Simona Ferraro*, Giacomo Biganzoli, Marco Bussetti, Silvana Castaldi, Elia Mario Biganzoli* and Mario Plebani

# Managing the impact of inter-method bias of prostate specific antigen assays on biopsy referral: the key to move towards precision health in prostate cancer management

https://doi.org/10.1515/cclm-2022-0874
Received September 5, 2022; accepted September 23, 2022; published online November 2, 2022

**Abstract**

**Objectives:** We assessed the inter-method bias of total (tPSA) and free (fPSA) prostate-specific antigen (PSA) immunoassays to establish if tPSA-based risk thresholds for advanced prostate cancer (PCa), obtained from one method (Roche) can be converted into the corresponding concentrations assayed by other methods. Then we evaluated the impact of the bias of tPSA and fPSA on the estimation of the %f/tPSA ratio and performed a re-calibration of the proposed thresholds for the %f/tPSA ratio according to the assay used.

**Methods:** tPSA and fPSA were measured in 135 and 137 serum samples, respectively by Abbott Alinity i, Beckman Access Dxl, Roche Cobas e801, and Siemens Atellica IM analytical platforms. Scatterplots, Bland-Altman diagrams, Passing-Bablok (PB) were used to inspect and estimate the systematic and proportional bias between the methods. The linear equations with confidence intervals of the parameter estimates were used to transform the tPSA risk thresholds for advanced PCa into the corresponding concentrations measurable by the other analytical methods. To construct a correction coefficient for converting the %f/tPSA ratio from one method to the other, PB and non-parametric boostrapping were used.

**Results:** The inter-method bias is not constant but strictly linear allowing the conversion of PSA results obtained from Roche into the other assays, which underestimate tPSA vs. Roche. Siemens and Abbott vs. Roche and Beckman assays, being characterized by a positive and a negative proportional bias for tPSA and fPSA measurements, tend to overestimate the %f/tPSA ratio.

**Conclusions:** There is a consistent risk to miss advanced PCa, if appropriate conversion factors are not applied.

**Keywords:** cancer; harmonization; immunoassay; predictive value; risk.

## Introduction

The clinical value of a laboratory test has to be ascertained on the basis of its impact on patient management [1–3]. When clinicians use biomarker results from different assays whose harmonization appears to be suboptimal, as it has been reported for total and free prostate specific antigen (tPSA, fPSA) measurements, the surveillance of the analytical performances in standard quality control programs is not enough and further steps should be moved to pursue the appropriate interpretation of the test results [3, 4]. This issue becomes relevant when one considers that urologists are often unaware of using heterogeneous tPSA and fPSA results to offer biopsy to outpatients at increased risk of high-grade prostate cancer (PCa), to promote active surveillance of slow-growing PCa, and to run effective diagnostic programs [5]. In such a situation, it should be relevant to actually define whether the reported

Elia Mario Biganzoli and Mario Plebani contributed equally to this work.

***Corresponding authors: Simona Ferraro**, Endocrinology Laboratory Unit, "Luigi Sacco" University Hospital, Università degli Studi di Milano, Via GB Grassi 74, 20157, Milan, Italy; and Newborn Screening and Genetic Metabolic Diseases Unit, "V. Buzzi" Children's Hospital, Milan, Italy, E-mail: simona.ferraro@asst-fbf-sacco.it; and
**Elia Mario Biganzoli**, Medical Statistics Unit, Department of Biomedical and Clinical Sciences, "Luigi Sacco" University Hospital, Università degli Studi di Milano, Milan, Italy, E-mail: elia.biganzoli@unimi.it
**Giacomo Biganzoli,** Medical Statistics Unit, Department of Biomedical and Clinical Sciences, "Luigi Sacco" University Hospital, Università degli Studi di Milano, Milan, Italy
**Marco Bussetti,** Immunoematologia e Medicina trasfusionale Ospedale Castelli, Verbania, Italy
**Silvana Castaldi,** Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; and Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milan, Italy
**Mario Plebani,** Department of Medicine-DIMED, University of Padova, Padova, Italy. https://orcid.org/0000-0002-0270-1711

inter-method bias of tPSA and fPSA measurements might be tolerated or the method-dependency of tPSA and fPSA results may contribute to an unfavorable risk-benefit ratio of the PSA based-screening strategy as it has been endorsed by recent Clinical Practice Guidelines (CPG) [2, 5]. Notably, the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) endorsed the recommendation that for medical tests that "have a central role in the decision-making of a specific disease and where cutoff/decision limits are established," assay specifications should be based on the effect of analytical performance on the clinical outcome (termed Model 1), as opposed to basing specifications on biological variation (BV) (Model 2) [6]. The use of BV data to assure the maintenance of the inter-assay equivalence of the results within objectively defined analytical performance specifications (APS) has been longer emphasized and exploited in the framework of tumor markers [7]. Currently, two types of studies have been suggested to set specifications under Model 1: (a) direct outcome studies (i.e., analyses based solely on empirical data, such as clinical trials evaluating the impact of varying analytical procedures on outcomes) and (b) indirect outcome studies (i.e., analyses using non empirical approaches, such as decision analytic modeling, to determine the impact of varying procedures on outcomes) [8]. The indirect methods are expected to play a dominant role in this context since the direct outcome studies are often unfeasible for several reasons including financial constrains [9].

There are however several unsolved criticisms concerning the determination of tPSA and fPSA and the interpretation of the results for decision making on biopsy referral. First, there is not a consensus on the range of tPSA value (2.0–10.0 or 4.0–10.0 μg/L) that should trigger the use of reflex testing of fPSA to allow the estimation of %f/tPSA ratio as second level test for recommending biopsy referral [10]. Second, a wide range of threshold levels for the %f/tPSA ratio (i.e., between <10 and <25%) to rule-in for biopsy referral has been reported to be used in the clinical practice, as it has emerged from the meta-analysis of the clinical studies evaluating the predictive capability of %f/tPSA ratio in those patients with a tPSA result falling either in the range 4.0–10.0 or 2.0–10.0 μg/L [11, 12]. Noteworthy, Roddam et al. have reported that a threshold value <10% implies a positive predictive value (PPV) of 55% (obtained by maximizing the specificity at 80%) and a ratio between PCa of any grader missed and unnecessary biopsies avoided of ~1:3.3 in the range of tPSA 4.0–10.0 μg/L (no data available in the range 2–4 μg/L) [11].

Both in patients between 2–4 and 4–10 μg/L the %t/fPSA decision threshold of <20% (most used in clinical practice to rule-in for biopsy referral) implied a similar PPV (~40%) and a quite similar ratio between PCa of any grade missed and unnecessary biopsies avoided (1:5.5 and 1:5.8 in the lower and higher range of tPSA results, respectively) [11].

