Assessing the Clinical Impact of Prostate-Specific Antigen Assay Variability and Nonequimolarity: Simulation Study based on the Population of the United Kingdom

ANDREW W. RODDAM,1* CHRISTOPHER P. PRICE,2 NAOMI E. ALLEN,1 and ANTHONY MILFORD WARD3 on behalf of the National Health Service Prostate Cancer Risk Management Programme

Background: Prostate-specific antigen (PSA) is the most widely used serum biomarker to differentiate between malignant and benign prostate disease. Assays that measure PSA can be biased and/or nonequimolar and hence report significantly different PSA values for samples with the same nominal amount. This report investigates the effects of biased and nonequimolar assays on the decision to recommend a patient for a prostate biopsy based on age-specific PSA values.

Methods: A simulation model, calibrated to the distribution of PSA values in the United Kingdom, was developed to estimate the effects of bias, nonequimolarity, and analytical imprecision in terms of the rates of men who are recommended to have a biopsy on the basis of their assay-reported PSA values when their true PSA values are below the threshold (false positives) or vice versa (false negatives).

Results: False recommendation rates for a calibrated equimolar assay are 0.5–0.9% for analytical imprecision between 5% and 10%. Positive bias leads to significant increases in false positives and significant decreases in false negatives, whereas negative bias has the opposite effect. False-positive rates for nonequimolar assays increase from 0.5% to ~13% in the worst-case scenario, whereas false-negative rates are almost always 0%.

Conclusions: Biased and nonequimolar assays have major detrimental effects on both false-negative and false-positive rates for recommending biopsy. PSA assays should therefore be calibrated to the International Standards and be unbiased and equimolar in response to minimize the likelihood of incorrect clinical decisions, which are potentially detrimental for both patient and healthcare provider.

© 2004 American Association for Clinical Chemistry

Prostate-specific antigen (PSA)4 is the most widely used serum biomarker to differentiate between malignant and benign prostate disease and is currently the subject of numerous large-scale randomized trials to assess its potential as a screening method for early prostate cancer detection (1). PSA circulates in human serum in two main forms, either free (“unbound”) or complexed (“bound”) (2); complexed PSA refers to the fraction of total PSA (tPSA) that is bound to any of the proteinase inhibitors, whereas free PSA (fPSA) is the fraction not bound to any protein. Generally, tPSA concentrations in healthy men are low (<4 µg/L) and are increased in men with prostate disease (3); men with prostate cancer also tend to have a lower fraction of fPSA and a higher fraction of complexed PSA than men with benign disease (4, 5).

Assays for the measurement of PSA used by diagnostic laboratories, the majority of which are now commercially available assays, have considerable differences in their epitope recognition, which can lead to significantly different tPSA values for the same serum sample (6, 7). Therefore, samples containing effectively the same

4 Nonstandard abbreviations: PSA, prostate-specific antigen; fPSA, free PSA; and tPSA, total PSA.
amount of tPSA but with different proportions of fPSA can produce very different values for both tPSA and fPSA. These differential response assays are usually referred to as skewed-response or nonequimolar assays. This has led to the development of International Standards for the purposes of assay calibration for both fPSA and tPSA (8).

The clinical relevance of using an equimolar assay for tPSA remains controversial (9, although two recent studies (10, 11) have shown the superior diagnostic ability of equimolar response assays in differentiating between benign prostatic hyperplasia and prostate cancer. However, there has been no investigation on the impact of assays that are either nonequimolar or biased (i.e., not numerically calibrated to the International Standards) on the decision to recommend a patient to undergo a prostate biopsy for cancer detection. This is a relevant clinical question because a biopsy is a costly and invasive procedure that is often associated with negative psychologic consequences.

Whereas the majority of commercial assays are equimolar and unbiased (12), there remain some that are not. In this study we used a computer-simulation-based approach to investigate the effects of bias, nonequimolarity, and analytical imprecision on the decision to recommend a patient for a prostate biopsy.

