BRIEF COMMUNICATION

Prostate-Specific Antigen Screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer–Rotterdam

Elisabeth M. Wever, Gerrit Draisma, Eveline A. M. Heijnsdijk, Monique J. Roobol, Rob Boer, Suzie J. Otto, Harry J. de Koning

Correspondence to: Elisabeth M. Wever, MSc, Department of Public Health, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, the Netherlands (e-mail: e.m.wever@erasmusmc.nl).

Dissemination of prostate-specific antigen (PSA) testing in the United States coincided with an increasing incidence of prostate cancer, a shift to earlier stage disease at diagnosis, and decreasing prostate cancer mortality. We compared PSA screening performance with respect to prostate cancer detection in the US population vs in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC–Rotterdam). We developed a simulation model for prostate cancer and PSA screening for ERSPC–Rotterdam. This model was then adapted to the US population by replacing demography parameters with US-specific ones and the screening protocol with the frequency of PSA tests in the US population. We assumed that the natural progression of prostate cancer and the sensitivity of a PSA test followed by a biopsy were the same in the United States as in ERSPC–Rotterdam. The predicted prostate cancer incidence peak in the United States was then substantially higher than the observed prostate cancer incidence peak (13.3 vs 8.1 cases per 1000 man-years). However, the actual observed incidence was reproduced by assuming a substantially lower PSA test sensitivity in the United States than in ERSPC–Rotterdam. For example, for nonpalpable local- or regional-stage cancers (ie, stage T1M0), the estimates of PSA test sensitivity were 0.26 in the United States vs 0.94 in ERSPC–Rotterdam. We conclude that the efficacy of PSA screening in detecting prostate cancer was lower in the United States than in ERSPC–Rotterdam.

Prostate-specific antigen (PSA) testing was introduced in the United States in 1986 to monitor prostate cancer progression. The test was rapidly adopted for the early detection of prostate cancer, and as a consequence, the incidence of prostate cancer has increased rapidly since 1988, peaking in 1992 (1). The benefits and harms of PSA testing depend on its performance in detecting prostate cancers and on the benefits of consequent early treatment. The performance of PSA testing as a screening test depends on the cutoff level for recommending a biopsy, the compliance to a biopsy recommendation, and the diagnostic accuracy of the biopsies that are performed.

Differences between PSA screening performance with respect to the detection of prostate cancer in a trial and in a population are crucial information for translating the results of a prostate cancer screening trial to a population setting. In this study, we compared PSA screening performance for detecting prostate cancers in the US population with that in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC–Rotterdam). Because PSA screening performance in the US Prostate, Lung, Colorectal and Ovarian (PLCO) trial may be comparable with PSA screening performance in the US population (2), the results of this analysis could also provide quantitative ex-

planations for the different mortality results of the ERSPC and PLCO trials (3,4).

For this analysis, we used the microsimulation screening analysis (MISCAN) model for prostate cancer (5,6), which simulates individual life histories and models the development of cancer in individuals as a sequence of tumor states. The model includes 18 detectable preclinical states in the natural history of prostate cancer that are derived from combinations of clinical T stage (T1, impalpable; T2, palpable, confined to the prostate; and T3+, palpable, with extensions beyond the prostatic capsule) (7), differentiation grade (well differentiated, Gleason score 2–6; moderately differentiated, Gleason score 7; and poorly differentiated, Gleason score 8–10) (8), and metastatic stage (local or regional [M0] and distant [M1]) (7). Cancer can progress from each preclinical state to the clinical disease state (ie, become diagnosed because of symptoms) (Supplementary Figure 1, available online). Preclinical cancers may be detected by PSA screening. Screen detection depends on the timing of PSA tests and on the test sensitivity. In the MISCAN model, the PSA test and a subsequent biopsy are modeled as a single test; therefore, PSA test sensitivity also depends on whether a positive test is followed by a biopsy. In the model, sensitivity is defined as the probability that a preclinical tumor is detected by a screening test at the time the test is taken. The parameters for PSA test sensitivity are stage specific because the sensitivity of a test primarily depends on the size of the tumor.

