ARTICLES

Lead Times and Overdetection Due to Prostate-Specific Antigen Screening: Estimates From the European Randomized Study of Screening for Prostate Cancer

Gerrit Draisma, Rob Boer, Suzie J. Otto, Ingrid W. van der Cruijsen, Ronald A. M. Damhuis, Fritz H. Schröder, Harry J. de Koning

Background: **Screening for prostate cancer advances the time of diagnosis (lead time) and detects cancers that would not have been diagnosed in the absence of screening (overdetection). Both consequences have considerable impact on the net benefits of screening.** ***Methods:*** **We developed simulation models based on results of the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC), which enrolled 42 376 men and in which 1498 cases of prostate cancer were identified, and on baseline prostate cancer incidence and stage distribution data. The models were used to predict mean lead times, overdetection rates, and ranges (corresponding to approximate 95% confidence intervals) associated with different screening programs.** ***Results:*** **Mean lead times and rates of overdetection depended on a man's age at screening. For a single screening test at age 55, the estimated mean lead time was 12.3 years (range = 11.6–14.1 years) and the overdetection rate was 27% (range = 24%–37%); at age 75, the estimates were 6.0 years (range = 5.8–6.3 years) and 56% (range = 53%–61%), respectively. For a screening program with a 4-year screening interval from age 55 to 67, the estimated mean lead time was 11.2 years (range = 10.8–12.1 years), and the overdetection rate was 48% (range = 44%–55%). This screening program raised the lifetime risk of a prostate cancer diagnosis from 6.4% to 10.6%, a relative increase of 65% (range = 56%–87%). In annual screening from age 55 to 67, the estimated overdetection rate was 50% (range = 46%–57%) and the lifetime prostate cancer risk was increased by 80% (range = 69%–116%). Extending annual or quadrennial screening to the age of 75 would result in at least two cases of overdetection for every clinically relevant cancer detected.** ***Conclusions:*** **These model-based lead-time estimates support a prostate cancer screening interval of more than 1 year. [J Natl Cancer Inst 2003;95:868–78]**

Whether asymptomatic men benefit from screening for prostate cancer is an unresolved question. Proponents of screening point to the fact that screen-detected cancers tend to have a favorable stage distribution (i.e., they are localized, and possibly curable, cancers) compared with clinically diagnosed cancers *(1,2)* and to decreasing prostate cancer mortality trends following the introduction of prostate-specific antigen (PSA) screening *(3,4).* Opponents of screening stress the fact that some men will be treated unnecessarily, because they would not have been diagnosed with prostate cancer without screening. In addition, opponents of screening suggest that the observed decrease in prostate cancer mortality might be due to better treatments or to misclassification of the cause of death *(5–7).* Moreover, they note that decreasing prostate cancer mortality trends have been registered in countries that do not have an active screening program *(8).* However, the effect of screening on prostate cancer mortality can only be established in a randomized clinical trial. Two such trials have been initiated: the European Randomized Study of Screening for Prostate Cancer (ERSPC) *(9–11)* and the U.S. National Cancer Institute-sponsored Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial *(12).* Results of those trials will not be available for several years.

Primary treatments for prostate cancer have adverse side effects. For example, among patients in the Rotterdam section of the ERSPC who underwent radical prostatectomy after a prostate cancer diagnosis, 80%–90% reported erectile dysfunction and 39%–49% reported urinary incontinence *(13).* Among the patients who underwent radiotherapy, 30%–35% reported bowel problems, 41%–55% reported impotence, and 6%–7% reported incontinence. Screening advances the time of diagnosis (lead time) and detects cancers that would not have been diagnosed in the absence of screening (overdetection). The impact of the adverse effects of primary treatment in the total balance of costs and benefits of screening will depend on the mean lead time and the rate of overdetection due to PSA screening.

Mean lead times due to PSA screening have been estimated in retrospective studies that used stored blood samples obtained from individuals who were later clinically diagnosed with prostate cancer; those estimates range from 5 to 10 years *(14–17).* The possibility of overdetection is indicated by the large percentage (20%–50%, depending on age) of men who died without prostate cancer symptoms but were found, at autopsy, to have prostate tumors *(18–20).* The rate of overdetection can be expressed in several ways (Fig. 1). Zappa et al. *(21)* measured overdetection as the estimated percent increase in the prostate cancer incidence rate in Italy caused by biennial screening. Etzioni et al. *(22)* calculated the amount of overdetection in the current PSA testing practice in the United States as a percentage

Affiliations of authors: G. Draisma, S. J. Otto, H. J. de Koning (Department of Public Health), I. W. van der Cruijsen, F. H. Schröder (Department of Urology), Erasmus MC, University Medical Center, Rotterdam, The Netherlands; R. Boer, Department of Public Health, Erasmus MC, University Medical Center, and Rand Health, Santa Monica, CA; R. A. M. Damhuis, Rotterdam Cancer Registry, Rotterdam.

Correspondence to: Gerrit Draisma, Ph.D., Dept. of Public Health (Rm. H-Ee20-87), Erasmus MC, University Medical Center Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands (e-mail: g.draisma@erasmusmc.nl).

See "Notes" following "References."

Journal of the National Cancer Institute, Vol. 95, No. 12, © Oxford University Press 2003, all rights reserved.

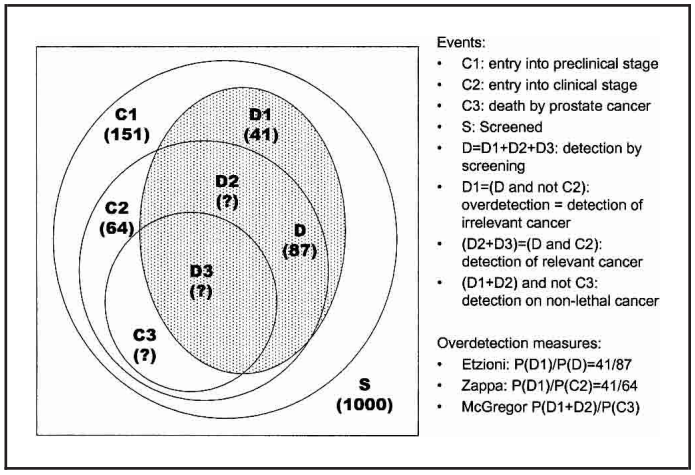

Fig. 1. Events in the natural history of prostate cancer and screening. All men are screened (S). Some of these men develop screen-detectable preclinical cancer (C1); a subset of C1 develops clinical cancer in the absence of screening (C2); a subset of C2 dies of prostate cancer (C3). The men with screen-detected cancer (D) are a subset of C1 and may be grouped into those who would not have been diagnosed with prostate cancer in the absence of screening (D1), those who would have been diagnosed with prostate cancer in the absence of screening but would not have died of prostate cancer (D2), and those who would have died of prostate cancer in the absence of screening (D3). The numbers represent the probability (P) of each event (×1000), as predicted by the basic model. Question marks indicate values that we have not estimated. The overdetection measures used by Zappa et al. *(21),* Etzioni et al. *(22),* and McGregor et al. *(23)* are shown at the lower right.

of total detection by considering the probability of overdetection to be equal to the probability of dying of other causes during the lead time. Finally, McGregor et al. *(23)* defined overdetection as the detection of nonlethal cancer.

We estimated mean lead times and overdetection rates associated with several different PSA screening programs with the simulation program MISCAN, an acronym for MIcrosimulation SCreening ANalysis *(24–26).* MISCAN models were validated against data from the Rotterdam section of the ERSPC trial, which enrolled 21 166 men in the control (i.e., unscreened) arm and 21 210 men in the screened arm and in which 1498 prostate cancers were diagnosed.

