The efficacy of PSA screening: impact of key components in the ERSPC and PLCO trial

Harry J de Koning* M.D.[1], Roman Gulati* M.S.[2], Sue M Moss Ph.D.[3], Jonas Hugosson M.D.[4], Paul F Pinsky Ph.D.[5], Christine D Berg M.D.[6], Anssi Auvinen M.D.[7], Gerald L Andriole M.D.[8], Monique J Roobol Ph.D.[9], E. David Crawford M.D.[10], Vera Nelen M.D.[11], Maciej Kwiatkowski M.D.[12], Marco Zappa M.D.[13], Marcos Luján M.D.[14], Arnauld Villers M.D.[15], Tiago M de Carvalho, M.Sc.[1], Eric J Feuer Ph.D.[16], Alex Tsodikov Ph.D.[17], Angela B Mariotto Ph.D.[16], Eveline AM Heijnsdijk* Ph.D.[1] and R. Etzioni* Ph.D.[2]

asterisks: equal contribution

1. Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands

2. Division of Public Health Sciences, Fred Hutchinson Cancer Research Institute, Seattle, Washington, US

3. Wolfson Institute, Queen Mary University of London, London, United Kingdom

4. Department of Urology, Sahlgrenska University Hospital, Göteborg, Sweden

5. Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland, US

6. Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins Medicine, Baltimore, US

7. School of Health Sciences, University of Tampere, Tampere, Finland

8. Division of Urologic Surgery, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri, US

9. Department of Urology, Erasmus Medical Center, Rotterdam, the Netherlands

10. Urologic Oncology, University of Colorado, Denver, Colorado, US

11. Provinciaal Instituut voor Hygiëne, Antwerp, Belgium

12. Department of Urology, Kantonsspital Aarau, Aarau, Switzerland

13. Unit of Epidemiology, Institute for Cancer Prevention, Florence, Italy

14. Servicio de Urología, Hospital Universitario Infanta Cristina, Universidad Complutense de Madrid, Parla, Madrid, Spain

15. Department of Urology, Centre Hospitalier Regional Universitaire, Lille, France

16. Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, US

17. Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, US

Corresponding author:

Dr. Eveline A.M. Heijnsdijk, Department of Public Health, Erasmus MC, PO BXO 2040, 3000 CA Rotterdam, the Netherlands, Email: e.heijnsdijk@erasmusmc.nl, Phone: +31107043446

Key words: PSA screening, prostate cancer, mortality reduction, modelling

Running head: Impact of components in ERSPC and PLCO

Funding: This work was supported by the National Cancer Institute [Grant Number U01CA157224] as part of the Cancer Intervention and Surveillance Modeling Network (CISNET). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

Disclosures: Dr. de Koning reports grants from CanCon, during the conduct of the study; grants from Beckman Coulter, outside the submitted work; Dr. Auvinen reports grants from Hybritech, during the conduct of the study; personal fees from Epid Research Inc, personal fees from MSD, outside the submitted work;Dr. Andriole reports grants from National Cancer Institute, during the conduct of the study; grants from Medivation (Clinical Research), grants from Progenics (Clinical Research), grants from Blue Earth Diagnostics (Clinical Research), grants from NIH-NIDDK, grants from Prostate Cancer Foundation, grants from Peter Michael Foundation, grants from St. Louis Men's Group Against Cancer, outside the submitted work; Dr. Berg reports personal fees from Medial Early Sign, LLC, personal fees from GRAIL, Inc, outside the submitted work; Dr. Kwiatkowski reports personal fees from Myriad, personal fees from Astellas, personal fees from Janssen, outside the submitted work; Dr. LUJAN reports grants from "Fondo de Investigación Sanitaria" (FIS): 93/0903, 96/0248, 96/1800, 99/0245, 02/0732, andÐ6/0831., grants from "Fundación para la Investigación en Urología" (FIU)., during the conduct of the study; Dr. Heijnsdijk report grants from Beckman Coulter, outside the submitted work; The other authors have nothing to disclosure.

Precis: In contrary to the results of meta-analytic pooling, we show that the ERSPC and PLCO trials are more consistent than published results suggest.

Acknowledgements: The authors thank the National Cancer Institute for access to the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial database. The interpretation and reporting of these data are solely the responsibility of the authors.

