Prostate-specific antigen velocity (PSAV) is one of the oldest concepts in PSA screening, yet today it is one of the most controversial. Publication of a wide range of studies with different designs, study populations, and results has fueled uncertainty about the best way to use PSAV and confused the issue of its utility in the early detection setting. Studies of disease prognosis suggest that PSAV is strongly associated with lethal cancers. However, prospective screening trials find that PSAV is at best a weak predictor of high-risk disease. In this commentary, we synthesize and reconcile the evidence about the value of PSAV in the early detection setting. We review recent studies of PSAV and determine a set of statistical considerations that we believe to be critical in study evaluation and interpretation. We explain why the association between PSAV and disease-specific survival does not necessarily imply that PSAV will be a useful screening tool. In addition, we argue that the standard concept of PSAV—the absolute change in PSA per year—confuses disease aggressiveness with the interval from disease onset to detection. We therefore recommend that other methods be explored to incorporate information about PSA kinetics that could ultimately improve and even transform—how we detect and treat prostate cancer.


Since the beginning of the prostate-specific antigen (PSA) screening era, PSA velocity (PSAV) has been investigated for its use in early detection. Over time, our understanding of PSAV has evolved, and our use of this summary of PSA growth has changed. Early on, the primary concern about the PSA test was its lack of specificity in older men and men with benign prostatic hypertrophy (BPH). Initial studies of PSAV focused on reducing false-positive tests in these populations. In a well-known analysis of data from stored serum samples in the Baltimore Longitudinal Study of Aging (BLSA), Carter et al. (1,2) showed that the PSA growth rate was higher in cancer patients than in men with BPH and normal prostate and proposed a PSAV threshold of 0.75 ng/mL per year to distinguish prostate cancer from BPH in men with elevated PSA levels. Thus, PSAV was initially used to restrict the definition of a positive PSA test so as to reduce the likelihood of unnecessary biopsies.

With the passage of time, overdiagnosis and overtreatment have emerged as the most pressing problems of the PSA screening era, and substantial effort is being devoted to identifying markers that will reliably distinguish indolent cancers from those that should be treated. Although several tissue markers are being studied closely for their use in guiding treatment decisions following diagnosis, PSAV is currently under scrutiny as a serum marker that might be useful even earlier—at the point of biopsy referral.

The intense focus on PSAV as a screening biomarker is due to two recent developments. First, results from the Prostate Cancer Prevention Trial (PCPT) show clearly that no single PSA cutoff separates men at high risk of prostate cancer or high-grade disease from men at low risk (3). Rather, there is a continuum of risk, and the frequency of high-grade disease at low PSA levels can be non-trivial. Second, several recent studies have shown that high prediagnosis PSAV is strongly associated with poor disease-specific survival following diagnosis and have suggested that PSAV may be useful in identifying those men with low PSA values who are at risk of harboring a potentially lethal tumor. For example, D’Amico et al. (4,5) found that PSAV greater than 2.0 ng/mL per year in the year before diagnosis was independently associated with a dramatically increased risk of prostate cancer death (adjusted hazard ratios [HRs] = 10 and 12) following treatment with radical prostatectomy or external beam radiation therapy. In a recent study of BLSA data, Carter et al. (6) also found a strong association between survival and higher PSAV as early as 10–15 years before diagnosis (adjusted HR = 4 per 1.0 ng/mL per year increase in PSAV). Based on this finding, they proposed that a PSAV threshold of 0.35 ng/mL per year be used in screening men with low PSA levels to increase detection of potentially lethal tumors during the window of curability.

In response to these studies, the most recent guidelines from the National Comprehensive Cancer Network (http://www.nccn.org/professionals/physician_gls/PDF/prostate_detection.pdf) recommend that biopsy be considered if PSA is elevated or PSAV is greater than 0.5 ng/mL per year. Thus, current use of PSAV expands the definition of a positive PSA test and increases the likelihood of referral to biopsy.

Affiliations of authors: Translational and Outcomes Research Group, Fred Hutchinson Cancer Research Center, Seattle, WA (RDE); University of Washington, Seattle, WA (NSW, LYT); University of Texas Health Sciences Center at San Antonio, San Antonio, TX (DPA, IMT).

Correspondence to: Ruth D. Etzioni, PhD, Translational and Outcomes Research Group, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave North, M2-B230, PO Box 19024, Seattle, WA 98109 (e-mail: retzioni@fhcrc.org).

See “Funding” and “Notes” following “References.”