The clinical impact is however different, since it has been observed that patients with aggressive cancer (i.e., International Society of Urological Pathology [ISUP] grade ≥3) have a higher likelihood to have tPSA levels higher than 4.0 μg/L. This is evident by comparing the 25th percentiles of the distribution of tPSA levels in patients with a PCa of ISUP grade ≥3 vs. those with of ISUP grade <3 (7.0 μg/L vs. 4.4 μg/L, respectively) [13]. Furthermore, this risk is far increased in patients older than 65 years [13]. This is relevant to emphasize since the recent recommendations endorse the stratification of PCa risk according to individual tPSA values and age, aiming to offer biopsy to patients at increased risk of high-grade disease [5, 14–16]. Accordingly, recent studies are developing PSA-based risk models as an aid in the personalized management of the diagnostic workup of PCa, to improve the risk-benefit ratio of the patient, reminding that the ratio of one cancer detected of any grade for every three performed biopsies is currently considered to be an acceptable clinical goal [5, 13].

Undoubtedly, the main criticism of the PSA-based risk algorithms is that tPSA and fPSA assays remain poorly harmonized also after fulfilling the recalibration to WHO International Standard (IS), and only for tPSA assays the inter-method bias appears to be acceptable in face of the minimum APS goal estimated according to BV data. First aim of this research was to inspect the patterns and assess the inter-method bias of tPSA and fPSA immunoassays to establish if tPSA-based risk thresholds for advanced PCa [12], obtained from one method used as reference (i.e., Roche) can be reliably converted into the corresponding concentrations assayed by the other methods. Second aim is to evaluate the impact of the bias of tPSA and fPSA on the estimation of the %f/tPSA ratio and perform a re-calibration of the proposed thresholds for the %f/tPSA ratio according to the assay used.

Roche assay was considered as reference method for two main reasons. First, since a recent study has defined tPSA thresholds for decision making on biopsy referral, by using well calibrated risk prediction models and tPSA results exclusively obtained by Roche assay [13]. Second, since this tPSA assay is a robust method, being both the assay design and analytical performances well characterized and reported [4, 5].

# Materials and methods

## Samples

Samples' preparation and measurements on the four analytical platforms Access Dxl (Beckman Coulter), Atellica IM (Siemens Healthcare Diagnostics), Alinity i (Abbott Diagnostics) and Cobas e801 (Roche Diagnostics) have been described elsewhere [4]. Also, the assay design and the characterization of the antibodies used were reported in the same paper [4].

Briefly, we selected 135 and 137 leftover serum samples, from different patients, as a representative clinical sampling for the measurement of tPSA and fPSA, respectively of the daily clinical routine, excluding from the collection haemolyzed, lipemic or icteric samples, according to the interference thresholds reported by PSA assays. The leftover serum specimens were immediately separated into a total of four 400 µL aliquots (one for each of the analytical platforms evaluated in the study) and stored at −80 °C until analysis. The study design followed the Clinical and Laboratory Standards Institute EP-09 protocol concerning the procedure of sample collection and measurements [17].

## Statistical methods

The first objective of the study was to investigate the inter-method bias between the Roche Cobas e801, which was taken as the reference method as aforementioned, and three other more widely commercialized methods used in clinical practice (Abbott Alinity i, Beckman Access Dxl, and Siemens Atellica) for tPSA measurement. As the study design included two runs for each tPSA sample for each method, four scatter plots were produced comparing the first run on the X-axis with the second run on the Y-axis. This was performed for all four methods to obtain an overview of the variability between the replicates. Subsequently, a matrix of graphs consisting of scatterplots and Bland-Altman diagrams showing the pairwise comparison for all methods was produced, to gain a direct insight into the existing bias between the methods. Since the Roche Cobas e801 was our reference method, we focused on comparing all methods with it (looking at the first column and first diagonal of the matrix). To obtain an estimate of the systematic and proportional bias existing between the tPSA measurement methods, we used Passing-Bablok regression first on each measurement run and then on the mean values of tPSA. The derived linear equations, consisting of an alpha intercept (the estimate of the systematic constant bias component) and a beta coefficient (the estimate of the proportional bias component), were provided and projected into the scatter plots of the Passing-Bablok regression line. The linear equations retrieved with confidence intervals of the parameter estimates were used to transform the tPSA risk thresholds for advanced PCa into the corresponding concentrations measurable by the other analytical methods. A table was produced to show the converted thresholds for each method.

Next, an in-depth study of the inter-method bias between the tests for measuring fPSA was carried out on the same analytical platforms described for tPSA. Since the existing bias between the methods for measuring fPSA was not the same as that found for measuring tPSA, the second objective of the study was to construct a correction coefficient for converting the ratio of fPSA to tPSA measurement from one method to the other, to obtain harmonized thresholds according to the assay used in clinical practice to measure these quantities. To obtain an unbiased estimate of the correction coefficient for each conversion method together with its 95% CI, we proceeded as follows:

(a) 1,000 non-parametric bootstrap samples of 137 mean tPSA values and 1,000 non-parametric bootstrap samples of 135 mean fPSA values were selected;
(b) Passing-Bablok regression analysis was applied for each sample to obtain an estimate of the beta coefficients for fPSA and an estimate of the beta coefficient for tPSA, in both cases for each comparison (Roche vs. Abbott, Roche vs. Beckman, Roche vs. Siemens);
(c) for each bootstrap sample and for each method comparison, a ratio of the beta coefficient for fPSA to the beta coefficient for tPSA was calculated; the distribution of the 1,000 beta ratios was visualized in three histograms;
(d) to obtain an unbiased average estimate of the ratio, we calculated the respective mean of the distribution, while to obtain the 95% Confidence Intervals (CIs) around the mean, we considered the quantile method as is usual in non-parametric bootstrapping.

All the statistical analyses were performed using R software (version 4.1.2).

# Results

From the scatter plots comparing the first set with the second analytical run of tPSA measurements for each assay (Supplementary Figure 1), we saw that the variance between the replicates increases as the tPSA value increases.

The scatterplots and Bland-Altman diagrams comparing Roche Cobas e801 with all other methods showed that the bias exists and that the difference is not constant but proportional. In fact, the beta regression coefficient test of the difference from the mean in the Bland-Altman plot was significant for all comparison methods considered ($p<0.05$). There is no statistical evidence of a constant bias component (i.e., intercept not significantly different from 0). In Figure 1, a matrix of graphs shows the inter-method bias existing between all tPSA assays. The lower half of the matrix shows the scatterplots, while the upper half of the matrix shows the Bland-Altman plots.

Looking at the first column of the matrix, all tests with respect to Roche underestimate the concentration of tPSA as its values increase. This can be seen because the points at increasing tPSA values deviate from the bisector line of the graph, falling further and further below the bisector line.