Subjects and Methods

ERSPC Trial, Rotterdam Section

The Rotterdam section of the ERSPC, one of eight participating centers in the ERSPC, was initiated in 1994 in the region of Rotterdam and was approved by the Erasmus MC medical ethics committee and the Dutch Ministry of Health *(10).* All men in the Rotterdam region aged 55–74 years were identified with the use of population registry data and invited to participate in the trial. Men who completed and returned the written informed consent form (response rate 49.2%) were randomly assigned to the screened arm or to the control (i.e., unscreened) arm of the trial. Prostate cancer cases among subjects in both arms of the trial were identified through systematic record linkage to the Rotterdam Cancer Registry. The Rotterdam Cancer Registry is part of the Netherlands Cancer Registry, which started in 1989. The data management office of the Rotterdam ERSPC center could access the medical files of participants who provided written informed consent to the study. In the January 2002 contribution of the Rotterdam center to the central ERSPC database, follow-up was considered complete up to July 2000 for subjects in both arms of the trial. By July 2000, trial enrollment was completed, with 21 166 men in the control arm and 21 210 men in the screened arm; 19 970 men (94%) in the screened arm had received a first screening test and 3545 (18%) of these men had received a second screening test. Two different screening protocols were used over the course of the Rotterdam ERSPC. The first 9766 men assigned to the screening arm received a digital rectal examination (DRE), a transrectal ultrasound (TRUS), and a PSA test. Those who had an abnormal DRE or TRUS, or whose serum PSA level was least 4 ng/mL, were referred for a prostate biopsy. (For the first 6894 men enrolled in the trial, the protocol indicated an early recall visit after 1 year for those with a negative biopsy result. The early recall visits were later dropped from the protocol.) The remaining 10 204 men invited for the first round and all men invited for the second round received a PSA test only; men who had a PSA level of 3 ng/mL or higher were referred for a standard sextant biopsy of the prostate *(27).* The recommended treatments for locally confined neoplasms were radical prostatectomy or radiotherapy.

By July 1, 2000, 1075 cases of prostate cancer had been detected in the first round of screening (including those found in the early recall visits), and 143 cases had been detected in the second round of screening. Only 23 cancers were diagnosed outside the screening program among men in the screened arm (interval cancers). A total of 221 prostate cancers were identified among men in the control arm. Thirty-six cases could not be classified because they never appeared for screening (n = 17) or they did not have a biopsy at the trial center following a positive PSA test (n = 19).

Basic MISCAN Prostate Cancer Model

MISCAN models are designed to evaluate cancer screening programs *(24–26).* MISCAN models use a Markov process of states and transitions to simulate and compare individual life histories in the presence and in the absence of a cancer screening program. Information about the epidemiology and natural history of the studied cancer, population characteristics, and screening modalities are used to set the key parameters of the model (see next section).

For this study, we constructed a model of the development of prostate cancer up to the time of detection by screening or clinical diagnosis. In the model, prostate cancers were characterized according to their stage and differentiation grade. Cancer stages in the model were localized cancer [corresponding to International Union Against Cancer (UICC) *(28)* tumor–node–metastasis (TNM) stages T1/2 N0/X M0/X], regional cancer (TNM stages T3+ or N+ with M0/X), and distant cancer (any TN and M1). The model distinguished three differentiation grades: Gleason score less than 7, Gleason score of 7, and Gleason score more than 7. Thus, the model distinguished nine different preclinical states. Screening was assumed to consist of a single screening test with a sensitivity that depended on the stage of the cancer. This simplification of the actual screening procedures seemed justifiable because the two screening protocols used in the ERSPC trial had similar detection rates (i.e., 55 cases per 1000 men screened for the initial protocol and 53 cases per 1000 men screened for the later protocol). A detailed description of the basic MISCAN model for prostate cancer is provided in the Appendix.

Model validation. Table 1 lists the data used to validate the model. Those data consisted of baseline (i.e., pre-screening) incidence of prostate cancer in The Netherlands in 1991 *(29),* baseline stage distribution of clinically diagnosed cancers [data for 1992 and 1993, supplied by the Rotterdam Cancer Registry *(30)*], and trial results up to July 1, 2000, including the number, stage, and grade distribution of screen-detected cancers and interval cancers among men in the screened arm and of cancers diagnosed among men in the control arm. First-round results included the cases detected during the early recall visits of the initial screening protocol.

Excluded from the analysis were all cases diagnosed since June 2000 (because of incomplete registry), cases diagnosed before randomization (i.e., prevalent cases), 17 cases in men who never appeared for screening, 10 cases in men older than 75 years, and 19 cases in men in the screened arm who had a positive screen test (PSA $\geqslant$3 ng/mL or positive DRE or TRUS) but did not undergo a biopsy at the trial center. Although most of the latter cases should probably be considered screen-detected cancers, we did not include them in this analysis because classification as screen-detected would not have affected results. We did not have complete information about the distributions of tumor stage and differentiation grade for 78 (35%) of 221 cases identified in the control arm and for 52 (4%) of the 1218 screen-detected cases, primarily because of missing Gleason scores; tumor stage was missing for two cases in the control arm and for 13 cases in the screened arm. We used only cases for which information was complete to fit stage and grade distribution in the model.

For each set of model parameters (cumulative incidence, transition probabilities, duration parameters, test sensitivities), the simulation model predicted the number of cases for each of the categories listed in Table 1. Model parameters were estimated by minimizing the difference between the observed and the predicted numbers of cancers as measured by the chi-square values ([observed – predicted]^2/predicted). Details on parameter estimation are given in the Appendix.

Variants of the basic model. We considered three variations of the basic model, which is shown in Fig. 2. The first alternative model included latent localized stage cancers (i.e., cancers with a very slow rate of development but detectable by screening as localized cancer), which would have had a negligible probability of clinical diagnosis. Such a model was suggested by the high detection rates in the trial (40–50 cases per 1000 men screened, compared with an incidence of 3.2 cases per 1000 man-years in the control arm). The second alternative model assumed that the duration of preclinical cancer stages followed exponential distributions rather than Weibull distributions. In this simplified model, the duration of the preclinical phase would have a larger variance than in the basic model. In the third alternative model, the differentiation grade (Gleason score) of a tumor could not change once it had become detectable by screening. Because the Gleason score is highly predictive for survival after treatment for prostate cancer, we wanted to verify that our data could rule out this possibility. All models were fit to baseline population data and trial data jointly and, separately, to trial data only, to investigate whether (self-) selection of the trial population affected trial results and our estimates of lead time and overdetection.

To correct for contamination, i.e., screening in the control arm, we included such screening in all models as a yearly screening test with a low attendance rate. In the models that were fitted to both baseline and trial data, we estimated the contamination rate jointly with the other model parameters. However, in models that were fitted to trial data alone, we used an estimated contamination rate of 20 tests per 1000 man-years (*see* Appendix). This estimate is based on the percentage of cases presenting without previous symptoms (30%). Assuming that these tests are due to screening and that the detection rates are the same as those in the screened arm, the contamination rate was calculated as 17 tests per 1000 man-years (using the first-round detection rate) or 22 tests per 1000 man-years (using the second-round detection rate). The estimated rate of PSA testing in the control arm was 71 tests per 1000 man-years *(31),* but the effective screening rate is less than this, because some of the tests were diagnostic tests rather than screening tests.

For all models, the goodness-of-fit is reported as the sum of the chi-square values for the trial data only. The corresponding *P* value was calculated with 42 degrees of freedom, which corresponded to the number of independent counts in the trial data shown in Table 1, and assumed independence between age-specific incidence or detection rates and the corresponding stage-distribution data. These *P* values should be interpreted as only approximate measures of the goodness-of-fit: the effective number of degrees of freedom is certainly less than 42, because the parameters that contributed to that number were estimated from the data and because of possible dependence between incidence and stage-distribution data.