Model parameters, including transition probabilities, mean dwelling times (the time from one preclinical state to another preclinical or clinical state), and stage-specific test sensitivities, are typically estimated as follows. A model is constructed for a specific situation, such as prostate cancer incidence in the United States or in both arms of the ERSPC–Rotterdam. Parameters are then estimated by numerical minimization of the deviance between observed numbers of cases and the number of cases predicted by the model. Deviances are calculated by assuming a Poisson likelihood for incidence data or by assuming a multinomial likelihood for stage distribution data.

CONTEXT AND CAVEATS

Prior knowledge
The benefits and harms of prostate-specific antigen (PSA) testing depend on its performance in detecting prostate cancers and on the benefits of consequent early treatment. The performance of PSA testing as a screening test depends on the cutoff level for recommending a biopsy, the compliance to a biopsy recommendation, and the diagnostic accuracy of the biopsies that are performed. Translating the results of a prostate cancer screening trial to a population setting requires a comparison of PSA screening performance for the detection of prostate cancer in these two situations.

Study design
A microsimulation screening analysis model for prostate cancer and PSA screening was developed for the European Randomized Study of Screening for Prostate Cancer (ERSPC)–Rotterdam trial and then adapted to the US population by replacing demography parameters with US-specific ones and the trial screening protocol with the frequency of PSA tests in the US population. The natural progression of prostate cancer and the sensitivity of a PSA test followed by a biopsy were assumed to be the same in the United States as in the trial.

Contribution
The model-predicted prostate cancer incidence peak in the United States was substantially higher than the observed prostate cancer incidence peak. However, the actual observed incidence was reproduced by assuming a substantially lower PSA test sensitivity in the United States than in ERSPC–Rotterdam.

Implications
PSA screening in the United States detected fewer prostate cancers than PSA screening in the European trial because of the lower sensitivity of PSA testing followed by biopsy.

Limitations
Other factors that differed between the US and ERSPC–Rotterdam populations and might influence the detection rates were not taken into account. The model used for the frequency of PSA testing included diagnostic tests. The model did not explain the steep drop in prostate cancer incidence in the United States after 1992. The reliability of the sensitivity estimates could not be determined.

From the Editors

In this study, we first developed an ERSPC model that simulated the prostate cancer progression and screening in ERSPC–Rotterdam. Estimates of natural history parameters and test sensitivities were obtained by using the observed detection rates, interval cancer rates, and stage distributions from ERSPC–Rotterdam (5,6). The parameter estimates of the model are presented in Supplementary Table 1 (available online), and the observed data used for the estimation are presented in Supplementary Table 2 (available online).

Next, to make the model results comparable with observed US data, the population in the model was adjusted to the US population by replacing the birth tables and life tables with US-specific tables, and the screening protocol of ERSPC–Rotterdam was replaced with the frequency of PSA testing in the US population. The frequency of PSA testing in the United States was modeled according to the approach described by Mariotto et al. (9). The frequency of a first PSA test and of repeat tests in the United States, as reproduced in the MISCAN model, is illustrated in Figure 1. On average, 80% of the screened men in the United States have a repeat PSA test within 2 years of the previous test.

We considered two US models. In model 1, we investigated the hypothesis that PSA screening in the United States is the same as in ERSPC–Rotterdam. In this

model, all prostate cancer–related parameters were the same as in the ERSPC model. In model 2, we investigated the hypothesis that the sensitivity of PSA screening in the United States is lower than that in ERSPC–Rotterdam. In this model, all prostate cancer–related parameters except for the test sensitivity parameters were the same as those in the ERSPC model. US-specific estimates of test sensitivities were obtained by using observed age-specific incidence and age-specific stage distribution (local or regional vs distant) in the US population. For estimation of the US-specific parameters, we used data from the Surveillance, Epidemiology, and End Results (SEER) registry for US men aged 50–84 years who were diagnosed with prostate cancer between January 1, 1985, and December 31, 2000. The data were based on the nine core catchment areas (SEER 9) of the SEER registry (http://seer.cancer.gov/). We used the test sensitivity parameter estimates of the ERSPC model as starting values for optimization of the estimates of the US model. The estimated test sensitivity parameters of the calibrated model are presented in Table 1 and the observed data used for calibrating the model are presented in Supplementary Table 3 (available online).

