Prostate-Specific Antigen and Free Prostate-Specific Antigen in the Early Detection of Prostate Cancer: Do Combination Tests Improve Detection?

Ruth Etzioni,[1] Seth Falcon,[1] Peter H. Gann,[2] Charles L. Kooperberg,[1] David F. Penson,[3] and Meir J. Stampfer[4]

[1]Fred Hutchinson Cancer Research Center, Seattle, Washington; [2]Feinberg School of Medicine, Northwestern University, Chicago, Illinois; [3]Department of Urology, University of Washington, Seattle, Washington; and [4]Harvard School of Public Health, Boston, Massachusetts

Abstract

Background: The combined use of free and total prostate-specific antigen (PSA) in early detection of prostate cancer has been controversial. This article systematically evaluates the discriminating capacity of a large number of combination tests. Methods: Free and total PSA were analyzed in stored serum samples taken prior to diagnosis in 429 cases and 1,640 controls from the Physicians' Health Study. We used a classification algorithm called logic regression to search for clinically useful tests combining total and percent free PSA and receiver operating characteristic analysis and compared these tests with those based on total and complexed PSA. Data were divided into training and test subsets. For robustness, we considered 35 test-train splits of the original data and computed receiver operating characteristic curves for each test data set. Results: The average area under the receiver operating characteristic curve across test data sets was 0.74 for total PSA and 0.76 for the combination tests. Combination tests with higher sensitivity and specificity than PSA > 4.0 ng/mL were identified 29 out of 35 times. All these tests extended the PSA reflex range to below 4.0 ng/mL. Receiver operating characteristic curve analysis indicated that the overall diagnostic performance as expressed by the area under the curve did not differ significantly for the different tests. Conclusions: Tests combining total and percent free PSA show modest overall improvements over total PSA. However, utilization of percent free PSA below a PSA threshold of 4 ng/mL could translate into a practically important reduction in unnecessary biopsies without sacrificing cancers detected. (Cancer Epidemiol Biomarkers Prev 2004;13(10):1640–5)

Introduction

What is the best prostate-specific antigen (PSA)–based test for the early detection of prostate cancer? This question has tantalized researchers since PSA was introduced. The sensitivity of the standard PSA-based test (positive if PSA > 4.0 ng/mL) is ~70% to 80% among men within 4 years prior to clinical diagnosis of prostate cancer, and the overall specificity is close to 90% (1). However, false-positive tests are not uncommon, particularly among older men and those with benign prostate conditions. This phenomenon argues for a more stringent, or specific, test. At the same time, several studies have established the presence of prostate cancer in some men with PSA levels below 4.0 ng/mL (2), suggesting a need for a more sensitive test (3).

Recent attempts to improve the performance of PSA have focused on the different molecular forms of PSA in serum: total PSA (TPSA), free PSA (not complexed to serum proteins), and complexed PSA (CPSA). Because the ratio of free PSA to TPSA (RPSA) tends to decline in men with prostate cancer, combination tests have generally used a threshold for RPSA within an interval of moderately elevated TPSA values, termed the reflex range. Early studies focused on a reflex range for TPSA of 4 to 10 ng/mL, with the goal of reducing false-positive rates (4, 5). Subsequent studies suggested that RPSA might be useful when TPSA is even lower than 4.0 ng/mL (2). A recent report by Gann et al. (6) observed that use of RPSA within a TPSA reflex range of 3 to 10 ng/mL could actually improve both specificity and sensitivity simultaneously relative to the conventional test. Uncertainty about the optimal reflex range has been compounded by recent results suggesting that CPSA may be preferable to both TPSA and RPSA (7). However, results concerning the utility of CPSA are also not consistent across studies (8). Current PSA guidelines do not, to our knowledge, provide any direction as to how free PSA and CPSA should be used in the early detection of prostate cancer.

In this article, we undertake a systematic analysis of the diagnostic performance of different strategies based on TPSA, CPSA, and tests combining free PSA with TPSA in banked plasma samples from the Physicians' Health Study (1, 6). This nested case-control study represents one of the earliest and most extensive sources of

Received 1/22/04; revised 4/1/04; accepted 5/3/04.