Estimating prostate cancer overdetection rates and lead times. We used the estimated natural history and screening parameters to simulate the effects of nine different screening programs on prostate cancer overdetection rates and lead times for a hypothetical cohort of 1 million men. The nine screening programs were single-screen tests at age 55, 60, 65, 70, or 75 years; annual screening from age 55 to 67; annual screening from age 55 to 75; screening at 4-year intervals, with tests at age 55, 59, 63, and 67 years; and screening at 4-year intervals, with tests at age 55, 59, 63, 67, 71, and 75 years.

The simulations considered the life history of each cohort member in the presence and absence of screening. Cancers detected by screening were divided into two categories: relevant cancers, which are cancers that would have been diagnosed within the person's lifetime in the absence of screening, and irrelevant cancers, which are cancers that would not have been diagnosed within the person's lifetime in the absence of screening. Overdetection was defined as the detection of irrelevant cancers. We calculated the amount of overdetection first as the relative increase, caused by screening, of the number of men with a cancer diagnosed during their lifetime, as in *(21),* and second as the fraction of irrelevant cancers of all detected cancers, as in *(22).* Lead time was defined as the amount of time, in years, between prostate cancer detection and either clinical diagnosis in the absence of screening or death by other causes. We report the mean lead times for all screen-detected cancers and for the screen-detected relevant cancers only. The first quantity measures the number of years patients live with a cancer diagnosis; the second quantity allows comparisons, with lead times estimated in retrospective studies.

We calculated 95% confidence intervals (CIs) for the mean lead times associated with screening at ages 55, 59, 63, and 67 by using the profile likelihood approach detailed in the Appendix. Because that approach was very computer intensive, instead of repeating the procedure for the other outputs (lead times in

Table 1. Model validation*

		Observed		Model prediction	
		n	Incidence	n	Incidence
		Baseline			
	Age group, y				
Incidence per 1000 man-years†	45–50	11	0.02	24.55	0.05
	50–55	58	0.14	66.35	0.16
	55–60	131	0.36	165.80	0.46
	60–65	387	1.19	368.34	1.13
	65–70	734	2.59	694.98	2.45
	70–75	950	4.50	947.44	4.49
	75–80	969	6.57	985.53	6.68
	80–85	668	7.98	696.96	8.33
	>85	431	8.52	390.11	7.71
Trial target population	55–75	2202	1.86	2176.56	1.91
Stage distribution, %‡	Stage		%		%
(n = 1637)	Localized		58.03		57.71
	Regional		18.81		19.44
	Distant		23.15		22.85
		Control arm			
	Age group, y				
Incidence per 1000 man-years	55–60	22	1.51	12.83	0.88
	60–65	29	1.59	34.03	1.87
	65–70	70	4.11	63.25	3.71
	70–75	80	5.96	82.35	6.13
	>75	20	4.60	33.83	7.77
Trial target population	55–75	201	3.18	192.46	2.96

Stage and grade distribution, %									
(n = 143)	Stage	G<7	G = 7	G>7	Total	G<7	G = 7	G>7	Total
	Localized	43.36	16.78	8.39	68.53	38.68	23.56	11.60	73.84
	Regional	5.59	9.79	9.79	25.17	4.16	6.16	6.52	16.85
	Distant	0.00	2.10	4.20	6.29	0.05	2.79	6.47	9.31
	Total	48.95	28.67	22.38	100.00	42.89	32.51	24.60	100.00

		Screened arm							
		Screening round 1		Screening round 2		Screening round 1		Screening round 2	
		n	Rate	n	Rate	n	Rate	n	Rate
	Age group, y								
Detection rate per 1000 men screened	55–60	176	27.65	7	26.02	144.43	22.69	5.61	20.87
	60–65	239	44.80	43	33.10	253.89	47.59	38.51	29.64
	65–70	365	76.60	46	42.63	371.33	77.93	50.29	46.61
	70–75	285	86.13	47	52.75	337.17	101.90	42.54	47.74
Trial target population	55–75	1065	53.86	143	40.42	1106.82	55.73	136.95	37.84

Stage and grade distribution of detected cancers, %									
	Round 1 (n = 1034)								
	Stage	G<7	G = 7	G>7	Total	G<7	G = 7	G>7	Total
	Localized	56.96	17.41	4.45	78.82	57.12	17.31	4.23	78.66
	Regional	7.93	9.77	2.90	20.60	8.01	9.73	3.05	20.79
	Distant	0.00	0.10	0.48	0.58	0.00	0.08	0.47	0.55
	Total	64.89	27.27	7.83	100.00	65.14	27.12	7.74	100.00
	Round 2 (n = 132)								
	Stage	G<7	G = 7	G>7	Total	G<7	G = 7	G>7	Total
	Localized	75.76	16.67	1.52	93.94	76.30	11.15	2.95	90.39
	Regional	1.52	3.03	1.52	6.06	4.33	4.09	1.09	9.52
	Distant	0.00	0.00	0.00	0.00	0.00	0.02	0.08	0.09
	Total	77.27	19.70	3.03	100.00	80.63	15.26	4.11	100.00

Interval cancers incidence per 1000 man-years in interval since last screen	Interval, y	n	Incidence	n	Incidence
	0–1	4	0.19	11.00	0.53
	1–2	4	0.25	10.04	0.62
	2–3	7	0.58	8.80	0.74
	3–4	6	0.85	5.64	0.80
	>4	2	0.91	2.92	1.33
	Total	23	0.40	38.40	0.66

*Observed counts used for validation together with counts as predicted by the base model. For stage and grade distribution, only complete cases are shown. Results from the European Randomized Study for Screening for Prostate Cancer (ERSPC) trial include all cases up to July 1, 2000. G = Gleason score.

†Data from the 1991 Netherlands Cancer Registry *(29)*.

‡Data from the Rotterdam Cancer Registry for 1992 and 1993 *(30)*.