Authors' contributions Conceptualization: HdK, RG, PP, AT, AB, EH and RE; Data curation and funding acquisition: SM, JH, CB, AA, GA, MR, EC, VN, MK, MZ, ML, AV**;** Formal analysis: RG, SM, AT, TdC, EH**;** Investigation: RG, TdC, EH**;** Methodology: HdK, RG, SM, PP, TdC, EF, AT, AB, EH and RE**;** Writing, original draft: HdK, RG, EH, RE ; Writing, review and editing: all authors

Abstract

Background: The European Randomized Study of Screening for Prostate Cancer (ERSPC) showed that prostate-specific antigen (PSA) screening significantly reduced prostate cancer mortality (rate ratio (RR)=0.79, 95%CI 0.69–0.91). The U.S. Prostate, Lung, Colorectal, and Ovarian (PLCO) trial showed no such reduction but had a wide confidence interval (RR for prostate cancer mortality=1.09, 95%CI 0.87–1.36). Standard meta-analyses are unable to account for key differences between trials that can impact the estimated effects of screening and the trials' point estimates.

Methods: We calibrated two micro-simulation models to individual-level incidence and mortality data from 238,936 men participating in the trials. A cure parameter for the underlying efficacy of screening was estimated by the models separately for each trial. We changed step-by-step major known differences in trial settings, including enrollment and attendance patterns, screening intervals, PSA thresholds, biopsy receipt, control arm contamination and primary treatment, to reflect a more ideal protocol situation and differences between the trials.

Results: Using the cure parameter estimated for the ERSPC, the models projected 19-21% and 6-8% prostate cancer mortality reductions in the ERSPC and PLCO settings, respectively. Using this cure parameter, the models projected a 37-43% reduction under annual screening with 100% attendance and biopsy compliance and no contamination. The cure parameter estimated for the PLCO was 0.

Conclusions: The observed cancer mortality reduction in screening trials is highly sensitive to trial protocol and practice settings. Accounting for these differences, the efficacy of PSA screening in the PLCO setting is not necessarily inconsistent with ERSPC results.

Introduction

The European Randomized Study of Screening for Prostate Cancer (ERSPC)[1-3] showed a significant prostate cancer mortality reduction of 21% for the PSA screening arm, while the US-based Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial did not show a difference in prostate cancer mortality between arms but had wide confidence intervals (prostate cancer mortality rate ratio 1.09, 95%CI 0.87-1.36).[4] A number of explanations for these seemingly inconsistent results have been debated.[5-8]

Selective trial populations, different protocols and practice settings, including differences in pre-trial screening, receipt of biopsies, control arm contamination, and primary treatments, may have influenced the trial results.

The ERSPC trial was conducted in seven centers in Europe with 162,243 men aged 55-69 at randomization. PSA testing was not common at the start of the trial and the estimated contamination in the control arm was lower than 15%.[7] Most centers used a screening interval of 4 years and a PSA threshold of 3.0 ng/ml for biopsy referral. Approximately 86% of the positive screens were followed by a biopsy.[1] The PLCO trial was conducted in 76,693 men aged 55-74 among whom prior screening was already common. At least 45% of the participants had at least one PSA test before randomization.[4] In addition, control arm participants were screened on average 2.7 times during the 6-year intervention phase of the trial.[9] Annual screening was used and the threshold for a positive PSA test was 4.0 ng/ml. Since in this trial, the biopsies were performed outside the study, only about 35% of participants with a positive screen received a biopsy.[10] Both trials involved variable use of digital rectal exam (DRE) testing.

Because of these differences, the results of the trials are not directly comparable. In standard meta-analyses, the results were simply pooled [11-13], suggesting that PSA screening has little effect on prostate

cancer mortality. Possible reasons for the apparent lack of consistency between the trials have not been evaluated formally to determine their quantitative impact on observed mortality reductions.

The aim of this study is to estimate the impact of trial population, protocols, contamination, and practice settings on the observed prostate cancer mortality reduction. We use two independently designed natural history models, which are informed using individual-level data from both trials, to systematically investigate the impact of these characteristics on the estimated efficacy of PSA screening.