The bias between methods highlighted in the scatter plots is reflected in the Bland-Altman plots (first row of the matrix). In fact, the difference between the methods to be compared with the Roche method and the Roche method has a clear negative linear trend as tPSA concentrations increase,

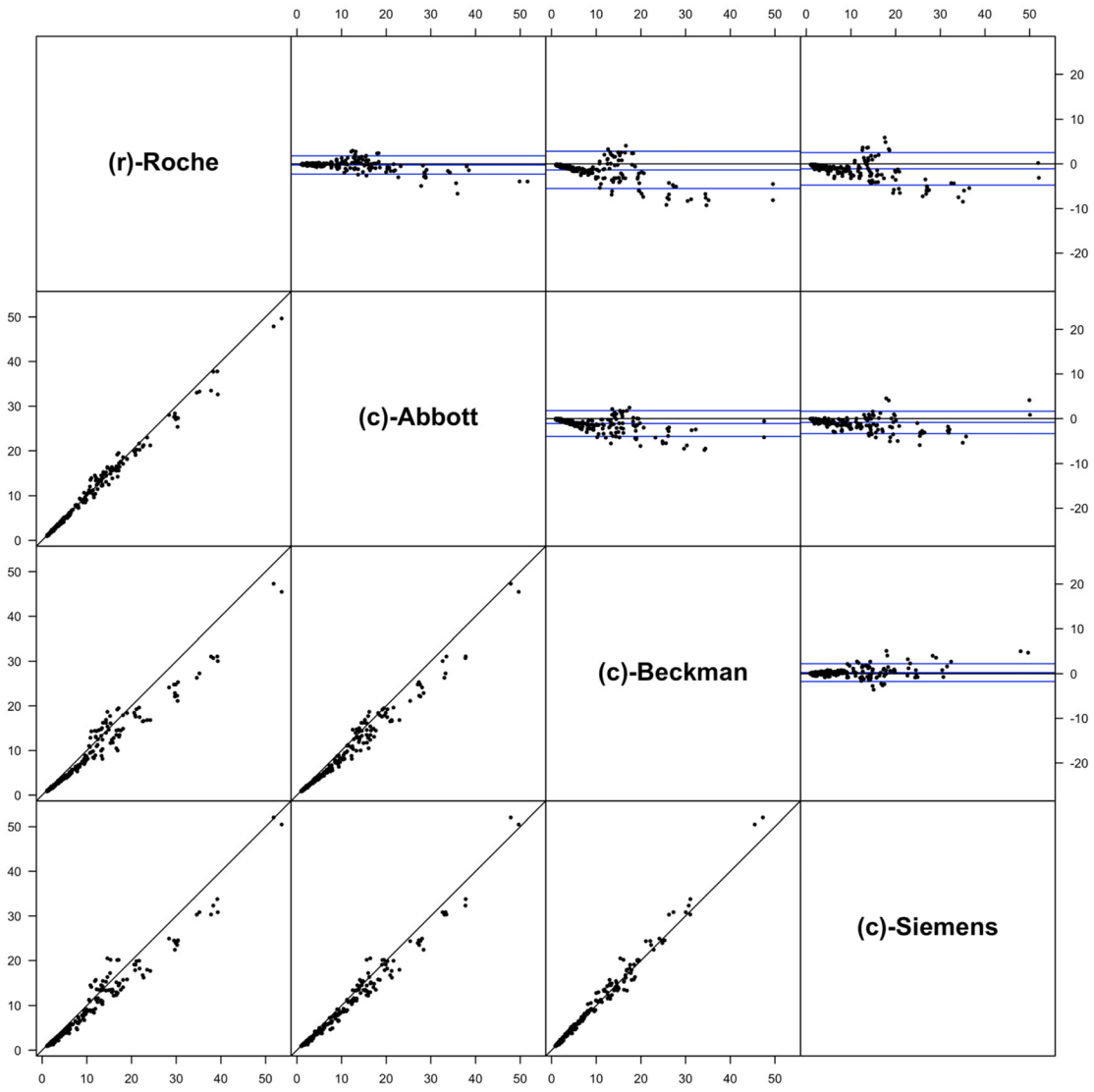


**Figure 1:** Matrix of graphs showing the inter-method bias existing between all tPSA assays. The lower half of the matrix shows the scatterplots, while the upper half of the matrix shows the Bland-Altman plots.

effectively indicating the underestimation of tPSA concentrations. Supplementary Table 1 shows the p-values of the hypothesis test that the beta coefficient of the regression of the difference from the mean is equal to 1.

## Conversion of the validated thresholds of tPSA

The intercept (systematic bias) and the slope (proportional bias) with their 95% CI. resulting from the Passing-Bablok regression analysis, performed first on each set of measurements and then on the mean values of tPSA, are shown in Supplementary Table 2. The comparisons between the methods considered were Roche vs. Abbott, Roche vs. Beckman, Roche vs. Siemens. In Figure 2, the Passing-Bablok regression lines on the mean tPSA values and the linear equations derived from the analysis are shown. The underestimation of tPSA concentration by the methods to be compared with the Roche assay is reflected by a negative intercept (systematic bias) and a Passing-Bablok regression coefficient (proportional bias) of less than 1.

Furthermore, in Table 1 we have reported the Roche tPSA risk thresholds and the associated PVs [13] together with the thresholds converted from Roche into the other tPSA assays. This conversion of results was performed by considering the systematic and proportional bias studied with Passing-Bablok regression models on mean values of tPSA.