Grant support: National Cancer Institute grants CA42182, CA58684, and CA57374 (Physicians' Health Study and the Prostate-Specific Antigen Substudy); GM54438 and CA97186 (R. Etzioni); and CA74841 (C.L. Kooperberg).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Ruth Etzioni, Program in Biostatistics, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M2-B230, Seattle, WA 98109-1024. Phone: 206-667-6561; Fax: 206-667-7004. E-mail: retzioni@fhcrc.org

information on serum PSA levels prior to diagnosis of prostate cancer.

Our analysis differs from prior studies in that we do not begin by selecting a specific reference range for TPSA or a threshold for RPSA within this range. Rather, our goals are to (*a*) systematically evaluate a wide range of clinically interpretable tests combining TPSA and RPSA and (*b*) determine whether this class of tests provides significant improvements in diagnostic performance relative to TPSA-based tests.

The ability to optimally combine information on multiple markers is important because single markers typically lack the sensitivity and specificity to be useful for mass screening. With genomic and proteomic studies yielding many novel markers for cancer detection (9), a statistically coherent framework will be needed to identify useful combination tests and to evaluate whether these tests provide statistically and clinically significant improvements over existing tests. The methods presented herein represent a broadly applicable framework that addresses this need.

Materials and Methods

The Physicians' Health Study. The Physicians' Health Study (6) was a randomized, placebo-controlled trial of aspirin and β-carotene among 22,071 U.S. male physicians ages 40 to 84 years in 1982. At enrollment, participants provided a blood sample, which was stored. The stored serum from 430 men later diagnosed with prostate cancer was subsequently reassayed for PSA and free PSA using the Tandem-R immunoradiometric assay (Hybritech, Inc., San Diego, CA). Cases were diagnosed up to 12 years after their serum had been sampled; most were diagnosed before the widespread dissemination of PSA screening in the population. TPSA and RPSA measurements were available for these cases and for 1,642 age-matched controls who had not been diagnosed with prostate cancer for up to 12 years of follow-up. A separate CPSA assay was not done at the time of the study; therefore, we approximate CPSA levels by the difference between TPSA and free PSA. In addition, information on other tests such as digital rectal exam was not available for the controls.

Statistical Analysis

Overview. Our analytic approach consists of two key components: (*a*) identification of potentially useful TPSA/RPSA combination tests and (*b*) statistical comparison of these tests with tests based on TPSA and CPSA. The second component is a comparison of receiver operating characteristic (ROC) curves (10) for the different types of tests. Statistical methods that extend this technique to tests using RPSA within intervals of TPSA have only recently been developed (11). In addition to estimating ROC curves for the three different types of tests considered (TPSA, CPSA, and TPSA/RPSA combination), we also evaluate the impact of time prior to diagnosis and subject age on the relative performance of the tests. This allows us to address, for example, whether tests that include information on percent free PSA can identify prostate cancer cases earlier than those based on TPSA.

Definition of Combination Tests. In combining information on TPSA and RPSA, we consider the set of *and-or* combinations of tests in each marker. We refer to tests of this type as *logic rules*. Logic rules are of particular interest because of their flexibility and clinical interpretability. The test which uses RPSA within a specified TPSA reflex range, is an example of a logic rule; however, the set of logic rules is far more general. In practice, we define a set of possible cutoffs for TPSA and RPSA and consider the collection of logic rules based on these cutoffs. For TPSA, we use cutoffs {1, 1.25, 1.5, 1.75, . . ., 9.75, 10} where all measurements are in nanogram per milliliter. For RPSA, we define cutoffs of {0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4}. Thus, the set of combination rules that we consider consists of all *and-or* combinations of threshold conditions in TPSA and RPSA using these cutoffs.

ROC Analysis. To construct the logic rule ROC curve, we use the methods described by Etzioni et al. (11). With a single marker, each point on the ROC curve represents the true-positive rate for a specific marker threshold versus the false-positive rate for that threshold. With multiple markers, every point on the ROC curve corresponds to a different rule (11). Each rule on the curve maximizes the true-positive rate given the corresponding false-positive rate; this collection of rules is ''optimal'' in the sense that for any rule that is not on the curve, there exists a rule on the curve that has higher true-positive and lower false-positive rates (12). The rules are selected by a classification algorithm called logic regression (13).