DOI: 10.1093/jnci/djm171
© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.
Despite findings linking prediagnosis PSAV with survival, the role of PSAV in early detection remains a matter of controversy, in large part due to results from prospective screening studies in which PSAV does not appear to add diagnostic value to the PSA level. In an analysis of PCPT data, Thompson et al. (3) found that when PSAV was used alone, it was a strong and independent predictor of the frequency of prostate cancer and high-grade disease; however, when PSAV was adjusted for PSA level and other predictors, it was no longer informative. Studies from the first two screening rounds of the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC) also found that PSAV did not improve accuracy when combined with PSA in the prospective setting (7,8). Finally, a recent analysis from the Prostate, Lung, Colon, and Ovarian (PLCO) cancer screening trial (9) shows that although PSAV was independently predictive of high-grade disease in this cohort, the inclusion of PSAV only slightly increased the area under the receiver operating characteristic curve (AUC) for prediction of high-grade tumors, from 0.626 to 0.646.

How can we reconcile these conflicting results? In this commentary, we highlight some key factors that must be considered when interpreting analyses of PSAV. We review a number of recently published high-profile articles on PSAV (3—9), distinguishing between prospective screening trials, such as the PCPT and ERSPC, and survival or prognostic studies, which may or may not include serial screening of all patients. We enumerate a set of statistical considerations that we believe to be critical in study evaluation and interpretation. We summarize our findings via six principles of interpretation.

Definitions
We define PSAV as the absolute increase in the PSA level from an earlier time ($t_0$) to a later time ($t_1$). Thus, if a man’s PSA level increases from 2.0 ng/mL at an initial test to 3.0 ng/mL 1 year later, his PSAV is 1.0 ng/mL per year. Other concepts of PSA change over time include the relative, or percent, change in PSA level and the PSA level doubling time. In this example, the percent change per year would be 50%, and under this relative change the doubling time would be 1.71 years.

Principles of Interpretation

PSA and PSAV are strongly correlated. This may seem obvious because the higher the absolute level of PSA, the greater the increase in PSAV. The PCPT (3), PLCO (9), and the recent study of Carter et al. (6) all found correlations in the vicinity of $r = .70$. This point can be illustrated by plotting PSAV versus PSA using representative data from the PCPT (Fig. 1). High PSAV is rare when PSA levels are low—only 3% of patients in the PCPT had PSAV values greater than 2.0 ng/mL per year. Thus, the question of whether PSAV provides independent information given PSA level is a valid one.

The strong correlation between PSA and PSAV arises because both increase as disease progresses (Fig 2). The three PSA trajectories in Fig. 2 have the same PSA doubling time, but at the point of diagnosis the associated PSAV values are quite different because the patients have prostate cancers that are at different points in their natural history. The figure shows that high PSAV can arise simply because disease has been progressing for some time. Indeed, PSAV may be an even more accurate reflection of the interval since disease onset than the absolute PSA level.

A high PSAV may be indicative of the mode of prostate cancer detection; the link between PSAV and mode of detection could partially explain results of prognostic studies. When cancers are heterogeneous in terms of their mode of detection...
were normal or only moderately elevated. Measurement, which could occur if the PSA levels in their first PSA test ascertained from stored serum or if they have more than one PSA measurement patients may still have observable PSAV if their PSA levels are diagnostic, and their survival includes only a short lead time. These latter patients may still have observable PSAV if their PSA levels are exponential before clinical diagnosis (1,12,13). During exponential growth, the percent change in PSA level and the corresponding doubling time are constant over time. In contrast, PSAV generally increases with time and, as previously mentioned, correlates with the absolute PSA level. Measuring PSA growth by doubling time or annual percent change would reduce the correlation between PSA level and the chosen measure of PSA change and would eliminate the confusion that results from not knowing whether a PSA level that is changing rapidly signifies an aggressive tumor or one that has been latent for a long time.

The lead time associated with PSA screening can be quite lengthy; Draisma et al. (10) have estimated a mean lead time of 11 years in the data from the Rotterdam section of the ERSPC. Observed survival among screen-detected cancers is the sum of the lead time and the survival from clinical diagnosis. A lengthy lead time will yield considerably higher observed survival time among screen-detected versus clinically detected cancers, resulting in an association between PSAV and life expectancy.

