# Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context

Gerrit Draisma, Ruth Etzioni, Alex Tsodikov, Angela Mariotto, Elisabeth Wever, Roman Gulati, Eric Feuer, Harry de Koning

**Background** The time by which prostate-specific antigen (PSA) screening advances prostate cancer diagnosis, called the lead time, has been reported by several studies, but results have varied widely, with mean lead times ranging from 3 to 12 years. A quantity that is closely linked with the lead time is the overdiagnosis frequency, which is the fraction of screen-detected cancers that would not have been diagnosed in the absence of screening. Reported overdiagnosis estimates have also been variable, ranging from 25% to greater than 80% of screen-detected cancers.

**Methods** We used three independently developed mathematical models of prostate cancer progression and detection that were calibrated to incidence data from the Surveillance, Epidemiology, and End Results program to estimate lead times and the fraction of overdiagnosed cancers due to PSA screening among US men aged 54–80 years in 1985–2000. Lead times were estimated by use of three definitions. We also compared US and earlier estimates from the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC) that were calculated by use of a microsimulation screening analysis (MISCAN) model.

**Results** The models yielded similar estimates for each definition of lead time, but estimates differed across definitions. Among screen-detected cancers that would have been diagnosed in the patients' lifetimes, the estimated mean lead time ranged from 5.4 to 6.9 years across models, and overdiagnosis ranged from 23% to 42% of all screen-detected cancers. The original MISCAN model fitted to ERSPC Rotterdam data predicted a mean lead time of 7.9 years and an overdiagnosis estimate of 66%; in the model that was calibrated to the US data, these were 6.9 years and 42%, respectively.

**Conclusion** The precise definition and the population used to estimate lead time and overdiagnosis can be important drivers of study results and should be clearly specified.

J Natl Cancer Inst 2009;101:374–383

Almost 20 years after its introduction, prostate-specific antigen (PSA) screening remains controversial. Randomized controlled trials are still ongoing in the United States and Europe, and it will be several years before efficacy results become available (1,2). Although prostate cancer mortality rates have declined in some countries with high use of PSA screening, such as the United States, mortality rates are also dropping in other countries with relatively low use of PSA screening, such as the United Kingdom (3). Other factors besides screening may be affecting mortality, including changes in treatment practices and early detection of recurrent disease.

As the debate about the benefits of PSA screening continues, there is growing recognition of its costs. One of the chief drivers of the costs of PSA screening is overdiagnosis—the detection of latent disease that would not have been diagnosed in the patient's lifetime in the absence of screening. Overdiagnosis is a particularly important issue in prostate cancer screening because the latent prevalence of disease, as estimated from autopsy studies, is much higher than its incidence in the absence of screening. Therefore, there is a large pool of silent cancers that could potentially be detected by screening. Because it is not usually clear whether a screen-detected cancer has been overdiagnosed, many overdiagnosed patients receive curative treatment (surgery or radiation therapy), which is associated with substantial costs and morbidity (4).

**Affiliations of authors:** Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, the Netherlands (GD, EW, HdK); Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA (RE, RG); Department of Biostatistics, University of Michigan, Ann Arbor, MI (AT); Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD (AM, EF).

**Correspondence to:** Gerrit Draisma, PhD, Department of Public Health, Erasmus MC, University Medical Center, PO Box 2040 3000 CA, Rotterdam, the Netherlands (e-mail: g.draisma@erasmusmc.nl).

**See** "Funding" and "Notes" following "References."

**DOI:** 10.1093/jnci/djp001

The frequency of overdiagnosis is associated with the time by which screening advances diagnosis, also called lead time. Because prostate cancer is often a slowly developing disease, PSA screening can be associated with lengthy lead times. The longer the lead time, the greater the likelihood of overdiagnosis. Thus, estimating the lead time is often a critical step in estimating the frequency of overdiagnosis.

Estimates of lead time and overdiagnosis due to PSA screening have been obtained from various sources. Several studies that used stored serum samples found mean lead time estimates ranging from 3 to more than 7 years (5–7); more recently, Tornblom et al. (8) estimated a median lead time of 11 years. Other studies estimated lead times on the basis of a comparison of detection rates in a population-based trial setting with baseline incidence, producing mean lead times between 5 and 12 years (9,10). Further investigations used models to explicitly link PSA screening frequencies with population trends in prostate cancer incidence as reported in the Surveillance, Epidemiology, and End Results (SEER) program (11–13) of the National Cancer Institute. In these studies, mean lead time estimates ranged from 3 to 7 years. Overdiagnosis estimates ranged from 25% to 84% of all screen-detected cancers (10,12–14).

It is clear that published lead time and overdiagnosis estimates vary considerably across studies. There are at least three reasons for this variability: 1) the context of the estimates, including population, epidemiology of the disease, and the way screening is practiced in those populations (eg, PSA level cutoffs and biopsy practices); 2) the definitions of lead time and overdiagnosis used; and 3) the methods used to calculate the estimates. The goal of this article was to explore each of these three factors as we investigate why different studies have yielded different results.

We estimated lead time and overdiagnosis within a specific population setting, namely, the US male population aged 50–84 years in 1985–2000. To investigate the influence of the definition of the lead times on the estimates, we considered three definitions of lead time (non-overdiagnosed, censored, and uncensored, as defined in "Methods").

The estimates presented were developed using three models that link PSA testing trends with population incidence rates: the model developed at the Fred Hutchinson Cancer Research Center (FHCRC) (11,12), the model developed at the University of Michigan (UMich) (13), and the microsimulation screening analysis (MISCAN) model developed at Erasmus MC in Rotterdam (10,15,16). The use of multiple models allowed us to produce robust results while exploring the influence of estimation methodology.

The FHCRC and UMich models were originally developed to study prostate cancer incidence and mortality in the United States. In contrast, the MISCAN model was originally based on baseline incidence in the Netherlands and results of the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC) (10,15). Thus, to enable comparisons with US data, the MISCAN model was calibrated to SEER incidence data.

This study was carried out in collaboration with the Cancer Intervention and Surveillance Modeling Network (CISNET; http://cisnet.cancer.gov/) of the National Cancer Institute. The primary goal of CISNET is to use modeling to quantify the roles of prevention, screening, and treatment in explaining cancer incidence and mortality trends. A key feature of the CISNET collaboration is that the models are developed independently, but modelers use standardized inputs and share details of model development to understand and explain any differences in model results.