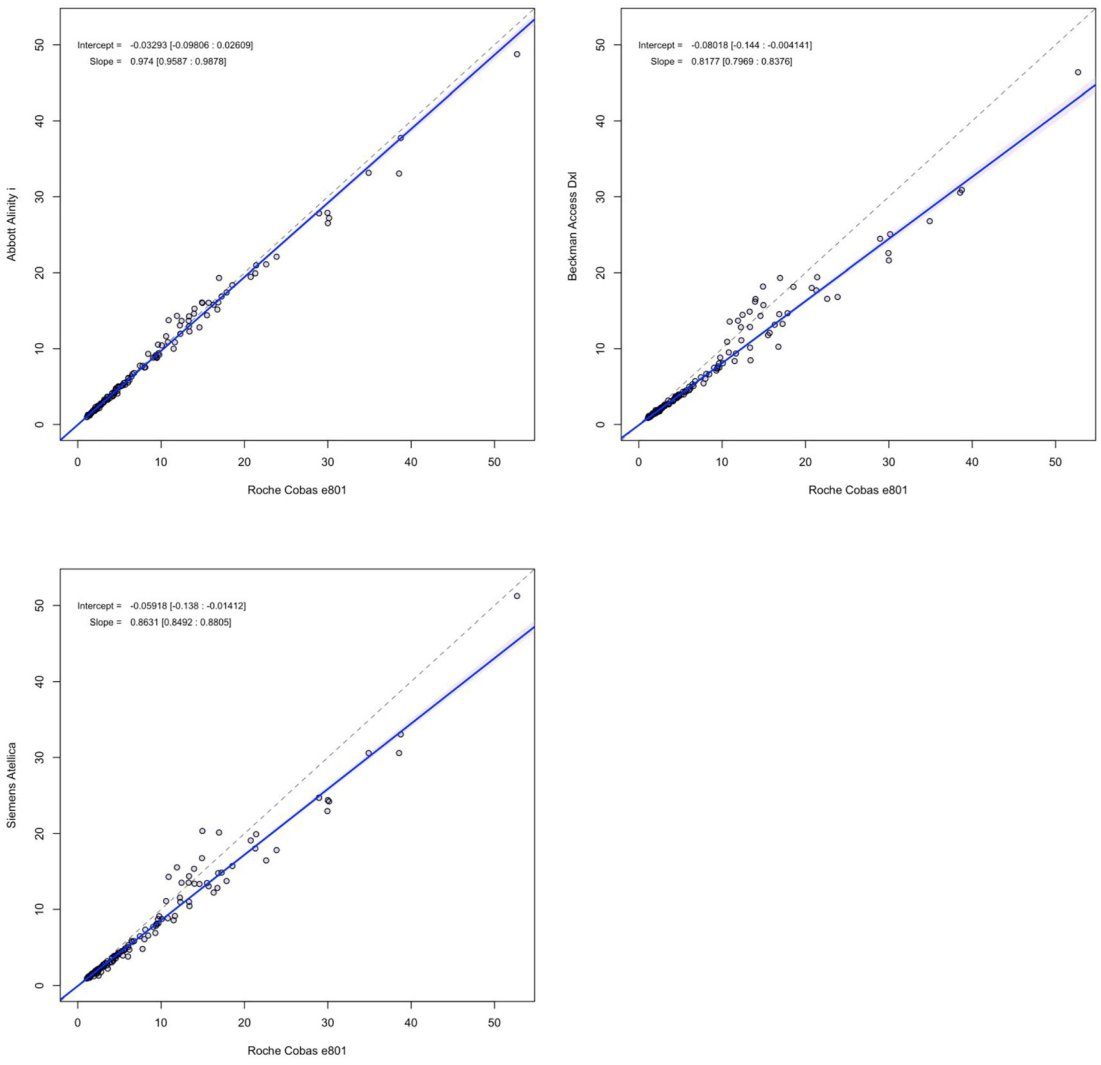


**Figure 2:** Passing-Bablok regression lines on the mean tPSA values and the linear equations derived from the analysis.

The Roche-based tPSA risk thresholds for advanced/aggressive PCa and associated PVs have been estimated in the clinical trial using exclusively tPSA Roche results, as explained in the introduction [13]. These thresholds were specifically estimated for men < and ≥65 years and asymptomatic for glandular inflammation, which may cause spurious increase of tPSA [13]. In the table we have further reported the PSA—based decision making for biopsy referral. Notably, the thresholds for %f/tPSA ratio have been reported as conditioned to the baseline tPSA value, in agreement with the results by of the metanalysis Roddam et al. [11]. In this table, %f/tPSA ratio have to be assumed as harmonized between assays.

In particular, the thresholds for %f/tPSA ratio shown in this table account for the sensitivity, specificity, PPV and ratio between cancer of any grade missed and unnecessary biopsies avoided, as estimated by Roddam et al. [11]. Thus the selection of the %f/tPSA ratio in this table may be assumed to represent the better compromise between risks and benefits of the patient. As aforementioned in the introduction, the risk of having a high grade PCa rises with the increase of PSA >4.4 µg/L and with age increase above 65 years [13].

For instance, the indication of a threshold value <10% for men younger than 65 years and of <20% for men older than 65 years, should account that there is a higher likelihood of missing a PCa of high grade in the latter group.

Noteworthy, in this table the %f/tPSA ratios must be assumed as harmonized between assays.

**Table 1:** Roche-based tPSA risk thresholds for advanced/aggressive PCa and associated PVs estimated according to the clinical trial using exclusively tPSA Roche results [13].

| Men <65 years asymptomatic for glandular inflammation | | | | | | |
|---|---|---|---|---|---|---|
| **Range of Roche-based risk thresholds for advanced (Gleason score ≥7)/ (ISUP≥3) PCa, µg/L** | **Abbott Diagnostics Alinity i, µg/L (95% CI)** | **Beckman Coulter Dxl Access, µg/L (95% CI)** | **Siemens Healthcare Diagnostics Atellica, µg/L (95% CI)** | **PVs** | **Further tests** | **Decision making on biopsy referral** |
| 2.8–4.1 | 2.69 (2.65, 2.73)–3.96 (3.91, 4.01) | 2.21 (2.17, 2.25)–3.27 (3.21, 3.33) | 2.36 (2.31, 2.38)–3.48 (3.43, 3.51) | Rule-out for PCa of Gleason score ≥7 (NPV~95%) | %f/tPSA<18%[a] active surveillance: PSA retesting after 2 years | tPSA absolute increase >2 µg/L & %f/tPSA<10%[b] |
| 4.1–4.9 | 3.96 (3.91, 4.01)–4.74 (4.67, 4.8) | 3.27 (3.21, 3.33)–3.93 (3.85, 4) | 3.48 (3.43, 3.51)–4.17 (4.12, 4.22) | Rule-out for PCa of ISUP ≥3 (NPV~97.5%) | | |
| 4.9–5.7 | 4.74 (4.67, 4.8)–5.52 (5.44, 5.6) | 3.93 (3.85, 4)–4.58 (4.48, 4.67) | 4.17 (4.12, 4.22)–4.86 (4.79, 4.92) | Rule-out for PCa of Gleason score ≥7 and of ISUP ≥3 (NPV~92% and NPV~95% respectively) | | %f/tPSA<10% |
| >5.7 | 5.52 (5.44, 5.6) | 4.58 (4.48, 4.67) | 4.86 (4.79, 4.92) | Rule-in for PCa of ISUP ≥3 (PPV~35%) | | %f/tPSA<10% |
| **Men ≥65 years asymptomatic for glandular inflammation** | | | | | | |
| **Range of Roche-based risk thresholds for advanced/ aggressive PCa, µg/L** | **Abbott Diagnostics Alinity i, µg/L (95% CI)** | **Beckman Coulter Dxl Access, µg/L (95% CI)** | **Siemens Healthcare Diagnostics Atellica, µg/L (95% CI)** | **PVs** | **Further tests** | **Biopsy referral** |
| 2.5–3.7 | 2.4 (2.36, 2.44)–3.57 (3.52, 3.61) | 1.96 (1.93, 2)–2.95 (2.9, 3) | 2.1 (2.05, 2.12)–3.13 (3.09, 3.16) | Rule-out for PCa of Gleason score ≥7 (NPV~90%) | %f/tPSA<18%[a] active surveillance: PSA retesting after 2 years | tPSA absolute increase >1 µg/L, & %f/tPSA<10%[b] |
| 3.8–5.3 | 3.67 (3.62, 3.71)–5.13 (5.06, 5.19) | 3.03 (2.98, 3.08)–4.25 (4.17, 4.33) | 3.22 (3.17, 3.25)–4.52 (4.45, 4.57) | Rule-out for PCa of Gleason score ≥7 (NPV~78.5%) | %f/tPSA<10% | Positive PI-RADS score (MRI) |
| 5.3–6.1 | 5.13 (5.06, 5.19)–5.91 (5.83, 5.98) | 4.25 (4.17, 4.33)–4.91 (4.81, 5.01) | 4.52 (4.45, 4.57)–5.21 (5.13, 5.27) | PPV>50% for PCa of Gleason score ≥7 | %f/tPSA<20% | |
| 6.1 | 5.91 (5.83, 5.98) | 4.91 (4.81, 5.01) | 5.21 (5.13, 5.27) | PPV>50% for PCa of ISUP ≥3 | | %f/tPSA<20%[c] |

