



HAL
open science

Prostate volume prediction on MRI: tools, accuracy and variability

Dimitri Hamzaoui, Sarah Montagne, Benjamin Granger, Alexandre Allera, Malek Ezziane, Anna Luzurier, Raphaëlle Quint, Mehdi Kalai, Nicholas Ayache, Hervé Delingette, et al.

► To cite this version:

Dimitri Hamzaoui, Sarah Montagne, Benjamin Granger, Alexandre Allera, Malek Ezziane, et al.. Prostate volume prediction on MRI: tools, accuracy and variability. *European Radiology*, 2022, 10.1007/s00330-022-08554-4 . hal-03409262

HAL Id: hal-03409262

<https://hal.science/hal-03409262v1>

Submitted on 4 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prostate volume prediction on MRI: tools, accuracy and variability

Dimitri Hamzaoui, MSc¹, Sarah Montagne, MD^{2,3,4}, Benjamin Granger, MD^{5,6}, Alexandre Allera, MD², Malek Ezziane, MD², Anna Luzurier, MD², Raphaëlle Quint², Mehdi Kalai², Nicholas Ayache, PhD¹, Hervé Delingette, PhD¹, Raphaële Renard Penna, MD, PhD^{2,3,4}

¹ Université Côte d'Azur, Inria, Epione Team, Sophia Antipolis.

² Academic Department of Radiology, Hôpital Pitié-Salpêtrière, Assistance Publique des Hôpitaux de Paris, Paris, France

³ Academic Department of Radiology, Hôpital Tenon, Assistance Publique des Hôpitaux de Paris, Paris, France

⁴ Sorbonne Universités, GRC n° 5, Oncotype-Uro, Paris, France

⁵ Department of Public Health, Pitié-Salpêtrière Academic Hospital, AP-HP, Sorbonne Universités, AP-HP, CIC-P 1421, Paris, France

⁶ Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, UMR 1136, CIC-1421, Hôpital Pitié-Salpêtrière, AP-HP, Paris, France

Dimitri Hamzaoui and Sarah Montagne contributed equally to this work.

Corresponding author: Dr Sarah Montagne, Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Academic Department of Radiology, Hôpital Pitié-Salpêtrière 47 bd de l'Hôpital 75013 Paris, France.

Tel: 0142176325; Email: sarah.montagne@aphp.fr

Prostate volume prediction on MRI: tools, accuracy and variability.

1
2 249 words
3

4 **Objective**

5
6 A reliable estimation of prostate volume (PV) is essential to prostate cancer management. The
7
8 objective of our multi-rater study was to compare intra and inter-rater variability of PV from
9
10 manual planimetry and ellipsoid formulas.
11

12 **Methods**

13
14 Forty treatment-naive patients who underwent prostate MRI were selected from a local
15
16 database. PV and corresponding PSA density (PSAd) were estimated on 3D T2-weighted MRI
17
18 (3T) by 7 independent radiologists using the traditional ellipsoid formula (TEF), the newer
19
20 biproximate ellipsoid formula (BPEF), and the manual planimetry method (MPM) used as
21
22 ground truth. Intra and inter-rater variability was calculated using the mixed model based
23
24 intraclass correlation coefficient (ICC).
25
26
27

28 **Results**

29
30 Mean volumes were 67.00 (± 36.61), 66.07 (± 35.03), and 64.77 (± 38.27) cm³ with the TEF,
31
32 BPEF, and MPM methods respectively. Both TEF and BPEF overestimated PV relative to
33
34 MPM, with the former presenting significant differences (+1.91 cm³, IQ=[-0.33 cm³, 5.07 cm³],
35
36 p-val=0.03).
37

38
39 Both intra (ICC>0.90) and inter-rater (ICC>0.90) reproducibility were excellent. MPM had the
40
41 highest inter-rater reproducibility (ICC=0.999). Inter-rater PV variation led to discrepancies in
42
43 classification according to the clinical criterion of PSAd>0.15 ng/mL for 2 patients (5%), 7
44
45 patients (17.5%), and 9 patients (22.5%) when using MPM, TEF, and BPEF respectively.
46
47

48 **Conclusion:**

49
50 PV measurements using ellipsoid formulas and MPM are highly reproducible. MPM is a robust
51
52 method for PV assessment and PSAd calculation, with the lowest variability. TEF showed a
53
54 high degree of concordance with MPM but a slight overestimation of PV. Precise anatomic
55
56 landmarks as defined with the BPEF led to a more accurate PV estimation, but also to a higher
57
58 variability.
59
60
61

Key Words:

Prostate, Magnetic Resonance Imaging, Volume, PSA, Segmentation

Key points:

Manual planimetry used for prostate volume estimation is robust, reproducible, with the lowest variability between readers.

Ellipsoid formulas are accurate, reproducible but with higher variability between readers.

Traditional ellipsoid formula tends to overestimate prostate volume.

Abbreviations

AP: Antero-Posterior

BPEF: Biproximate Ellipsoid Formula

CsPCa: Clinically Significant Prostate Cancer

DRE: Digital rectal exam

ICC: Intraclass Correlation Coefficient

IQ: Interquartile

MPM: Manual Planimetry Measurement

PCa: Prostate cancer

PSA: Prostate-specific antigen

PSAd: Prostate-specific antigen density (mpmPSAd, tefPSAd, and bpefPSAd are PSAd obtained using a volume estimated respectively by MPM, TEF and BPEF methods)

PV: Prostate volume

rSTD: Relative standard deviation

T: Tesla

T2W: T2-weighted

TEF: Traditional Ellipsoid Formula

TRUS: Transrectal Ultrasound

4 **Introduction:**

5
6 Accurate prostate volume (PV) measurement is necessary for the management of benign
7 prostate hypertrophy, prostate cancer (PCa) diagnosis and treatment planning. PSA density
8 (PSAd), one of the strongest predictors of clinically significant prostate cancer (CsPCa) in risk
9 models [1–3], is obtained by dividing the prostate specific antigen (PSA) level by PV. Hence,
10 it is highly dependent on accurate PV measurement.

11
12 PV can be estimated by digital rectal exam (DRE), transrectal ultrasound (TRUS) or MRI. MRI,
13 with a higher spatial resolution and a better soft tissue contrast, is superior to TRUS for
14 selecting outer boundaries and provides a more accurate and more reproducible volume
15 estimation [4-6]. The PI-RADS V2.1 stipulates that PV should always be reported on MRI and
16 should be determined either using manual or automated segmentation, or calculated using the
17 formula for a conventional prolate ellipse [7].

18
19 Manual segmentation is considered to give the closest volume estimation to pathologic
20 specimen volume [8–10]; however this approach is highly time consuming (up to 30 minutes)
21 [11]. Auto-segmentation software [12, 13] are promising tools, providing a quick and accurate
22 whole gland segmentation. However, because of technical factors and prostate shape
23 variability, PV calculation is still a challenging task for algorithms, which frequently require
24 significant segmentation adjustments by an expert reader.

25
26 The traditional ellipsoid formula (TEF) is very easy and quick, for clinical situations (only a few
27 minutes). However, it relies on geometric models that “approximate” the prostatic contour by
28 considering the prostate as a regular ellipse-like shape, whereas in reality it is usually irregular,
29 and often has an eccentrically enlarged median lobe. To enhance measurement consistency
30 and reduce intra and inter-rater variability in PV approximation, Wasserman et al. [14] recently
31 proposed a new ellipsoid formula, called the biproximate method (BPEF) based on very well
32 defined anatomic landmarks. BPEF is based on very well defined anatomic landmarks, and
33 includes measurement of intravesical prostatic protrusion (IPP), to locate prostate boundaries

with more precision.

Few studies have examined the precision and accuracy of ellipsoid and planimetry volumetric measurements. None of them evaluated the recently published BPEF [14], and most of them used one single reader segmentation as ground truth [15–19].

The purpose of our study was to evaluate intra and inter-rater variability in PV estimation, when using manual planimetry measurement (MPM), TEF and BPEF, on T2-weighted (T2W) images with a 3T MRI without endorectal coil, and with multiple raters of various levels of experience.

Dataset

This work was supported by the Health Data Center of the AP-HP (Assistance Publique - Hôpitaux de Paris) and was approved by our joint institutional review boards.

We compiled a cohort of forty patients randomly selected from one hundred fifty treatment-naive patients who underwent prostate MRI before the first round of biopsy (fulfilling the inclusion criterion for clinical indication of prostate MRI for PCa suspicion: elevated PSA, positive DRE, genetic susceptibility), between October 2013 and July 2019.

The one hundred fifty patients were divided in two groups for another work in progress: one (n=110) being a training set and the other (n=40) a test set. We used the latter for the current study, consisting of forty randomly selected patients whose ages and PV were representative of the larger dataset.

