Automated Computer-derived Prostate Volumes from MR Imaging Data: Comparison with Radiologist-derived MR Imaging and Pathologic Specimen Volumes

Julie C. Bulman, BA
Robert Toth, MS
Amish D. Patel, MD
B. Nicolas Bloch, MD
Colm J. McMahon, MB
Long Ngo, PhD
Anant Madabhushi, PhD
Neil M. Rofsky, MD

Purpose:
To compare prostate gland volume (PV) estimation of automated computer-generated multifeature active shape models (MFAs) performed with 3-T magnetic resonance (MR) imaging with that of other methods of PV assessment, with pathologic specimens as the reference standard.

Materials and Methods:
All subjects provided written informed consent for this HIPAA-compliant and institutional review board–approved study. Freshly weighed prostatectomy specimens from 91 patients (mean age, 59 years; range, 42–84 years) served as the reference standard. PVs were manually calculated by two independent readers from MR images by using the standard ellipsoid formula. Planimetry PV was calculated from gland areas generated by two independent investigators by using manually drawn regions of interest. Computer-automated assessment of PV with an MFA was determined by the aggregate computer-calculated prostate area over the range of axial T2-weighted prostate MR images. Linear regression, linear mixed-effects models, concordance correlation coefficients, and Bland-Altman limits of agreement were used to compare volume estimation methods.

Results:
MFA-derived PVs had the best correlation with pathologic specimen PVs (slope, 0.888). Planimetry derived volumes produced slopes of 0.864 and 0.804 for two independent readers when compared with specimen PVs. Ellipsoid formula–derived PVs had slopes closest to one when compared with planimetry PVs. Manual MR imaging and MFA PV estimates had high concordance correlation coefficients with pathologic specimens.

Conclusion:
MFAs with axial T2-weighted MR imaging provided an automated and efficient tool with which to assess PV. Both MFAs and MR imaging planimetry require adjustments for optimized PV accuracy when compared with prostatectomy specimens.

©RSNA, 2012
A determination of prostate gland volume (PV) facilitates an assessment of prostate disorders and, for prostate cancer, in conjunction with other parameters, can help predict the pathologic stage of disease, offer insights into the prognosis, and help predict treatment response (1,2). It can facilitate therapeutic planning for brachytherapy, cryotherapy, and minimally invasive benign prostatic hyperplasia therapy (3–7). Prostate-specific antigen levels have been modified to derive the prostate-specific antigen density by incorporating PV calculations to help guide clinical decisions (8,9). The clinical value of prostate-specific antigen density, however, is dependent on the quality of the PV estimate. The accuracy and variability of PV determinations pose limitations to its usefulness in clinical practice (5,10,11).

There are a variety of methods for PV estimation in terms of imaging modality and the strategy used to determine the volume (8,9,12–14). It is generally accepted that step-section planimetry yields better results compared with formula calculations and that magnetic resonance (MR) imaging can yield results that are better than those of transrectal ultrasonography (US) and computed tomography (CT) (4).

Planimetry with MR imaging has been shown to be the most accurate method with which to determine PV to date, but it is often too time consuming for efficient clinical practice; thus, alternative formula-derived methods are commonly used (13,15–17). While the ellipsoid formula is used as a practical clinical protocol for time-efficient PV estimation, it can yield substantial inaccuracies with two-dimensional transrectal US (11,18). Other formulas used to calculate prostate gland size are available, but to our knowledge there is no universally accepted approach that yields accurate results for the diverse sizes and shapes of prostate glands encountered in clinical practice or that helps to account for the influence of operator experience on volume determination (19).

In our study, we explored the use of fully automated multifeature active shape models (MFAs) for PV measurement with the goal of identifying a reliable, reproducible, and efficient tool for clinical practice. This MFA is based on a traditional segmentation system that uses active shape models. Prostate active shape models learn the shape of the prostate given a set of training studies and learn to define the prostate border by using the adjacent signal intensities. In addition, our MFA uses multiple statistical texture features to drive the appearance model, instead of simply using signal intensities, in an attempt to yield more accurate volume estimations. Active shape models have also been used to detect prostate cancer by using MR spectroscopy and T2-weighted ex vivo prostate MR imaging (20–23). The purpose of our study was, therefore, to compare MFA volume estimations to ellipsoid formula-derived and planimetry derived volumes, with pathologic specimens as the reference standard.