## CONTEXT AND CAVEATS

**Prior knowledge**

Estimates of lead time, which is the time that screening advances cancer diagnosis, and overdiagnosis, the detection by screening of cancers that would not be detected in the absence of screening, are highly variable for prostate cancer screening using prostate-specific antigen (PSA) testing.

**Study design**

Lead times and fractions of overdiagnosis for PSA testing of US men aged 54–80 years in 1985–2000 were estimated using three models of prostate cancer progression and detection calibrated to the Surveillance, Epidemiology, and End Results program. Estimates of lead times using different definitions were compared across models.

**Contributions**

Estimated lead times ranged from 5.4 to 6.9 years and were similar across models but different according to the definition used. Overdiagnosis ranged from 23% to 42% of all prostate cancers detected by PSA testing.

**Implications**

When reporting lead times in screening studies, the definition of lead time used can impact the outcome and thus should always be specified.

**Limitations**

A portion of the PSA screening tests included in the models was likely performed for diagnostic purposes after prostate cancer diagnosis. The estimates are imperfect, and it is unknown in which direction they may be biased.

*From the Editors*

## Methods

### Definitions of Lead Time and Overdiagnosis

The standard definition of lead time is the interval from screen detection to the time of clinical diagnosis, when the tumor would have surfaced without screening. However, patients with screen-detected cancers may die from other causes before the time of clinical diagnosis. This is called overdiagnosis. In this article, overdiagnosis is expressed as a percentage of all screen-detected cancers, unless otherwise specified. Because lead times are not directly observable, surrogate measures are often used, and as a consequence, estimates of lead time may refer to different quantities. Three variants exist for both lead time and the related concept of sojourn time—the time from disease onset to clinical diagnosis. Non-overdiagnosed lead times are calculated only for non-overdiagnosed cancers, that is, those for which the date of clinical diagnosis precedes the date of death (Figure 1, A). Censored lead times are calculated for both non-overdiagnosed cancers and overdiagnosed cancers, with lead times for overdiagnosed cancers censored at the date of death from other causes (Figure 1, B).

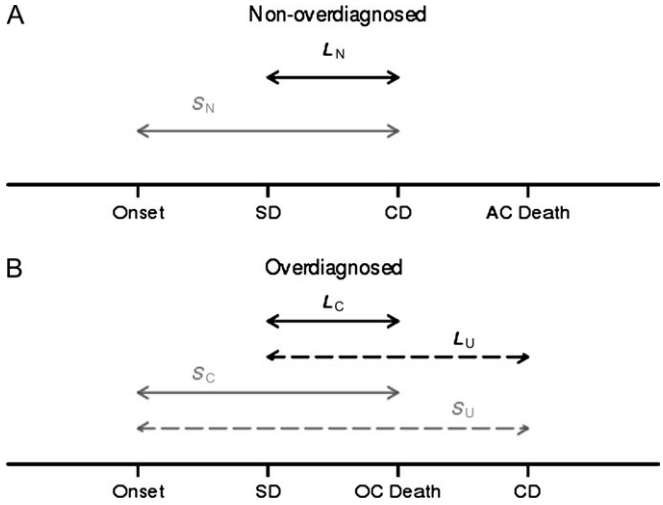


**Figure 1.** Lead time, sojourn time, and overdiagnosis. **A**) Non-overdiagnosed prostate cancers. **B**) Overdiagnosed prostate cancers. A non-overdiagnosed cancer patient is clinically diagnosed (CD) before dying from any cause (AC Death), whereas an overdiagnosed cancer patient dies of other causes (OC Death) before being clinically diagnosed. Lead times (*L*) begin at screen detection (SD), and sojourn times (*S*) begin at disease onset (Onset). Non-overdiagnosed lead and sojourn times ($L_N$ and $S_N$) are calculated only for non-overdiagnosed cancers. Censored lead and sojourn times ($L_C$ and $S_C$) are calculated for both non-overdiagnosed cancers and overdiagnosed cancers, with times for overdiagnosed cancers censored at OC death. Uncensored lead and sojourn times ($L_U$ and $S_U$) are also calculated for both non-overdiagnosed cancers and overdiagnosed cancers, with times for overdiagnosed cancers uncensored at OC Death. Note that the overdiagnosed cancers might include nonprogressive cancer or "insignificant" cancer, for which clinical diagnosis would "never" happen, even in the absence of other-cause mortality. CD = clinical diagnosis; SD = screen detection; AC Death = death from any cause; OC Death = death from other causes; Onset = disease onset time, when the tumor becomes detectable by screening.

Uncensored lead times are calculated for both non-overdiagnosed cancers and overdiagnosed cancers. The lead times for overdiagnosed cancers are not censored at the date of death from other causes (Figure 1, B). As might be expected, there is a major difference between lead times that are estimated only for non-overdiagnosed cancers and lead times that are estimated for both overdiagnosed cancers and non-overdiagnosed cancers. A drawback of many studies that estimate lead time is that the precise definition used is not made explicit.

To reconcile published estimates of lead time and overdiagnosis, we applied modeling approaches to estimate these three definitions of lead time. Before describing the models, however, it is useful to compare the definitions and to consider when each might be appropriate.

First, the mean lead times that are based on the three definitions are related—the mean non-overdiagnosed and mean censored lead times will always be shorter than the mean uncensored lead time. Thus, if a relatively high value is estimated by use of one definition, estimates that use the other definitions will also be high, in general. Second, of these three definitions, only the uncensored lead time is independent of age. Because of increasing mortality from other causes with age, both mean censored and mean non-overdiagnosed lead times decrease with age, whereas the risk of overdiagnosis increases.

Each definition of lead time is useful in the appropriate setting. The non-overdiagnosed lead time applies to the population of patients for whom screening is potentially beneficial and as such is valuable in designing screening schedules and in studies of potential or actual screening benefit. The censored lead time applies to the entire screen-detected population and reflects the extra time that patients must live with the knowledge that they have prostate cancer and the consequences of diagnosis and possibly treatment. Therefore, censored lead time is an important indicator of morbidity associated with screening and will be particularly relevant if the screening benefit is minimal or modest. The uncensored lead time is useful because it applies to death from the disease itself, in the absence of other causes. The uncensored lead time is closely linked to overdiagnosis because overdiagnosis may be defined as corresponding to the occurrence of other-cause death within the uncensored lead time.