To obtain an assessment of comparative predictive performance that is not overly optimistic, the logic rules are identified using a training data set, consisting of a randomly selected two-thirds of the original sample. We use a test data set, consisting of the remaining one-third, to evaluate the corresponding true-positive and false-positive rates and construct the ROC curve. We also construct ROC curves for CPSA and TPSA on the test data. For robustness, we implement analyses for 35 different runs, each corresponding to a different test-train split. In general, we present results from all the runs; where necessary (e.g., in plots of the ROC curves), we present results for the run in which the area under the logic rule ROC curve (AUC) is the median over all the runs.

Estimating the AUC. To test whether apparent differences in the ROC curves are statistically significant, we compare the AUCs. The AUC is a general measure of diagnostic performance, interpretable as an average true-positive rate over the full range of false-positive rates (10). For multiple markers, the AUC is interpretable as the probability that any pair of case-control observations will be correctly classified by at least one rule on the curve.

Because the AUC is interpretable as a probability, logistic regression may be used to determine whether it differs according to test type (TPSA, CPSA, logic rule; refs. 11, 14). For example, to compare TPSA with CPSA, an indicator of test type (1 for TPSA, 0 for CPSA) would be entered as an independent variable in the appropriate logistic regression model (11, 14). SEs for the regression coefficients can be determined by bootstrapping (14). To evaluate whether independent variables such as age and time from test to diagnosis affect the relative diagnostic

performance of the different tests, we include interactions between these factors and indicators of test type in the regression models.

In practice, we conduct three separate logistic regression analyses, the first comparing TPSA with CPSA, the second comparing TPSA with the TPSA/RPSA combination, and the third comparing CPSA with the combination rule. Each test-train split of the data yields a different set of results for each analysis. We summarize results by reporting mean coefficient estimates as well as the number of times for which coefficients of interest are statistically significant. A result that is consistently significant across runs indicates a robust association of the corresponding covariate with the AUC. All statistical significance tests are conducted at the (two-sided) 0.05 level.

Results

Table 1 summarizes key characteristics of cases and controls. TPSA was significantly higher among cases ($P < 0.01$, Wilcoxon rank sum test), as was CPSA ($P < 0.01$). RPSA was significantly lower among cases ($P < 0.01$). These differences were observed in spite of the median time from test to diagnosis being 8 years.

Figure 1 provides a scatter plot of the TPSA and RPSA results for the cases and controls in the study. For display purposes, we have plotted data from a random 30% of controls and have cut off the horizontal axis at a TPSA value of 20 ng/mL; 17 cases and 8 controls had TPSA values above 20 ng/mL. All of these cases and six of the eight controls also had RPSA values below 0.2. The plot shows that a substantial proportion of cases have TPSA values below the conventional cutoff of 4.0 ng/mL and that several controls have TPSA values above this cutoff. However, we note that the cases with TPSA below 4.0 ng/mL tend to have longer time intervals between testing and diagnosis than those with PSA above 4.0 ng/mL (7.5 versus 9.1 years on average; $P < 0.001$) and the controls with TPSA above 4.0 ng/mL are older than those with lower TPSA values (60.2 versus 65.1 on average; $P < 0.001$).

Figure 1 provides a graphical illustration of the difference between the standard PSA-based test, a logic rule [the one identified by Gann et al. (6)], and a CPSA-based test (CPSA > 3.75 ng/mL). Each test splits the marker space into a test-positive region and a test-negative region. The TPSA-based test is a vertical line, the logic rule is a step-shaped line, and the CPSA-based test is an arc. Different cutoffs for CPSA move this arc across the plot, redefining the test-positive and test-negative regions. For both the CPSA-based test and the logic rule, points with moderately elevated TPSA and high RPSA values are classified as negative.