[a]According to Roddam et al. [11] in this range of values, a %f/tPSA<18 the ratio between cancer missed and unnecessary biopsies avoided is ~1:5.5 (specificity=78%). [b]According to Roddam et al. [11] in this range of values, a %f/tPSA<10% the ratio between cancer missed and unnecessary biopsies avoided is ~1:3.3 (specificity=80%). [c]According to Roddam et al. [11] in this range of values, a %f/tPSA<20% the ratio between cancer missed and unnecessary biopsies avoided is ~1:5.5 (specificity=42%; sensitivity=85%), but there is a high likelihood that the missed PCa could be of high grade in the age ≥65 years. NPV, negative predictive value; PPV, positive predictive value; PCa, prostate cancer; ISUP, International Society of Urological Pathology; MRI, multiparametric magnetic resonance imaging; PD-RAS, prostate imaging-reporting and data system. These thresholds were reported for men < and ≥65 years asymptomatic for glandular inflammation [13]. The Roche-based risk thresholds are shown together with the thresholds converted from Roche into the other tPSA assays. We have further reported the PSA—based decision making. Notably, the thresholds for %f/tPSA ratio have been reported in agreement with the results reported by Roddam et al. in the metanalysis, conditioned to the baseline tPSA value [11]. In this Table %f/tPSA ratio have to be assumed as harmonized between assays.

## Study of a coefficient of correction for the ratio between free-PSA and total-PSA

To obtain a correction coefficient for the conversion of the fPSA/tPSA ratio between methods, considering Roche as the reference test, we first studied the inter-method bias of fPSA using the same approach as for tPSA. In Supplementary Figure 2, scatter plots show the relationship between the first and second run for each test. As in the case of tPSA, as the values of the fPSA concentration increase, the variance between the replicates increases.

The scatterplots and Bland-Altman plots comparing the Roche Cobas e801 measurements with all other methods showed that, even for fPSA, there is a bias, and the difference is not constant. Again, the beta regression coefficient test of the difference from the mean in the Bland-Altman plot was significant for all comparison methods considered ($p<0.05$). In Figure 3, a matrix of graphs shows the inter-method bias existing between all methods for the measurement of fPSA. The lower half of the matrix shows the scatterplots, while the upper half of the matrix shows the Bland-Altman plots.

Looking at the first column of the matrix, it can be seen that the Abbott and Siemens assay overestimates the concentration of fPSA as its values increase. However, this is not true for the Beckman assay, which underestimates fPSA concentration as its values increase.

The bias between the methods shown in the scatterplots is reflected in the Bland-Altman plots (first row of the matrix). In fact, the difference between the other methods and Roche has a clear positive linear trend as fPSA concentration increases for the Abbott vs. Roche and Siemens vs. Roche comparisons, and a clear negative trend for the Beckman vs. Roche comparison. In Figure 4 we have

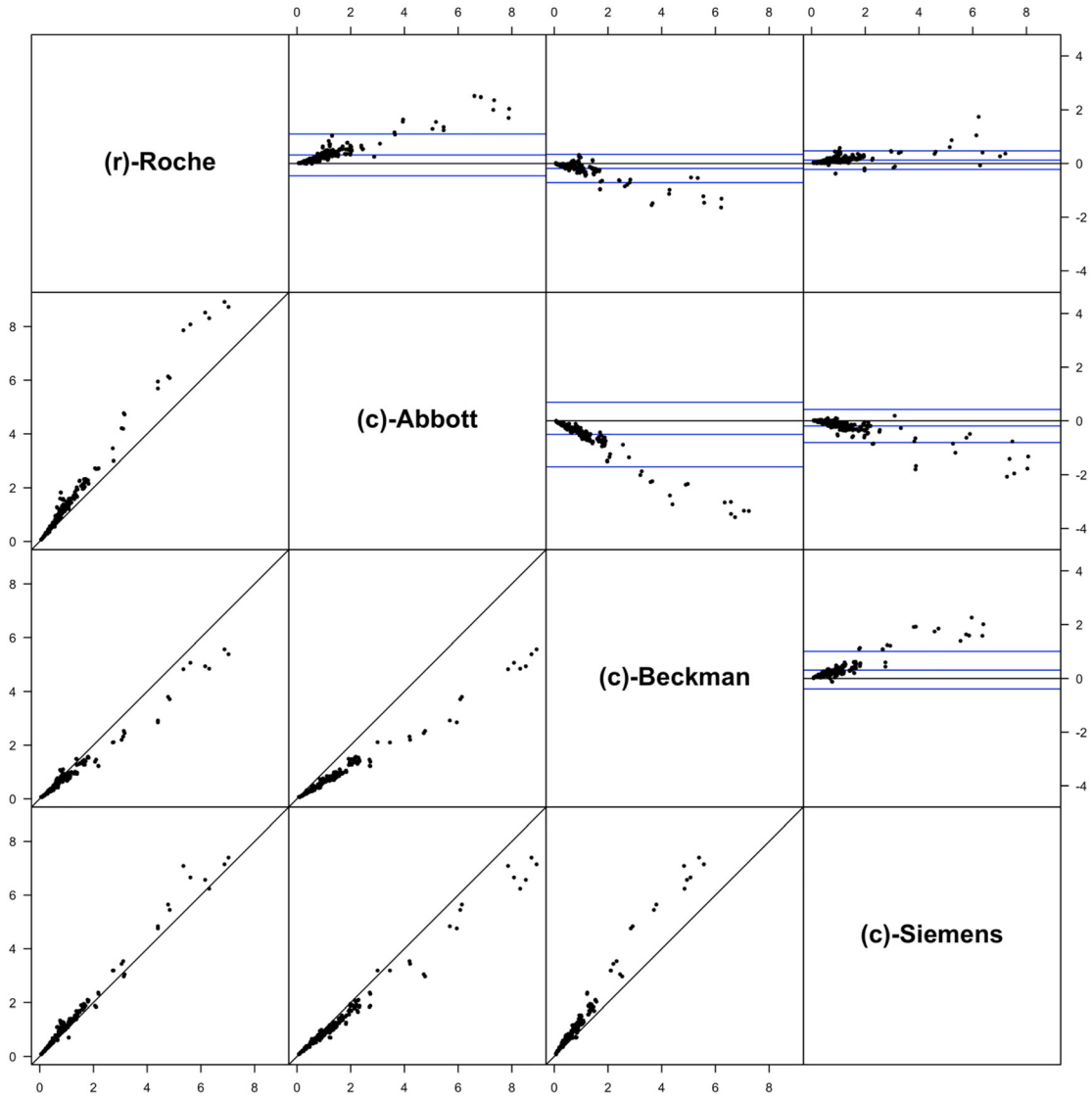


**Figure 3:** Matrix of graphs showing the inter-method bias existing between all methods for the measurement of fPSA. The lower half of the matrix shows the scatterplots, while the upper half of the matrix shows the Bland-Altman plots.