Methods

Data

Individual data from both trials were obtained on: age at randomization, trial arm, screening center, screening test dates and results, occurrence of biopsy, prostate cancer incidence, mode of detection (screen or interval cancer), clinical TNM-stage and Gleason score at diagnosis, primary treatment and date and cause of death. The median follow-up was 11 years for ERSPC[2] and 13 years for PLCO.[4]

Modeling the trials

Two multistate disease course models of the Cancer Intervention and Surveillance Modeling Network (CISNET), the Erasmus MC-MIcrosimulation Screening Analysis (Erasmus-MISCAN) model and the Fred Hutchinson Cancer Research Center (FHCRC) model, were used to simulate the trials. The models were independently developed to describe the natural history of prostate cancer and to investigate prostate cancer progression, screening sensitivity, detection, and improvement in prognosis given screening and primary treatment. The two models have been described extensively[14-17] (https://resources.cisnet.cancer.gov/registry). In short, in the Erasmus-MISCAN model, disease progresses through a sequence of states defined by stage and grade. In each state, there is a probability

of clinical detection and, dependent on the screen sensitivity and attendance, a probability of screen detection.[17,18] In the FHCRC model, PSA growth is externally estimated using results of serial PSA tests from the Prostate Cancer Prevention Trial. The risk of onset of a preclinical screen-detectable tumor increases with age and the risks of progression to metastasis and to disease detection in the absence of screening increases with PSA levels.[15] Detailed descriptions of the models are provided in Supplementary Material 1.

Calibration

Each model was calibrated to the ERSPC and PLCO trials separately. Disease progression rates (for the Erasmus-MISCAN model also the PSA test sensitivity) were calibrated against the incidence and stage distributions of clinically-detected cancers in both control arms and the screen-detected and interval cancers in the screened arms (Supplementary Material 2). We used enrollment patterns, screen attendance, and receipt of biopsies by age and PSA-level to model the number of screens and biopsies in the screened arms of the trials (Table 1). Screening before, during and after the intervention period (contamination) in the PLCO was simulated using a model described previously.[19] Briefly, we assumed that before the trial participants followed screening patterns previously reconstructed for the US population[20], which they also followed after the 6-year intervention phase. We assumed control arm participants had a 20% higher intensity of screening than the general US population during the 6-year invention period to match the estimated average 2.7 screens in this period.[9] For the ERSPC, we assumed a contamination rate of 5% of US population screening patterns, leading to a comparable number of screened men as estimated in several centers.[21-23]

Survival

Both models generated prostate cancer survival from clinical diagnosis in the absence of screening or localized treatment benefits. Prostate cancer survival was estimated using a common proportional hazards regression model with piecewise constant hazards[24] fit to Surveillance, Epidemiology, and End Results (SEER) data for untreated cases diagnosed in 1983-1986, just prior to the advent of PSA screening. This baseline survival was improved for localized cases who received radical prostatectomy, or radiation therapy in combination with hormone therapy, using a hazard ratio of 0.62 and for non-metastatic cases who received radiation monotherapy using a hazard ratio of 0.7.[25] Distributions of treatments depending on age, Gleason score, and stage were based on separate multinomial regression models fit to trial data (Supplementary Material 3). Other-cause survival was generated using US and European life tables.

Modeling screening benefit

The mortality benefit of PSA screening was modeled as a cure probability that depended on the lead time (years by which detection of the cancer is advanced by screening compared to the clinical situation) and was implemented only for screen-detected, non-metastatic, and non-overdiagnosed cases as cure probability = 1 – exp (-cure parameter x lead time). Thus the probability of cure increases with lead time, with diminishing incremental benefit for longer lead times. In the models, cured men were assigned to die at their independently generated date of other-cause death. Men who were not cured died at the same time they would have died if they had not been screened.

In a previous study modeling the PLCO trial, the models substantially over-projected observed prostate cancer mortality despite closely reproducing incidence and stage and grade patterns.[19] Therefore, we included a baseline survival hazard ratio to improve the baseline survival, reasoning there have been improvements in disease management since the period 1983-1986 beyond screening or primary

treatment. In this study, we jointly calibrated this hazard ratio with the cure parameter to the observed prostate cancer mortality data for both trials separately.