Because most survival/prognostic studies do not provide information on mode of detection, it is difficult to determine how much of a role it may have played in published results regarding the link between PSAV and survival. However, we note that there are major differences in the PSA and PSAV distributions across survival studies and prospective screening studies, suggesting that there may be corresponding differences in sample composition. For example, in the study of D’Amico et al. (5) of survival following radiation therapy, 91% of patients had PSA values greater than 4.0 ng/mL and 42% of patients had PSAV greater than 2.0 ng/mL per year. In contrast, in the prospectively screened PCPT cohort, 25% of patients had PSA values greater than 4.0 ng/mL and only 3% had PSAV greater than 2.0 ng/mL per year.

Some of the variation between studies can be explained by known differences in study design. For example, the PCPT excluded men with PSA levels greater than 3.0 ng/mL, had about 65% compliance with biopsy recommendations, and conducted an end-of-study biopsy. In contrast, the PLCO trial (9), in which 75% of patients had PSA values greater than 4.0 ng/mL at baseline, did not exclude participants based on PSA level and recorded only 40% receiving biopsy within 1 year (11). Little is known about the reasons for diagnosis in the cohorts in the D’Amico study (4,5), but the recent study of Carter et al. (6) included patients who were diagnosed between 1973 and 1996, and the BLSA instituted prospective PSA screening in 1991. Thus, the latter study likely included a mix of clinically detected and screen-detected cancers, with many of the deaths arising in the clinically detected subgroup.

We emphasize that even if the PSAV–survival association is partly due to mode of detection or lead-time bias, this does not invalidate the results of prognostic studies. However, it does raise the question of whether patients with higher PSAV have poorer observed survival because their cancers are detected at a later time point or because their tumors are truly biologically more aggressive. The answer to this question will ultimately determine whether PSAV is likely to be useful in the early detection setting.

How PSAV is measured may affect study results. Many different methods have been proposed for computing PSAV. Thompson et al. (3) consider 20 different formulations that include expressions for both absolute and relative change in PSA level. We prefer PSA doubling time or percent change for the following reason. Studies that use data gathered from stored serum samples have shown repeatedly that PSA levels in prostate cancer increase exponentially before clinical diagnosis (1,12,13). During exponential growth, the percent change in PSA level and the corresponding doubling time are constant over time. In contrast, PSAV generally increases with time and, as previously mentioned, correlates with the absolute PSA level. Measuring PSA growth by doubling time or annual percent change would reduce the correlation between PSA level and the chosen measure of PSA change and would eliminate the confusion that results from not knowing whether a PSA level that is changing rapidly signifies an aggressive tumor or one that has been latent for a long time.

To estimate the percent change, we would suggest that a linear regression model be fit to the logarithm of the observed PSA values. The percent change is given by 100 × (exp[β]−1), in which β is the coefficient in the linear regression equation. The corresponding doubling time is given by ln(2)/ln(β). The doubling time can be computed only if the observed PSA levels show an increasing trend over time.

Perhaps as important as the method for estimating PSAV is the interval over which PSAV is computed (14,15). As Pinsky et al. (9) note, this decision must take into account the bias–variance trade-off; that is, a long interval will yield estimates that are stable but that may be attenuated or biased. In practice, the interval is often selected by convenience. The variation across subjects may be systematic in studies that include participants with both clinically detected and screen-detected cancers. If these studies require multiple PSA measurements per individual, as is typical when evaluating PSAV, then it is highly likely that patients with cancers that are detected by screening will have more observations with which to construct a PSAV measurement than patients with clinically detected cancer. The extent and impact of such systematic variation in the calculation of PSAV have not been formally studied.

Even if PSAV differs dramatically between individuals with and without prostate cancer or between those with lethal and
nonlethal cancers, this does not automatically imply that PSAV will be useful in early detection of prostate cancer. This principle is partly due to the correlation that exists between PSA level and PSAV. But there is also a very real difference between association studies that use regression techniques to quantify dependence between a marker and an outcome and classification studies that aim to quantify the sensitivity, specificity, and diagnostic accuracy of the marker in detecting disease (16). Even if a marker such as PSAV is independently associated with disease risk or with the risk of high-grade cancer, it may not substantially increase diagnostic accuracy. For example, Pinsky et al. (9) showed that PSAV was an independent predictor of the risk of high-grade (Gleason 7–10) disease in the PLCO study (odds ratio = 1.23 per unit increase in PSAV, \( P = .003 \)). However, the combination of PSAV and PSA value yielded an AUC of 0.646, which is very similar to the AUC of 0.626 given by the PSA value alone. A similar result was noted by Fall et al. (17), who studied the prognostic value of PSAV based on postdiagnosis PSA measurements in prostate cancer patients who were managed by active surveillance. As Pepe et al. (16) show, substantive improvement in the AUC can be very difficult to achieve via marker combinations and requires an extremely strong independent association between the added markers and disease risk.