**Materials and Methods**

A computer-based simulation was used to model the effects of different assays in a population of men with PSA concentrations representative of the general population. For each scenario, 1000 random samples, each of a size of 30,000, were drawn from a population designed to represent true tPSA values. For each of the 1000 samples, the observed (assay-reported) tPSA values were calculated using the simulated true tPSA values and models representing a biased or nonequimolar assay along with variation corresponding to laboratory imprecision and sample handling. All calculations were carried out in the data analysis language R (13).

**Population-based PSA concentrations**

The study population consisted of males in the United Kingdom 45–84 years of age because this age group represents the population most likely to undergo a PSA test. On the basis of a previous report of PSA values recorded during 1999 in the population of the United Kingdom (14), it was possible to construct approximate age-specific log-normal PSA distributions for men 45–54, 55–69, and 70–84 years of age. Each simulation sample size of 30,000 was achieved by equally sampling 10,000 men from the three age-specific PSA distributions.

**Laboratory variation**

For each simulation the impact of laboratory and sample-handling variability was modeled by assuming that the overall assay CV was 5%, 7.5%, or 10%. Including this variation in the simulated age-specific PSA distributions required the use of a simple statistical model, where:

\[
\text{Observed PSA} = \text{true PSA} + \text{error}
\]

and error was assumed to be a gaussian distribution with mean 0 and SD equal to \( CV \times \text{true PSA} \). This corresponds to the situation of a correctly calibrated equimolar assay.

**Assay inaccuracy (bias)**

It was assumed that a biased assay will proportionately over- or underreport the tPSA concentration depending on the underlying tPSA concentration in the sample. The amount of bias can be calculated using the following equation:

\[
\text{Bias amount} = \text{true PSA} \times \text{bias}
\]

where bias is allowed to vary from −25% to 25%. At the extremes, this was equivalent to assuming a bias of 1 \( \mu g/L \) at a PSA concentration of 4 \( \mu g/L \) (a commonly used PSA cutpoint); a variation of ±1 \( \mu g/L \), irrespective of the actual tPSA concentration, was also investigated. The impact of bias was modeled by assuming that the gaussian error distribution used to generate the laboratory variation had a mean equal to the bias rather than 0 but had the same SD. This ensured that the amount of laboratory variation remained constant but that observed values were systematically smaller or larger than the true values depending on the amount of bias.

**Nonequimolar assay**

The impact of a nonequimolar assay, as demonstrated by Wians et al. (15), is that the assay responds disproportionately to the amount of tPSA in a sample at a rate proportional to the amount of fPSA. The value of the assay-reported PSA can be calculated according to the following equation:

\[
\text{Assay-reported PSA} = \text{true PSA} + \text{fPSA} \times r
\]

where \( r \) is the rate at which the assay responds to the amount of fPSA. This assay-reported PSA value is then substituted for “true PSA” in the laboratory variation equation described above. Scenarios investigated proportions of fPSA of 10%, 20%, and 50%, along with assay response rates of 10% and 25%, as estimated from Wians et al. (15).

**Outcome measurements**

Because age-specific population distributions of tPSA were calibrated according to the population of the United Kingdom, age-related PSA reference intervals in the United Kingdom were adapted, such that men 45–54, 55–69, and 70–84 years of age were recommended for a biopsy if their PSA was >3, 4, or 5 \( \mu g/L \), respectively.

To measure the effect of bias and nonequimolarity on the recommendation to carry out a prostate biopsy, for each of the 1000 simulations the following were recorded for each age group separately:

- The proportion of men whose true PSA was below the age-related cutoff but whose observed PSA was above it (a false-positive); and
• The proportion of men whose true PSA was above the age-related cutoff but whose observed PSA was below it (a false-negative).