In model 1, both the predicted and observed incidence peaks occurred in 1992. However, the predicted prostate cancer incidence peak in the United States was
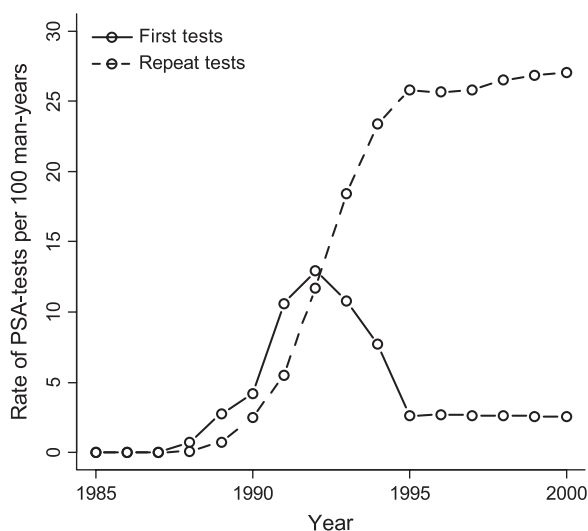

Figure 1. Frequency of first prostate-specific antigen (PSA) tests and repeat tests in the US population as generated by microsimulation screening analysis (MISCAN). The frequencies are for men aged 50–84 years.

substantially higher than the observed prostate cancer incidence peak (13.3 vs 8.1 cases per 1000 man-years), which suggests a lower detection of prostate cancer in the United States than in ERSPC–Rotterdam (Figure 2). In model 2, the predicted incidence peak was the same size as the observed incidence peak (Figure 2). However, estimates of test sensitivity were lower in the United States than in ERSPC–Rotterdam. For example, for nonpalpable local- or regional-stage cancers (ie, stage T1M0), the estimates of PSA test sensitivity were 0.26 in the United States vs 0.94 in ERSPC–Rotterdam (Table 1).

The lower sensitivity of PSA screening in the United States compared with ERSPC–Rotterdam in model 2 could be due to a higher PSA cutoff level for recommending biopsy in the United States, a lower biopsy compliance rate in the United States, or a lower sensitivity of the biopsies in the United States. The latter possibility is unlikely because more biopsy cores are generally taken in the United States than were taken in ERSPC–Rotterdam. The other two possibilities might explain the lower sensitivity of PSA screening in the United States. A higher PSA cutoff level for recommending biopsy in the United States could follow from the fact that the recommended PSA cutoff level in the United States is 4 ng/mL, whereas the PSA cutoff level in ERSPC–Rotterdam was 3 ng/mL. A lower biopsy compliance rate in the United States could, for instance, indicate that some physicians in the United States might have used a higher PSA cutoff level than recommended (ie, higher than 4 ng/mL) or might have advised a confirmatory PSA test if the first PSA level was elevated. Confirmatory PSA tests would lower the biopsy compliance rate because men with a PSA level higher than the cutoff level at the first test but with a PSA level lower than the cutoff level at the confirmatory test would probably be advised to not have a biopsy; therefore, some men with a PSA level higher than the cutoff level at the first PSA test would not have a biopsy. Pinsky et al. (2) reported a biopsy compliance rate in the PLCO trial of 41% within 1 year of a positive PSA test. They suggested that this biopsy compliance rate is representative of US screening practice, given that men with a positive PSA test in the PLCO trial were referred to their
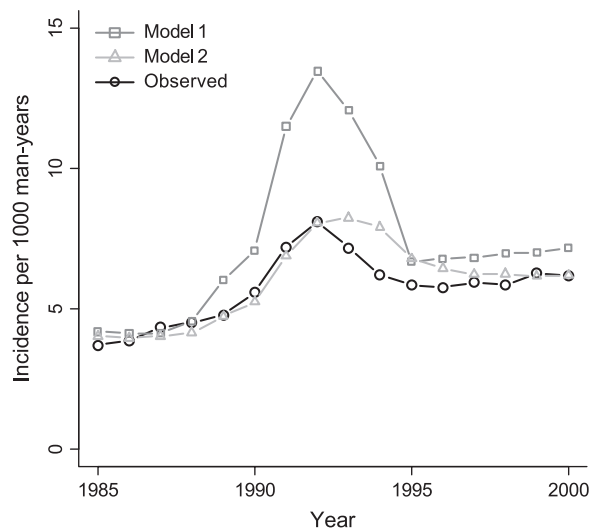