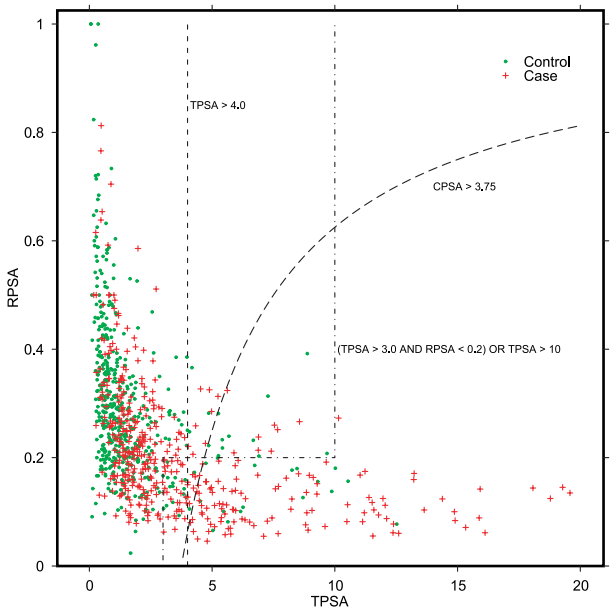

Figure 1. Scatter plot of the TPSA/RPSA data for study participants together with the conventional TPSA-based rule, the curve CPSA = 3.75, and the rule identified by Gann et al. (6). For display purposes, we have plotted data from a random 30% of controls and have cut off the horizontal axis at TPSA = 20 ng/mL. A total of 17 cases and 8 controls had TPSA values above 20 ng/mL. All of these cases and six of the eight controls also had RPSA values below 0.2.

Table 1. Characteristics of cases and controls

	Cases (n = 429)	Controls (n = 1,640)
Mean (SD) age at test (y)	60.3 (7.22)	60.7 (7.23)
Mean (SD) TPSA at test (ng/mL)	5.50 (11.08)	1.84 (2.68)
Number with TPSA > 4.0 (ng/mL; %)	147 (34.3)	144 (8.8)
Number with TPSA > 10.0 (ng/mL; %)	49 (11.4)	21 (1.3)
Mean (SD) RPSA level at test	0.20 (0.12)	0.30 (0.15)
Number with RPSA < 0.2 (%)	247 (57.6)	403 (24.5)
Mean (SD) CPSA level	4.8 (10.33)	1.4 (2.36)
Number with CPSA > 3.75 (%)	131 (30.5)	117 (7.1)
Mean time from test to diagnosis (y)	8.57 (2.62)	NA

NOTE: Data from cases and controls with both TPSA and RPSA values available from serum drawn at the time of enrollment.

Figure 2A plots the TPSA, CPSA, and logic rule ROC curves for the test data from a representative run, namely, one in which the AUC was closest to ''average'' across the 35 runs. We considered the following thresholds when plotting the ROC curves for the TPSA-based and CPSA-based rules: {0.25, 0.5, 0.75, 1, 1.25, 1.5, . . ., 10}. Because high specificity is important in cancer screening studies, Fig. 2B also shows the ROC curves restricted to false-positive rates below 20%. Table 2 lists the logic rules on which Fig. 2B is based.

In Fig. 2B, the test TPSA > 4.0 ng/mL has a false-positive rate of 10.1% and a true-positive rate of 36%. Divergence of the ROC curves in this region indicates that there exist combination tests with improved diagnostic performance relative to the standard TPSA-based test. Figure 2B shows that two logic rules (rules 6 and 7 in Table 2) and one CPSA-based rule (CPSA > 3.5 ng/mL) have both lower false-positive rates and higher true-positive rates than the standard TPSA-based test. In 29 of 35 runs (83%), we identified logic rules for which both sensitivity and specificity were at least as high as the standard test, which had true-positive rates ranging from 24% to 41% across runs and false-positive rates