## Reference Standard
Ninety-one prostatectomy specimens (prostate gland with attached seminal vesicles), which were removed during radical prostatectomy, were weighed by a pathologist when fresh. As demonstrated by Rodriguez et al (24), there

### Materials and Methods
This retrospective study was an arm of a prospective study that was approved by the institutional review board of Beth Israel Deaconess Medical Center and was compliant with the Health Insurance Portability and Accountability Act. All subjects provided written informed consent.

From August 2007 to May 2009, 96 consecutive subjects with whole-mounted specimens from radical prostatectomy who had undergone preoperative pelvic MR imaging were included in our study. Two subjects were excluded for lack of access to MR images. One subject was excluded because MR imaging was performed with a 1.5-T magnet. One subject was excluded for lack of endorectal coil use. One subject was excluded because pathologic specimen weight was unavailable. A total of 91 patients (mean age, 59 years; age range, 42–84 years) were included in this data set (Fig 1).

### Advantages of Knowledge
- Multifeature active shape models (MFAs) offer a computerized and automated alternative with which to determine in vivo prostate gland volumes (PVs).
- MFA distribution of PV estimates is similar to that of planimetry, has a high concordance correlation coefficient with planimetry, and thus, may be an acceptable method with which to determine PV.

### Implication for Patient Care
- Consistently and accurately determined PVs obtained by using an MFA may facilitate appropriate clinical decisions that are based on PV determination.
is a 0.997 correlation between prostate weight and displaced water volume in milliliters; Varma and Morgan (25) found a similar agreement. Therefore, the specimen weight is used as the true volume of the prostate gland. Rodriguez et al (24) also showed 3.8 g to be the average weight of seminal vesicles, and this number was subtracted from specimen weight to compensate for the attached seminal vesicles. This calculated weight multiplied by 1.05 g/mL (specific gravity of prostatic tissue) served as the reference standard for our study (26). MR Imaging Technique Ninety-one prostates were analyzed with MR images acquired with a 3-T whole-body imager (GE Healthcare Technologies, Waukesha, Wis) and an endorectal coil (Medrad, Pittsburgh, Pa) inflated with 60 mL of a 100% weight per volume barium sulfate suspension for improved spatial resolution (27–29). T2-weighted MR images were acquired in the axial, coronal, and sagittal planes (repetition time msec/echo time msec, 3300–6250/155–165; echo train length, 20–21; four signals acquired; field of view, 16 × 16 cm [coronal and sagittal] or 14 × 14 cm [axial]; section thickness, 2.2–3.0 mm with no gap). The section thickness varied according to the size of the prostate gland to maintain a consistent imaging time and in-plane resolution.

PV Estimation with Planimetry To establish a surrogate in vivo ground truth, planimetry was performed by two independent readers (reader 3: A.D.P., with less than 1 year of experience; reader 4: C.J.M., with more than 6 years of experience) who did not have prior exposure to the cases. T2-weighted axial MR images of the prostate were analyzed by using a workstation (Advantage 4.0; GE Medical Systems). The prostatic capsule was manually traced by drawing a region of interest (ROI) with the cursor on each two-dimensional MR image of the series, which took 5–10 minutes per case. ROIs were initially drawn around the midgland and were then drawn section by section up to the base and down to the apex. Upon completing the ROI tracing, the software displays the area for each section in square centimeters. The volume was estimated by multiplying the sum of these areas by the section thickness in centimeters.