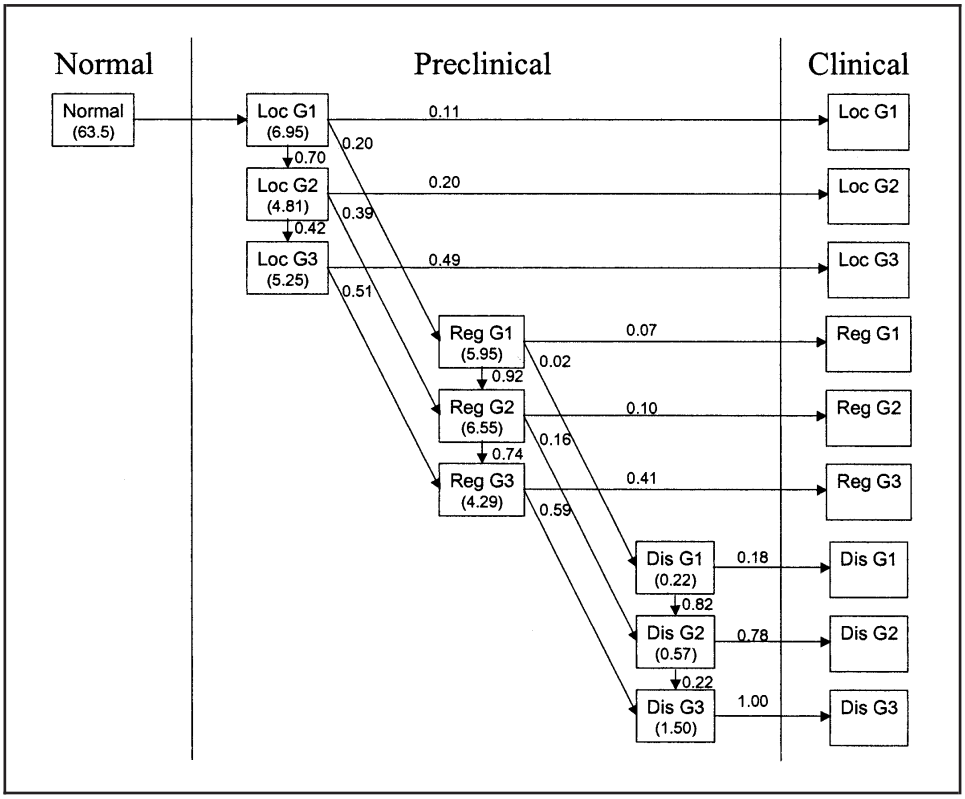

Fig. 2. The MISCAN model of the history of prostate cancer up to clinical diagnosis. The model distinguishes between cancer tumor–node–metastasis (TNM) stages (normal, localized [Loc: T1/2, N0/X, M0/X], regional [Reg: T3/4 or N+ and M0/X], and distant [Dis: any TN, M1]) and between differentiation grades (G1: Gleason score <7; G2: Gleason score = 7; and G3: Gleason score >7). Screening may detect cancer in one of the preclinical stages. The course of events may be interrupted by death from other causes. Key parameters of the basic model, fitted to all data, are indicated in the diagram. Transition probabilities are indicated next to the **arrows;** mean dwelling times in years are indicated in parentheses. Other parameters are cumulative incidence (the probability of ever getting prostate cancer) (0.19); Weibull shape parameters for dwelling times in the normal (10.7), localized (5.3), and regional or distant stages (5.0); sensitivities of the screening test for localized (0.64), regional (0.91), and distant stages (0.97); and contamination (30 tests per 1000 man-years). Time of death by causes not related to prostate cancer was obtained from the standard male life table (Statistics Netherlands, 1991–1995).

other programs, overdetection rates), we obtained ranges that corresponded to approximate 95% CIs in the following way. The profile likelihood approach generated a range of models, one of which corresponded to the lower limit of the 95% CI for mean lead time, and another that corresponded to the upper limit. We used the predictions of these two models as approximate limits of the 95% CIs for the other outputs. This procedure was justified by the close relation between the mentioned outputs (i.e., between lead time in the age 55–67 program and lead times in other programs and/or overdetection rates).

RESULTS

ERSPC Trial Results

Table 1 summarizes trial results and baseline prostate cancer incidence information. Prostate cancer detection rates were high in both screening rounds of the trial: Approximately 54 cases were detected per 1000 men (aged 55–75 years) in the first round and approximately 40 cases were detected per 1000 men in the second round. The first-round detection rate in the screened arm was nearly 30 times the baseline incidence in the 55–75 year age group (1.86 cases per 1000 man-years) and 17 times the incidence in the control arm (3.18 cases per 1000 man-years). Prostate cancer incidence in the control arm was 1.7 times the baseline incidence. The number of interval cancers diagnosed (23) was very small compared with the number of cancers detected in the screened arm (1208) and the number of cancers diagnosed in the control arm (221).

Screen-detected cancers had a more favorable stage distribution than those clinically diagnosed in the pre-screening era, when 23% of the diagnosed cancers were distant-stage cancers (*see* Table 1 and Fig. 3, A). In the trial, only 0.6% of the cancers detected in the first round of screening and none of the cancers detected in the second round were metastatic. The percentage of localized cancers changed from 58% at baseline to 79% of cancers detected in the first round and 94% of cancers detected in the second round of screening. The cancers diagnosed in the control arm had an intermediate distribution: 6.3% were distant and 69% were localized.

Screen-detected cancers were also more well-differentiated than clinically diagnosed cancers. For instance, only 8% of the cancers detected in the first round of screening and 3% of those detected in the second round were poorly differentiated (Gleason score >7), compared with 22% of the cancers found in the control arm of the trial (Fig. 3, B). Gleason scores were not available for cancers detected in the pre-screening era.

Goodness of Fit of Model Predictions

When fitted to both baseline and trial data, the basic model accurately predicted the main characteristics of the data: the baseline age-specific incidence (crude incidence = 1.91 cases per 1000 man-years predicted versus 1.86 cases observed in the 55–75 year age group; *see* also Fig. 4), the overall incidence in the control arm (2.96 cases per 1000 man-years predicted versus 3.18 cases observed in the 55–75 year age group), and the overall detection rate in both the first (55.73% predicted versus 53.86% observed) and second screening rounds (37.84% predicted versus 40.42% observed) of the trial (Table 1). The tumor stage and grade distributions predicted by the model were also similar to the observed distributions. The estimated test sensitivities in the model were 64%, 91%, and 98% for local, regional, and distant cancers, respectively. In the model, the difference between incidence in the control arm and baseline incidence was the result of screening in the control arm at a rate of 30 tests per 1000 man-years.

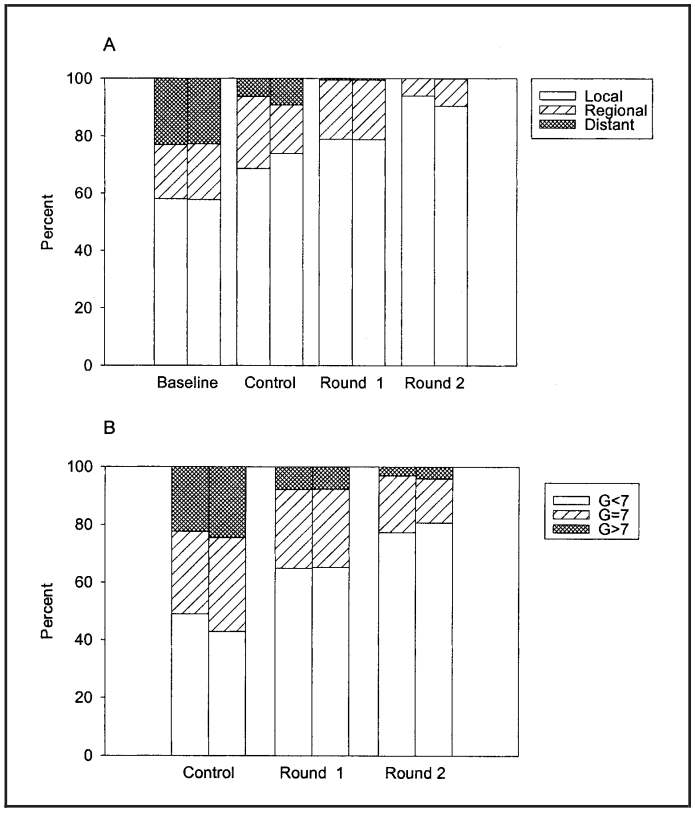

Fig. 3. Distribution of tumor stage and differentiation grade of cancers at clinical diagnosis or detection as observed (**left bar of each pair**) and as predicted by the basic model (**right bar of each pair**). **A**) Tumors were categorized as localized (T1/2, N0/X, M0/X), regional (T3/4 or N+, M0/X), or distant (any TN, M1) cancers. Shown are the baseline stage distribution [data obtained from the 1992–1993 Rotterdam Cancer Registry *(30)*] and the stage distributions of cancers in the control arm of screen-detected cancers from the first and second screening rounds of the Rotterdam European Randomized Study of Screening for Prostate Cancer trial. **B**) Shown are the Gleason score (G) distributions of cancers in the control arm and of screen-detected cancers from the first and second screening rounds of the Rotterdam European Randomized Study of Screening for Prostate Cancer trial.