The perceived value of PSAV in early detection will depend on how it will be used as a screening test and on the comparison (non–PSAV based) test. There are many different ways to use PSAV as part of an early detection strategy, and it is not always obvious what test is being evaluated in a given study. For example, Schröder et al. (8) found that PSAV did not improve diagnostic performance in the ERSPC trial, but they considered the test (PSA > 4.0 ng/mL and PSAV > \( c \)), in which \( c \) is a specified threshold for PSAV above which biopsy is recommended, and compared this with the test (PSA > 4.0 ng/mL). By contrast, Carter et al. (6,18) compared the performance of PSAV with the performance of PSA when the PSA level was less than 4 ng/mL. Thus, this study effectively evaluated the test (PSA < 4 ng/mL and PSA > \( c \)) for referral to biopsy against the test (PSA < 4 ng/mL and PSAV > \( c \)). The PLCO (9) and the PCPT (3) considered linear combinations of PSA and PSAV, i.e., tests of the form \( a \text{PSA} + b \text{PSAV} > c \), for which the coefficients \( a \) and \( b \) were determined via logistic regression from the study–specific datasets, i.e., they were different for the two studies. Both studies also included other predictors of risk (e.g., age, family history) in their linear combinations. Finally, the current guidelines of the National Comprehensive Cancer Network (http://www.nccn.org/professionals/physician_gls/PDF/prostate_detection.pdf) propose that PSAV be used in combination with the PSA level via an “or” rule (e.g., PSAV > \( c \) or PSA > \( d \)). The specific values used in the current guidelines are \( c = 2.5 \) ng/mL and \( d = 0.5 \) ng/mL per year. The relative performance of this combination test will be different depending on whether it is compared with the standard PSA value cutoff of 4.0 ng/mL for referral to biopsy or a lower cutoff, such as 2.5 ng/mL.

No published study of PSA or PSAV answers the key questions that must be addressed to determine whether PSAV is going to be useful in population screening for prostate cancer. Acknowledging that both benefit and cost must be considered when introducing a new screening test, our version of these questions is as follows:

If we use PSAV at low PSA levels rather than the standard PSA cutoff of 4.0 ng/mL, how many more men will be detected within the “window of curability”? In addition, what will this strategy cost in terms of false-positive tests and overdiagnosis?

No PSAV study conducted to date has directly addressed these questions. Many studies focus on association (is PSAV correlated with high-risk disease?) rather than on classification (does PSAV enable us to reliably distinguish high-risk from low-risk patients and individuals with prostate cancer from those without the disease?). Even those studies that do provide classification measures fall short of quantifying the likely incremental benefits and costs associated with adding PSAV as a screening marker. Knowing, for example, from the PLCO that the (linear) combination of PSA value and PSAV will yield an AUC of 0.646 rather than the AUC of 0.626 associated with PSA value does not tell us how many more patients will be cured by using the combination test and how many more will be falsely declared positive or overdiagnosed. Studies of PSAV and survival generally ignore the issue of false-positive tests, but PSAV can be falsely elevated, for example, in patients with acute prostatitis.

Of the survival studies conducted to date, the study of Carter et al. (6) is the most suggestive that PSAV at low PSA levels may be useful in early detection of lethal prostate cancers. Although this study had only 20 prostate cancer deaths (12 in the primary analyses), it yielded a hazard ratio for prostate cancer death of 3.97 (\( P = .02 \)) associated with a 1 ng/mL per year increase in PSAV 10–15 years before clinical diagnosis. The association between prediagnosis PSAV and survival remained when the analysis was restricted to patients who were diagnosed before PSA screening became routine in the BLSA (HR = 3.75). Yet this result still does not tell us how many more cancers will be cured under an early detection strategy that recommends biopsy when PSAV exceeds 0.35 ng/mL per year at low PSA levels (the threshold recommended by the study investigators), and it does not address the potential costs of such a strategy in terms of the expected increase in false-positive tests and potential increase in overdiagnosis. It is likely that these benefits and costs will depend not only on the specific PSAV threshold used but also on the frequency of screening and compliance to biopsy recommendations in the population. In addition, the cost–benefit profiles may be very different for different age and ethnic groups.