### Modeling Population Incidence for Inference About Lead Time and Overdiagnosis

The pattern of disease incidence in a population undergoing screening for the first time is well established (17). Initial dissemination of the screening test leads to an increase in disease incidence; as use of the test stabilizes, incidence declines. The height and width of the incidence peak following the introduction of screening provide information about the lead time associated with the test and, together with the trend in incidence following the peak, also provide information about the frequency of overdiagnosis (18). However, extracting information about lead time and overdiagnosis from population incidence trends requires knowledge of trends in population screening and a quantitative mechanism that links screening in the population with disease incidence patterns. In this analysis, we used common data sources and three different models to estimate lead time and overdiagnosis associated with PSA screening in the United States.

Although the three models have been independently developed, each builds on a concept of the natural history of the disease that includes onset, progression, and diagnosis in the absence of screening. The natural history models are described below. The parameters of the natural history models are estimated so that the incidence of disease that is projected by the model matches the incidence observed in the SEER population. This estimation process is termed model calibration. The calibrated models are then used to produce estimates of mean lead time and overdiagnosis, either analytically or via simulation. For validation purposes, each model also projects the number of screening tests and the total incidence of prostate cancer among men aged 50–84 years in 1985–2000.

### Prostate Cancer Incidence and Screening Frequencies

The models are calibrated to the incidence of prostate cancer by age, stage, and calendar year. These data were obtained from the nine core catchment areas (SEER 9) of the SEER registry (http://seer.cancer.gov/). For the dissemination of PSA screening, we used the results of Mariotto et al. (19), who retrospectively constructed PSA screening histories in the population by use of survey data from the 2000 National Health Interview Survey (20) and claims data from the linked SEER-Medicare database (http://healthservices.cancer.gov/seermedicare/). PSA screening started in the late 1980s and increased to a level of 30% of the male population aged 50–84 years by the year 2000. The frequency of the first PSA tests peaked

in 1992, when approximately 12% of the male population aged 50–84 years had their first test (Figure 2).

All models use the maximum likelihood method for estimating some (MISCAN and FHCRC) or all (UMich) parameters. Specifically, the models predict the counts of cancers by calendar year, 5-year age group, and stage (local–regional vs distant) from 1985 through 2000, ages 50–84 years, in the SEER 9 registries. Parameters are estimated by maximization of the likelihood of these observed counts, assuming each count to be Poisson distributed with the predicted count as mean. This is equivalent to minimization of the deviance between observed and predicted counts. A common assumption in our models is that observed incidence trends from 1985 through 2000 can be explained by the dissemination of PSA screening; that is, in the absence of screening, the models assume flat incidence rates at 1985–1987 levels. Also all models use standard US life tables that have been corrected for prostate cancer mortality to calculate mortality from causes other than prostate cancer.

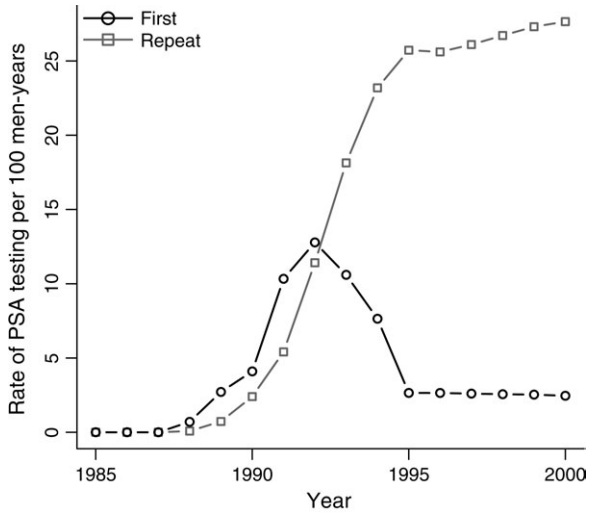


**Figure 2.** Dissemination of PSA screening. The graph presents the age-adjusted [to the US standard million (21)] frequency of first PSA tests and repeat tests in men aged 50–84 years. Rates are based on the results of Mariotto et al. (19). PSA = prostate-specific antigen.

### The MISCAN Model

The MISCAN prostate model is a microsimulation model that simulates individual life histories. In such models, lead time and overdiagnosis estimates are obtained by simply tallying the relevant events. For example, the overdiagnosis frequency was estimated by counting the proportion of patients who have a date of screen-detected prostate cancer whose date of other-cause death would have preceded the date of clinical diagnosis if there had been no screening. Cancer development is modeled as a semi-Markov process, generating transitions from one state to the next. In addition to the healthy state, there are nine states in the natural history of prostate cancer that are derived from combinations of clinical stage (T1, T2, and T3) and Gleason grade (well, moderately, and poorly differentiated) (10,15,16). Most parameters in the MISCAN model were based on results of the Rotterdam ERSPC trial (15,16). For calibration to SEER 9 incidence from 1985 through 2000, several parameters were changed and estimates specific for the US population were obtained via maximum likelihood estimation. The final calibrated model differed from the Rotterdam in two aspects: we assumed and estimated a lower sensitivity of PSA screening in the United States, and we added and estimated an extra stage-specific risk of clinical diagnosis, implying an earlier diagnosis of prostate cancer in the absence of screening in the United States. Note that in the MISCAN model, the PSA test and subsequent biopsy are modeled as a single test, with stage-specific sensitivities. We also converted the disease stages from T1–T3 to the SEER local–regional and distant stages and reestimated the stage- and grade-specific risks of transition from local–regional to metastatic disease.

### The FHCRC Model

The FHCRC model explicitly links individual PSA levels and prostate cancer progression events, including disease onset, metastasis, and clinical presentation (22). We assume that individual PSA levels increase linearly (on the natural logarithmic scale) with age and that the slope of the increase changes after the time of disease onset. The link between the rate of increase of the PSA concentration and disease progression formalizes the intuitive notion that an individual whose PSA level is increasing very slowly is likely to have a longer interval before his disease spreads beyond the prostate. This approach is similar to models in which the risk of disease progression is assumed to depend on tumor size (23,24), but the tumor size variable is replaced with an individual-specific marker trajectory (22,25–27). Ages at disease onset, transition from localized to metastatic disease, and clinical presentation are controlled by corresponding hazard functions. We assumed the hazard of disease onset to be proportional to age, the hazard of transition from a local–regional-stage tumor to a distant-stage tumor to increase with the PSA level, and the hazard of transition from a preclinical state to clinical diagnosis also to increase with PSA level and greater when the disease becomes metastatic.