PV Estimation with MFA Computer automated assessment of PV was determined by using noncommercial software that analyzed the aggregate computer-calculated prostate area over the range of the 91 test set axial T2-weighted MR imaging cases. This volume calculation is a single-click automated process that requires a maximum of 200 seconds per prostate gland. An extension to traditional active shape models (30) was implemented, which used statistical texture features (mean, variance, gradient magnitude) instead of the signal intensities of the image to determine the location of the prostate, which was called an MFA. The method was similar to that used by Toth et al (31), except it was performed in three dimensions. A single click in the center of the prostate on a single midgland T2-weighted axial section was used to initialize the MFA for propagation through the entire data set, and the locations that were determined to lie on the surface of the prostate were detected from the texture features. The number of voxels enclosed by this surface was thus used to estimate the volume of the prostate (Fig 2). Thirty-one cases obtained with 3-T T2-weighted endorectal coil MR imaging (mean PV, 52.1 mL; range, 25–112 mL), which were distinct from the test set, were used to train the MFA once in an offline setting. The MFA was trained by placing a series of equally spaced landmarks on the surface of the organ. Principal component analysis was then used to capture shape variations, such that the segmentation can be constrained to only valid prostate shapes (prostate shapes similar to the training set). To train the appearance of the prostate border, a Gaussian distribution of the texture features surrounding each landmark was created. Hence, each landmark has a
distinctive appearance model, allowing for variations of appearance at different locations of the prostate surface. To segment the prostate in a new image, the same set of textural features is extracted, and the locations with the highest probability of lying on the surface are detected. However, the locations for which the textural features yield the highest probability might not all be accurate, or the resulting surface might be jagged. The trained statistical shape model is then used to constrain and smooth the segmentation to only valid prostate shapes; thus, the prostate is segmented. This method was previously used for successful volume estimations of the prostate by Toth et al (32), where the MFA-derived volumes were compared with manually estimated volumes obtained by planimetric estimation of the prostate capsule derived from manual segmentations by an expert radiologist on T2-weighted MR images.

Statistical Analysis
The Wilcoxon signed rank test was used to compare the mean PV between any two methods. There were 15 pairwise comparisons of interest (Table 1); therefore, type I error was adjusted by using the Bonferroni multiple comparison adjustment (α level of .05 divided by 15 comparisons, yielding an adjusted α of .003). The P values were compared with multiple comparison–adjusted type I error. To assess concordance between two methods, linear regression was used to obtain the individual slope and its 95% CI. To test for the difference between any two of these slopes (there were 10 comparisons of interest of five slopes [Table 1]), linear mixed-effects models (33) with linear contrasts with compound symmetry structure for the variance-covariance were used. In addition, the concordance correlation coefficient and its 95% CI were computed. The percentage of measurements whose between-methods differences were within the limits of agreement from the Bland-Altman plot was also reported. All analyses were performed with SAS software (version 8; SAS Institute, Cary, NC).

Results

Distribution of PVs
Table 2 shows the mean ± standard deviation, median, and range for PV with each of the six methods. The distributions show asymmetry, with the medians consistently smaller than the means, indicating slight skewness of the right tail. The mean and median PVs with the four MR imaging methods and MFA are consistently smaller than those with the pathologic reference standard. This indicates a potential underestimation of the PV with the MR imaging and MFA methods. The variability in the distribution of these five methods is similar to that of the pathologic reference standard. The P values from Table 1 for comparison of these five methods with the pathologic reference standard indicate that the underestimation is indeed significant (15.8%, P = .0001). The first column in Table 1 also shows the mean amount and variability of underestimation.

PV Estimates with the Ellipsoid Formula
The ellipsoid formula as performed by reader 1 was the least correlated with the pathologic reference standard (slope, 0.805; 95% CI: 0.707, 0.903). The ellipsoid formula as performed by reader 2 had a slope of 0.864 (95% CI: 0.786, 0.942) when compared with pathologic reference standard. The ellipsoid formula as performed by reader 1 had slopes of 0.910 and 0.987 when compared with planimetry as performed by readers 3 and 4, respectively. When compared with each other, calculations by readers 1 and 2 had a slope of 0.897 (95% CI: 0.810, 0.983). The underestimation of the PV as compared with the pathologic reference standard was similar for readers 1 and 2 (−9.56 vs −7.38). Reader 2 also had a higher concordance correlation coefficient (0.837 vs 0.779) and percentage of measurements within the limits of agreement (95.6% vs 92.3%).

PV Estimates with Planimetry
This analysis was performed to assess whether planimetry was an appropriate surrogate ground truth for PV obtained with the pathologic reference standard. The slope of the regression model in comparison to the pathologic specimens was 0.864 for reader 3 and 0.804 for reader 4. When the two planimetry data sets were compared, the slope of the line was 1.074 (95% CI: 1.046, 1.101). Reader 4 had a higher concordance correlation coefficient (0.897 vs 0.843) and a similar percentage of measurements within the Bland-Altman limits of agreement (95.6% vs 94.5%) compared with reader 3. The mean amount of underestimation of the pathologic PV was smaller in reader 4 than in reader 3 (−5.57 vs −9.26).