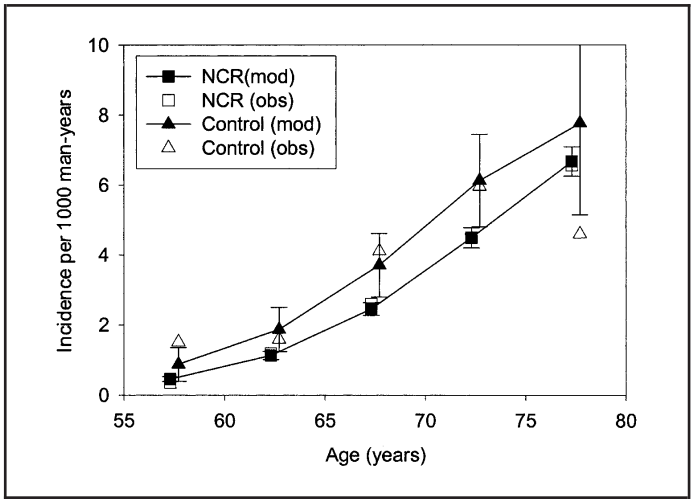

Fig. 4. Age-specific prostate cancer incidence rates in The Netherlands, according to the 1991 Netherlands Cancer Registry [NCR; *(29)*], and in the control arm of the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer trial as observed (obs), and as predicted by the basic model (mod). In the model, opportunistic screening in the control arm occurs at a rate of 3% per year. Model predictions are represented by **solid symbols** and connected by **lines. Vertical bars** represent 95% prediction intervals. Corresponding observed values are indicated by **open symbols.**

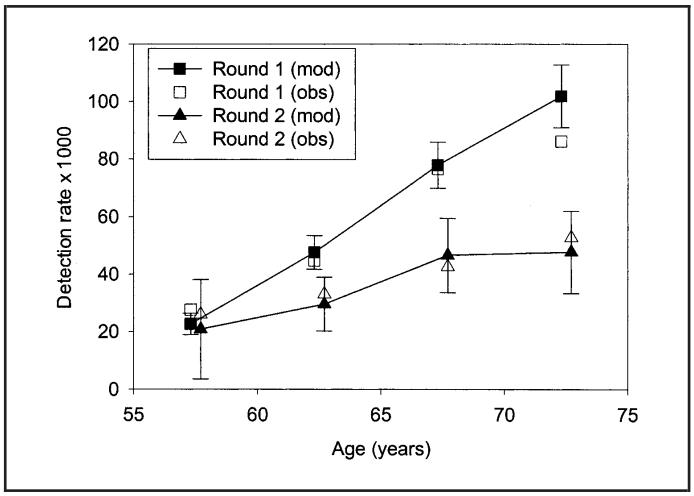

Fig. 5. Age-specific detection rates in the first and second screening rounds of the Rotterdam European Randomized Study of Screening for Prostate Cancer trial, as observed (obs) and as predicted by the basic model (mod). Model predictions are represented by **solid symbols** and connected by **lines. Vertical bars** represent 95% prediction intervals. Corresponding observed values are indicated by **open symbols.**

However, the basic model did not accurately predict age-specific incidence and detection rates in the trial population. In the control arm, observed incidence for the youngest age group (55–60 years) was statistically significantly higher than that predicted by the model, whereas the observed incidence for the oldest age group (75+ years) was statistically significantly lower than predicted (Fig. 4). In addition, the observed detection rates for the first round of screening were slightly higher in the youngest age group (55–60 years) and lower in the oldest age group (70–75 years) than those predicted by the model (Fig. 5). The model also predicted more interval cancers than were observed, especially in the 2-year period directly following screening (21 cancers predicted versus eight cancers observed). The overall incidence of interval cancers predicted by the model was 0.66 per 1000 man-years; the observed incidence was 0.40 per 1000 man-years.

None of the alternative models performed better than the basic model when fitted to all available data, i.e., baseline and trial data. However, we could obtain a statistically acceptable fit for all four models when fitting to the trial data only (Table 2). Notably, the models predicted fewer (25–31) interval cancers in this case. In general, chi-square values obtained for the basic and latent-stage cancer models were similar and were substantially lower than those obtained for the fixed-grade or exponential models.

Lead Time and Overdetection in Prostate Cancer Screening

Table 3 summarizes the predictions from the basic model of mean lead time and rates of overdetection associated with various screening programs. We assumed that each screening program had 100% attendance. The lower and upper limits of the ranges in the table are based on predictions of short and long lead time variants of this model, respectively.

Table 2. Sensitivity analysis*

Source of data for fitting model	Model	Fit		Lifetime risk per 1000 men		Mean lead time, y		Overdetection	
		χ^2†	P‡	Any cancer	Relevant cancer	All cases	Relevant cases	% of detection	% increase in lifetime risk
Trial + baseline	Latent-stage	68.8	.006	153.5	63.7	10.9	11.6	47	62
	Basic	63.0	.020	150.5	63.5	11.2	12.3	48	65
	Exponential	126.4	.000	172.7	62.2	11.3	10.8	54	71
	Fixed-grade	93.2	.000	167.8	63.3	11.4	11.9	51	70
Trial only	Latent-stage	36.7	.702	149.8	52.1	12.2	9.5	59	93
	Basic	37.5	.670	148.6	55.8	12.0	12.8	51	77
	Exponential	44.1	.382	153.7	49.5	12.7	11.6	60	98
	Fixed-grade	48.2	.237	147.7	51.6	12.3	11.7	57	92

*Estimates of lead times and overdetection rates in a screening program with a 4-year interval starting at age 55, as predicted by four different models. Each model was fitted to baseline and trial data or to trial data only.

†As presented in Table 1, control arm and screened arm sections.

‡Calculated with 42 degrees of freedom.

Table 3. Predictions of mean lead time and overdetection rates associated with screening from the basic model*

	Type of cancer		
	Any	Relevant	Irrelevant
Lifetime risk per 1000 men†	151‡ (145 to 166)‡	64	87 (80 to 103)
Mean sojourn time§, y† (range)	12.7 (12.1–14.2)	15.4	10.8 (10.0–12.5)

Screening program	Age, y	Mean lead time‖, y (range)		Detection per 1000 men			Overdetection (range)	
		All cases	Relevant cases	All cases	Relevant cases	Irrelevant cases	% of detection	% increase lifetime risk
Single	55	12.3 (11.6–14.1)	12.8 (12.0–14.6)	15	11	4	27 (24–37)	6 (5–9)
	60	11.0 (10.4–12.4)	11.5 (11.0–13.0)	31	19	12	38 (34–47)	18 (15–25)
	65	9.5 (9.0–10.5)	10.0 (9.6–11.0)	52	28	24	47 (43–55)	38 (33–49)
	70	7.7 (7.4–8.3)	8.1 (7.9–8.7)	64	30	34	53 (50–60)	54 (49–60)
	75	6.0 (5.8–6.3)	6.2 (6.0–6.6)	54	24	30	56 (53–61)	47‡
Interval	Every y, 55–67	12.3 (11.8–13.3)	13.7 (13.3–14.7)	103	52	51	50 (46–57)	80 (69–116)
	Every y, 55–75	11.6 (11.1–12.6)	13.4 (13.0–14.4)	140	61	79	56 (54–61)	124 (111–153)
	Every 4 y, 55–67	11.2 (10.8–12.1)‡	12.3 (11.9–13.2)	87	45	41	48 (44–55)	65 (56–87)
	Every 4 y, 55–75	10.3 (9.9–11.2)	11.7 (11.3–12.5)	123	57	66	54 (51–59)	105 (95–124)

*In the model, the indicated screening programs were applied to a hypothetical cohort of 1 million men, assuming 100% attendance. Detected cancers were classified as relevant (i.e., those that would have been diagnosed in the absence of screening) or irrelevant (i.e., those that would not have been diagnosed in the absence of screening). The rate of overdetection is expressed as a percentage of total detection and as the percent increase in the lifetime risk of a prostate cancer diagnosis.