PSA concentration curves and within- and between-individual variances were estimated by use of data from the Prostate Cancer Prevention Trial (28), which conducted annual screening of 18 000 men for up to 7 years. To project disease incidence, we simulated a population of natural histories and superimposed PSA screening tests according to schedules that were projected by the results of Mariotto et al. (19). A biopsy is recommended for men with a PSA level of 4.0 ng/mL or greater. The rate of compliance with the recommendation is based on data from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (29), a 23-year trial including PSA screening for 37 000 men. Finally, biopsy sensitivity improves across calendar years based on a literature review of biopsy schemes. Given individual PSA trajectories, screening schedules, and biopsy compliance and sensitivity patterns in the population, the hazard rate parameters were estimated from SEER 9 incidence and stage distribution by use of maximum likelihood methods. After calibration, estimates of mean lead time and overdiagnosis were computed via simulation.

### The UMich Model

The UMich model is a statistical mixed model that was specifically designed to allow estimation of its parameters directly from cancer

registry data representing population incidence (13). The natural history of the disease is taken to consist of healthy, preclinical, and clinical or diagnosed states. If screening schedules were known in the population, the UMich model would be similar to classical statistical models of cancer screening (30,31). But because individual screening schedules are unknown, incidence rates are calculated by averaging across the distribution of screening schedules and the distribution of random natural histories. The model parameters include the sensitivity of the screening test, the distribution of age at tumor onset, and the distribution of the sojourn time in the preclinical state.

Test sensitivity is assumed to be an increasing function of the time since tumor onset, and Weibull distributions are assumed for the distribution of age at tumor onset and the sojourn time distribution. Given age at onset, cancer detection by screening is assumed to represent an independent risk that competes with the sojourn time. However, because the tumor onset and the screening schedule are unobservable, the observed risks become mutually dependent.

A multiplicative secular trend in calendar time was introduced to model the increasing incidence pattern observed in the 1980s before PSA testing was introduced. The trend settles at a plateau in the PSA era, leaving the description of the dynamics in the PSA era to utilization patterns of the test. To make the model amenable to population data, a two-stage point-process model of random PSA screening schedules in the population was built and specified to reproduce the observed patterns of test utilization by age and calendar time (13). The model was fit to SEER 9 incidence and stage distribution by maximizing the likelihood for observed population rates. After the model was calibrated, lead time and overdiagnosis estimates were derived analytically on the basis of expressions for the sojourn time distribution and probabilities of related events (13).

### Web Supplements

Detailed supplemental descriptions of the FHCRC and UMich models are available at http://cisnet.cancer.gov/profiles/.

## Results

### Prostate Cancer Incidence

After the introduction of the PSA test for the early detection of prostate cancer in 1987–1988, its use rapidly rose and reached a steady rate of about 30 tests per 100 men-years in 1996 (Figure 2). The number of men receiving their first test peaked in 1992. In the same period, prostate cancer incidence rose from approximately 400 per 100 000 men-years in 1987 to 600 per 100 000 in 1996, with a distinct peak of 800 per 100 000 men-years in 1992 (Figure 3), coinciding with the peak in first PSA tests. Observed incidence was reasonably well reproduced by all three models. The MISCAN and UMich models slightly underpredicted, and the FHCRC model slightly overpredicted incidence in the late 1980s. Both MISCAN and FHCRC model projections lagged behind the observed incidence peak. In the UMich model, incidence after 1996 was lower than that in the other models. The models estimated that 47%–58% of prostate cancers were screen-detected in 2000.

Observed and predicted incidence of local–regional prostate cancer closely followed the overall incidence pattern (Figure 4, A). The pattern for distant-stage incidence was different. In the nine core SEER catchment areas, distant-stage disease incidence dropped from 68 to 34 per 100 000 between 1990 and 1995, gradually declining to 24 per 100 000 in 2000, a decline of 65%. This pattern was imperfectly reproduced by all three models, underestimating distant-stage disease incidence before 1990 and overestimating it thereafter, predicting smaller and more gradual declines—from 40% in the MISCAN model to 50% in the UMich model (Figure 4, B).

### Lead Time and Overdiagnosis

In the SEER 9 database, 235 112 prostate cancers were registered during 1985–2000 in men aged 50–84 years. The total number of life-years was 42.3 million, implying a crude incidence rate of 555 per 100 000 men. The MISCAN, FHCRC, and UMich models predicted that 239 000, 244 000, and 230 000 men, respectively, were diagnosed with prostate cancer, respectively (Table 1). The models projected that 7.9 (MISCAN), 7.8 (FHCRC), and 7.4 (UMich) million PSA tests were conducted in the same period and age group. An estimated 44% (MISCAN), 42% (FHCRC), and 38% (UMich) of prostate cancers were detected by PSA screening, and an estimated 42% (MISCAN), 28% (FHCRC), and 23% (UMich) of the screen-detected cancers were overdiagnosed. In the MISCAN and FHCRC models, approximately 19% and 12%, respectively, of total incidence was overdiagnosed, whereas the UMich model estimated this to be 9%.

As expected, the estimated mean uncensored lead times were greater than the mean censored and non-overdiagnosed lead times. The uncensored estimates ranged from 7.2 to 10.0 years, the censored estimates ranged from 5.7 to 7.8 years, and the non-overdiagnosed lead times ranged from 5.4 to 6.9 years. The estimates from the MISCAN model were consistently higher than those from the FHCRC and UMich models, but the range across models was quite narrow for each definition of lead time.

### MISCAN Model: Comparison of Results With ERSPC Data vs Results With SEER Data

In 2003, the MISCAN group reported a mean lead time for non-overdiagnosed cancers of 13.4 years associated with annual screening from ages 55 to 75 years, with more than 50% of all screen-detected cancers being overdiagnosed (10). The estimates were obtained from a model that was based on incidence in the Netherlands before the PSA era (1991) and the cancer detection and diagnosis rates in ERSPC Rotterdam. We calculated incidence predictions from this model applied to the US situation, with only screening patterns changed. The model predicted a far more pronounced incidence peak than that observed (Figure 5). Following calibration, which involved allowing lower sensitivities of the screening test in the United States than in the trial situation in Rotterdam and higher hazards of preclinical prostate cancer being diagnosed in the United States than in the Netherlands, we obtained the fitted model predictions shown in Figures 3 to 5. Of course, estimated lead time and rate of overdiagnosis were affected by the calibration (Table 2). With the original Netherlands–Rotterdam parameters, the mean non-overdiagnosed lead time would have been 7.9 years and the overdiagnosis frequency would