MRI protocol

MRI exams were performed using 3 Tesla (3T) clinical systems (SIGNA™ Architect, GE Healthcare, and MAGNETOM™ Skyra, Siemens Healthcare) using a 32-channel phased-array torso coil. Patients were advised to perform bowel preparation before the exam and to empty their bladder; 1mg glucagon was administered intra muscularly to reduce peristaltic motion. All MRI protocols included 3D T2W images (Supplementary Table 1).

Prostate volume evaluation:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Seven radiologists of different levels of experience evaluated independently the forty PV on the MRI: three experts (GR1, >1000 prostate MRI interpreted), two seniors (GR2, 500 prostate MRI) and three juniors (GR3, <100 prostate MRI). A training meeting was organized before the start of the study to review anatomic boundary definitions, in order to standardize PV evaluation. We tested three evaluation methods: whole gland segmentation with computation of the extracted volume (MPM), and two estimations by ellipsoid formulas (TEF and BPEF).

1. **Manual planimetry measurement (MPM)**

First, the readers were asked to segment the whole prostate gland (using a polygon segmentation), with the open-source MedInria software [20]. This software enables the user to define prostate contours as accurately as possible while checking the result in axial, sagittal, and coronal views. Prostate boundaries were manually drawn in axial plan on some slices of the 3D T2W, and the whole volume was obtained through interpolation. The interpolated contours were then corrected when needed. The mask volumes estimations (n=280, 40x7) were computed using the toolbox Simple ITK [21,22] (<http://www.simpleitk.org>) (Supplementary Fig. 1).

MPM was considered as the ground truth [8–10, 15].

2. **Traditional ellipsoid formula (TEF)**

According to the PI-RADS V2.1 recommendations [7], the ellipsoid formula was based on: (maximum antero-posterior (AP) dimension) x (maximum longitudinal dimension) [both placed on the mid-sagittal T2W image] x (maximum transverse dimension) [placed on an axial T2W image] x 0.52 (Fig. 1A, 1B).

3. **Biproximate ellipsoid formula (BPEF)**

This method was described by Wasserman et al. [14], and is based on the same formula as TEF but with differences on axes measurements.

1 Length measurement was made on the mid-sagittal plane. Transverse and AP measurements
2 were made on the axial plane showing maximal diameter, and were drawn from the inside
3 border of external prostatic capsule (Fig. 1C, 1D).
4
5
6
7

8 **PSA density calculation**

9
10 PSA density calculation was given by the ratio between PSA serum concentration (provided at the
11 time of MRI) and MPM, TEF, and BPEF volumes, to obtain mpmPSAd, tefPSAd, and
12
13
14
15
16
17
18
19
20
21 bpefPSAd.

22 **Statistical analysis**

23 The paired Wilcoxon signed-rank test was used for paired samples comparisons. P-values
24 from multiple tests were corrected with the Holm-Bonferroni method. All statistical tests were
25 two-sided. A p-value < 0.05 after correction was considered indicative of a statistically significant
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
The paired Wilcoxon signed-rank test was used for paired samples comparisons. P-values
from multiple tests were corrected with the Holm-Bonferroni method. All statistical tests were
two-sided. A p-value < 0.05 after correction was considered indicative of a statistically significant
difference. To assess the inter-rater and the intra-rater variability we used the relative standard
deviation (rSTD), defined for an element with multiple measures X_1, X_2, \dots, X_n as $\frac{\sigma(X)}{\mu(X)}$, and
the intraclass correlation coefficient (ICC) derived from a two-way mixed, average measures,
absolute agreement model. We used the Python module Pingouin (version 0.3.10,
<https://pingouin-stats.org>) to compute those statistical tests.

Empirical statistical power was computed with the R package MKpower (version 0.5,
<https://cran.r-project.org/web/packages/MKpower/index.html>), using the mclust package [23]
to estimate distributions with a Gaussian mixture model.

We also used the R package “lme4” (version 1.1-26, <https://cran.r-project.org/web/packages/lme4/index.html>) to fit a linear mixed-effect model, considering the
raters and methods effects as fixed effects and the subject impact as a random effect. Inspired

by McGraw and al.[24] we defined ICC_{rater} as $\frac{\sigma_{\text{method}} + \sigma_{\text{Sujet}}}{\sigma_{\text{Residual}} + \sigma_{\text{method}} + \sigma_{\text{Sujet}} + \sigma_{\text{Rater}}}$ and ICC_{method} as

$\frac{\sigma_{\text{Rater}} + \sigma_{\text{Sujet}}}{\sigma_{\text{Residual}} + \sigma_{\text{method}} + \sigma_{\text{Sujet}} + \sigma_{\text{Rater}}}$ to estimate the overall impact of raters and methods on the

variability.

No statistical tests were made for experience level.

To assess the impact of PV variability on PSAd, we estimated the number of cases that would lead to a clinical disagreement between the PSAd scores computed by raters using the same PV method. Specifically, we identify cases in which there was no unanimous consensus on whether the PSAd fell above or below the classical threshold of 0.15ng/mL for CsPCa suspicion [25, 26]. We also computed specificity, sensitivity, and Area Under the Curve (AUC), taking as PV for a given patient and a given method the mean of the volumes obtained by the seven raters [27].

Results

Study population

Median age at MRI was 64 years [range 45-76 years], mean PSA level was 8.4±5.6ng/mL.

Among the forty patients, seventeen (42.5%) were classified with a PI-RADS≥3. Patients' demographic characteristics are summarized in Table 1.

Prostate volume measurements, MPM, TEF, BPEF

Mean volumes were 67.00 (±36.61), 66.07(±35.03), and 64.77(±38.27) cm³ with TEF, BPEF, and MPM, respectively. Median PV measurements for each technique and each rater are given in Table 2 and illustrated in Figure 2. Figure 3 shows detailed results for each rater and each method.

While considering the volume distribution, taking as PV for a given patient and a given method the mean of the volumes obtained by the seven raters, we observed that the median difference of calculated volumes compared to the reference (MPM) was significant for TEF with a slight overestimation of PV of 1.91cm³ (IQ=[-0.33cm³, 5.07cm³], p-val=0.03, power=0.71) but not for BPEF (1.45 cm³, IQ=[-1.07cm³, 5.63cm³], p-val=0.43, power=0.28). Detailed results for those comparisons are listed on Table 3.

1 No statistical difference was found between BPEF and TEF (median difference=-0.58 cm³,
2 IQ= [-3.32cm³, 2.56cm³], p-val=0.15, power=0.27).
3

4 **PSAd measurements**

5
6
7 The median values were 0.117 (IQ= [0.079, 0.193]), 0.127 (IQ= [0.082, 0.202]) and 0.119 (IQ=
8 [0.078, 0.197]) for tefPSAd, mpmPSAd, and bpefPSAd (Fig. 2B). As seen for PV, there was a
9 significant difference between mpmPSAd and tefPSAd (p-val=0.01).
10
11
12
13

14 **Intra-reader variability**

15 Volumes

16
17
18 For each reader, the ICC between the three PV methods was above 0.90. Detailed results are
19 presented in Table 4. No substantial differences were observed according to experience.
20
21
22
23

24 **Inter-rater variability**

25 Volumes

26
27
28 Using ICC to assess the inter-rater variability, the highest ICC was obtained by MPM
29 (ICC=0.999, CI95%= [0.997, 0.9995]), followed by TEF (ICC=0.988, CI95%= [0.978, 0.994])
30 and BPEF (ICC=0.984, CI95%= [0.968, 0.992]).
31
32
33
34
35
36
37

38 MPM's ICC is significantly higher than other methods' ICC, while rSTD for MPM was
39 significantly lower than those of ellipsoid methods (Fig. 4A, 4B). In addition, the 1st quartile,
40 median, and 3rd quartile of PV measurements are more consistent between raters using MPM
41 than with the ellipsoid methods, as illustrated in Fig. 3.
42
43
44
45
46
47

48 Similar results were obtained on PSAd (Fig. 4C, 4D).
49
50

51 **Inter-rater agreement for axes measurement with BPEF**

52
53
54 Length measurements is the main source of variability (compared to width or AP
55 measurements), with a mean rSTD and ICC of 0.13±0.04 and 0.943 for length against 0.04
56
57
58
59
60
61
62
63
64
65

1
2 ± 0.03 and 0.986 for transverse diameter and 0.08 ± 0.05 and 0.969 for AP diameter. The
3 differences on rSTD are significant when comparing all three axes (Fig. 5).
4

5 Evaluation of variability using a linear mixed-effect mode

6
7 The linear mixed-effect models comparing both ellipsoid methods with MPM returned an ICC
8 of 0.956 (CI95%: [0.923 – 0.969]) for TEF and 0.932 (CI95%: [0.888 – 0.953]) for BPEF. Once
9 again, we found a statistically significant difference between MPM and TEF ($p=0.01$), but not
10 between MPM and BPEF ($p=0.116$). ICC_{Rater} was 0.956 for TEF and 0.911 for BPEF, and
11 ICC_{Method} was 0.955 for TEF and 0.910 for BPEF.
12
13
14
15
16
17
18
19
20