To minimize the effects of sampling errors, these estimated rates were averaged over the 1000 simulations in each age group for each combination of assay variability, bias, and nonequimolarity. The effect of differing amounts of bias and nonequimolarity on the average estimated false-positive and -negative biopsy recommendation rates was compared with that of an unbiased equimolar response assay. These were translated into proportions of men undergoing a PSA test who would have incorrect biopsy recommendations based on the size of the population of the United Kingdom in 2001 (16) and national PSA testing rates, estimated in 1999 (14). All PSA tests were assumed to be performed with the same assay.

**Sensitivity Analysis**

To assess the sensitivity of the results to changes in the age-specific PSA distributions, we performed several sensitivity analyses in which the relative proportions of men in each of the three age groups who would be considered test positive were changed. To assess the impact of using age-specific PSA cutoffs, we performed sensitivity analyses using single PSA cutoffs of 4 and 3 μg/L for each age group. In each sensitivity analysis, all combinations of nonequimolar and biased responses were used, and outcomes were recalculated.

**Results**

Using the estimate of the population of the United Kingdom in 2001 from the Office for National Statistics and age-specific PSA testing rates of 1.5%, 4.2%, and 5.3% for men 45–54, 55–69, and 70+ years, respectively, in 1999, we determined that ~325 000 PSA tests are carried out annually in the United Kingdom. On the basis of the assumed age-specific distributions, ~9%, 16%, and 26% of the PSA tests in the 45–54, 59–60, and 70–84 age groups, respectively, will be over the age-specific PSA cutoffs, leading to ~60 000 biopsy recommendations per year (an overall biopsy rate of 18.5%).

If a calibrated and equimolar assay was used, laboratory and sample handling errors of 5%, 7.5%, and 10% produced false-positive and -negative biopsy recommendation rates of 0.5%, 0.7%, and 0.9%, respectively, which is equivalent to 1600, 2300, and 3000 false-positive biopsies and an equivalent number of false-negative biopsies.

The effects of biased and nonequimolar assays on PSA values in terms of the proportions of men that would be wrongly recommended or failed to be recommended for a biopsy for a constant laboratory and sample handling CV of 7.5% are shown in Table 1. Other values of laboratory CV (5% or 10%) produce similar results when some degree of bias or nonequimolarity is present (data not shown). In the case of an equimolar response assay, as the proportion of bias increases, the false-negative rate falls to 0% very rapidly, whereas the false-positive rate steadily increases to 5.6% with a positive bias of 25%. Negative

---

**Table 1. Effects of assay bias and nonequimolarity on incorrect biopsy recommendations assuming laboratory error of 7.5%.

<table>
<thead>
<tr>
<th>Bias</th>
<th>fPSA:tPSA ratio</th>
<th>False-positive rate, %</th>
<th>False-negative rate, %</th>
<th>False-positive rate, %</th>
<th>False-negative rate, %</th>
<th>False-positive rate, %</th>
<th>False-negative rate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>10%</td>
<td>25%</td>
<td>25%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>-25%</td>
<td>0.10</td>
<td>0.0</td>
<td>5.8</td>
<td>0.0</td>
<td>3.2</td>
<td>0.0</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.0</td>
<td>5.8</td>
<td>0.4</td>
<td>1.1</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.0</td>
<td>5.8</td>
<td>6.7</td>
<td>0.0</td>
<td>8.3</td>
<td>0.0</td>
</tr>
<tr>
<td>-10%</td>
<td>0.10</td>
<td>0.0</td>
<td>2.4</td>
<td>0.8</td>
<td>0.6</td>
<td>1.0</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.0</td>
<td>2.4</td>
<td>2.8</td>
<td>0.1</td>
<td>3.4</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.0</td>
<td>2.4</td>
<td>9.8</td>
<td>0.0</td>
<td>11.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Unbiased</td>
<td>0.10</td>
<td>0.7</td>
<td>0.7</td>
<td>2.6</td>
<td>1.0</td>
<td>2.9</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.7</td>
<td>0.7</td>
<td>4.9</td>
<td>0.0</td>
<td>5.6</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.7</td>
<td>0.7</td>
<td>11.8</td>
<td>0.0</td>
<td>13.3</td>
<td>0.0</td>
</tr>
<tr>
<td>10%</td>
<td>0.10</td>
<td>2.4</td>
<td>0.1</td>
<td>4.7</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>2.4</td>
<td>0.1</td>
<td>7.1</td>
<td>0.0</td>
<td>7.7</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>2.4</td>
<td>0.1</td>
<td>13.8</td>
<td>0.0</td>
<td>15.2</td>
<td>0.0</td>
</tr>
<tr>
<td>25%</td>
<td>0.10</td>
<td>5.6</td>
<td>0.0</td>
<td>7.9</td>
<td>0.0</td>
<td>8.3</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>5.6</td>
<td>0.0</td>
<td>10.2</td>
<td>0.0</td>
<td>10.8</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>5.6</td>
<td>0.0</td>
<td>16.6</td>
<td>0.0</td>
<td>18.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