Figure 2. Observed and predicted age-adjusted incidence per 1000 man-years for men aged 50–84 years in the US models. In model 1, prostate-specific antigen (PSA) screening in the US population is the same as in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC–Rotterdam). In model 2, the sensitivity of PSA screening is lower in the US population than in ERSPC–Rotterdam.

personal physician for follow-up. In the screening arm of ERSPC–Rotterdam, biopsies were administered by the screening center at no charge to the subject, and reminders for biopsy appointments were sent if necessary, resulting in a biopsy compliance rate of approximately 90%. In model 2, the detection rates at first PSA screening and at repeat PSA screening were 18 and 12 per 1000 screened men, respectively (Table 1), which are comparable with the detection rates at the first round of screening (16 per 1000 screened men) and

repeat screening (11 per 1000 screened men) in the PLCO trial (10).

This study has four limitations. First, we did not take into account other factors, such as race, that differed between the US and ERSPC–Rotterdam populations and might influence the detection rates. Approximately 10% of the US population is black, whereas nearly 100% of the ERSPC–Rotterdam population was white. Because the incidence of prostate cancer was higher among black men than among white men during the study period (11),

Table 1. Estimates of sensitivity, detection rate, and deviance for the two US models*

Item	Model 1	Model 2
Sensitivity by stage†		
T1M0	0.94	0.26
T2M0	0.94	0.26
T3M0	1.00	0.27‡
T1M1	0.96	0.84
T2M1	0.97	0.84
T3M1	1.00	0.84
Detection rate per 1000 screened men		
At first PSA test	62	18
At repeat PSA test	13	12
Deviance	44 727	23 438

* PSA = prostate-specific antigen.

† T1, T2, and T3 are the three clinical T stages (T1, nonpalpable; T2, palpable, confined to the prostate; and T3, palpable, with extensions beyond the prostatic capsule); M0 is the local or regional stage; and M1 is the distant stage.

‡ The range of plausible values is 0.24–0.29. The range of plausible values indicates a range in which the 95% confidence interval will be with near certainty, see Supplementary Figure 2 (available online). Because of restrictions on the sensitivities (sensitivity increases with clinical T stage and metastatic state), this range cannot be calculated for the other parameters.

these racial differences might explain the different detection rates estimated for the two populations. However, the incidence of prostate cancer among whites in the US population was similar to the overall incidence (11), which indicates that the effect of black men on the overall observed incidence was small, as was their effect on the outcomes of this study.

Second, we assumed that the model that we used for the frequency of PSA testing (9) would apply to screening tests. During the construction of that model, all follow-up PSA tests taken after diagnosis as well as PSA tests occurring within 3 months of a previous PSA test were eliminated. However, a fraction of the remaining tests might be diagnostic tests that were used to confirm a suspicion for prostate cancer. The size of this fraction is unknown, but including this fraction of diagnostic tests as screening tests would imply that the actual screening rate is lower than in the model.

