, Smith W. Tables of spherical codes. Tech. rep. AT&T, Shannon Lab.; 1994, www.research.att.com/njas/.

[25] Terris MK, Stamey TA. Determination of prostate volume by transrectal ultrasound. J Urology 1991;145:984–7.

[26] Tong S, Cardinal HN, McLoughlin RF, Downey DB, Fenster A. Intra- and interobserver variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging. Ultrasound Med Biol 1998;24(5):673–81.

[27] Weiss BE, Wein AJ, Malkowicz SB, Guzzo TJ. Comparison of prostate volume measured by transrectal ultrasound and magnetic resonance imaging: is transrectal ultrasound suitable to determine which patients should undergo active surveillance? Urol Oncol Semin Orig Invest 2012;(1078–1439), http://dx.doi.org/10.1016/j.urolonc.2012.03.002 http://www.sciencedirect.com/science/article/pii/S107814391200083X

[28] Wolff JM, Boeckmann W, Mattelaer P, Handt S, Adam G, Jakse G. Determination of prostate gland volume by transrectal ultrasound: correlation with radical prostatectomy specimens. Eur Urol 1995;28:10–2.