* Biopsies incorrectly recommended.

* Biopsies incorrectly not recommended.
amounts of bias have the opposite effect, such that an assay that systematically underreports PSA values leads to large increases in the false-negative rates and decreases in the false-positive rate. However, a constant bias across all PSA concentrations rather than a bias proportionate to PSA concentration has little impact on the estimated false-negative or -positive rates in Table 1 (results not shown).

In the case of an unbiased, but nonequimolar assay, the false-positive rates increase to 11.8% and 13.3% for assay response rates of 10% and 25%, respectively, as the fraction of tPSA in the sample increases to 50% (Table 1). Nonequimolar assays will always overestimate the amount of tPSA in a sample and therefore reduce the number of false-negative recommendations; in most cases examined in this analysis, the false-negative recommendation rates were 0%.

Use of both a biased and a nonequimolar assay greatly inflates the rate of false-positive biopsy recommendation rates from ~0.7% to 18% in the most extreme case (Table 1). In all scenarios, the false-negative rate rapidly decreases to 0, meaning that biased and nonequimolar assays rarely underestimate the tPSA in a sample and will tend to overestimate the observed tPSA concentrations, leading to a false recommendation for biopsy.

Sensitivity analyses were conducted in which 7%, 10%, and 20% of the PSA tests in the 45–54, 59–60, and 70–84 age groups, respectively, were assumed to be over the age-specific PSA cutoffs. Even with this considerable change to the proportions of men regarded as test-positive, the false-positive and -negative biopsy recommendation rates remained very similar (within 1–2% of those shown in Table 1; results not shown). Although use of a single PSA cutoff (either 3 or 4 µg/L) rather than age-specific cutoffs produced different estimates of the false-positive and -negative rates, the patterns seen in Table 1 remained (data not shown).

Discussion
The results from this simulation study show that small deviations in bias and nonequimolarity can lead to significant increases in the number of false-positive biopsy recommendations; only assays that have a negative bias increase the false-negative rate. The results therefore demonstrate the importance of using a calibrated, unbiased equimolar assay to measure serum PSA values. The pattern of rates seen in Table 1 was shown to be robust to changes in key model assumptions by several sensitivity analyses. Although the simulations were carried out under the assumption that all PSA assays would use the same assay method, the estimated rates can be applied to the number of tests carried out with a particular assay to estimate the numbers of false-positive and -negative recommendations that it would produce. The results of this study are consistent with two previous reports that showed that the use of an equimolar assay can reduce the number of patients having an unnecessary biopsy by up to 10% (10, 11). Moreover, it has been shown that reported estimates of diagnostic ability are dependent on assay choice (17); thus, combining results from different assays can make generalizations difficult (18).

From a clinical viewpoint it is important to minimize the effects of positive bias and nonequimolar responses to reduce the number of false-positive recommendations because this will lead to a lower cost burden for the healthcare provider and will reduce unnecessary stress and discomfort for patients. Assays that are negatively biased will substantially increase the number of false-negative biopsy recommendations and will lead to many men not being recommended for a biopsy when their true PSA is above the cutoff; in this case, many cancers will potentially be missed.