Model runs

Each model projected the mortality rate ratio for each trial by year of follow-up. Then, using the cure parameter calibrated to the ERSPC (because the published effect of screening was positive), the models systematically varied key characteristics of the trials. We first replaced observed characteristics (control arm contamination, attendance patterns, receipt of biopsies) in the ERSPC setting in a cumulative way with idealized versions of no control arm contamination, perfect attendance, perfect compliance with biopsy recommendations, then substituted the idealized ERSPC setting with the idealized PLCO setting, and finally inserted observed PLCO characteristics (Supplementary Material 4). In each run, the numbers of prostate cancer cases and prostate cancer deaths, and corresponding person-years of follow-up were projected, and the prostate mortality rate ratio was calculated. We quantified stochastic uncertainty around mortality rate ratio point estimates using ranges across 100 simulations and examined sensitivity to estimates of the cure parameter.

Results

Calibration results

Both calibrated models approximated the observed patterns of prostate cancer incidence, grade and stage distributions, and mortality in both arms of both trials (Figure 1 and Supplementary Material 5 and 6). The corresponding lead times are shown in Figure 2 for men aged 60-65 at screen detection and in Supplementary Material 7 for all age groups. The estimated cure parameter was 0.22 (Erasmus-MISCAN) and 0.18 (FHCRC) for the ERSPC. The corresponding cure probability by lead time is shown in Figure 3. Cancers detected early by screening were detected substantially earlier in both trials. For the

PLCO trial, we estimated hazard ratios to improve baseline survival of 0.40 (Erasmus-MISCAN) and 0.31 (FHCRC) and for the ERSPC of 0.82 (Erasmus-MISCAN) and 0.77 (FHCRC), illustrating important differences in background risk for men enrolled in the two trials. Because there were more prostate cancer deaths in the screening arm than in the control arm of the PLCO, the estimated cure parameter for that trial was 0 for both models. Consequently, we examined sensitivity of the mortality reduction and PSA screening efficacy to trial population, protocols, and practice settings using the cure parameter estimated for the ERSPC.

Prostate cancer mortality reduction adjusted for different trial characteristics

Starting with the observed prostate cancer mortality reduction in the ERSPC trial of 21% (95%CI 9%-32%) after 11 years of follow-up (run 0, Erasmus-MISCAN: 21%, FHCRC: 19%), the projected mortality reduction increased as the settings became more idealized (Figure 4). The largest screening effect in ERSPC was predicted under no contamination, 100% attendance, 100% receipt of biopsy for positive screens, and annual screening, with mortality reductions of 43% (Erasmus-MISCAN; uncertainty range 34%-52%) and 37% (FHCRC; uncertainty range 16%-59%) after 11 years of follow-up (run 5). Sensitivity analyses using the 95%CI of the point estimate of the ERSPC for fitting the cure parameter, indicated a 20%-64% prostate cancer mortality reduction in run 5 (Supplementary Material 8). Sensitivity analyses of uncertainty in the joint estimation of the cure parameter and improvement in baseline prostate cancer survival indicated a 16%-65% prostate cancer mortality reduction in run 5 (Supplementary Material 9).

The projected reduction diminished substantially as the idealized PLCO setting was systematically replaced with observed characteristics, to 8% (Erasmus-MISCAN) and 6% (FHCRC) under observed settings for all characteristics after 13 years of follow-up (run 12). These projections approach the

published ratio in PLCO (9% increase; 95%CI 13% reduction to 36% increase). When a cure parameter of 0 was used, an increase in prostate cancer mortality was found (run 13, Erasmus-MISCAN 3% and FHCRC 5%). Both models found that infrequent receipt of biopsies (runs 9 vs 10) and high contamination (runs 11 vs 12) increased the prostate cancer mortality rate ratio considerably. Although the models generally agreed, different effects were predicted for some trial characteristics, especially for 100% receipt of biopsy in the ERSPC and for the PSA threshold of 4 ng/ml in the PLCO.

Discussion

Efficacy is the extent to which a specific intervention produces a beneficial result under ideal conditions. In practice, true efficacy is rarely estimated as such. Randomized controlled trials, the gold standard for assessing screening interventions, can only assess efficacy limited by the circumstances of the implementation. Our results show that, by explicitly accounting for differences in implementation and settings between ERSPC and PLCO, it is possible to partially reconcile their seemingly different results. In particular, the infrequent receipt of biopsies after a positive test and the high contamination rate in the control arm of the PLCO are the main factors explaining why, even in the presence of a screening benefit such as that observed in the ERSPC, the PLCO could have yielded a negative result.