†Estimates refer to situation without screening.

‡Prediction from basic model and range supported by the data (approximate 95% confidence interval). Ranges are presented only for those predictions that have a close relation to mean lead time in the 4-year interval, age 55–67, screening program, for which the range is a proper 95% confidence interval.

§Sojourn time = duration of the screen-detectable preclinical phase of the disease.

‖Lead time = time that diagnosis is advanced by screening.

With a single screening test, the estimated mean lead time decreased with screening age from 12.3 years (range = 11.6–14.1 years) at age 55 to 6.0 years (range = 5.8–6.3 years) at age 75. The estimated mean lead times associated with screening programs that have regular screening intervals ranged from 9.9 to 13.3 years. The estimated mean lead times associated with annual screening were approximately 1 year longer than those associated with 4-year interval screening. These lead-time estimates apply to all cases, both relevant cases that would have been diagnosed without screening and irrelevant cases that would not. Mean lead-time estimates for relevant cases only were slightly higher than they were for all cases: up to 0.5 year longer in screening with a single test or 1–2 years longer in interval screening.

The basic model predicted that the lifetime risk of developing screen-detectable cancer was 151 per 1000 men (range = 145–166 per 1000 men) (Table 3). Only 64 per 1000 men would be clinically diagnosed with prostate cancer in the absence of screening (relevant cases). To illustrate how overdetection is calculated, consider screening with a single test at age 65. The test was predicted to detect 28 relevant cases and 24 irrelevant cases per 1000 men screened; thus, screening would detect 28 of 64 (44%) relevant cancers, overdetection would occur in 24 of 52 (47%) detected cases, and the lifetime risk of a cancer diagnosis would rise from 64 to 88 per 1000 men, an increase of 38%. With a single screening test at age 70, the largest fraction of the clinically relevant cancers, 47% (30 of 64), would be detected. However, the test would detect irrelevant cancers in 34

of 64 detected cases, corresponding to an overdetection rate of 53% (range = 50%–60%), increasing the lifetime risk of a cancer diagnosis from 64 to 98 per 1000 men, an increase of 54% (range = 49%–60%).

The model predicted that screening with a 4-year interval from age 55 to 67 would detect 70% (45 of 64) of all clinically relevant cancers. However, irrelevant cancers would be found in 41 per 1000 men; this number corresponds to an overdetection rate of 48% (range = 44%–55%) and an increase in the lifetime prostate cancer risk from 64 to 106 per 1000, a relative increase of 65% (range = 56%–87%). Annual screening from age 55 to 67 was predicted to increase detection of relevant cancers to 81% (52 of 64), with an overdetection rate of 50% (range = 46%–57%) and an 80% increase in lifetime risk (range = 69%–116%). Extending interval screening programs to the age of 75 may detect up to 95% (61 of 64) of all clinically relevant cancers. However, such extensions would detect at least two irrelevant cancers for every relevant cancer, resulting, for annual screening, in an overdetection rate of 56% (range = 54%–61%) and a relative increase of 124% (range = 111%–153%) in the lifetime risk of prostate cancer.

The alternative models, when fitted to both baseline and trial data, predicted lead times and overdetection rates for the 4-year interval screening program from age 55 to 67 that generally were similar to the predictions from the basic model, i.e., within the ranges presented in Table 3. By contrast, models fitted to trial data only (Table 2) predicted 1-year-longer lead times (12.0–12.7 years), 5% higher rates of overdetection (51%–60%), and a larger increase in the lifetime risk of prostate cancer (77%–98%).

Discussion

Our results suggest that regular screening for prostate cancer may advance diagnosis by at least 10 years. Approximately half of the screen-detected cancers in our models would not have been diagnosed in the absence of screening; in men screened between age 70 and 75, two of three cancers detected by screening would not have been diagnosed without screening. The introduction of screening would lead to a 60%–90% higher incidence of prostate cancer. How valid are these estimates, and how do they compare with other published estimates?

The basic model, fitted to baseline and trial data, reproduced the essential characteristics of the observed data on clinical incidence, detection rates, and tumor stage and Gleason score distributions. Contamination (i.e., opportunistic screening among men in the control group) could explain both the higher cancer incidence and the more favorable stage distribution of cancers in the control arm compared with those at baseline in the pre-screening era. The model-estimated rate of contamination (30 tests per 1000 man-years) was somewhat higher than the estimate of 17–22 tests per 1000 man-years that was based on the fraction of cases in the control arm among men who reported having no symptoms before diagnosis (*see* Appendix).

However, observed incidence and detection rates in the older age groups of the trial were statistically significantly lower than predicted by the basic model (Figs. 4 and 5). In contrast, models that were fitted to trial data only predicted lower incidence of prostate cancer than was observed in the baseline population registry data in these age groups. These results suggest a (self-) selection effect in the older age groups (i.e., that the older participants in the trial were healthier than average). Because the participation rate in the trial was just under 50%, this possibility cannot be ruled out.

Selection might also be responsible for the difference between observed (23) and predicted (38 in the basic model) numbers of interval cancers. Again, fitting to trial data only resulted in lower predicted numbers of interval cancers in all models (i.e, 25–31 interval cancers).

Data on the sensitivity of PSA screening in combination with sextant biopsies are scarce. Probably the sensitivity is largely determined by the biopsy part of the screening. For sextant biopsies, a sensitivity of approximately 70% can be inferred from the literature *(32–34).* In our models, estimated test sensitivities ranged from 64% to 79% for localized cancer, from 77% to 94% for regional cancer, and from 84% to 99% for distant cancers. Because most screen-detected cancers are localized, our estimates are compatible with the biopsy sensitivity estimates above.

Our estimates are based on the 1991 data on prostate cancer incidence in The Netherlands and on the results of the Rotterdam section of the ERSPC. Consequently, our lead-time and overdetection estimates apply specifically to the 1991 situation in The Netherlands with respect to clinical detection of prostate cancer and to the screening test used in the Rotterdam section (PSA level $\geqslant$3 ng/mL and a positive biopsy). The intensive screening protocol, its changes over time, and the possibility of selection in the trial population should also be kept in mind when generalizing our results to other situations. Finally, our results apply to screening programs with 100% attendance.

Other lead-time estimates have come from retrospective studies, in which blood samples were obtained from healthy men who later developed clinical prostate cancer. For example, Gann et al. *(17)* estimated a mean lead time of 5.5 years for prostate cancer patients whose mean age at baseline was 63.9 years, using a PSA test with a cutoff level of 4 ng/mL (366 cases). Hugosson et al. *(14)* estimated a mean lead time for increased PSA (i.e., PSA $\geqslant$3 ng/mL) of 7 years in a cohort of men who were aged 67 years when their blood samples were collected in 1980 (52 cases). For prostate cancer patients who had PSA levels between 3 and 10 ng/mL, the estimated mean lead time was 9.2 years *(14).* Stenman et al. *(16)* estimated that PSA levels exceeded 4 ng/mL an average of 9.2 years before diagnosis (44 cases). Pearson et al. *(15,35)* used serial PSA measurements to estimate that PSA levels begin to increase exponentially approximately 7.3 years before diagnosis in men with localized or regional cancers and 9.2 years before diagnosis in men with metastatic distant cancers (18 cases in total). Our lead-time estimates (i.e., 8–10 years for a single screening test at age 65 or 70) are higher than those of Gann et al. *(17),* comparable to those of Hugosson et al. *(14),* Stenman et al. *(16),* and Pearson et al. *(15,35),* and substantially higher than the 3-year lead time used in analyses of the recent decline in prostate cancer mortality following the introduction of PSA screening in the United States *(6).*

Lead-time estimates from retrospective studies should be used with caution for two reasons. First, they may be limited by the length of the follow-up period in the study. For instance, our basic model predicts that 40% of the lead times exceed 15 years for a single test at age 67. Second, as illustrated in Table 3, mean lead time varies with a man's age at screening. Consequently, estimates that are based on screening at age 65 or over cannot be used to model screening at age 55.