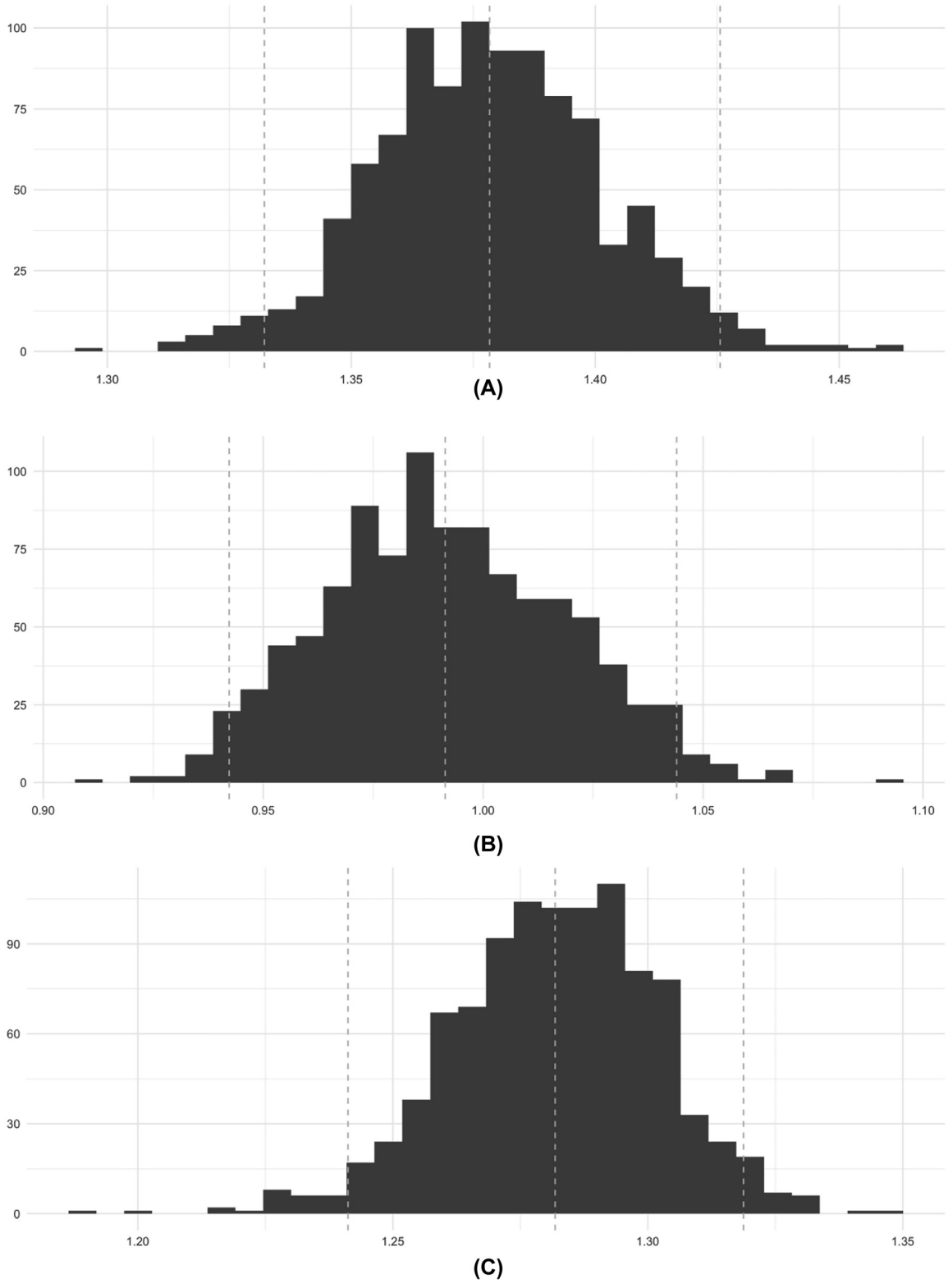


**Figure 4:** Distribution of the 1,000 bootstrap coefficient of corrections for the ratio for each method comparison with Roche. (A) Distribution of the coefficient correction derived from 1,000 bootstrap sample for the method comparison Roche vs. Abbott, (B) Distribution of the coefficient of correction derived from 1,000 bootstrap sample for the method comparison Roche vs. Beckman, (C) Distribution of the coefficient of correction derived from 1,000 bootstrap sample for the method comparison Roche vs. Siemens.

**Table 2:** Estimation of f/tPSA% ratio starting by assuming Roche tPSA assay as reference.

| Established thresholds by Roche assay | Abbott Diagnostics Alinity i, % 95% CI | Beckman Coulter Dxl Access, % 95% CI | Siemens Healthcare Diagnostics Atellica, % 95% CI |
|---|---|---|---|
| 20% | 27.56 (26.64–28.52) | 19.82 (18.84–20.88) | 25.64 (24.82–26.38) |
| 18% | 24.80 (23.98–25.67) | 17.84 (16.96–18.79) | 23.08 (22.34–23.74) |
| 10% | 13.78 (13.32–14.26) | 9.91 (9.42–10.44) | 12.82 (12.41–13.19) |

Abbott = Roche*1.378 (95% CIs: 1.332; 1.426), Beckman = Roche*0.991 (95% CIs: 0.991; 1.044), Siemens = Roche*1.282 (95% CIs: 1.241; 1.319). These ratio were considered as the main used in clinical decision making, as reported by Roddam et al. [11].

reported the distribution of the bootstrap coefficient of corrections for the ratio for each method comparison with Roche.

Table 2 reports the conversions of the %f/tPSA ratios assumed to be estimated by Roche into values derived from other assays. Coefficient of corrections with the 95% CI are reported in the footnote of the table.

## Discussion

The studies investigating the state of harmonization of tPSA and fPSA measurements and comparing the derived %f/tPSA ratios, after the recalibration of tPSA and fPSA assays to the WHO ISs 96/670 and 96/668 respectively, agree on the low interchangeability of the results obtained from these assays [4, 18, 19]. Some authors further advise that the % inter-method bias might be tolerated only for tPSA and not for fPSA, facing on the APS goals established according to BV data [4]. This information is however not considered by the panel of experts involved in the planning of CPGs and of clinical trials, who discard the method dependency of tPSA and fPSA results and the consequent effects on patients' outcome [5, 20].

No study has currently evaluated how the method-dependency of tPSA and fPSA results may affect patients' outcomes (i.e., increase the rate of undue biopsies, of slow-growing cancers treated, of patients undergoing rescreening programs), causing an unfavorable risk-benefit ratio of the PSA based-screening strategy fulfilling current recommendations [3, 5].The large majority of the clinical trials report about the poor capability of tPSA to predict PCa risk and about the modest additional contribution of %f/tPSA ratio to improve the rule-in for biopsy referral, discarding the analytical issues [5, 21–23]. Consequently, recent CPGs have emphasized the urgent need to improve the PSA-based decision-making for prostate biopsy referral in order to finally reduce the harms of overdiagnosis and overtreatment [5, 14–16]. As a result, CPGs appear to have scaled down the role of serum PSA measurements for deciding the need for biopsy, releasing, however, sparing and heterogeneous indications about the selection of patients who most likely will benefit from PSA testing and about the decision thresholds [5]. Clinical prediction models in this framework have been considered as a valuable aid to support healthcare professionals and patients in making decision about therapeutic interventions and further diagnostic testing [13, 24, 25]. An important aspect of prediction models is the extent to which estimated risks are "calibrated" to the observed outcomes, and the use of different PSA immunoassays can greatly contribute to the miscalibration of models to predict PCa in patients referred for biopsy after an "abnormal" PSA result [5, 24, 26]. Our study has considered to assess the pattern and the estimate of the bias between the most used commercially available tPSA and fPSA methods vs. Roche assays, to allow the conversion of PSA results and ratios obtained from Roche assays into the other methods.