In addition to allowing us to examine differences between the trials, the models also afford insights into the mortality benefit that might potentially result from an ideal screening regimen. If all men in the ERSPC were screened annually (ignoring selection effects), received a biopsy after a positive test, and there was no contamination, the models predict that the prostate cancer mortality reduction due to screening would have been about 40% after 11 years. Extrapolating this to the European population setting suggests that 1 screen at age 55 could lead to 6,657 (5%) fewer prostate cancer deaths annually and biennial screening for ages 55-69 to 62,529 (44%) fewer deaths annually (Supplementary Material 10).

Earlier studies investigated explanations for the apparently different results of the trials.[7,26-28] We previously found that contamination in the PLCO substantially lowered its power.[19] Questions have been raised about possible differences in treatment men received in the screening and control arm of the ERSPC.[29] However, after correcting for age and tumor stage, no significant differences in treatment were found.[30] Our analysis shows that, if all patients in the control arm received treatment according to the frequencies (by age, tumor stage and grade) observed in the screening arm, the prostate cancer mortality reduction would remain unchanged. A similar result holds in the PLCO.

The level of contamination in the ERSPC has not been systematically reported and therefore had to be estimated from earlier published studies, which showed contamination ranging from 7-40% per year across centers.[21-23,31] The only study investigating the level of screening before the start of the ERSPC is a study of the Finnish center.[22] In this study, 10% of the men in the intervention arm had been screened before. However, both pre-trial and contamination estimates include PSA tests conducted because of symptoms, which could have accounted for up to half of the PSA tests performed.[21,23] Also, not all PSA tests were followed by a biopsy. For example, in the Rotterdam control arm, only 8% of positive opportunistic PSA tests were followed by biopsy.[23] We did not assess the influence of other less important characteristics separately, for example, population size, age distribution and enrollment patterns, other cause mortality, DRE testing, or biopsy sensitivity. However, we believe we have accounted for the characteristics most likely to be influential.

Using the cure parameter estimated for the ERSPC in the PLCO setting, we obtained a prostate cancer mortality reduction of 6-8%. This means that if PSA testing in the PLCO had been as efficacious as in the ERSPC, the circumstances of its implementation (e.g., infrequent receipt of biopsies, high contamination, healthy screenee effect) would likely have resulted in a modest reduction in prostate cancer mortality. This result is consistent with our prior study, in which we showed that contamination increased the

mortality rate ratio and decreased the power of the trial to detect a mortality difference from 40–70% to 9–25%.[19]

Initially, we planned to consider a symmetric approach, by also starting from the PLCO cure parameter and working towards more ideal situations, and back to the ERSPC. However, the best fit cure parameter for the PLCO was 0, and when there is no benefit, it is impossible to examine how benefit depends on the circumstances of implementation. A limitation is that this result depends on how much of the lower-than-expected mortality is attributed to changes in baseline survival relative to the pre-PSA era (e.g., due to improvements in care) rather than screening benefit in both arms. We feel our approach and prediction is valid, in the situation that one trial has shown an effect of earlier treatment of screen-detected lesions, and that the other trial has been underpowered.

In assessing the efficacy of any screening test, it is important to recognize that results will depend on how the test is implemented. If we start with a cure parameter estimated for the ERSPC, then under idealized circumstances (no control arm contamination, perfect attendance, perfect compliance with biopsy recommendations, run 5), the models predicted about a 40% mortality reduction after 11 years, which is greater than the 21% observed. However, under real-world circumstances of control arm contamination, and less-than-perfect attendance and biopsy compliance as in the PLCO trial, the models predicted a much reduced mortality reduction, of the order of 6-8%. Thus, the trials are likely less inconsistent than their results suggest. Further, the benefit of PSA screening under idealized circumstances is likely more than the trial results suggest. It could be as high as 40% which has previously been reported to imply a net benefit and a reasonably favorable tradeoff when accounting for the main harms of PSA screening.[16,32] However, specialized methods will be required to extract an estimate of what this idealized benefit might be based on the data from both trials.