21 Impact of volume methods on PSA measurement and linked diagnosis.

22
23 In addition to variability estimation using ICC and rSTD – which gives similar results to those
24 obtained with volumes, we also estimated the number of disagreements arising from PSA
25 differences. There were two patients (5%), seven patients (17.5%), and nine patients (22.5%)
26 with disagreements when using MP, TEF and BPEF, respectively.
27
28
29
30
31
32

33 At the clinical threshold of 0.15ng/mL, bpefPSAd had a sensitivity of 59% and a specificity of
34 83% against 65% and 78% for tefPSAd, and 65% and 73% for mpmPSAd. AUC from the three
35 PSA were similar (Fig. 6).
36
37
38
39
40

41 **Discussion**

42
43 In our study we evaluated the inter and intra-rater variability of PV measurement on T2W MRI
44 (3T) by seven independent readers using 3 different methods: MPM, TEF, and the recently
45 published BPEF [14].
46
47
48
49

50 Despite the difficulty of delineating the prostate gland, in particular at its extremities, we found
51 MPM to be the most reproducible method ($ICC=0.999$, $CI95\%=[0.997, 0.9995]$). Compared to
52 planimetry, we found a slight overestimation of PV with both ellipsoid formulas, significant with
53 TEF ($p\text{-val}=0.03$) but not with BPEF ($p\text{-val}=0.43$) with a median difference of 1.91cm³ and
54 1.485cm³. Empirical powers are coherent with those results. Nevertheless, supplementary
55
56
57
58
59
60
61
62
63
64
65

tests with more subjects are necessary to confirm them with a better statistical power.

Several studies have looked into the accuracy of PV estimation with ellipsoid formulas on 3T MRI with different type of PV estimation method, in particular by measuring the AP dimension in the axial vs sagittal plane. They found high levels of concordance between ellipsoid formulas and reference (manual planimetry or prostatectomy specimen), with either a slight overestimation [16, 17], or an underestimation [9, 19], probably due to variations in the measurements and image interpretation.

Sosna et al.[16] compared values from ellipsoid formula among 6 different data sets and found that the best estimate was obtained using two diameters from the sagittal plane multiplied by the right-left diameter of the axial plane as recommended in PIRADS V2.1[7].

These authors argued that measuring the AP dimension from sagittal rather than axial images could result in more precise estimate of PV since the shape of the prostate is more oval or ellipsoid in the sagittal plane. However, measurement of the AP dimension in the sagittal plane may lead to an overestimation because of the inclusion of peri-capsular veins and/or thick anterior fibromuscular stroma. This has been shown by Ghafoor et al. [17], who compared the gland volume measurement between the TEF as defined in the PIRADS V2.0 and V2.1[7, 28] (AP measurement in the sagittal vs axial plane). They found a slight but significant overestimation ($p < 0.001$) of gland volume with AP measurement in the sagittal plane by 2.6mL compared to the reference.

Turkbey et al.[18] compared the accuracy of fully automated segmentation, manual segmentation and ellipsoid volumetric measurement using post-operative prostate specimens as “ground truth”. Authors found a strong positive correlation between true PV, PV derived from the ellipsoid formula ($R=0.86-0.90$, $p<0.0001$), and manual segmentations ($R=0.89-0.91$, $p< 0.0001$). The strongest correlation was between true PV and manual segmentations.

Bezinque et al. [15] found an excellent correlation between the ellipsoid formula and MRI R3D (automatic prostate segmentation with manual adjustments by an experienced radiologist) measurement ($ICC=0.90$), showing that MRI using the ellipsoid formula provides accurate estimates of PV for most patients.

1 Wasserman et al. [14] found an excellent inter and intra-rater reliability (precision of 0.95 and
2 0.98, respectively) of their updated method (BPEF), with the length measurement being the
3 most common cause of variation between readers. Our results are concordant with an
4 excellent correlation between BPEF and reference (median difference of PV of 1.45 cm³, IQ=
5 [-1.07 cm³, 5.63cm³], p-val =0.43). However inter-rater variability was the highest with this
6 method (ICC=0.984, CI95%= [0.968, 0.992]), mainly due to length measurements with a mean
7 rSTD of 0.13±0.04 against 0.04 ±0.03 and 0.08±0.05 for width and AP, respectively. This
8 underlies the difficulty of delineating precise lower and upper landmark in the mid sagittal plane
9 while measuring length.

10 Very few studies have examined precision, accuracy and agreements of ellipsoid and
11 planimetry volumetric measurements with MRI, and most of them are limited by not taking into
12 account the level of rater experience in the analysis, and the absence of inter and intra-rater
13 variability evaluation.

14 Ghafoor et al.[17], found an excellent inter-rater agreement between four readers for TEF (ICC
15 > 0.90), however only one reader provided the reference (whole gland manual segmentation).

16 Bulman et al.[9] compared results from pathologic standard, planimetry by two different
17 readers, ellipsoid formula by two other readers, and results obtained by an automated method,
18 with overall good to excellent agreement between the different methods and readers.

19 Other works only considered one reference segmentation with no analysis on either inter or
20 intra-rater variability [15, 16, 18, 19].

21 Evaluating them and determining their sources are essential to provide accurate gland
22 measurements, in order to obtain the lowest variability impact on PSA_d calculation. In our
23 study, we found very low intra and inter-rater variability between planimetry and the two
24 ellipsoid derived formulas with an ICC>0.90 for all readers with no impact of level of
25 experience, although no statistical tests confirmed this last result. Ellipsoid methods may be a
26 timesaver for expert radiologists and allow young radiologists without experience in prostate
27 MRI to safely perform accurate prostate volumetry, an essential step in learning prostate MRI.
28 However, the MPM method was significantly more reproducible than the ellipsoid-based

1 methods, but also the most time consuming. Until fast reliable automated or manually adjusted
2 MRI software are available, ellipsoid formula methods are appropriate for routine clinical work
3 with a high degree of concordance. Although both TEF and BPEF differed from the reference,
4 overestimation was higher with TEF due to less defined anatomical boundaries, but BPEF was
5 less reproducible probably because of a new definition of these landmarks.
6
7
8
9

10 The observed high level of concordance between the measurements translates into a high
11 level of concordance for PSA_d risk classifications. However, considering 0.15ng/mL as
12 threshold, disagreement of volumetry-based PSA_d levels was lower with MPM (only 5% of
13 disagreements) compared to TEF and BPEF (17.5 % and 22.5 % disagreement). These
14 individual case errors were not apparent in statistical analysis, but highlighted the relevance of
15 accurate volume estimation for PSA_d measurements as it may affect biopsy strategy whether
16 for tumor detection or as part of active surveillance.
17
18
19
20
21
22
23
24
25

26 Our study has several limitations. First, due to the prohibitive time cost, the sample size was
27 small. However, this limit was partially offset by the number of readers (seven radiologists)
28 who each provided a planimetry for all forty MRIs of the dataset.
29
30

31 Second, we chose manual planimetry as the “ground truth” measure for total prostatic volume
32 rather than the pathological specimen, which may be considered as a limitation. However, it
33 has been shown that the mean prostate volume was significantly smaller ex vivo than in vivo
34 with an average change in volume of 19.5% because of loss of vascularity. In addition, tissue
35 shrinkage during specimen processing is one of the factors that may significantly affect the
36 accuracy of prostate volume measurement [29, 30]. MRI volume imaging eliminates these
37 variables, and it has been argued by multiple authors that MRI volumetric measurements in
38 the living patient should replace post-mortem measurement as the “gold standard” [4, 14, 31].
39
40
41
42
43
44
45
46
47
48
49
50
51 We accepted in this study that manual planimetry could be considered to have the highest
52 level of accuracy and should be considered as ground truth.
53
54
55
56
57

58 **Conclusion:**

59 Manual planimetry is a robust and reproducible method for PV measurement and PSA_d
60
61
62
63
64
65

1 calculation, with the lowest variability between readers. Volumes computed with the traditional
2 ellipsoid formula showed a high degree of agreement with those estimated by planimetry but
3 with a slight overestimation of PV. Delineation of clear anatomical boundaries as defined in the
4 biproximate ellipsoid method leads to a more accurate assessment of PV but with a slight
5 decrease in reproducibility. This underlines the importance of developing efficient and
6 reproducible automatic segmentation tools in prostate MRI in the future.
7
8
9
10
11
12
13
14

15 **Acknowledgements:**

16 We thank Julien Castelneau, Inria software engineer for his help in the development of
17 MedInria Software (MedInria - Medical image visualization and processing software by Inria
18 <https://med.inria.fr>-RRID:SCR_001462). This work has been supported by the French
19 government, through the 3IA Côte d’Azur Investments and UCA DS4H Investments in the
20 Future project managed by the National Research Agency (ANR) with the reference numbers
21 ANR-19-P3IA-0002 and ANR-17-EURE-0004. Data were extracted from the Clinical Data
22 Warehouse of the Greater Paris University Hospitals (Assistance Publique – Hôpitaux de
23 Paris).
24
25
26
27
28
29
30
31
32
33
34