PV Estimates with MFA
Pathologically determined PV was underestimated by a mean −7.76 ± 7.69
with MFA. Volume estimates with MFA yielded the slope closest to one when compared with the pathologic reference standard (slope, 0.888; [95% CI: 0.800, 0.976]; concordance correlation coefficient, 0.867) (Fig 3). MFA PVs were compared with planimetry PVs; the results were not significant for readers 3 or 4 ($P = .265$ and $.027$, respectively) with the adjusted type I error level of $.004$. The concordance correlation coefficient between MFA and the pathologic reference standard was .867, with 93.4% of the measured differences within the Bland-Altman limits of agreement.

### Discussion

The computer segmentation method used in our study, MFA, brings automation and accuracy to volume determinations.

Manual placement of ROIs and cursors is inherently user dependent and, thus, introduces variability into manually generated volume determinations. The expected improved consistency from automated volume determinations offers the potential to improve clinical usefulness of PVs.

MR imaging yields strong soft-tissue contrast and a more accurate assessment...
of PV than does transrectal US or CT (12,13,34). MR imaging reduces inra-
observer inconsistency by a factor of 3.5, at a minimum, compared with CT (35), while avoiding exposure to ionizing radiation.

In our study, the best estimates of PV were obtained with MFA and plani-
metry as performed by reader 4 when imaging-based determinations were com-
pared with volumes derived from pathology specimens. Furthermore, the MFA-
and planimetry-determined volumes showed substantial overlap in slopes and CIs, suggesting MFA as an appropriate surrogate for planimetry. In addi-
tion, MFA eliminates the time required to draw ROIs on the numerous sections (approximately 15–20 sections) for each case. The active shape model takes approximately 180 seconds to segment each three-dimensional study, with a maximum allowed time of 200 seconds. Thus, the MFA algorithm may be a preferable clinical tool on the basis of both accuracy and efficiency.

Reader 1 was a group of seven radiologists who prospectively and inde-
dependently interpreted prostate MR imaging during their clinical interpretation sessions. This category included measurements generated by residents in training with limited experience in reading prostate MR images, especially with respect to delineating the apex and base of the prostate for measuring ellipsoid diameters. Reader 2 was an experienced reader who retrospectively reviewed all 91 cases for gland dimensions. Though readers 1 and 2 showed strong agreement with one another (slope, 0.897; 95% CI: 0.810, 0.983), the differing slopes suggest that ellipsoid formula–derived volume measurements have a user dependence. It is also possible that there is dependency on the interpretive cir-
cumstance, since reader 1 PVs were generated during clinical interpretations, whereas reader 2 PVs were generated in a dedicated research mode, isolated from the distractions encountered in clinical practice. It is interesting that the ellipsoid formula PVs showed a stronger relationship to planimetry PVs than to pathologically determined PVs, whereas MFA PVs showed a stronger relationship to pathologically determined PVs. When the ellipsoid formula–derived PVs were compared with the in vivo ground truth of planimetry PVs, the regression line slopes were closer to one, indicating a stronger relationship with in vivo images than with pathology specimens.

It is encouraging that all calculated volume strategies performed reasonably well. There seem to be reasonable choices available to the radiologist when con-
sidering volume determinations. Since the reference standard for comparison is uncertain (32), it remains a choice whether to prefer a method that aligns better with pathologic or in vivo plani-
metry findings.

The correlation between results speaks to the overall precision of volume es-
timation methods. However, the lower bound slope of the best-performing vol-
ume estimation method (MFA) versus the pathologic reference standard can be as low as 0.800, indicating up to a 20% underestimation of PV. All methods showed a similar and consistent under-
estimation of PV when compared with the pathologic reference standard. The clinical implications of this underesti-
matation and inaccuracy should be fur-
ther investigated. On the other hand, it may be useful to use the equations determined in our study on a prospec-
tive series to see whether the attendant adjustments yield improved accuracy.

We performed our imaging compar-
isons as part of our routine approach to prostate MR imaging with an endorectal coil. We recognize that there may be volume changes that occur in vivo in MR imaging with an endorectal coil; how-
ever, there are also potential volume changes in the ex vivo specimens that occur prior to weight measurement. Both of these may be implicated in the in-
creased strength of association between MR imaging ellipsoid formula–generated PVs and planimetry-generated PVs.