We have considered Roche assay as reference method for several reasons. First, because a recent study has defined tPSA thresholds for identifying or excluding advanced PCa, as an aid to personalize management of the diagnostic workup, by using well calibrated risk prediction models and tPSA results exclusively obtained by Roche assay [13]. Noteworthy, most of evidence, supporting the new CPGs, resorts to clinical trials which have blended PSA results obtained by different methods [21–23]. Second, tPSA assay is a robust method, whose analytical performances have been largely investigated, and based on a well characterized assay design (i.e., clear identification of the epitopes recognized by the monoclonal antibodies) [4, 27].

According to our data, the inter-method bias is not constant on the explored range of measurement but strictly linear and this allows the conversion of the results obtained from Roche into the other assays, which underestimate with respect to the reference method (systematic negative proportional bias of all assays vs. Roche). The use of tPSA Roche and Abbott assays might be considered substantially interchangeable for the clinical classification of the patient, since the slight underestimate exhibited by Alinity Abbott (–3.5%) overlaps with the within laboratory imprecision declared by the manufacturer for this range of values. Notably, both Abbott and Roche assays are the ones

using an antibodies of Group 6 (International Society of Oncology and Biomarkers (ISOBM) classification), recognizing the aminoterminal region of PSA whereas the other assays use antibodies pairs different from those employed by Roche [4, 27]. Both Group 6 antibodies used by Abbott and Roche sandwiches have been recognized to cross-react with human glandular kallikrein 2 (hK2), which is greatly increased in high grade PCa and hK2 has been proposed in addition to %f/tPSA to reduce unnecessary biopsies without missing an undue number of tumors [4, 28, 29]. This is evident from the bias plot since for tPSA values >10 μg/L, where there is a highest likelihood to find advanced PCas [13], there is a marked increasing trend for underestimating tPSA concentrations by Siemens and Beckman with respect to Roche and Abbott assays.

Accordingly, the average underestimate exhibited by Beckman and Siemens assays vs. Roche (−20 and −16%, respectively) in the clinical practice may imply the underuse of the second level tests and/or of biopsy referral with the risk of missing advanced PCa (i.e., although the underestimate far increases with the increasing of tPSA >10 μg/L), if the thresholds of risk are not properly converted.

This is the first experimental evidence that may support previous data reporting that some commercial methods (i.e., Roche), although professed to be calibrated against IS 96/670, continued to exhibit a better alignment to the Hybritech IS calibration, resulting in approximately 20–25% higher tPSA values than other assays equally recalibrated to the WHO IS 96/670 [19]. Historically, a serum tPSA concentration of 4.0 μg/L as a cutoff point for biopsy referral was defined using immunoassays calibrated against the Hybritech IS, and accordingly some CPGs shifted the decision cutoff to 3.1 μg/L, recalculated to achieve similar diagnostic efficacy when the recalibration to WHO IS was introduced [5, 20]. This was obviously promoted independently of the employed assays, by assuming that the common calibration to IS 96/670 had *per se* improved enough the inter-assay harmonization [5].

The second aim of this study was to predict the impact of the method-dependent estimated %f/tPSA ratio on biopsy referral and to consider the feasibility to obtain a correction coefficient for the conversion of the %f/tPSA ratio between methods, by assuming again Roche as the reference test. This is a further important step for improving the management of PCa, since the %f/tPSA ratio is the most available and used second level test to increase the PSA-based rule-in for biopsy [10, 13]. A systematic positive proportional bias of fPSA Abbott and Siemens assays and a systematic negative proportional bias of fPSA Beckman vs. Roche were evident, and this allowed to perform the conversion of the ratios obtained by Roche assays into those obtained by the other methods. Notably, fPSA methods Beckman, Roche, Siemens were calibrated against the same IS (96/668 IS) whereas fPSA Abbott showed a more pronounced positive bias vs. the other assays likely being calibrated vs. 96/670 IS.

Furthermore, in the ratio the negative proportional bias of fPSA and tPSA Beckman tend to cancel each other, and so the ratios derived by PSA Roche and Beckman assays may be assumed to equally perform in the decision making for biopsy referral. According to our results, with respect to Roche and Beckman assays, Siemens and Abbott assays, being characterized by a positive and a negative proportional bias for tPSA and fPSA measurements, tend to overestimate the %f/tPSA ratio. This means that there is a consistent risk to miss advanced PCa, if appropriate conversion factors are not applied.

We should finally emphasize that this is the first study boosting for the clinical implementation of method-dependent PSA thresholds of risk for advanced and aggressive PCa, allowing to pragmatically improve the decision-making for prostate biopsy referral and reduce patient harms due to overdiagnosis and overtreatment, fulfilling the recommendations of recent CPGs [5]. Furthermore this study shows how the interplay between laboratory and biostatistical competence, that is part of the Health Technology Assessment model, might ensure that appropriate data and methods are used to identify and properly approach those aspects that heavily affect decision-making. The final purpose is to inform decision-making in order to promote an equitable, efficient, and high-quality health system [30].

In conclusion, assessing the effect of the inter-method bias on the clinical outcome is undoubtedly a valuable option to restore the central role of PSA tests in the CPGs, by considering that, at present, no other markers may replace PSA in decision making for biopsy referral and its predictive power is continuously investigated to be further improved [31, 32].

This is the first study using the assessment of the inter-method bias of tPSA and fPSA assays to predict the impact on the rule-in of patients for biopsy referral and to pragmatically release conversion factors, that applied to the results obtained by different assays, may improve the accuracy and precision of PSA-based risk prediction models and algorithms.

There is still need for better integration and use of laboratory tests in care pathways also for those markers as PSA testing that are considered to cover a consolidated role in decision making [33–35]. Using the information on the analytical performances of the tests in a more pragmatic way to improve patient outcome, as in this case, is part of

"what should be done now and in the future for enhancing value in laboratory medicine" [33]. In this framework, we have demonstrated that we may translate the information on the accuracy of the PSA methods in daily practice to move towards precision health in prostate cancer detection. For this goal it will be important to involve the medical and surgical staff who should be able to correctly understand the results of this test and take safe and effectiveness decisions for their patients [30].

**Research funding:** None declared.

**Author contributions:** All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved. S. Ferraro and G. Biganzoli, were involved in the provision of study material and conceptualization; M. Bussetti was involved in the provision of study material; E. Biganzoli, M. Plebani, S Castaldi revised and corrected and approved the final draft. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest.