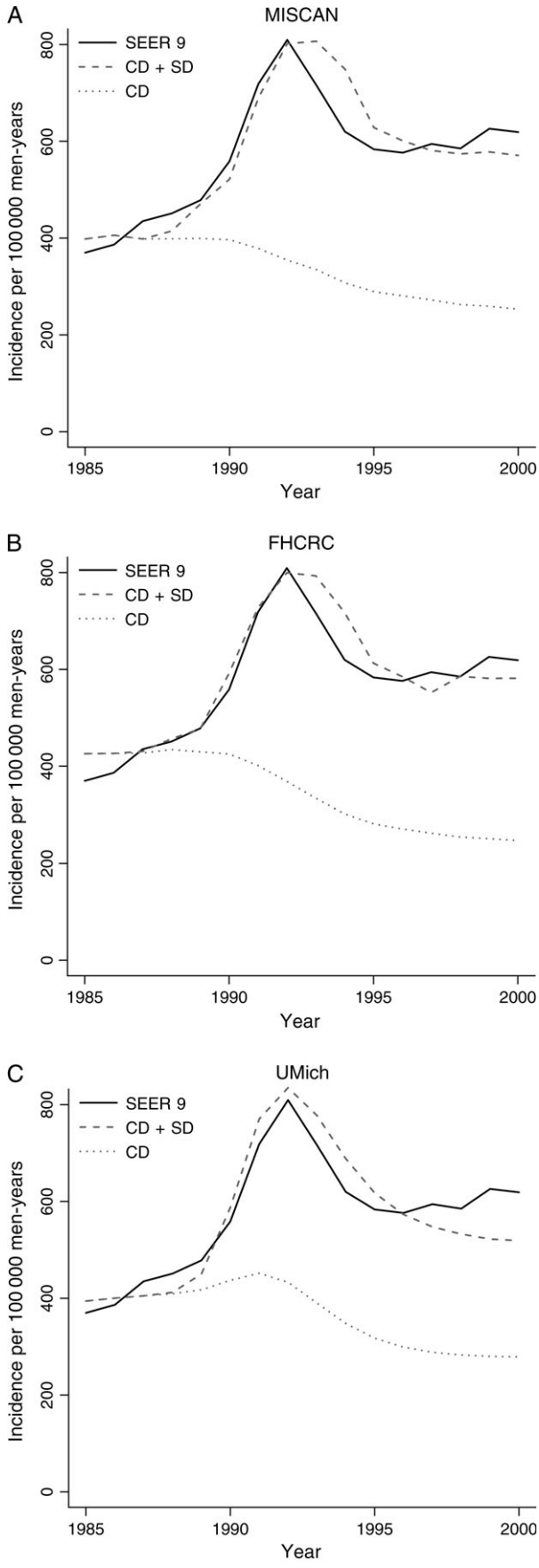


**Figure 3.** Age-adjusted [to the US standard million (21)] incidence of prostate cancer in men aged 50–84 years as observed in SEER 9 and predicted by each of the models. Predicted incidence is separated into screen-detected (SD) and clinically diagnosed (CD) components. **A**) The MISCAN model. **B**) The FHCRC model. **C**) The UMich model. SEER 9 = the nine core catchment areas in the Surveillance, Epidemiology, and End Results program of the National Cancer Institute; MISCAN = microsimulation screening analysis model based on ERSPC Rotterdam, calibrated to SEER 9 incidence; FHCRC = microsimulation model developed at the Fred Hutchinson Cancer Research Center, explicitly linking PSA levels and prostate cancer development; UMich = analytic model developed by A. Tsodikov (University of Michigan); PSA = prostate-specific antigen; ERSPC = European Randomized Study of Screening for Prostate Cancer.

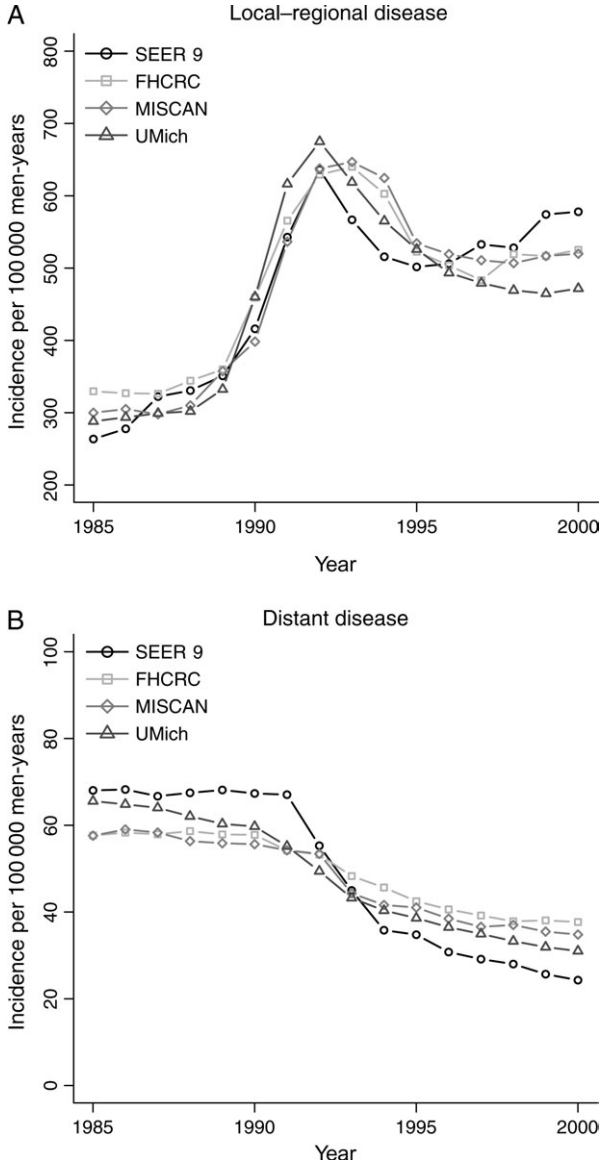


**Figure 4.** Age-adjusted [to the US standard million (21)] incidence in men aged 50–84 years of local–regional-stage and distant-stage prostate cancer as observed (SEER 9) and predicted by each of the models. **A**) Local–regional-stage prostate cancer. **B**) Distant-stage prostate cancer. SEER 9 = the nine core catchment areas in the Surveillance, Epidemiology, and End Results program of the National Cancer Institute; MISCAN = microsimulation screening analysis model based on ERSPC Rotterdam, calibrated to SEER 9 incidence; FHCRC = microsimulation model developed at the Fred Hutchinson Cancer Research Center, explicitly linking PSA levels and prostate cancer development; UMich = analytic model developed by A. Tsodikov (University of Michigan); PSA = prostate-specific antigen; ERSPC = European Randomized Study of Screening for Prostate Cancer.

have been 66% of screen-detected cancers; in the calibrated model, the mean non-overdiagnosed lead time was 6.9 years and the overdiagnosis frequency was 42%.