Third, a weakness of our model is that it fails to explain why prostate cancer incidence in the United States dropped so steeply after 1992 (Figure 2). In our model, the cancers detected in repeat tests led to a slower decline of incidence after 1992 than what was observed. However, the frequency of repeat PSA testing remained at a level of 30% (Figure 1), and it is unclear why these tests detected so little cancer in the US population.

Fourth, we could not compute 95% confidence intervals for the sensitivity parameters: Because of random noise in the simulated predictions and restrictions on the sensitivities (sensitivity increases with clinical T stage and metastatic state), formal 95% confidence intervals are difficult to obtain when using the microsimulation model. However, for fixed values of other model parameters, the range of plausible values for test sensitivity for a local or regional stage tumor in clinical stage T3 (ie, in state T3M0) was narrow (0.24–0.29). The range of plausible values contains with near certainty a standard computed 95% confidence interval. The calculation of the range of plausible values is presented in Supplementary Figure 2 (available online).

In conclusion, PSA screening in the United States did not detect as many prostate cancers as PSA screening in ERSPC–

Rotterdam because of the lower sensitivity of PSA testing followed by a biopsy. The consequence of this lower test sensitivity is that the effects of PSA screening in the United States are likely to be different from those observed in the ERSPC–Rotterdam. For example, Draisma et al. (12) noted that the lead time (time by which screening advances diagnosis) and the frequency of overdiagnosis were smaller in the United States than in ERSPC–Rotterdam (mean non-overdiagnosed lead time: 6.9 vs 7.9 years; overdiagnosis frequency: 42% vs 66%), indicating that the harms of PSA testing in the United States, although still substantial, are likely to be less than those in the ERSPC–Rotterdam. The benefits of PSA screening in the United States are also likely to be different from those in ERSPC–Rotterdam. The ERSPC trial has shown that screening for prostate cancer by using PSA tests can reduce prostate cancer mortality (4); however, we cannot directly translate these mortality reductions to the US population because of differences between the two populations, such as the lower sensitivity of PSA screening in the United States. Finally, this analysis also shows quantitatively that it is likely that there is a difference in the sensitivity of the PSA screening (PSA test followed by a biopsy) in the ERSPC and PLCO trials, which is likely to have contributed to the different outcomes of the trials.

References

1. Legler JM, Feuer EJ, Potosky AL, Merrill RM, Kramer BS. The role of prostate-specific antigen (PSA) testing patterns in the recent prostate cancer incidence decline in the United States. *Cancer Causes Control.* 1998;9(5):519–527.
2. Pinsky PF, Andriole GL, Kramer BS, et al. Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. *J Urol.* 2005;173(3):746–750; discussion 750–751.
3. Andriole GL, Grubb RL III, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med.* 2009;360(13):1310–1319.
4. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med.* 2009;360(13):1320–1328.
5. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst.* 2003;95(12):868–878.
6. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* 2006;119(10):2366–2371.
7. Schroder FH, Hermanek P, Denis L, Fair WR, Gospodarowicz MK, Pavone-Macaluso M. The TNM classification of prostate cancer. *Prostate Suppl.* 1992;4:129–138.
8. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep.* 1966;50(3):125–128.
9. Mariotto AB, Etzioni R, Krapcho M, Feuer EJ. Reconstructing PSA testing patterns between black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer.* 2007;109(9):1877–1886.
10. Grubb RL III, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. *BJU Int.* 2008;102(11):1524–1530.
11. Ries LAG, Harkins D, Krapcho M, et al. *SEER Cancer Statistics Review, 1975-2003.* Bethesda, MD: National Cancer Institute; 2006. http://seer.cancer.gov/csr/1975_2003/ Accessed July 3, 2009.
12. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst.* 2009;101(6):374–383.

Affiliations of authors: Department of Public Health (EMW, GD, EAMH, RB, SJO, HJdeK) and Department of Urology (MJR), Erasmus Medical Center, Rotterdam, the Netherlands.