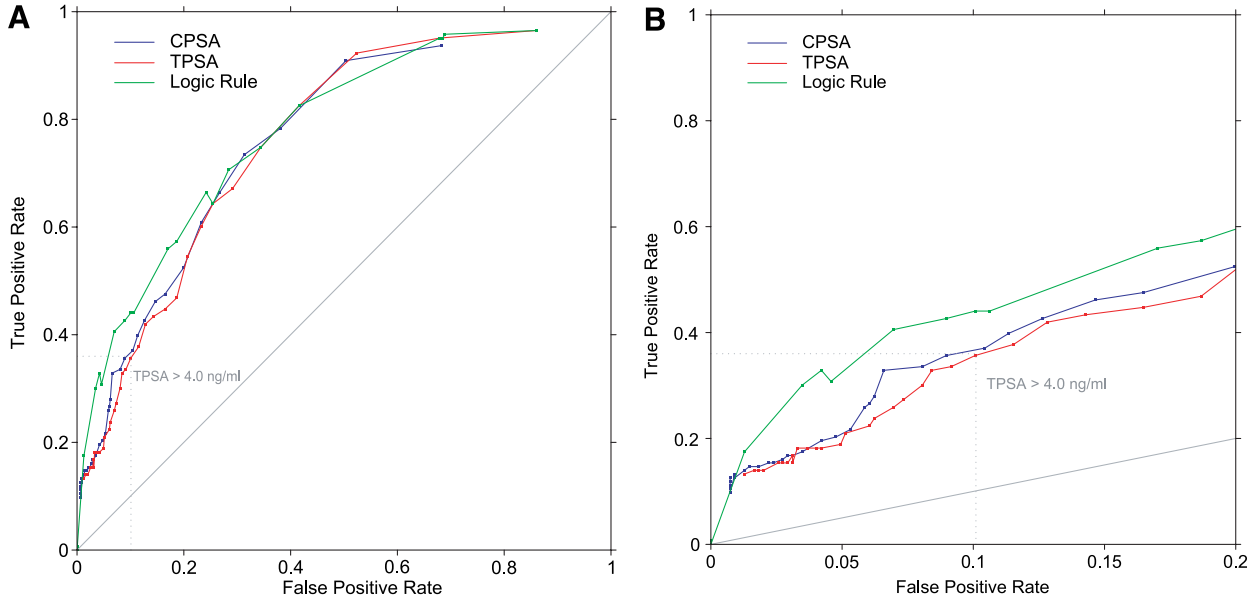

Figure 2. ROC curves for TPSA, CPSA, and the logic combination rule (test data). Results are based on the test-train split under which the AUC for the logic rule was the median across all 35 test-train splits. **A.** False-positive rates range from 0 to 1. **B.** False-positive rates range from 0 to 0.2. The test TPSA > 4.0 ng/mL has false-positive rate of 10.1% and true-positive rate of 36%. Thus, two logic rules (rules 6 and 7 in Table 2) and one CPSA-based rule (CPSA > 3.5 ng/mL) have both lower false-positive rates and higher true-positive rates than the standard TPSA-based test.

ranging from 6% to 11%. On average, these logic rules led to a 2.3% decrease in the false-positive rate and a 3% increase in the true-positive rate relative to TPSA > 4.0 ng/mL. The logic rules with higher sensitivity and specificity than TPSA > 4.0 ng/mL all extended the TPSA reflex range to below 4.0 ng/mL, with RPSA thresholds in varying between 0.1 and 0.25. Similarly, in 18 of 35 runs (51%), we identified tests based on CPSA with sensitivity and specificity at least as high as TPSA > 4.0 ng/mL. The CPSA thresholds for these rules ranged from 3.0 to 3.5 ng/mL. On average, these rules led to a 0.6% decrease in the false-positive rate and a 1.1% increase in the true-positive rate relative to TPSA > 4.0 ng/mL. Consistent with previous studies (e.g., ref. 6), we also identified several combination tests that, by slightly lowering the TPSA reflex range, substantially reduced false-positive rates (up to 50%) with small losses in sensitivity.