Even when an unbiased, equimolar assay is used, sample handling and laboratory variability can lead to misclassification rates of 5–10%. Such random variability will mean that some men whose true values are close to the cutoff will be incorrectly classified. Although we were able to quantify the effects of this source of random variation on biopsy recommendation rates, other sources of variation may also have significant effects on both true and observed PSA values. PSA values are known to vary within an individual over time, and a recent metaanalysis suggested that intraindividual variation is associated with a CV of ~12% (19). Although this additional source of variation is clearly significant, to isolate the effects of using biased or nonequimolar assays, we assumed that the simulated value of PSA represents an individual’s true PSA at the time of sampling. For the same reasons, this study did not attempt to account for the effects of other potential diagnostic tools in reaching a final recommendation for biopsy. In practice, results of a digital rectal exam, symptomatology of the patient, or a strong family history of disease may provide additional evidence to support or not support the recommendation for a biopsy.

This study used PSA testing rates based on a survey conducted in general practice in the United Kingdom in 1999 (14). However, there was no information about the nature of the population undergoing tests, and it is possible that these men were more likely than the general population to be symptomatic. Indeed, 16% of men 55–69 of age had a PSA >4 µg/L compared with a mean of 11% (range, 7–15%) in the worldwide prostate cancer screening trials, which recruited men >55 years of age (1). In the general population in the United Kingdom with a higher proportion of asymptomatic men, the PSA distribution would certainly be expected to be lower. However, a more recent survey of PSA testing in general practices in the United Kingdom suggests that the proportions of men 50–59 and 60–69 years of age with a PSA concentration >4 µg/L are 9.7% and 18.8%, respectively; therefore, the PSA distribution in the 55–69 age group is unlikely to have changed significantly (20). These revised estimates fall within the bounds of the sensitivity analyses considered here and suggest that any changes in the numbers of PSA tests performed will lead to an almost proportionate increase
increase in the absolute numbers of false-positive and -negative recommendations.

In conclusion, this analysis has shown the effect that using biased and nonequimolar PSA assays can have on false-positive and -negative biopsy recommendation rates. It strongly supports the assertion that assays for PSA and its fractions must be calibrated to the International Standards and be unbiased and equimolar in their response. These results highlight the importance of minimizing laboratory variation; although the false-recommendation rates are low in the case of an unbiased equimolar assay, even small increases in assay CVs can lead to significant increases in false-negative and -positive rates. Although this analysis is specific to PSA testing, these conclusions are relevant to other diagnostic tests. Efforts to reduce assay CV and to ensure accurate calibration against relevant International Standards are essential to minimize the likelihood of making incorrect clinical decisions that are potentially detrimental for both the patient and healthcare provider.

Appendix

This work was commissioned by the PSA Isoform Working Group, which reports to the Scientific Reference Group of the NHS Prostate Cancer Risk Management Program:

Dr. Naomi E. Allen, Cancer Research UK Epidemiology Unit, University of Oxford, Oxford, United Kingdom.
Dr. Michael J. Duffy, St. Vincent’s University Hospital, Dublin, United Kingdom.
Professor Freddy C. Hamdy, Department of Urology, University of Sheffield, Sheffield, United Kingdom.
Dr. Anthony Millford Ward, University of Sheffield, Sheffield, United Kingdom.
Julietta Patnick, National Health Service Cancer Screening Programs.
Professor Christopher P. Price, Bayer HealthCare and University of Oxford, Oxford, United Kingdom.
Janet Rimmer, National Health Service Cancer Screening Programs.
Dr. Andrew W. Roddam, Cancer Research UK Epidemiology Unit, University of Oxford, Oxford, United Kingdom.
Dr. Cathie Sturgeon, Royal Infirmary, Edinburgh, United Kingdom.
Peter White, United Kingdom National External Quality Assessment Scheme for Immunochemistry & Immunology.

References