35 We also thank Dr Hari Sreedhar for a thorough proofreading of this paper.
36
37
38
39

40 **Reference list:**

41 **Reference list**

- 42 1. Benson MC, Whang IS, Pantuck A, et al (1992) Prostate specific antigen density: a
43 means of distinguishing benign prostatic hypertrophy and prostate cancer. J Urol 147:815–
44 816
- 45 2. Seaman E, Whang M, Olsson CA, et al (1993) PSA density (PSAD). Role in patient
46 evaluation and management. Urol Clin North Am 20:653–663
- 47 3. Distler FA, Radtke JP, Bonekamp D, et al (2017) The Value of PSA Density in
48 Combination with PI-RADS™ for the Accuracy of Prostate Cancer Prediction. J Urol 198:575–
49 582
- 50 4. Rahmouni A, Yang A, Tempany CMC, et al (1992) Accuracy of In-Vivo Assessment of
51 Prostatic Volume by MRI and Transrectal Ultrasonography. Journal of Computer Assisted
52 Tomography 16:935–940
- 53 5. Lee JS, Chung BH (2007) Transrectal ultrasound versus magnetic resonance imaging
54
55
56
57
58
59
60
61
62
63
64
65

in the estimation of prostate volume as compared with radical prostatectomy specimens.
Urol Int 78:323–327

6. Paterson NR, Lavallée LT, Nguyen LN, et al (2016) Prostate volume estimations using magnetic resonance imaging and transrectal ultrasound compared to radical prostatectomy specimens. *Can Urol Assoc J* 10:264–268

7. Turkbey B, Rosenkrantz AB, Haider MA, et al (2019) Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. *Eur Urol* 76:340–351

8. Garvey B, Türkbeý B, Truong H, et al (2014) Clinical value of prostate segmentation and volume determination on MRI in benign prostatic hyperplasia. *Diagn Interv Radiol* 20:229–233

9. Bulman JC, Toth R, Patel AD, et al (2012) Automated computer-derived prostate volumes from MR imaging data: comparison with radiologist-derived MR imaging and pathologic specimen volumes. *Radiology* 262:144–151

10. Jeong CW, Park HK, Hong SK, et al (2008) Comparison of prostate volume measured by transrectal ultrasonography and MRI with the actual prostate volume measured after radical prostatectomy. *Urol Int* 81:179–185

11. Mahdavi SS, Chng N, Spadinger I, et al (2011) Semi-automatic segmentation for prostate interventions. *Med Image Anal* 15:226–237

12. Aldoj N, Biavati F, Michallek F, et al (2020) Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. *Scientific Reports* 10:14315

13. Meyer A, Rakr M, Schindele D, et al (2019) Towards Patient-Individual PI-Rads v2 Sector Map: Cnn for Automatic Segmentation of Prostatic Zones From T2-Weighted MRI. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, Venice, Italy, pp 696–700

14. Wasserman NF, Niendorf E, Spilseth B (2020) Measurement of Prostate Volume with MRI (A Guide for the Perplexed): Biproximate Method with Analysis of Precision and Accuracy. *Sci Rep* 10:575

15. Bezinque A, Moriarity A, Farrell C, et al (2018) Determination of Prostate Volume: A Comparison of Contemporary Methods. *Acad Radiol* 25:1582–1587

16. Sosna J, Rofsky NM, Gaston SM, et al (2003) Determinations of prostate volume at 3-Tesla using an external phased array coil: comparison to pathologic specimens. *Acad Radiol* 10:846–853

17. Ghafoor S, Becker AS, Woo S, et al (2020) Comparison of PI-RADS Versions 2.0 and 2.1 for MRI-based Calculation of the Prostate Volume. *Acad Radiol*.
<https://doi.org/10.1016/j.acra.2020.07.027>

18. Turkbey B, Fotin SV, Huang RJ, et al (2013) Fully automated prostate segmentation on MRI: comparison with manual segmentation methods and specimen volumes. *AJR Am J Roentgenol* 201:W720-729

19. Mazaheri Y, Goldman DA, Di Paolo PL, et al (2015) Comparison of prostate volume measured by endorectal coil MRI to prostate specimen volume and mass after radical prostatectomy. *Acad Radiol* 22:556–562

20. Vichot F, Cochet H, Bleuzé B, et al (2012) Cardiac Interventional Guidance using Multimodal Data Processing and Visualisation: medInria as an Interoperability Platform

21. Lowekamp BC, Chen DT, Ibáñez L, Blezek D (2013) The Design of SimpleITK. *Front Neuroinform* 7:45

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
22. Yaniv Z, Lowekamp BC, Johnson HJ, Beare R (2018) SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *J Digit Imaging* 31:290–303
 23. Scrucca L, Fop M, Murphy TB, Raftery AE (2016) mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* 8:289–317
 24. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1:30–46
 25. Mottet N, van den Bergh RCN, Briers E, et al (2021) EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* 79:243–262
 26. Rozet F, Mongiat-Artus P, Hennequin C, et al (2020) [French ccAFU guidelines - update 2020-2022: prostate cancer]. *Prog Urol* 30:S136–S251
 27. Eri LM, Thomassen H, Brennhovd B, Håheim LL (2002) Accuracy and repeatability of prostate volume measurements by transrectal ultrasound. *Prostate Cancer Prostatic Dis* 5:273–278
 28. Weinreb JC, Barentsz JO, Choyke PL, et al (2016) PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol* 69:16–40
 29. Jonmarker S, Valdman A, Lindberg A, et al (2006) Tissue shrinkage after fixation with formalin injection of prostatectomy specimens. *Virchows Arch* 449:297–301
 30. Orczyk C, Taneja SS, Rusinek H, Rosenkrantz AB (2014) Assessment of change in prostate volume and shape following surgical resection through co-registration of in-vivo MRI and fresh specimen ex-vivo MRI. *Clinical Radiology* 69:e398–e403
 31. Haas M, Günzel K, Miller K, et al (2017) Is the Ellipsoid Formula the New Standard for 3-Tesla MRI Prostate Volume Calculation without Endorectal Coil? *Urol Int* 98:49–53

33 **Tables and figures legends**

34
35
36
37

38 Fig. 1 Example of 3D T2W MRI showing manual prostate measurement: measures are made
39 in the axial plan showing the biggest prostate width (Fig. 1A, 1C) and the midsagittal plan (Fig.
40 1B, 1D). Fig. 1A, 1B show the 3 axis used to determine prostate volume by the TEF, and Fig.
41 1C, 1D the ones used for the BPEF. In Fig. 1D, the line joining the vesico-prostatic angles and
42 vesico-prostatic line are showed in green dotted line. Prostate length is calculated by summing
43 both red lines (gland length + median lobe length).
44
45
46
47
48
49
50

51 Fig. 2 Subject-wise mean prostate volumes (Fig. 2A) and PSA_d (Fig. 2B) for each method. The
52 dotted line in Fig. 1B represents the 0.15ng/mL clinical threshold.
53
54

55 Fig. 3 Volume estimations for the 7 raters and the 3 methods. Each color corresponds to one
56
57
58
59
60
61
62
63
64
65

rater.

1
2 Fig. 4 Inter-rater variability for prostate volume measurement (4A, 4B) and PSAd (4C, 4D),
3 depending on the estimation method (p -value <0.05 for all 3 distributions). Fig. 4A and 4C
4 shows relative standard deviation (rSTD); Fig. 4B and 4D shows Intraclass Correlation (ICC).

5
6
7 Fig. 5 BPEF axis measure variability. Fig. 5A shows mean measures for each axis (length in
8 brown, width in pink and antero-posterior in grey). Fig. 5B shows the rSTD variability
9 distribution for each axis. rSTD distribution are significantly different for each axis (p -
10 val <0.001). Fig. 5C shows the ICC distribution for each axis.

11
12
13 Fig. 6 ROC curves and AUC for PSAd determination when prostate volume is estimated by
14 the three methods (TEF in red, BPEF in purple, MPM in green).

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Supplementary Fig. 1 Example of a mask obtained by manual segmentation, in axial (Fig. 1A)
and sagittal (Fig. 1B) view. The grey area contours the whole prostate gland, used to calculate
prostate volume. The white one is the transitional zone, used for other studies.