Legends of figures

Figure 1: Observed and predicted cumulative percentage of prostate cancer incidence (left) and prostate cancer mortality (right) in the ERSPC (A) and PLCO (B) by year of follow-up.

Figure 2: Lead time distribution of screen-detected cases in the models, using the base ERSPC model (A) or the PLCO model (B), for men aged 60-65 at prostate cancer diagnosis. This is defined as the time from detection (screen and interval) until clinical detection before age 100 in the absence of death from other causes. In the Erasmus-MISCAN and FHCRC models, 31% and 20%, respectively, of cases were clinically detected and therefore had a lead time of 0 (and corresponding cure probability of 0). Results for other ages at diagnosis were similar.

Figure 3: The cure probability for screen-detected cases by lead time in the ERSPC trial as estimated by the two models. In the models, cured men were assigned to die at their independently generated date of other-cause death. Men who were not cured died at the same time they would have died if they had not been screened. So, for example, 60% (FHCRC) to 70% (Erasmus-MISCAN) of men with a lead time of 5 years will not die from prostate cancer and the remaining 30% to 40% will die at the same time and from the same cause as if they had not been screened.

Figure 4: Step-by-step prostate cancer mortality rate ratios and simulation-based uncertainty ranges for the Erasmus-MISCAN and FHCRC models. The changes in the models are cumulative. In run 13 a cure parameter of 0 is used, in all other runs, the ERSPC-based cure parameter is used (0.22 for Erasmus-MISCAN, 0.18 for FHCRC). Supplementary Material 9 provides intervals that incorporate variability in the estimated cure rate parameter (FHCRC model). For each run 0-13, 100 simulations of a single ERSPC or PLCO trial population were performed to generate sample mortality rate ratios; the bracketed line

(uncertainty range) and dot represent, respectively, the range and mean of the sample mortality rate ratios observed over the 100 simulations.

In run 0-5 a follow-up of 11 years is used, in run 6-13 the follow-up is 13 years. In each step the listed implementation change is added to the previous step.

References

1. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med 2009;360:1320-8.
2. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012;366:981-90.
3. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014;384:2027-35.
4. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. J Natl Cancer Inst 2012;104:125-32.
5. Barry MJ. Screening for prostate cancer--the controversy that refuses to die. N Engl J Med 2009;360:1351-4.
6. La Rochelle J, Amling CL. Prostate cancer screening: what we have learned from the PLCO and ERSPC trials. Curr Urol Rep 2010;11:198-201.
7. Schröder FH, Roobol MJ. ERSPC and PLCO prostate cancer screening studies: what are the differences? Eur Urol 2010;58:46-52.
8. Pinsky PF, Black A, Parnes HL, et al. Prostate cancer specific survival in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Cancer epidemiology 2012;36:e401-6.
9. Pinsky PF, Blacka A, Kramer BS, Miller A, Prorok PC, Berg C. Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Clin Trials 2010;7:303-11.
10. Grubb RL, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. BJU Int 2008;102:1524-30.
11. Chou R, Croswell JM, Dana T, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2011;155:762-71.
12. Djulbegovic M, Beyth RJ, Neuberger MM, et al. Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. BMJ (Clinical research ed 2010;341:c4543.
13. Ilic D, O'Connor D, Green S, Wilt TJ. Screening for prostate cancer: an updated Cochrane systematic review. BJU Int 2011;107:882-91.
14. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative PSA-based prostate cancer screening strategies. Ann Intern Med 2013;158:145-53.
15. Gulati R, Inoue L, Katcher J, Hazelton W, Etzioni R. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. Biostatistics 2010;11:707-19.
16. Heijnsdijk EA, de Carvalho TM, Auvinen A, et al. Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. J Natl Cancer Inst 2015;107:366.
17. Wever EM, Draisma G, Heijnsdijk EA, et al. Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. J Natl Cancer Inst 2010;102:352-5.
18. Draisma G, Postma R, Schröder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. Int J Cancer 2006;119:2366-71.
19. Gulati R, Tsodikov A, Wever EM, et al. The impact of PLCO control arm contamination on perceived PSA screening efficacy. Cancer Causes Control 2012;23:827-35.