A surrogate in vivo ground truth, planimetry, was used in our study to ap-
proximate changes that may occur ex vivo and for comparison with the vol-
ume estimates derived from MR images with an endorectal coil. Heijmink et al (36) showed a significant difference in PV (mean 18% decrease) when MR images were acquired with an endorec-
tal coil compared with those acquired with a body-array coil. In that study, the anteroposterior dimension was, on average, reduced by 15.7% with an endorectal coil. Such shape and volume changes may be influenced by water loss or physical vasoconstriction from the pressure of the inflated endorectal bal-
loon (36). Our results show an average underestimation across the five volume measurements of 15.8% when com-
pared with prostatectomy specimens,
which is consistent with the findings of Heijmink et al. As a point in favor of imaging with an endorectal coil, the signal-to-noise benefit enables higher spatial resolution for a given imaging time and, therefore, may yield a more accurate delineation of gland borders. Furthermore, most studies indicate that the use of an endorectal coil has diagnostic advantages (27,28).

The small differences in MR imaging planimetry results between readers 3 and 4 may be related to experience in determining the prostate gland borders. The results of reader 3 likely benefited from the fact that reader 4 designated the inferior-most and superior-most sections for reader 3. In general, inaccuracies in planimetry may be related to volume averaging and difficulty in delineating prostate borders, especially at the apex.

The segmentation algorithm used in our study is fully automated and efficient with use of MFAs. To optimize precision in volume determinations, the MFA technology must be trained. An important aspect of our study was that training was a one-time process, as opposed to other studies in which training is an iterative process done multiple times. In our data, we trained the MFA on a series of T2-weighted MR prostate images, but it should be noted that this method is generalizable to other organs or modalities, provided that a sufficiently representative training population is available for generating accurate statistical shape and appearance models for the MFA. A clear application would be in the kidney, where distinct boundaries and image contrast exist between the retroperitoneal fat and the soft tissue of the renal parenchyma.

In our study, the algorithm used to generate test set volumes was trained on 31 images that were not included in the test set. With smaller training set sizes (like our training set), it is critically important to include a sample that employs a wide range of prostate shapes and sizes (mean PV, 52.1 mL; range, 25–112 mL). While improved performance of the algorithm is expected to result from employing a larger training set, it would be helpful to investigate the effects of training set content and size on algorithm accuracy.

We subtracted 3.8 g from each specimen weight to adjust for seminal vesicle contributions to PV. This produced a shift in data and did not affect the correlation. Despite the strong correlation between specimen weight and true volume, seminal vesicle size can vary substantially between patients, resulting in a small source of error in our study. We recognize concerns about the use of prostatectomy specimens as the reference standard considering the potential ex vivo blood loss and the inclusion of periprostatic tissue (37). Future prospective studies with more precise pathologic analysis (removal of seminal vesicles and periprostatic tissue, immediate weight measurement) may yield additional information on the accuracy of in vivo volume estimates with MFA and MR imaging.

Additionally, it would be useful to investigate a larger data set that included a greater number of glands, both smaller and larger, than in our series (smaller than 30 mL and larger than 60 mL) to assess the accuracy of MFA-derived PVs specifically in these outliers, where other methods have frequently fallen short.

We note that no significant differences between methods were observed; therefore, it may require larger sample sizes to reveal such differences. However, the trend points toward MFA as having the closest value to one compared with the pathologic reference standard; therefore, it seems reasonable that MFA can perform at least as well as other methods and should be considered on the basis of its expected practicality in clinical practice. While our implementation of this MFA is not available commercially, our study points toward the feasibility of an automated approach for in vivo PV determinations.

In conclusion, PV estimates with MFAs on axial T2-weighted MR images yield strong approximations of prostatectomy specimen–determined PVs and can serve as a surrogate for MR imaging planimetry determined PVs, offering the prospect for accurate volume determinations in clinical practice.

Disclosures of Potential Conflicts of Interest:
J.C.B. No potential conflicts of interest to disclose. A.D.P. No potential conflicts of interest to disclose. B.N.R. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: is a consultant for Medrad and has provided expert testimony for Medrad. Other relationships: none to disclose. C.I.M. No potential conflicts of interest to disclose. L.N. No potential conflicts of interest to disclose. A.M. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: holds various patents; is the president and co-founder of Bibris; is the majority stock holder in Bibris and Vascuris. Other relationships: none to disclose. N.M.R. No potential conflicts of interest to disclose.

References