Tables

Table 1 Demographic and clinical characteristics of study participants (n=40)

Table 2 Distribution of prostate volume estimations for each method and rater

Table 3 Mean difference between estimated volumes for each rater when using MPM versus
ellipsoid methods (TEF or BPEF)

Table 4 Intra-rater agreement on volume estimation (evaluated by ICC) for each rater

Supplementary Table 1 MRI acquisition specificities

Supplementary Table 2 Distribution of estimated volumes difference for each rater

Table 1 Demographic and clinical characteristics of study participants (n=40)

Variable	Value
Age (years) ^a	64 [45-76]
PSA (ng/mL) ^b	8.4 (± 5.6)
MRI Equipment	
3T SIGNA™ Architect, General Electrics	11 (27%)
3T MAGNETOM™ Skyra, Siemens Healthcare	29 (73%)
PI-RADS	
PI-RADS 1-2	23 (57.5%)
PI-RADS 3	4 (10%)
PI-RADS 4	6 (15%)
PI-RADS 5	7 (17.5%)

^aMedian [range]^bMean (±STD)

Table 2 Distribution of prostate volume estimations for each method and rater

Group		TEF ¹	BPEF ¹	MPM ¹
Group 1 (Experts)	Reader 3	49.95/62.33/104.68	57.82/69.52/97.73	36.31/55.26/85.98
	Reader 6	32.48/49.86/68.70	32.78/50.51/72.55	35.56/57.92/83.58
	Reader 7	43.75/58.5/85.75	38.32/59.41/81.20	36.85/56.46/87.50
Group 2 (Seniors)	Reader 1	39.90/64.53/81.76	35.65/59.18/77.41	30.08/51.14/79.35
	Reader 2	41.00/58.98/85.55	46.36/66.74/87.01	38.11/60.20/91.52
Group 3 (Juniors)	Reader 4	33.50/51.4/76.63	32.94/52.02/79.68	34.03/54.34/81.27
	Reader 5	35.25/61.35/87.43	36.19/52.16/69.78	36.38/56.11/84.86

¹Q1/Median/Q3

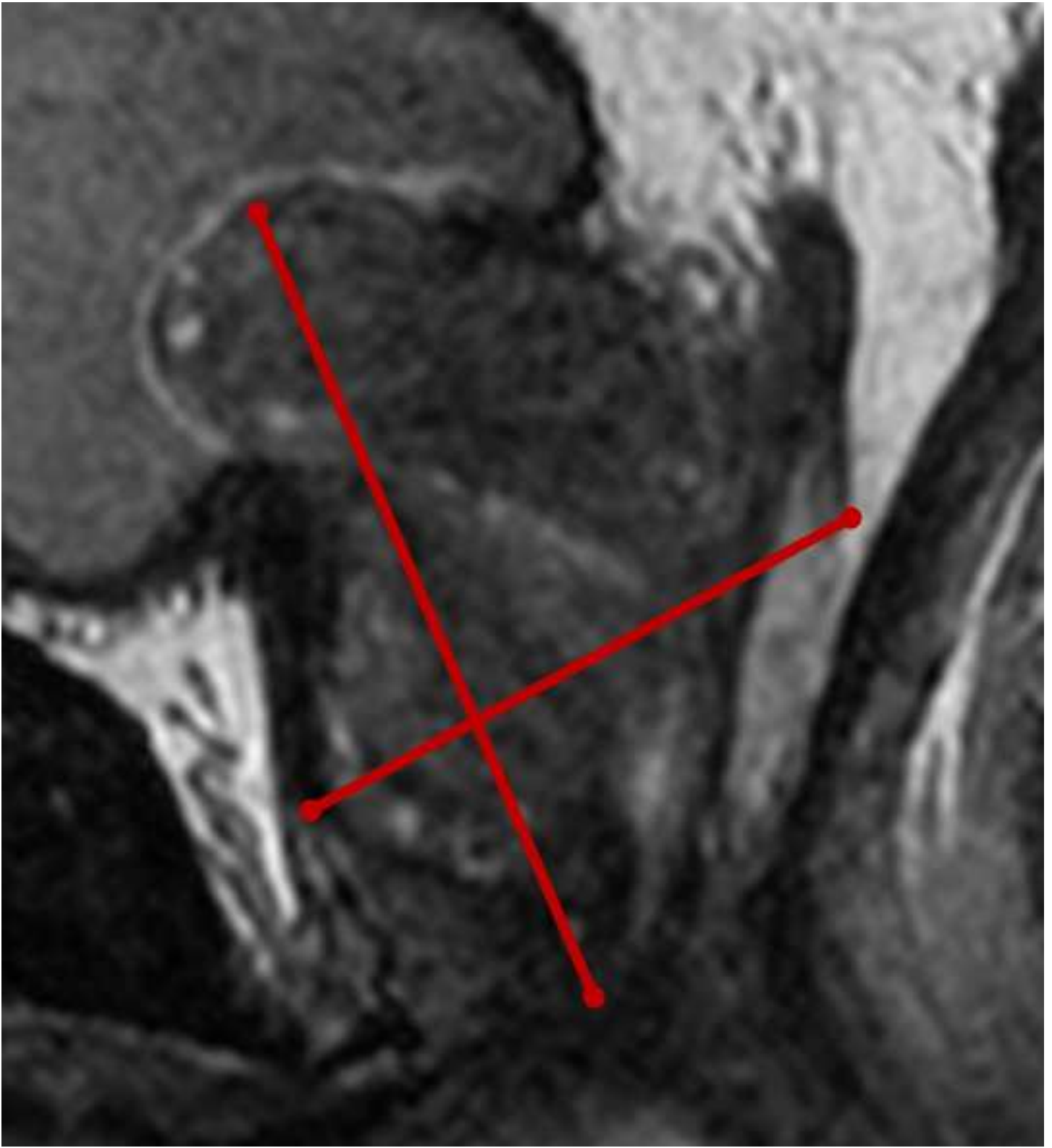
Table 3 Mean difference between estimated volumes for each rater when using MPM versus ellipsoid methods (TEF or BPEF)

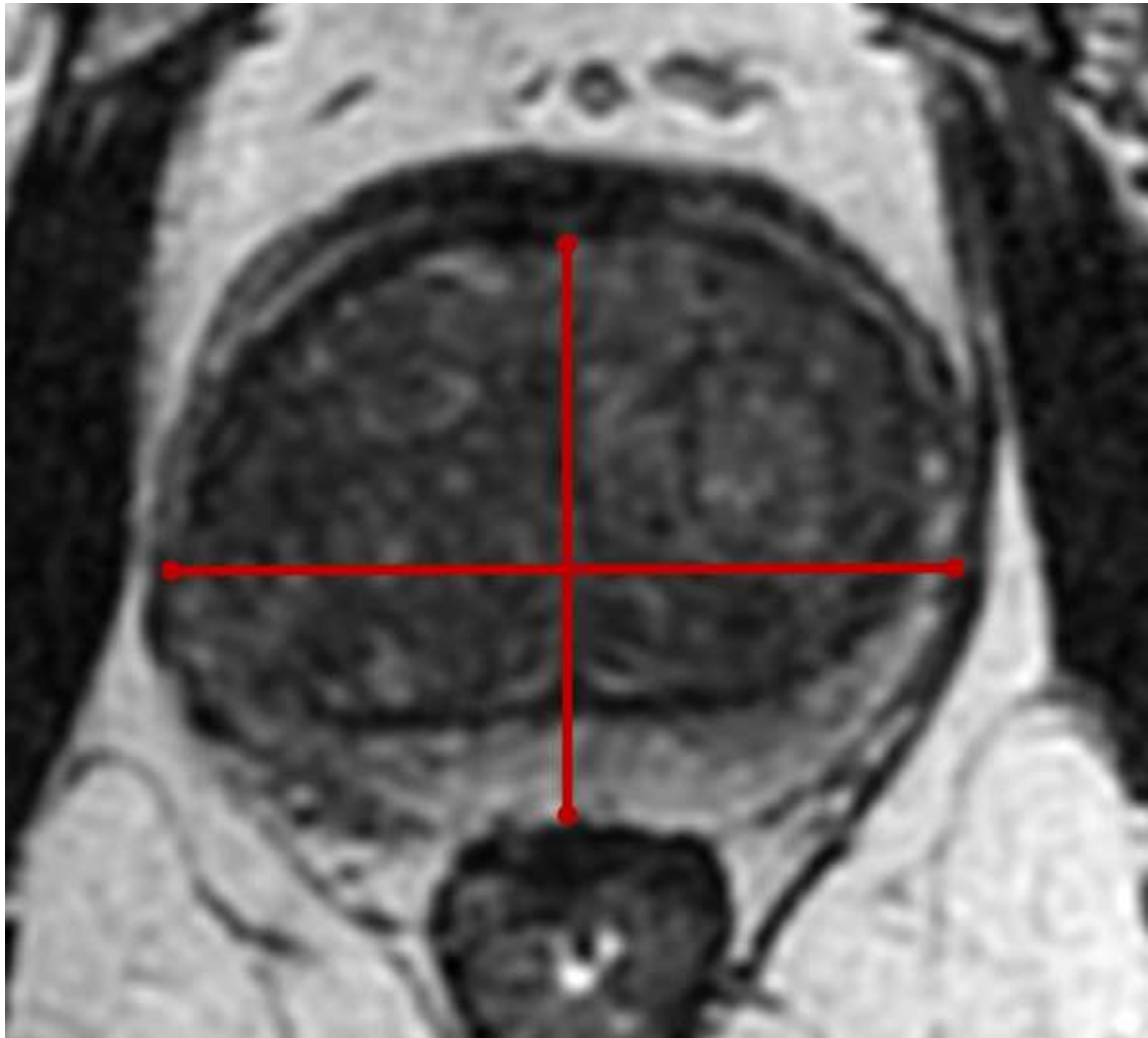
Group	Rater	TEF versus MPM	<i>p-val</i>	BPEF versus MPM	<i>p-val</i>
Group 1 (Experts)	Reader 3	13.317 ± 12.226	<0.001	16.517 ± 15.270	<0.001
	Reader 6	-7.966 ± 9.985	<0.001	-6.545 ± 12.584	0.02
	Reader 7	0.846 ± 11.935	0.62	1.493 ± 12.728	0.62
Group 2 (Seniors)	Reader 1	9.111 ± 9.932	<0.001	4.637 ± 13.235	0.22
	Reader 2	0.473 ± 7.503	0.69	2.374 ± 14.218	0.15
Group 3 (Juniors)	Reader 4	-3.195 ± 9.350	0.62	-2.664 ± 10.156	0.62
	Reader 5	3.000 ± 9.420	0.0498	-6.720 ± 12.869	0.004