Our results on overdetection should be compared with the results of Zappa et al. *(21)* and Etzioni et al. *(22),* who used the same definition of overdetection that we did, i.e., the detection of cancer that would not have been found in the absence of screening. Zappa et al. *(21)* expressed overdetection as the relative increase in cancer incidence caused by screening. They estimated the cumulative cancer incidence in a cohort of 10 000 men over a period of 10 years in a situation without screening and with screening. For biennial screening programs with five invitations, starting at age 60 or 65, their estimates of the relative increase in cancer incidence were 25%–51% and 65%–93%, respectively. Our estimates were based on a much larger number of cancers (1452 versus 58) but were similar to their estimates. However, when we applied their method for estimating overdetection to our data and used a screening program with a 4-year interval (for men aged 55–67), we obtained a higher estimate of overdetection: screening raised the lifetime risk of a prostate cancer diagnosis by 107% (*see* Appendix).

Etzioni et al. *(22)* used a competing risk model to estimate overdetection in the United States. In that model, the probability of overdetection for a screen-detected case equals the probability of dying of other causes during the lead time. For each age, this probability can be obtained from the standard survival table, assuming a suitable distribution function for lead time. Overall overdetection can be calculated as a weighted sum of these probabilities. Etzioni et al. used mean lead times of 3, 5, and 7 years to obtain estimated overdetection probabilities of 15%, 25%, and 35%, respectively. These estimates apply to screening programs that use 4 ng/mL as the cutoff value for PSA. In their model, only the 5-year lead time was compatible with U.S. incidence trends. Our estimates of mean lead time are much higher than those of Etzioni et al. and are associated with higher rates of overdetection. The difference between these estimates is probably related more to the bias in lead-time estimates referred to above than to the lower cutoff value for PSA used in our trial (3 ng/mL).

In conclusion, these lead-time estimates, which are the first to be based on results of a large-scale screening trial for prostate cancer in a population-based setting, support a screening interval of more than 1 year. Screening for prostate cancer is likely to advance diagnosis considerably and to be associated with considerable overdetection. The net balance of favorable and unfavorable effects of screening remains to be established.

APPENDIX

The MISCAN Model

The MISCAN program was designed at our institution and has been used extensively for the analysis and surveillance of screening programs *(24–26).* In MISCAN, individual life histories are first simulated in the absence of screening. A life history consists of a sequence of states and the time spent in those states (dwelling times). These states and times are generated by a semi-Markov process: at each step a next state is generated with probabilities determined by the present state. The distribution of the dwelling time in the present state is also determined by the present state. Optionally, transition probabilities and dwelling time distributions can be made age-dependent. Death from other causes is generated independently using a standard life table.

Fig. 2 shows the states in the prostate cancer model up to the moment of clinical detection. It specifies all relevant parameter values used in the basic model: cumulative incidence (the probability of ever getting cancer), transition probabilities and mean dwelling times, the Weibull shape parameters that determine the variance of these dwelling times, and the estimated test sensitivities. We used the Dutch male life table (Statistics Netherlands, 1991–1995) for generating the time of death from other causes of each individual.

Screening is superimposed on the life histories in the absence of screening. Preclinical cancers may be detected by screening. Detection depends on attendance and the sensitivity of the screening test for the specific preclinical state. Attendance rates were obtained directly from the trial database.

Parameter Estimation

All parameters shown in Fig. 2 were estimated from the data presented in Table 1 with the use of a model that included the trial population characteristics (e.g., age distributions and attendance rates). For each set of parameter values, the model generates life histories and counts the cases in each of the categories in Table 1. These counts are considered predictions from the model. Although the predictions are random, increasing the sample size in the simulations reduces variance. Parameters are estimated by minimizing the difference between observed and predicted counts, measured as the sum of the chi-square quantities using an adapted version of the simplex optimization method of Nelder and Mead, as outlined in *(36).* Adaptations to the random nature of the objective function are a shrinking value of 0.9 instead of 0.5 and recalculation of all function values at a simplex shrink step *(37).* Optimization was initiated with small sample sizes (i.e., 10 000) and repeated with larger sample sizes (i.e., up to 1 million) when optimization progress was no longer statistically significant. For all models, we tried a number of different starting values and kept the best result.

Confidence Intervals

We used the basic model and a profile likelihood approach to obtain a 95% CI for mean lead time in the 4-year interval screening program for individuals aged 55–67 years. We used penalized optimization to obtain a range of models with different lead times, each of them optimized for the results in Table 1. The penalty added to the objective function (sum of the chi-square values calculated from the difference between observed and predicted numbers in Table 1) was the squared difference between mean lead time predicted by the model and a target lead time chosen from a range (8, 9, . . ., 14 years) of lead times. We considered the sum of chi-square values to be a random variable from a chi-square distribution with 53 degrees of freedom (the number of independent cells) and used it to obtain a likelihood for each model lead time. A 95% CI for mean lead time was calculated from the resulting profile likelihood. The models that correspond to the lower and upper limit of this CI were used to calculate ranges supported by the data for lead time and rates of overdetection in the other programs.

Estimating Contamination

From the medical files of the men in the control arm who were diagnosed with prostate cancer, we found that 30% did not have symptoms before diagnosis. Therefore, we assumed that they had been diagnosed because of opportunistic screening. Assuming further that the sensitivity of screening in the control arm equaled the sensitivity in the screened arm, we calculated the rate of testing using the formula: test rate $= 0.3 \times$ (control arm incidence/screened arm detection rate) $= 0.3 \times (3.18/53.9)$ or $0.3 \times (3.18/40.4) = 17$ or 23 tests per 1000 man-years, depending on whether first-round or second-round detection rates were used. This test rate was less than the rate of 71 PSA tests per 1000 man-years reported in the control arm of the Rotterdam trial, because the latter rate also included diagnostic PSA tests. No direct estimate of the proportion of PSA tests used for screening purposes was available from the data *(31).*

Estimating Overdetection by the Method of Zappa et al. *(21)*

The probability of a man getting prostate cancer at age t can be approximated by the product of the age-dependent cancer incidence $I(t)$

and the probability of still being alive at age *t,* the standard survival function *S(t).* The lifetime risk of getting a cancer diagnosis is approximated by the sum of the products $I(t) \times S(t)$ for all ages *t.* We used the data of The Netherlands Cancer Registry for 1991 and the male Dutch life table to calculate this risk, which was 68 cases in 1000 men. For screening at age 55, 59, 63, and 67 years, we assumed detection rates of 28 cases per 1000 men in the first screen (the observed first-round detection at age 55 in Table 1), and 33, 43, and 53 cases per 1000 men (the observed second-round detection rates in Table 1) in the following screens. Interval cancers appear at a rate of 0.4 cases per 1000 man-years (also from Table 1) in each year up to the age of 70. Thereafter we assumed that incidence returned to the normal age-dependent level of eight cases per 1000 man-years at age 85. Multiplying as before the number of cases detected or clinically diagnosed per 1000 men at age *t* with the probability, *S(t),* of being alive and adding the products resulted in a total of 141 prostate cancer cases per 1000 men. Thus, screening would raise incidence by 107% (141/68–1). This estimate is higher than our model-based estimates of 62%–93% (Table 3). The difference may relate to the assumption that all repeat screenings would have the same detection rate as the second round did.