The AUCs for TPSA across the 35 runs ranged from 0.70 to 0.78, with a mean of 0.74. The average AUC for the logic rule and CPSA was 0.76. In the logistic regressions, interaction terms were rarely statistically significant, with the exception of the TPSA-CPSA comparison where the test type: age interaction term was significant in 14 runs, suggesting that any improvements in diagnostic performance associated with CPSA might be restricted to older men. Results are presented in Table 3, which indicates similar diagnostic performance (as measured by the AUC) for the three types of tests. For example, in the comparison of TPSA with the logic rule, the coefficient for test type was statistically significant in only 3 of 35 runs; similarly, for CPSA and the logic rule, the indicator of test type was statistically significant in only 10 runs. The coefficient estimates for the CPSA-TPSA comparison suggest a slight degradation in diagnostic performance associated with the use of CPSA in younger men and a corresponding improvement in older men.

Table 2. Logic rules identified from the training data corresponding to the logic rule ROC curve in Fig. 2B (i.e., false-positive rates ≤ 0.2) together with estimates of sensitivity and specificity from the test data

	Logic rule	False-positive rate (%)	True-positive rate (%)
1	TPSA > 2.75 and RPSA ≤ 0.05	0	0.7
2	TPSA > 2.25 and RPSA ≤ 0.1	1.28	17.5
3	TPSA > 3.5 and RPSA ≤ 0.15	3.48	30.1
4	(TPSA > 3.5 and RPSA ≤ 0.15) or (1.5 < TPSA ≤ 3.5 and RPSA ≤ 0.1)	4.21	32.9
5	TPSA > 3 and RPSA ≤ 0.15	4.58	30.8
6	(TPSA > 3.5 and RPSA ≤ 0.2) or (1.5 < TPSA ≤ 3.5 and RPSA ≤ 0.1)	6.96	40.6
7	(1.5 < TPSA ≤ 8.5 and RPSA ≤ 0.15) or TPSA > 8.5	8.97	42.7
8	(1.5 < TPSA ≤ 6.75 and RPSA ≤ 0.15) or TPSA > 6.75	10.07	44.1
9	(2 < TPSA ≤ 2.25 or TPSA > 6.75) or (2.25 < TPSA ≤ 6.75 and RPSA < 0.15)	10.62	44.1
10	TPSA > 2 and RPSA ≤ 0.25	17.03	55.9
11	(2 < TPSA ≤ 4.25 and RPSA ≤ 0.25) or TPSA > 4.25	18.68	57.3

NOTE: Results are based on the test-train split under which the AUC for the logic rule was the median across all 35 test-train splits. By comparison, on this test data set, the standard TPSA-based rule (positive if TPSA ≥ 4.0 ng/mL) had a false-positive rate of 0.10 and a true-positive rate of 0.36.

Table 3. Results of logistic regression analyses comparing AUCs for different rules (test data)

Independent variable	Coefficient estimate [Average (interquartile range*)]	*Z* statistic [No. runs with $P < 0.05$]
A: TPSA compared with logic rules		
Test type		
TPSA	Baseline	
Logic	0.21 (0.14, 0.28)	3
Time from test to diagnosis (y)	−0.11 (−0.14, −0.08)	21
Age at time of test		
≤60	Baseline	
>60	−0.13 (−0.27, 0.02)	4
B: TPSA compared with CPSA		
Test type		
TPSA	Baseline	
CPSA	−0.11 (−0.14, −0.08)	21
Time from test to diagnosis (y)	−0.11 (−0.14, −0.08)	20
Age at time of test		
≤60	Baseline	
>60	−0.21 (−0.37, −0.05)	8
Test type × age	0.11 (0.07, 0.15)	14
C: CPSA compared with logic rules		
Test type		
CPSA	Baseline	
Logic	0.27 (0.20, 0.35)	10
Time from test to diagnosis (y)	−0.11 (−0.15, −0.08)	21
Age at time of test		
≤60	Baseline	
>60	−0.07 (−0.21, 0.08)	2

NOTE: Results from 35 test-train splits are presented. SDs for coefficient estimates are estimated based on 250 bootstrap samples. A positive coefficient estimate indicates that an increase in the independent variable is associated with an increase in the AUC (i.e., an improvement in diagnostic performance). The coefficient values for the test type variable are interpretable as follows: Exp(coefficient) gives the amount by which the AUC odds [AUC/(1 − AUC)] are increased for the given test type relative to the baseline test type (14). Interaction terms in A and C were rarely significant, so only models with main effects are presented. A positive (negative) coefficient implies that increasing values of the corresponding covariate are associated with better (poorer) diagnostic performance.