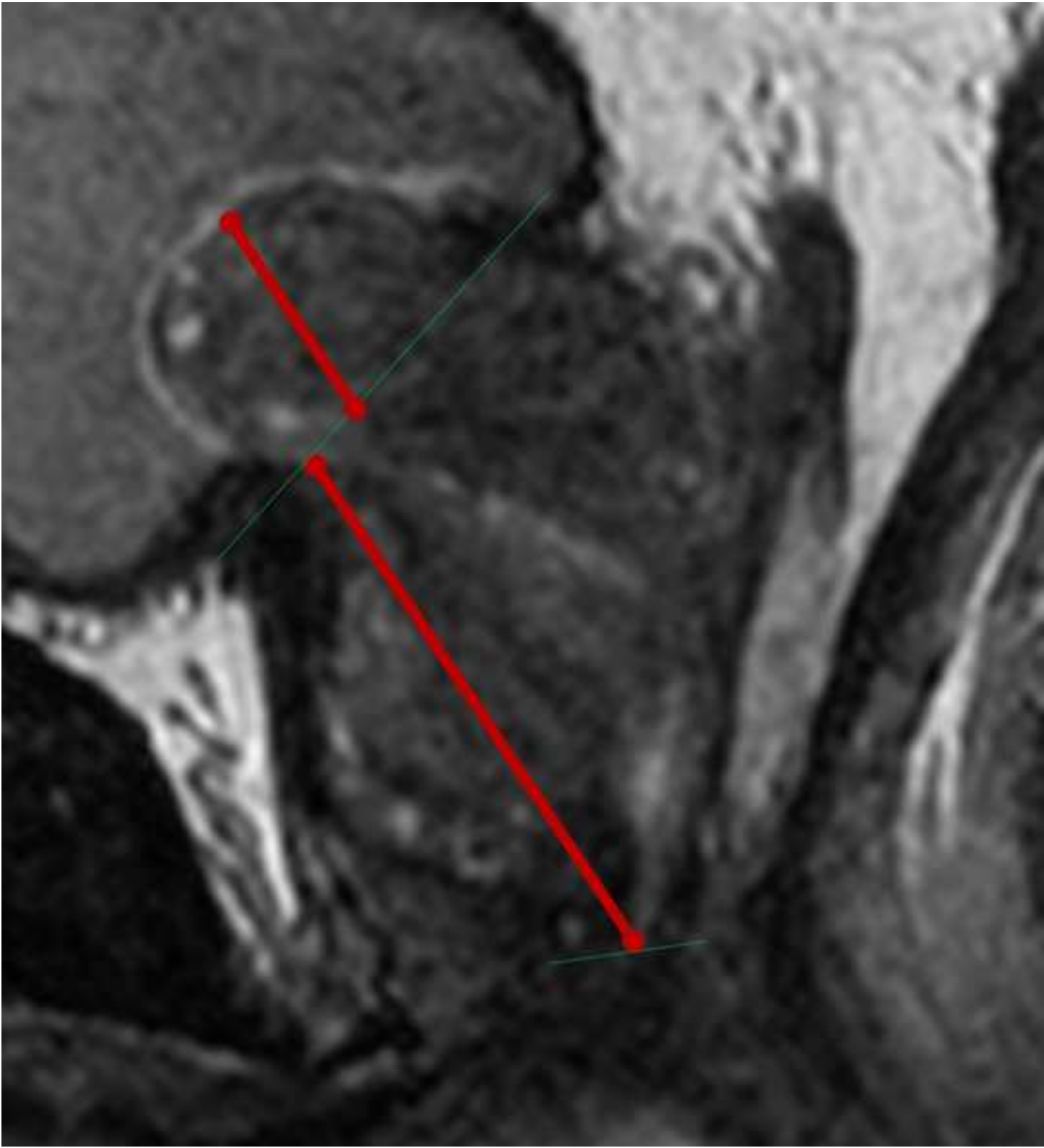
Table 4 Intra-rater reproducibility of volume estimation (evaluated by ICC) for each rater

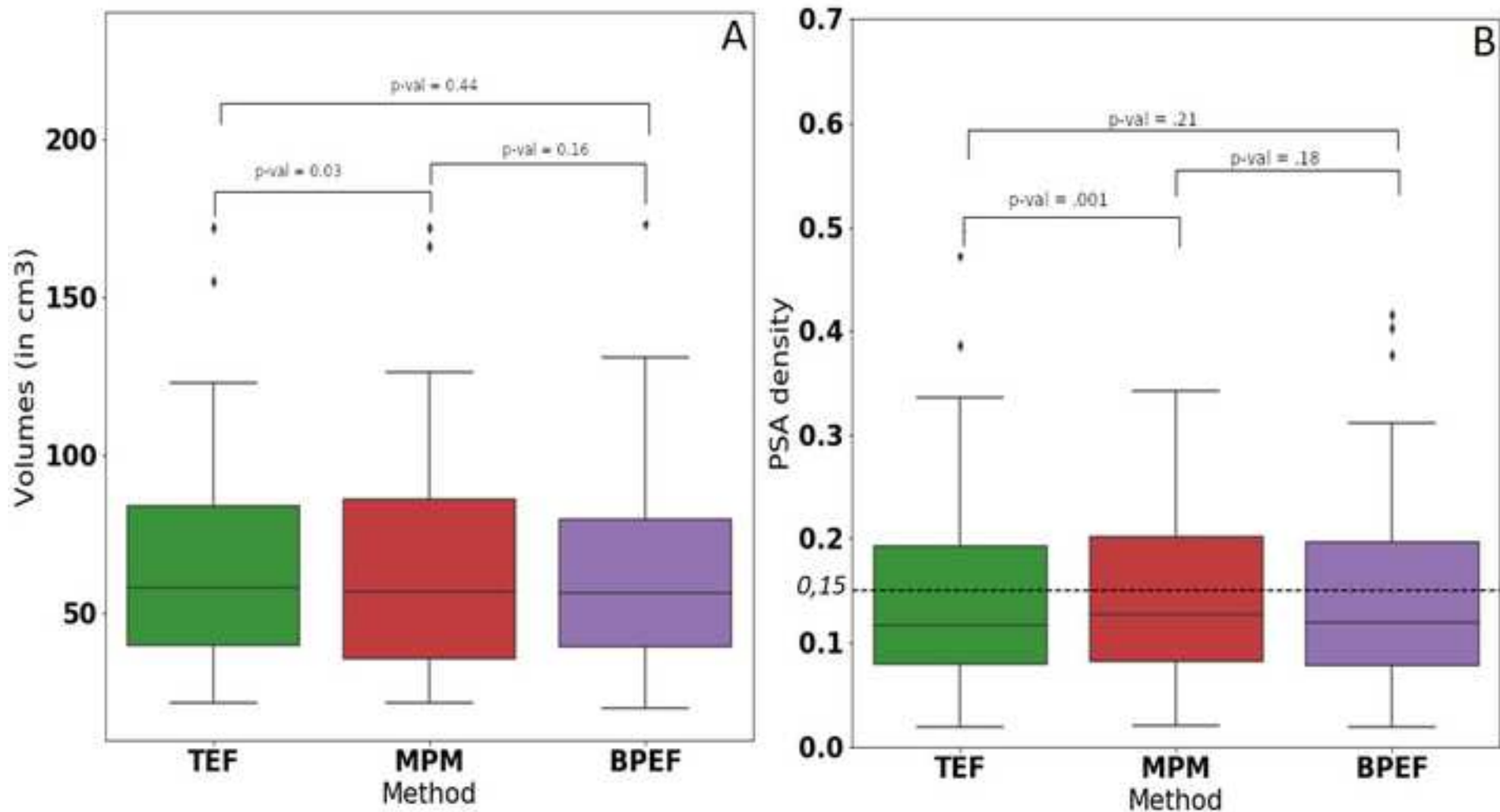
Group	Rater	ICC	CI95%
Group 1 (Experts)	Reader 3	0.960	[0.885,0.982]
	Reader 6	0.981	[0.960,0.990]
	Reader 7	0.978	[0.963,0.988]
Group 2 (Seniors)	Reader 1	0.981	[0.961,0.990]
	Reader 2	0.982	[0.969,0.990]
Group 3 (Juniors)	Reader 4	0.987	[0.978,0.993]
	Reader 5	0.977	[0.948,0.999]