REFERENCES

(1) Catalona WJ, Smith DS, Ratliff TL, Basler JW. Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. JAMA 1993;270:948–54.

(2) Candas B, Cusan L, Gomez JL, Diamond P, Suburu RE, Levesque J, et al. Evaluation of prostatic specific antigen and digital rectal examination as screening tests for prostate cancer. Prostate 2000;45:19–35.

(3) Hankey BF, Feuer EJ, Clegg LX, Hayes RB, Legler JM, Prorok PC, et al. Cancer surveillance series: interpreting trends in prostate cancer–part I: evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. J Natl Cancer Inst 1999;91:1017–24.

(4) Bartsch G, Horninger W, Klocker H, Reissigl A, Oberaigner W, Schoenitzer D, et al. Prostate cancer mortality after introduction of prostate-specific antigen mass screening in the federal state of Tyrol, Austria. Urology 2001;58:417–24.

(5) Feuer EJ, Merrill RM, Hankey BJ. Cancer surveillance series: interpreting trends in prostate cancer–part II: cause of death misclassification and the recent rise and fall in prostate cancer mortality. J Natl Cancer Inst 1999; 91:1025–32.

(6) Etzioni R, Legler JM, Feuer EJ, Merrill RM, Cronin KA, Hankey BJ. 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; 91:1033–9.

(7) Newschaffer CJ, Otani K, McDonald MK, Penberthy LT. Causes of death in elderly prostate cancer patients and in a comparison nonprostate cancer cohort. J Natl Cancer Inst 2000;92:613–21.

(8) Oliver SE, May MT, Gunnell D. International trends in prostate-cancer mortality in the "PSA ERA". Int J Cancer 2001;92:893–8.

(9) Auvinen A, Rietbergen JB, Denis LJ, Schroder FH, Prorok PC. Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group. J Med Screen 1996;3:97–104.

(10) de Koning HJ, Auvinen A, Berenguer Sanchez A, Calais da Silva F, Ciatto S, Denis L, et al. Large-scale randomized prostate cancer screening trials: program performances in the European Randomized Screening for Prostate Cancer trial and the Prostate, Lung, Colorectal and Ovary cancer trial. Int J Cancer 2002;97:237–44.

(11) de Koning HJ, Liem M, Baan CA, Boer R, Schröder FH. 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;98:268–73.

(12) Gohagan JK, Prorok PC, Hayes RB, Kramer BS. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. Control Clin Trials 2000; 21:251S–72S.

(13) Madalinska JB, Essink-Bot ML, de Koning HJ, Kirkels WJ, van der Maas PJ, Schroder FH. Health-related quality-of-life effects of radical prostatectomy and primary radiotherapy for screen-detected or clinically diagnosed localized prostate cancer. J Clin Oncol 2001;19:1619–28.

(14) Hugosson J, Aus G, Carlsson S, Eriksson H, Lilja H, Lodding P, 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;85:1078–84.

(15) Pearson JD, Carter HB. Natural history of changes in prostate specific antigen in early stage prostate cancer. J Urol 1994;152(5 Pt 2):1743–8.

(16) Stenman UH, Hakama M, Knekt P, Aromaa A, Teppo L, Leinonen J. Serum concentrations of prostate specific antigen and its complex with alpha 1-antichymotrypsin before diagnosis of prostate cancer. Lancet 1994; 344:1594–8.

(17) Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. JAMA 1995;273:289–94.

(18) Franks LM. Latent carcinoma of the prostate. J Pathol Bacteriol 1954;68: 603–16.

(19) Carter HB, Piantadosi S, Isaacs JT. Clinical evidence for and implications of the multistep development of prostate cancer. J Urol 1990;143:742–6.

(20) Watanabe M, Nakayame T, Shiraishi T, Stemmermann GN, Yatani R. Comparative study of prostate cancer in Japan versus the United States; a review. Urol Oncol 2000;5:274–83.

(21) Zappa M, Ciatto S, Bonardi R, Mazzotta A. Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence Screening Pilot Study. Ann Oncol 1998;9:1297–300.

(22) Etzioni R, Penson DF, Legler JM, di Tommaso D, Boer R, Gann PH, et al. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. J Natl Cancer Inst 2002;94:981–90.

(23) McGregor M, Hanley JA, Boivin JF, McLean RG. Screening for prostate cancer: estimating the magnitude of overdetection. CMAJ 1998;159:1368–72.

(24) de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J Natl Cancer Inst 1995;87:1217–23.

(25) Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79–93.

(26) Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. Comput Biomed Res 1999;32:13–33.

(27) Beemsterboer PM, Kranse R, de Koning HJ, Habbema JD, Schröder FH. Changing role of 3 screening modalities in the European randomized study of screening for prostate cancer (Rotterdam). Int J Cancer 1999;84:437–41.

(28) International Union against Cancer (UICC). TNM classification of malignant tumours. 4th ed. 2nd revision. Berlin (Germany): Springer Verlag; 1992. p. 141–4.

(29) Visser O, Coebergh JW, Schouten LJ. Incidence of cancer in the Netherlands, 1991. Utrecht (The Netherlands): The Netherlands Cancer Registry; 1994. p. 13, 45.

(30) Spapen SJ, Damhuis RA, Kirkels WJ. Trends in the curative treatment of localized prostate cancer after the introduction of prostate-specific antigen: data from the Rotterdam Cancer Registry. BJU Int 2000;85:474–80.

(31) Otto SJ, van der Cruijsen IW, Liem MK, Korfage IJ, Lous JJ, Schröder FH, et al. Effective PSA contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. Int J Cancer 2003;105:394–9.

(32) Djavan B, Ravery V, Zlotta A, Dobronski P, Dobrovits M, Fakhari M, et al. Prospective evaluation of prostate cancer detected on biopsies 1, 2, 3 and 4: when should we stop? J Urol 2001;166:1679–83.

(33) Djavan B, Mazal P, Zlotta A, Wammack R, Ravery V, Remzi M, et al. Pathological features of prostate cancer detected on initial and repeat prostate biopsy: results of the prospective European Prostate Cancer Detection study. Prostate 2001;47:111–7.

(34) Vis AN, Boerma MO, Ciatto S, Hoedemaeker RF, Schroeder FH, van der Kwast TH. Detection of prostate cancer: a comparative study of the diagnostic efficacy of sextant transrectal versus sextant transperineal biopsy. Urology 2000;56:617–21.

(35) Pearson JD, Morrell CH, Landis PK, Carter HB, Brant LJ. Mixed-effects regression models for studying the natural history of prostate disease. Stat Med 1994;13:587–601.

(36) Bazaraa MS, Sherali HD, Shetty CM. Nonlinear programming: theory and algorithms. New York (NY): John Wiley and Sons; 1993. p. 353.
(37) Barton RR, Ivey JS. Nelder-Mead simplex modifications for simulation optimization. Manage Sci 1996;42:954–73.

NOTES

Supported by grants EUR-94-869 and EUR-98-1757 (to Erasmus University, Rotterdam) from the Dutch Cancer Society and grants 002-22820 and 2000-2-1016 (to Erasmus University, Rotterdam) from The Netherlands Organization for Health Research and Development (ZONMw), by 5th Framework program grant QLRI-2000-01741 (to Erasmus University, Rotterdam) from the European Union, and by Europe Against Cancer.

We thank Monique Roobol and Mark Wildhagen of the Rotterdam Center of ERSPC for providing data and answers to our questions. Mark Wildhagen is credited for constructing the first MISCAN models for prostate cancer. René Eijkemans of our department has been helpful with the statistical analysis. We thank reviewers and editors for their critical appraisal of the manuscript, which resulted in considerable improvements.