## Discussion

The lead time and the likelihood of overdiagnosis are quantities that are critical in the assessment of the likely benefits and costs of any screening test; yet, in the case of PSA screening, results have been variable and confusing. This article is the first, to our knowledge, to closely examine the reasons for discrepancies across studies. Our results clearly show that the context or population used to derive the estimates, the definition of lead time used, and the estimation methodology all have important roles.

We considered three definitions of lead time that have been used in previous publications and showed that results differ depending on the definition used. The uncensored definition yields the longest estimated lead times and the non-overdiagnosed definition the shortest. We feel strongly that for future studies to be correctly interpreted, analysts should specify the definition used in their publications. Other definitions have also been reported. For example, McGregor et al. (14) defined overdiagnosis as the detection by screening of disease that would not have led to prostate cancer death. Because the majority of prostate cancer patients do not die of the disease (32,33), the estimates of overdiagnosis due to PSA screening reported by McGregor et al. were considerably higher than ours, exceeding 80%.

The definition of lead time may be constrained or even dictated by the study design. In studies that use stored serum samples, for example, mean lead time is estimated empirically as the average time from the first abnormal PSA test result to prostate cancer diagnosis among the cancer patients with serum samples in the repository. Gann et al. (5) used this method to estimate a mean lead time of 5 years that was based on one serum sample per patient, and Pearson et al. (34) estimated a mean lead time of 3 years by use of serial serum samples. Note that the lead times estimated in these studies refer to patients who were clinically diagnosed during the study (excluding overdiagnosed cancers), that is, corresponding to non-overdiagnosed lead time as shown in Figure 1. However, this approach has some deficiencies. First, the estimates could be seriously affected by the limited follow-up time, for example, 10 years in Gann et al. (5). Tornblom et al. (8), for example, studied prostate cancer incidence in Gothenburg (Sweden) in a cohort of men aged 67 years in 1980 and who had a blood sample taken in 1980. They estimated a median lead time of 7.8 years with 12 years of follow-up and 10.7 years with 20 years of follow-up for PSA levels of 3 ng/mL and greater. Second, this approach assumes that cancer would have been identified by biopsy examination at the time of the abnormal PSA test.

There are also different definitions of overdiagnosis. From an epidemiological or public health perspective, the standard definition is the one that we used in this analysis, namely, the event of other-cause death before the date of clinical diagnosis. However, the clinical literature has suggested an alternative definition, namely, the detection of "clinically insignificant" disease—tumors smaller than 0.2 $cm^3$, organ confined, and with Gleason score less than 7 (35). By this definition, the frequency of overdiagnosis is substantially lower than that reported in the present article (36). However, autopsy studies have shown that tumors that are clinically significant in this sense have a considerable chance of going undiagnosed during a lifetime, as recently reviewed (37). Therefore, we argue that this alternative definition of overdiagnosis, although potentially useful in the future, is likely premature now.

Regarding the issue of context, comparing the results from the MISCAN ERSPC and MISCAN SEER models is revealing. Lead time and overdiagnosis estimates from the original model that was

**Table 1.** PSA screening and the diagnosis of prostate cancer in the SEER 9 population aged 50–84 years during 1985–2000, as predicted by the three models*

| Group | Item | MISCAN | FHCRC | UMich |
|---|---|---|---|---|
| A | No. of screening tests | 7919110 | 7769666 | 7433518 |
| B | No. of men diagnosed with PC† | 238720 | 243565 | 230449 |
| C | No. of screen-detected cancers | 106061 | 103058 | 86975 |
| | Percentage of group A | 1.3 | 1.3 | 1.2 |
| | Percentage of group B | 44.4 | 42.3 | 37.8 |
| D | No. of overdiagnosed cancers | 44499 | 28874 | 19872 |
| | Percentage of group B | 18.6 | 11.9 | 8.6 |
| | Percentage of group C | 42.0 | 28.0 | 22.9 |
| E | Lead time, y | | | |
| | Non-overdiagnosed, mean | 6.9 | 5.9 | 5.4 |
| | Censored, mean | 7.8 | 5.9 | 5.7 |
| | Uncensored, mean | 10.0 (median) | 7.2 | 8.8 |

* PSA = prostate-specific antigen; PC = prostate cancer; SEER 9 = the nine core catchment areas in the Surveillance, Epidemiology, and End Results program of the National Cancer Institute; MISCAN = the microsimulation screening analysis model based on ERSPC Rotterdam, calibrated to SEER 9 incidence; FHCRC = the microsimulation model developed at the Fred Hutchinson Cancer Research Center, explicitly linking PSA levels and prostate cancer development; UMich = the analytic model developed by Dr Tsodikov (University of Michigan); ERSPC = European Randomized Study of Screening for Prostate Cancer.

† Observed number of men diagnosed with prostate cancer = 235112.

based on the Rotterdam data were comparable with those published for PSA screening in the Netherlands (10). Clearly, prostate cancer and PSA screening in the US population seem to be different from the trial setting in Rotterdam (see also Figure 5). Two sets of parameters were changed: In the SEER model, the sensitivity of the screening test was lower than that in the ERSPC model, and the hazard of clinical diagnosis higher, implying an earlier diagnosis in the absence of PSA screening. The lower sensitivity is justified by the lower PSA cutoff at 3 ng/mL in Rotterdam vs 4 ng/mL in the 1990s in the United States, and probably more important, by the higher biopsy compliance rate (90%) in the ERSPC Rotterdam study than in the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial (approximately 40%) (29), which is supposedly representative of US practice. Partially counterweighting these differences may be adherence to the less sensitive sextant biopsy scheme in ERSPC Rotterdam, whereas US biopsy practices gradually adopted extended-core schemes. For the assumed earlier diagnosis in the absence of PSA screening, there is less evidence, but it allowed a higher predicted incidence rate in 1985–1987 without raising incidence over the entire study period. Because lead time and overdiagnosis are defined relative to clinical diagnosis, this assumption also resulted in lower estimates, consistent with the other models. This exercise shows that baseline clinical incidence and the intensity of screening follow-up, both of which may differ across populations, may be important drivers of reported estimates of lead time and overdiagnosis in different studies.