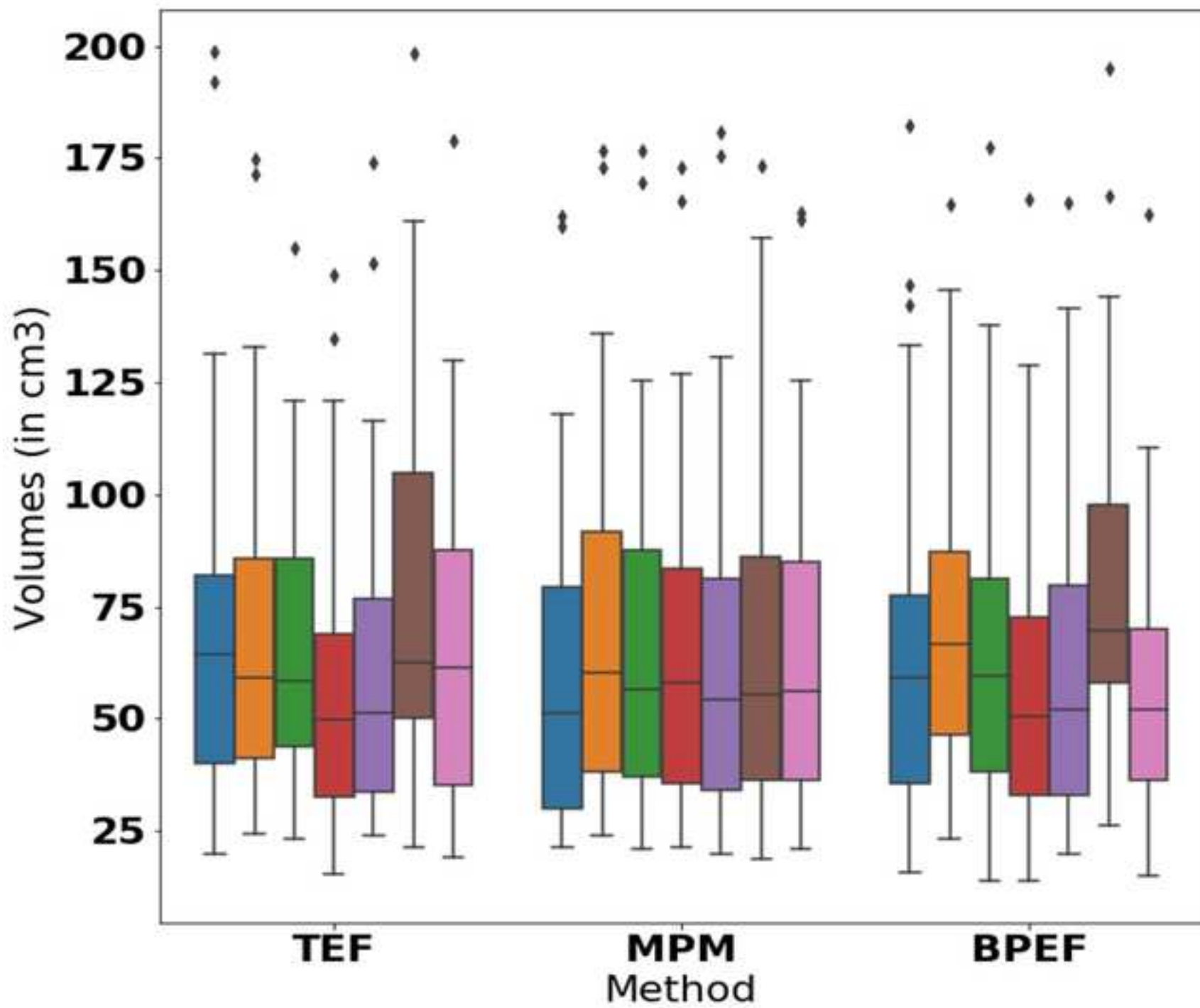


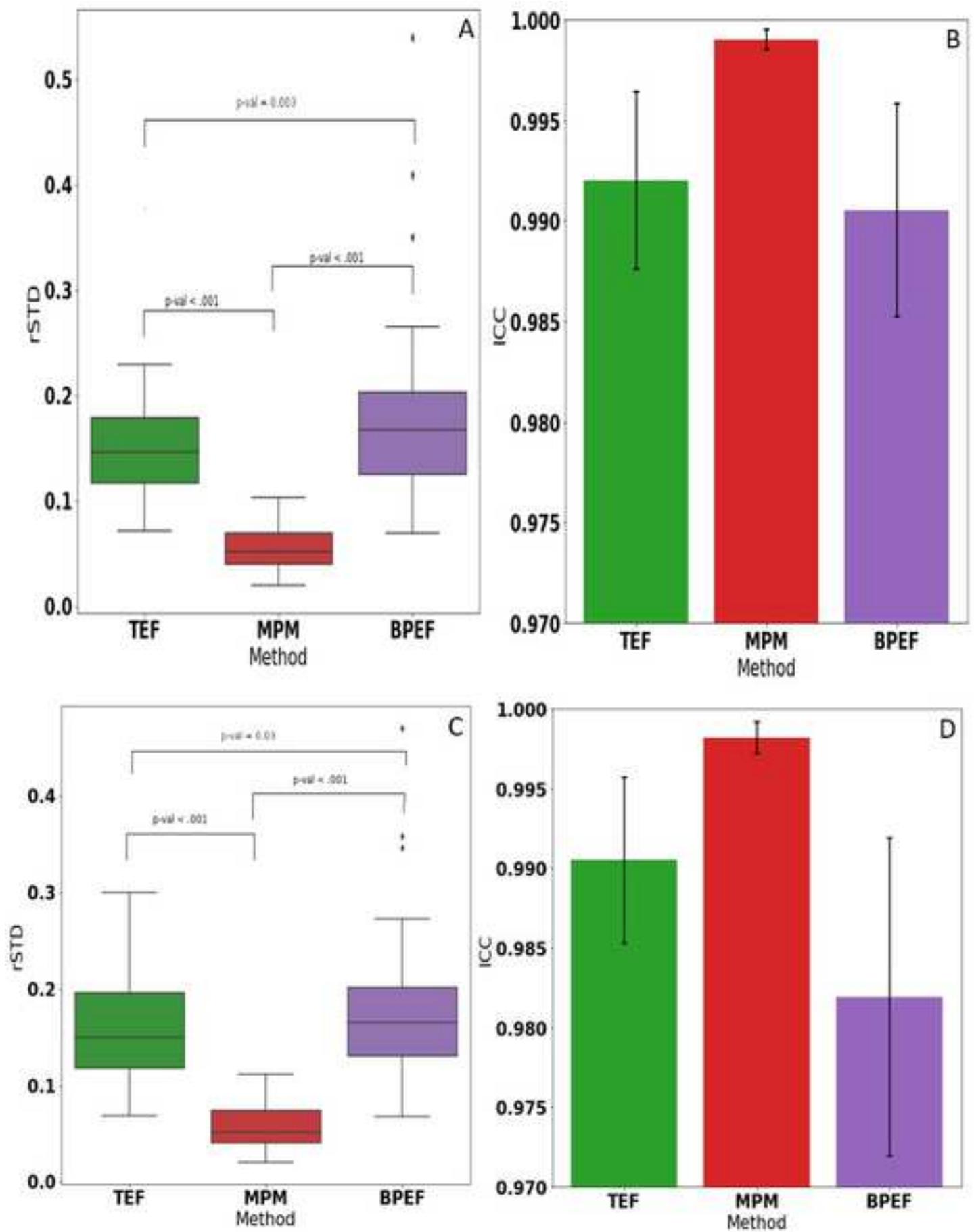


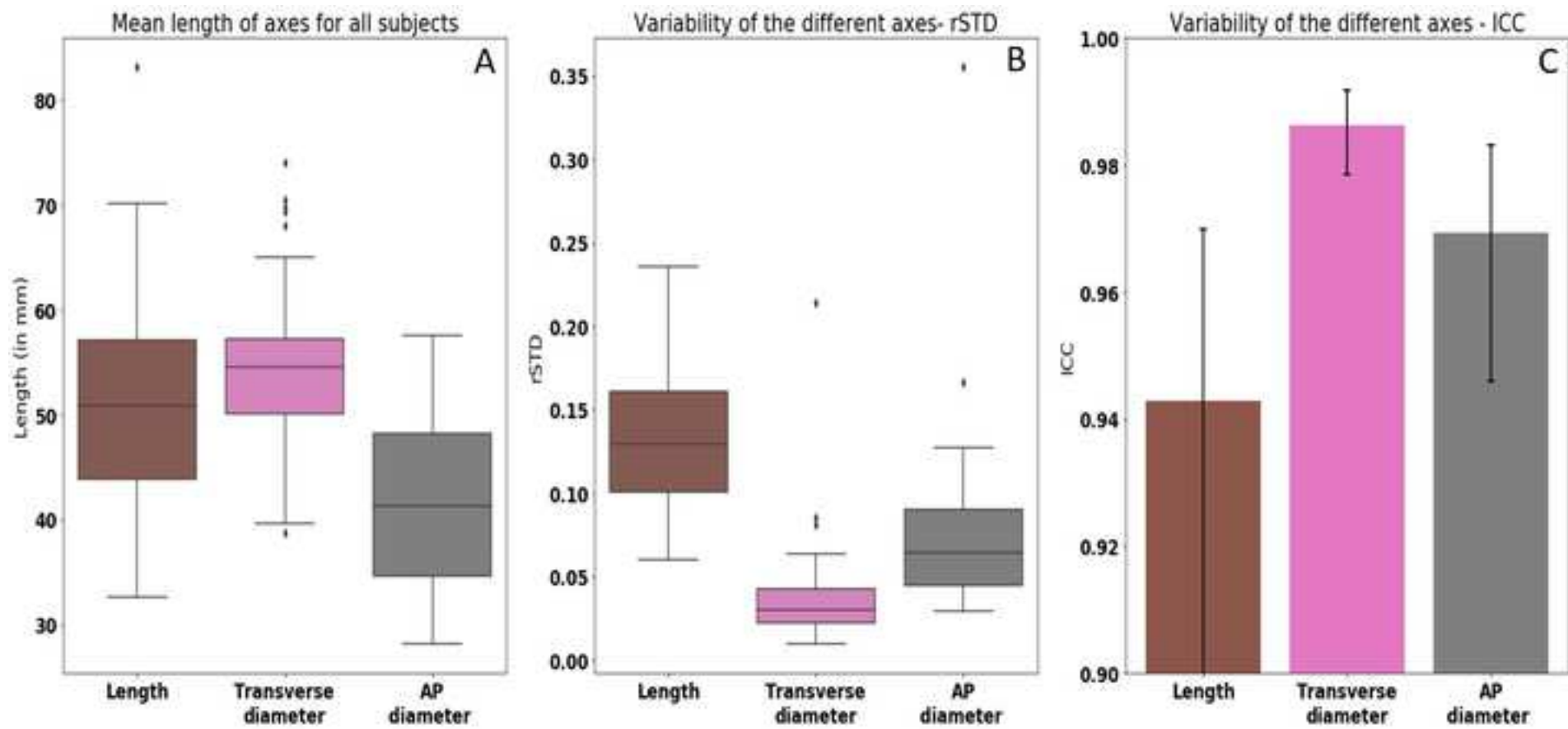


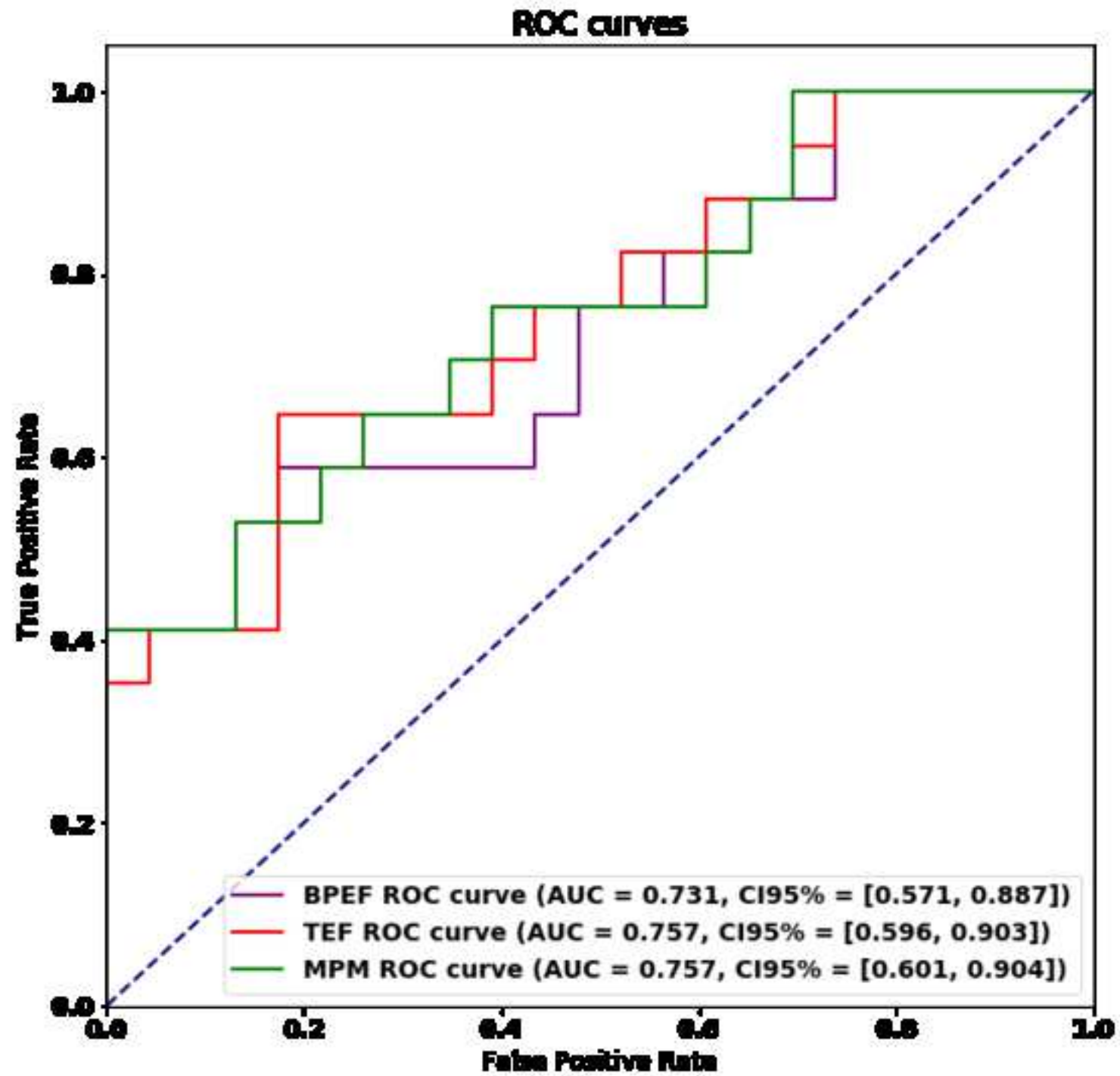












1. [Acknowledgements]

We thank Julien Castelneau, software Engineer Inria, for his help in the development of MedInria Software. This work has been supported by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002". Data were extracted from the Clinical Data Warehouse of the Greater Paris University Hospitals (Assistance Publique – Hôpitaux de Paris).

2. Funding

The authors state that this work has not received any funding.

Compliance with Ethical Standards

3. Guarantor:

The scientific guarantor of this publication is Pr Raphaële Renard Penna.

4. Conflict of Interest:

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

5. Statistics and Biometry:

Dr Benjamin Granger kindly provided statistical advice for this manuscript. One of the authors also has significant statistical expertise: Dimitri Hamzaoui.

6. Informed Consent:

Written informed consent was obtained from all subjects (patients) in this study.

7. Ethical Approval:

Institutional Review Board approval was obtained.

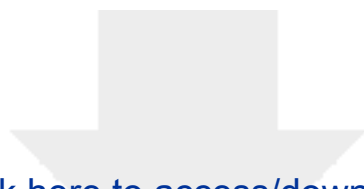
8. Study subjects or cohorts overlap:

Our study subjects haven't been previously reported in another study.

9. Methodology

Methodology:

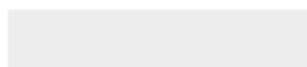
- retrospective
- observational
- performed in 2 institutions



Click here to access/download

Supplementary Material

Supplementary Tables_BLINDED.docx







Click here to access/download
Supplementary Material
Supplementary figure_1B.jpg