*The elements of the tuple are the 25th and 75th quintiles, respectively.

Discussion

In this study, we have presented a systematic evaluation of the performance of tests based on TPSA, CPSA, and the combination of TPSA and percentage free PSA. Our results suggest that tests using information on RPSA (whether as logic rules or as CPSA-based tests) do not provide substantially better discriminating power *in general* than tests based solely on TPSA. Our findings concerning the utility of CPSA versus TPSA mirror those of Partin et al. (15) and Okihara et al. (16) but differ from those of Brawer et al. (7), who compared 385 men with negative biopsies and 272 men with biopsy-proven prostate cancer and found that the AUC for CPSA was significantly greater than for TPSA.

Although the different rules showed similar overall diagnostic performance, the ROC curves indicated that specific combination tests could provide improvements over the standard TPSA-based test. Across test-train splits, we consistently identified logic combination tests with lower false-positive and higher true-positive rates than TPSA > 4.0 ng/mL. Given the wide prevalence of PSA testing in the population, use of these tests could translate into a practically important reduction in unnecessary biopsies without sacrificing cancers detected (6).

All of the TPSA/RPSA combinations with higher sensitivity and specificity than TPSA > 4.0 ng/mL extended the TPSA reflex range to below 4.0 ng/mL. Combination tests that improved specificity with only small losses in cancers detected also were of this form. This is consistent with several prior studies of TPSA and RPSA (6, 17, 18) as well as studies that have identified disease cases with TPSA levels below 4.0 ng/mL (2, 3, 17). Of note, these combination tests all used RPSA at low TPSA levels, indicating that simply lowering the threshold for TPSA, as has been recently suggested (3), may not be an optimal approach. If detection of cases with low PSA levels is important, but limiting false-positive tests is a priority, then our results suggest that a lowering of the TPSA threshold should also be accompanied by a threshold criterion on RPSA (or some other discriminating marker); otherwise, false-positive rates could become prohibitively high.

We found that lowering the TPSA threshold to 2.5 ng/mL, as has been suggested (3), led to an average false-positive rate of 18.9% and a corresponding true-positive rate of 50.5%. Assuming that the prevalence of latent, biopsy-detectable prostate cancer is 25%, this translates into 2.13 biopsies per cancer detected. In contrast, the logic rules that lowered the TPSA threshold but used RPSA in this range had false-positive rates of 6.91% and true-positive rates of 36.06% on average, which translates into 1.57 biopsies per cancer detected—a 26% reduction. Of note, the standard TPSA > 4.0 ng/mL rule led in our data set to average false-positive and false-negative rates of 10% and 36%, respectively, which translates into 1.83 biopsies per cancer detected. Thus, the logic rules that used RPSA within a lower TPSA reflex range reduced false-positive rates by 30% on average and could result in practice in a 37% reduction in the number of biopsies per cancer detected.

A key advantage of the Physicians' Health Study data set is that the majority of prostate cancers are clinically significant in the sense that they were at some point diagnosed prior to the PSA era, within the lifetime of the patient. The design of the present study (nested, case-control) contrasts with that of prospective screening studies (e.g., ref. 3), in which prostate cancer cases consist of men with a positive PSA and biopsy-detectable disease. The differences between the case populations in case-control and prospective screening studies lead to different definitions of sensitivity in the two types of studies, which may account for differences between study results. For example, the sensitivity of the test TPSA > 4.0 ng/mL among participants in the Physicians' Health Study within 4 years prior to diagnosis was estimated by Gann et al. (1) to be 73%; however, Punglia et al. (3) estimated sensitivity among prospectively screened cases to be only 19% for men younger than 60 and 35% for men over 60. In that our estimates of sensitivity pertain to cases whose disease will become apparent during their lifetimes (non-overdiagnosed cases), these estimates may be more relevant for clinical practice.