Another source of variation could be caused by model parameterization. In the multiparameter MISCAN model, it is likely that different combinations of parameter values might fit the data equally well, which might impact on lead time and overdiagnosis estimates. By contrast, in the more parsimonious UMich model, parameters are well identified and have narrow confidence intervals (13). However, the impact of this source of variation is likely to be much smaller than that of model structure and assumptions. In this respect, the UMich model differs from both the MISCAN and FHCRC models in that its parameter estimates are based on SEER incidence only, whereas in the other models, data from other sources were also used for parameter estimation.

Finally, we discuss the role of the methods used to estimate lead times and overdiagnosis. In the present investigation, the specific model used plays a relatively minor role. The models yielded lead time and overdiagnosis estimates that were fairly consistent. It is important to note that these estimates depend on a common assumption in all three models—the dissemination of PSA screening is assumed to be the main causal factor of incidence trends since 1985. Although the models do reproduce overall incidence trends, the fit is not perfect. For example, the observed reduction

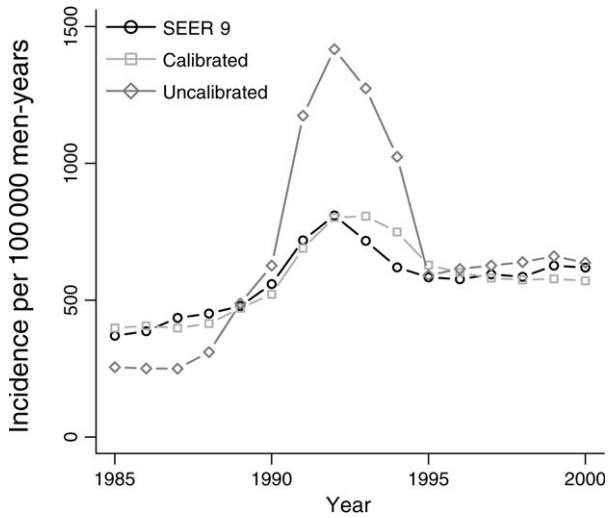


**Figure 5.** Age-adjusted [to the US standard million (21)] incidence of prostate cancer in men aged 50–84 years as observed in SEER 9 and predicted by the calibrated* and uncalibrated† MISCAN models. *Model calibrated to SEER 9 incidence, with risk of clinical diagnosis and test sensitivity estimated from SEER 9 incidence. †Original, uncalibrated model with parameters estimated from incidence in the Netherlands in 1991 and cancer rates observed in the Rotterdam section of ERSPC. SEER 9 = the nine core catchment areas in the Surveillance, Epidemiology, and End Results program of the National Cancer Institute; ERSPC = European Randomized Study of Screening for Prostate Cancer; MISCAN = microsimulation screening analysis.

**Table 2.** PSA screening and the diagnosis of prostate cancer in the SEER 9 population aged 50–84 years from 1985 to 2000 inclusive, as predicted by the calibrated (SEER) and uncalibrated (ERSPC) MISCAN models*

| Group | Item | MISCAN SEER† | MISCAN ERSPC‡ |
|---|---|---|---|
| A | No. of screening tests | 7919110 | 7799540 |
| B | No. of men diagnosed with PC§ | 238720 | 276615 |
| | No. of screen-detected cancers | 106061 | 208763 |
| | % of group A | 1.3 | 2.7 |
| C | % of group B | 44.4 | 75.5 |
| | No. of overdiagnosed cancers | 44499 | 137903 |
| | % of group B | 18.6 | 49.9 |
| D | % of group C | 42.0 | 66.1 |
| | Lead time, y | | |
| | Non-overdiagnosed cancers, mean | 6.9 | 7.9 |
| | Censored, mean | 7.8 | 9.4 |
| E | Uncensored, median | 10.0 | 19.0 |

* PSA = prostate-specific antigen; PC = prostate cancer; SEER 9 = the nine core catchment areas in the Surveillance, Epidemiology, and End Results program of the National Cancer Institute; ERSPC = European Randomized Study of Screening for Prostate Cancer; MISCAN = microsimulation screening analysis.

† The MISCAN model calibrated to SEER 9 incidence from 1985 (before the PSA era) through 2000. This model uses rates of clinical diagnosis and test sensitivities estimated from SEER 9 incidence.

‡ The uncalibrated MISCAN model with parameters estimated from cancer incidence in the Netherlands before the PSA era (1991) and ERSPC Rotterdam trial results.

§ Observed number of men diagnosed with prostate cancer = 235112.

of distant disease incidence is only partially reproduced by the models, replicating results of Etzioni et al. (38), who, using a different model (not calibrated to stage-specific incidence), also found that the model-projected decline in distant-stage incidence was less extreme than that observed in SEER. Also, the estimates of the mean uncensored lead time and overdiagnosis frequency are higher than those reported by Telesca et al. (39). Assuming observed incidence to be the sum of a smooth incidence trend in the absence of screening and an excess incidence that is a function of screening patterns and exponentially distributed lead times, they obtained estimates of mean uncensored lead times of 6.34 years for whites and 7.67 years for blacks. Telesca et al. (39) also showed that their estimates, which were based on population incidence, are sensitive to assumptions about background incidence. Thus, the specific modeling approach used can be influential, although our experience suggests that context and lead time definition are probably more important in explaining the heterogeneity of published lead time and overdiagnosis estimates across studies.

This study has several limitations. The estimates depend on the following assumptions: 1) All incidence trends since 1985 are due to PSA screening, which amounts to assuming an unobserved flat incidence rate in the absence of screening. This assumption may be reasonable, but we do not have independent evidence to support it. 2) We assumed that Mariotto's model of PSA testing practice (19), which we used, is about screening tests. In the construction of her model, all follow-up PSA tests taken after diagnosis were eliminated as well as PSA tests occurring within 3 months of a previous PSA test. 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 it would imply that the screening rate is lower than we assumed. Finally, it is clear that these models were not perfect in predicting observed incidence. Incidence as predicted by the models show a lag of 1 or 2 years with respect to observed incidence, and the models fail to explain fully the decline in distant disease. Consequently, the estimates of mean lead time and overdiagnosis rate will not be perfect either, although it is not clear in what direction they might be biased.

In conclusion, we have presented estimates of lead time and overdiagnosis from three models with different natural history descriptions and estimation strategies, but all applied to the US (SEER 9) population and used common inputs for PSA screening trends and pre-PSA clinical incidence. We have highlighted the critical roles of lead time definition, population context, and estimation methodology. We propose that future studies of lead time clearly define the specific measure used (non-overdiagnosed, censored, and uncensored) and describe key inputs (background incidence, screening protocols, biopsy compliance and sensitivity) that might differ across populations and hence might explain differing estimates of lead time and overdiagnosis associated with PSA screening. We hope that our findings will help explain the substantial variability in the reported estimates of these important measures.