**Informed consent:** Not applicable.

**Ethical approval:** Not applicable.

## References

1. Plebani M. Clinical laboratory: bigger is not always better. Diagnosis 2018;5:41–6.
2. Smith AF, Shinkins B, Hall PS, Hulme CT, Messenger MP. Toward a framework for outcome-based analytical performance specifications: a methodology review of indirect methods for evaluating the impact of measurement uncertainty on clinical outcomes. Clin Chem 2019;65:1363–74.
3. Ferraro S, Biganzoli EM. The clinical value of assessing the intermethod bias: the lesson from prostate specific antigen measurement. Clin Chem Lab Med 2021;60:149–51.
4. Ferraro S, Bussetti M, Rizzardi S, Braga F, Panteghini M. Verification of harmonization of serum total and free prostate-specific antigen (PSA) measurements and implications for medical decisions. Clin Chem 2021;67:543–53.
5. Ferraro S, Bussetti M, Panteghini M. Serum prostate specific antigen (PSA) testing for early detection of prostate cancer: managing the gap between clinical and laboratory practice. Clin Chem 2021;67:602–9.
6. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. Clin Chem Lab Med 2017;55:189–94.
7. Sturgeon CM, Duffy MJ, Stenman U-H, Lilja H, Brunner N, Chan DW, et al. National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. Clin Chem 2008;54:e11–79.
8. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53:833–5.
9. Horvath AR, Bossuyt PM, Sandberg S, St John A, Monaghan PJ, Verhagen-Kamerbeek WD, et al. Setting analytical performance specifications based on outcome studies—is it possible? Clin Chem Lab Med 2015;53:841–8.
10. Ferraro S, Caruso S, Panteghini M. Reflex testing of free prostate-specific antigen as effective health care policy. Arch Pathol Lab Med 2019;143:1045.
11. Roddam AW, Duffy MJ, Hamdy FC, Ward AM, Patnick J, Price CP, et al. NHS Prostate Cancer Risk Management Programme. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2–10 ng/ml: systematic review and meta-analysis. Eur Urol 2005;48:386–99.
12. Huang Y, Li ZZ, Huang YL, Song HJ, Wang YJ. Value of free/total prostate-specific antigen (f/t PSA) ratios for prostate cancer detection in patients with total serum prostate-specific antigen between 4 and 10 ng/mL: a meta-analysis. Medicine 2018;97:e0249.
13. Ferraro S, Bussetti M, Bassani N, Rossi RS, Incarbone GP, Bianchi F, et al. Definition of outcome-based prostate-specific antigen (PSA) thresholds for advanced prostate cancer risk prediction. Cancers 2021;13:3381–95.
14. Gandaglia G, Albers P, Abrahamsson PA, Briganti A, Catto JWF, Chapple CR, et al. Structured population based prostate-specific antigen screening for prostate cancer: the European Association of Urology Position in 2019. Eur Urol 2019;76:142–50.
15. Wolf AMD, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, Volk RJ, et al. American Cancer Society PCa Advisory Committee. American cancer society guideline for the early detection of prostate cancer: update 2010. CA A Cancer J Clin 2010;60:70–98.
16. Carroll PR, Parsons JK, Andriole G, Bahson RR, Carlsson S, Castle EP, et al. NCCN clinical practice guidelines in oncology. Prostate cancer early detection. Version 2.2019 – May 31; 2019. Available from: https://www.nccn.org/professionals/physician_gls/pdf/prostate_detection.pdf [Accessed Oct 20, 2020].
17. Clinical and Laboratory Standards Institute (CLSI). EP09-A3—measurement procedure comparison and bias estimation using patient samples; approved guideline, 3rd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2013.
18. Stephan C, Klaas M, Mulleer C, Schnorr D, Loening S, Jung K. Interchangeability of measurements of total and free prostate-specific antigen in serum with 5 frequently used assay combinations: an update. Clin Chem 2006;52:59–64.

19. Foj L, Filella X, Alcover J, Augé JM, Escudero JM, Molina R. Variability of assay methods for total and free PSA after WHO standardization. Tumor Biol 2014;35:1867–73.
20. Filella X, Albaladejo MD, Allué JA, Castano MA, Morell-Garcia D, Ruiz MÀ, et al. Prostate cancer screening: guidelines review and laboratory issues. Clin Chem Lab Med 2019;57:1474–87.
21. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014;384: 2027–35.
22. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow H. Prostate-specific antigen-based screening for prostate cancer: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2018;319:1914–31.
23. Pinsky PF, Parnes HL, Andriole G. Mortality and complications after prostate biopsy in the Prostate, Lung, Colorectal and Ovarian Cancer Screening (PLCO) trial. BJU Int 2014;113:254–9.
24. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making 2015;35:162–9.
25. Carlsson S, Assel M, Vickers A. Letter to the editor concerning 'do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis'. Ann Oncol 2015;26:1031.
26. Ferraro S, Marano G, Ciardi L, Vendramin C, Bongo AS, Bellomo G, et al. Impact of calibration fitting models on the clinical value of chromogranin A. Clin Chem Lab Med 2009;4:1297–303.
27. Stenman U-H, Paus E, Allard WJ, Andersson I, Andrès C, Barnett TR, et al. Summary report of the TD-3 workshop: characterization of 83 antibodies against prostate specific antigen. Tumor Biol 1999;20(Suppl):1–12.
28. Leinonen J, Leinimaa M, Zhang W-M, Piironen T, Pettersson K, Lilja H, et al. Reactivity of anti-PSA monoclonal antibodies with recombinant human kallikrein-2. Tumor Biol 1999;20:35–7.
29. Braun K, Sjoberg DD, Vickers AJ, Lilja H, Bjartell AS. A four-kallikrein panel predicts high-grade cancer on biopsy: independent validation in a community cohort. Eur Urol 2016;69:505–11.
30. Ferraro S, Biganzoli EM, Castaldi S, Plebani M. Health technology assessment to assess value of biomarkers in the decision-making process. Clin Chem Lab Med 2022;60:647–54.
31. Oldenburg J, Bjerner JL, Lilja H, Aas K, Fossa SD, Mueller C, et al. Long-term predictive value of serum PSA values obtained in clinical practice: results from the Norwegian Prostate Cancer Consortium (NPCC). J Clin Oncol 2022;40(16 Suppl):5021.
32. Ferraro S, Biganzoli EM. Association between total prostate-specific antigen (tPSA), free/tPSA and prostate cancer mortality. BJU Int 2022;129:418.
33. Plebani M, Laposata M, Lippi G. Driving the route of laboratory medicine: a manifesto for the future. Intern Emerg Med 2019;14: 337–4.
34. Lippi G, Plebani M. Personalized medicine: moving from simple theory to daily practice. Clin Chem Lab Med 2015;53:959–60.
35. Kim EH, Andriole GL. Prostate-specific antigen-based screening: controversy and guidelines. BMC Med 2015;13:61–4.

**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2022-0874).