In this article, we have focused on diagnostic properties of PSA-based tests and not on the value of PSA testing in terms of its benefits—or costs. Given that some of the controversy about PSA testing centers on morbidity of false-positive tests, improving false-positive rates is clearly worthwhile—although there may be some cost implications associated with the additional tests. However, the value of improving true-positive rates is not clear, particularly in light of concerns about overdiagnosis associated with PSA screening. Although we identify tests that seem to provide modest improvements in sensitivity, our results pertain only to non-overdiagnosed cases. It is not clear whether these tests will increase the likelihood of overdiagnosis in a prospective screening setting, nor whether any such increases will be outweighed by the survival benefits that may accrue as a result of improved sensitivity.

To summarize, our findings indicate that discrimination between asymptomatic prostate cancer cases and controls may be enhanced by the use of information on the different molecular forms of PSA. The specific combination rules that outperform the standard TPSA-based rule in terms of both sensitivity and specificity all lower the reflex range for TPSA but use a threshold criterion for RPSA within this range. Our approach illustrates how use of multiple markers can be guided by systematic consideration of a wide range of combination tests coupled with a coherent statistical framework for evaluating and comparing diagnostic performance.

References

1. Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. JAMA 1995;273:289–94.
2. Catalona WJ, Smith DS, Ornstein DK. Prostate cancer detection in men with serum PSA concentrations of 2.6 to 4.0 ng/mL and benign prostate examination. Enhancement of specificity with free PSA measurements. JAMA 1997;277:1452–5.
3. Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM. Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. N Engl J Med 2003;349:335–42.
4. Catalona WJ, Smith DS, Wolfert RL, et al. Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. JAMA 1995;274:1214–20.
5. Partin AW, Catalona WJ, Southwick PC, Subong EN, Gasior GH, Chan DW. Analysis of percent free prostate-specific antigen (PSA) for prostate cancer detection: influence of total PSA, prostate volume, and age. Urology 1996;48:55–61.
6. Gann PH, Ma J, Stampfer MJ. Strategies combining total and percent free prostate specific antigen for detecting prostate cancer: a prospective evaluation. J Urol 2002;167:2427–34.
7. Brawer MK, Cheli CD, Neaman IE, et al. Complexed prostate specific antigen provides significant enhancement of specificity compared with total prostate specific antigen for detecting prostate cancer. J Urol 2000;163:1476–80.
8. Stamey TA, Yemoto CE. Examination of the 3 molecular forms of serum prostate specific antigen for distinguishing negative from positive biopsy: relationship to transition zone volume. J Urol 2000; 163:119–26.
9. Etzioni R, Urban N, Ramsey SD, et al. The case for early detection. Nat Rev Cancer 2003;3:243–52.
10. Pepe MS. The statistical evaluation of medical tests for classification and prediction. New York: Oxford University Press; 2003.
11. Etzioni R, Kooperberg C, Pepe MS, Smith R, Gann PH. Combining biomarkers to detect disease with application to prostate cancer. Biostatistics 2003;4:523–38.
12. Baker SG. Identifying combinations of cancer markers for further study as triggers of early intervention. Biometrics 2000;56:1082–7.
13. Ruczinski I, Kooperberg C, LeBlanc ML. Logic regression. J Comput Graph Stat 2003;12:475–511.
14. Dodd LE, Pepe MS. Semiparametric regression for the area under the receiver operating characteristic curve. J Am Stat Assoc 2003;98: 409–17.
15. Partin AW, Brawer MK, Subong EN, et al. Prospective evaluation of percent free-PSA and complexed-PSA for early detection of prostate cancer. Prostate Cancer Prostatic Dis 1998;1:197–203.
16. Okihara K, Cheli CD, Partin AW, et al. Comparative analysis of complexed prostate specific antigen, free prostate specific antigen and their ratio in detecting prostate cancer. J Urol 2002;167:2017–23; discussion 23–4.
17. Catalona WJ, Partin AW, Slawin KM, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. JAMA 1998;279:1542–7.
18. Reissigl A, Klocker H, Pointner J, et al. Usefulness of the ratio free/total prostate-specific antigen in addition to total PSA levels in prostate cancer screening. Urology 1996;48:62–6.