## References


1. De Koning HJ, Liem MK, Baan CA, Boer R, Schroder FH, Alexander FE. Prostate cancer mortality reduction by screening: power and time frame with complete enrollment in the European Randomised Screening for Prostate Cancer (ERSPC) trial. *Int J Cancer.* 2002;98(2):268–273.
2. Gohagan JK, Prorok PC, Kramer BS, Cornett JE. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial of the National Cancer Institute [see comments]. *J Urol.* 1994;152(5 pt 2): 1905–1909.
3. Oliver SE, May MT, Gunnell D. International trends in prostate-cancer mortality in the "PSA era". *Int J Cancer.* 2001;92(6):893–898.
4. Korfage IJ, Essink-Bot ML, Borsboom GJ, et al. Five-year follow-up of health-related quality of life after primary treatment of localized prostate cancer. *Int J Cancer.* 2005;116(2):291–296.
5. Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *JAMA.* 1995;273(4):289–294.
6. Hugosson J, Aus G, Carlsson S, et al. Would prostate cancer detected by screening with prostate-specific antigen develop into clinical cancer if left undiagnosed? A comparison of two population-based studies in Sweden. *BJU Int.* 2000;85(9):1078–1084.

7. Stenman UH, Hakama M, Knekt P, Aromaa A, Teppo L, Leinonen J. Serum concentrations of prostate specific antigen and its complex with alpha1-antichymotrypsin before diagnosis of prostate cancer. *Lancet.* 1994; 344(8937):1594–1598.
8. Tornblom M, Eriksson H, Franzen S, et al. Lead time associated with screening for prostate cancer. *Int J Cancer.* 2004;108(1):122–129.
9. Auvinen A, Maattanen L, Stenman UH, Tammela T, Ranniko S, Aro J. Lead-time in prostate cancer screening (Finland). *Cancer Causes Control.* 2002;13(3):279–285.
10. 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.
11. Etzioni R, Legler JM, Feuer EJ, Merrill RM, Cronin KA, Hankey BF. Cancer surveillance series: interpreting trends in prostate cancer—part III: quantifying the link between population prostate-specific antigen testing and recent declines in prostate cancer mortality. *J Natl Cancer Inst.* 1999; 91(12):1033–1039.
12. Etzioni R, Penson DF, Legler JM, et al. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Natl Cancer Inst.* 2002;94(13):981–990.
13. Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. *Stat Med.* 2006;25(16):2846–2866.
14. McGregor M, Hanley JA, Boivin JF, McLean RG. Screening for prostate cancer: estimating the magnitude of overdetection [see comments]. *Can Med Assoc J.* 1998;159(11):1368–1372.
15. Draisma G, de Koning HJ. MISCAN: estimating lead-time and over-detection by simulation. *BJU Int.* 2003;92(suppl 2):106–111.
16. 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.
17. Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol.* 1983;118(6):865–886.
18. Feuer EJ, Wun LM. How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization? A dynamic population model approach. *Am J Epidemiol.* 1992;136(12):1423–1436.
19. 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.
20. Centers for Disease Control and Prevention. *National Health Interview Survey 2000.* Hyattsville, MD: National Center for Health Statistics; 2002.
21. National Cancer Institute. *Surveillance Epidemiology and End Results: US Standard Populations*, 1940-2000. Bethesda, MD: NCI; 2008.
22. Inoue LYT, Etzioni R, Morrell C, Muller P. Modeling disease progression with longitudinal markers. *J Am Stat Assoc.* 2008;203(March):259–270.
23. Kimmel M, Flehinger BJ. Nonparametric estimation of the size-metastasis relationship in solid cancers. *Biometrics.* 1991;47(3):987–1004.
24. Bartoszynski R, Edler L, Hanin L, et al. Modeling cancer detection: tumor size as a source of information on unobservable stages of carcinogenesis. *Math Biosci.* 2001;171(2):113–142.
25. Morrell CH, Pearson JD, Carter HB, Brant LJ. Estimating unknown transition times using a piecewise nonlinear mixed effects model in men with prostate cancer. *J Acoust Soc Am.* 1995;90:45–53.
26. Whittemore AS, Lele C, Friedman GD, Stamey T, Vogelman JH, Orentreich N. Prostate-specific antigen as predictor of prostate cancer in black men and white men. *J Natl Cancer Inst.* 1995;87(5):354–360.
27. Inoue LY, Etzioni R, Slate EH, Morrell C, Penson DF. Combining longitudinal studies of PSA. *Biostatistics.* 2004;5(3):483–500.
28. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. *N Engl J Med.* 2003;349(3): 215–224.
29. 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.
30. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics.* 1984;40(1):1–14.
31. Zelen M, Feinleib M. On the theory of screening for chronic diseases. *Biometrika.* 1969;56(3):601–661.
32. Albertsen PC, Hanley JA, Gleason DF, Barry MJ. Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer [see comments]. *JAMA.* 1998;280(11): 975–980.
33. American Cancer Society. *Cancer Facts & Figures 2007.* Atlanta, GA; 2007.
34. Pearson JD, Luderer AA, Metter EJ, et al. Longitudinal analysis of serial measurements of free and total PSA among men with and without prostatic cancer. *Urology.* 1996;48(6A suppl):4–9.
35. Epstein JI, Walsh PC, Carmichael M, Brendler CB. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. *JAMA.* 1994;271(5):368–374.
36. Loeb S, Gonzalez CM, Roehl KA, et al. Pathological characteristics of prostate cancer detected through prostate specific antigen based screening. *J Urol.* 2006;175(3 pt 1):902–906.
37. O'Donnell H, Parker C. What is low-risk prostate cancer and what is its natural history? *World J Urol.* 2008;26(5):415–422.
38. Etzioni R, Gulati R, Falcon S, Penson DF. Impact of PSA screening on the incidence of advanced stage prostate cancer in the United States: a surveillance modeling approach. *Med Decis Making.* 2008;28(3): 323–331.
39. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics.* 2008;64(1):10–19.

## Funding

Cancer Intervention and Surveillance Modeling Network, through a National Cancer Institute cooperative agreement mechanism (U01-CA88160 to Erasmus MC [H.K]).

## Notes

The authors are solely responsible for the study design, the collection and analysis of the data, the interpretation of the results, the preparation of the manuscript, and the decision to publish the manuscript.

Manuscript received July 18, 2008; revised December 9, 2008; accepted December 31, 2008.