Conclusions

In prospective screening trials, attention generally focuses on the link between prediagnosis changes in PSA and high-risk disease, with the definition of risk based on clinical or pathologic characteristics of the tumor that are known to be associated with prognosis. In survival or prognostic studies, attention generally focuses on disease-specific survival from detection, with the mode of detection typically unspecified. Although interpretation issues differ for screening trials and prognostic studies, we argue that to date, PSAV-based analyses have generally not addressed the key questions that must be answered to determine whether PSA change over time is likely to be a valuable marker in early detection of prostate cancer.

The value of PSAV in prostate cancer detection has been addressed by a wide range of studies with different populations, designs, and analytic approaches. One of the main goals of this
commentary has been to reconcile some of the inconsistencies across studies by highlighting features of study design and potential sources of bias that might explain why different types of studies have produced differing results.

Although studies of the association between PSAV and survival have been compelling, there is no question that the associations found in these studies are considerably weaker or even absent in prospective screening trials. We have proposed an explanation for this phenomenon that calls into doubt the applicability of at least some survival results to the prospective screening setting. This explanation does not negate the value of PSAV in prognosis and treatment decisions; PSAV may be a useful tool after diagnosis, but its value in screening remains to be proven.

As we have noted, the value of PSAV in screening will depend on the baseline screening policy to which PSAV is being added. If the baseline screening policy refers men to biopsy when their PSA level exceeds the standard 4.0 ng/mL cutoff, then the use of PSAV at low PSA levels could generate false-positive tests and levels of overdiagnosis in excess of those already associated with the standard 4.0 ng/mL cutoff, and costs that far outweigh any benefits. In this observation, our concerns parallel those of Hoffman (20) and Welch et al. (21) who argued against expanding the definition of a positive test by lowering the PSA threshold. However, if current practice continues to trend toward referral to biopsy at ever lower PSA levels, then the addition of a selection process such as one based on PSAV or on a disease risk score (3) could help to prevent untenable numbers of unnecessary biopsies and overdiagnosed cancers.

Even if one interprets the existing data as showing that PSAV may have an important role in the early detection setting, the fact remains that, to date, no published study of PSAV actually answers the key questions that must be addressed to confidently integrate PSAV into mass screening. What would it take to address these questions? First, we need a better understanding of the “window of curability” in prostate cancer. Carter et al. (22) have suggested that this window ends when PSA reaches a level of 4.0–5.0 ng/mL, but even in their data, 70% of patients with PSA values greater than 5.0 ng/mL were “curable” (organ-confined or capsular penetration with Gleason score below 7 and negative margins). Thus, the window of curability could extend to relatively high levels of PSA in many cases. Second, we need models and methods that can quantify the expected incremental costs and benefits of incorporating PSAV-based policies in early detection. This is a tall order because such models will need to link PSA growth with disease progression from a localized, curable state to a more advanced, disseminated condition. To date, we know of only one model that explicitly considers the link between PSA growth and natural history (23). This type of approach could be calibrated to the population setting and ultimately used to compare the positive and negative outcomes of screening policies that do and do not use PSAV.

Finally, the metric that is typically used to summarize PSA kinetics must be reevaluated. We have argued that PSAV is faulty because it confuses disease aggressiveness with duration and clouds the interpretation of studies linking PSA growth and disease-specific survival. There is clearly a need for markers other than PSA to improve diagnostic performance in early detection of prostate cancer, and many markers are currently under investigation. Perhaps the most important lesson to take from the PSAV debate is that we should be thinking about marker kinetics in a broader way. As has been shown, for example, in ovarian cancer (24), the serial marker record is a rich source of information that could potentially be mined to achieve far greater diagnostic accuracy—and potential benefit—that can be attained via a single-number summary.

References


(22) Carter HB, Epstein JI, Chan DW, Fozard JL, Pearson JD. Recommended prostate-specific antigen testing intervals for the detection of curable prostate cancer. JAMA 1997;277:1456–60.


**Funding**

Early Detection Research Network Data Management and Co-ordinating Center, National Cancer Institute (U01-201146 to Z. Feng).

**Notes**

We thank Drs H. Ballantine Carter and Paul Pinsky for helpful comments on an earlier draft of the manuscript and Chen Chi for assistance with Fig. 1. I. M. Thompson is a consultant for Veridex, which studies prostate cancer biomarkers. The authors are solely responsible for the content, and the sponsors had no role in the preparation of the manuscript or the decision to submit the manuscript for publication.

Manuscript received March 30, 2007; revised July 26, 2007; accepted August 28, 2007.