20. 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;109:1877-86.
21. Bokhorst LP, Bangma CH, van Leenders GJ, et al. Prostate-specific antigen-based prostate cancer screening: reduction of prostate cancer mortality after correction for nonattendance and contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. Eur Urol 2014;65:329-36.
22. Ciatto S, Zappa M, Villers A, Paez A, Otto SJ, Auvinen A. Contamination by opportunistic screening in the European Randomized Study of Prostate Cancer Screening. BJU Int 2003;92 Suppl 2:97-100.
23. Otto SJ, van der Cruijsen IW, Liem MK, 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.
24. Friedman M. Piecewise Exponential Models for Survival Data with Covariates. Ann Statist 1982;10:101-13.
25. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited : Treatment changes and prostate cancer mortality declines. Cancer 2012;118:5955-63.
26. Croswell JM, Kramer BS, Crawford ED. Screening for prostate cancer with PSA testing: current status and future directions. Oncology (Williston Park) 2011;25:452-60, 63.
27. Studer UE, Collette L. What can be concluded from the ERSPC and PLCO trial data? Urologic oncology 2010;28:668-9.
28. Palma A, Lounsbury DW, Schlecht NF, Agalliu I. A System Dynamics Model of Serum Prostate-Specific Antigen Screening for Prostate Cancer. American journal of epidemiology 2016;183:227-36.
29. Haines IE, Gabor Miklos GL. Prostate-specific antigen screening trials and prostate cancer deaths: the androgen deprivation connection. J Natl Cancer Inst 2013;105:1534-9.
30. Wolters T, Roobol MJ, Steyerberg EW, et al. The effect of study arm on prostate cancer treatment in the large screening trial ERSPC. Int J Cancer 2010;126:2387-93.
31. Lujan M, Paez A, Pascual C, Angulo J, Miravalles E, Berenguer A. Extent of prostate-specific antigen contamination in the Spanish section of the European Randomized Study of Screening for Prostate Cancer (ERSPC). Eur Urol 2006;50:1234-40; discussion 9-40.
32. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med 2012;367:595-605.

Table 1. Inputs of the models for each trial. Most inputs are age-, stage-, and/or center-specific. The average value is presented for comparison between the arms and trials.

	ERSPC		PLCO	
	Screen arm	Control arm	Screen arm	Control arm
Sample size	72,891	89,352	38,343	38,350
Age at randomization	55-69	55-69	55-74	55-74
Screen attendance	On average 82% MISCAN: By center and round FHCRC: By center	Not applicable	By age and round on average 85%	Not applicable
Screen protocol	2 years interval for Sweden, 4 years for other centers (7-years interval between rounds 1 and 2 for Belgium) Screening from age 55 to 69/71/74 depending on center MISCAN: PSA threshold of 3 ng/ml for all centers FHCRC: PSA threshold and DRE testing by center	Not applicable	1 year interval for 6 years, PSA threshold of 4 ng/ml Screening from age 55-74 FHCRC: also DRE testing	Not applicable
Biopsy compliance	On average 86% MISCAN: By age, center and round	On average 86% MISCAN: 86% for all	By age and round, on average 35% FHCRC: also by PSA	On average 35%

	FHCRC: By age, PSA, and center	FHCRC: By age, PSA, and center		
Biopsy sensitivity	80%	80%	Increasing from 70% in 1990 to 93% in 2000	Increasing from 70% in 1990 to 93% in 2000
Contamination	Pretrial screening: about 3-5% of participants had a PSA test No contamination during trial	Pretrial screening: about 3-5% of participants had a PSA test During trial: about 17,000 tests	Pretrial screening: about 50% of participants had a PSA test During trial: no contamination Post-trial screening: US population screening	Pretrial screening: about 50% of participants had a PSA test During trial: about 2.7 test per participant [9] Post-trial screening: US population screening
Treatment of local-regional cases	By age, stage, and grade, on average 47% radical prostatectomy * 21% radiation therapy 32% conservative management or active surveillance	By age, stage, and grade, on average 53% radical prostatectomy * 17% radiation therapy 30% conservative management or active surveillance	By age, stage, and grade, on average 59% radical prostatectomy * 22% radiation therapy 19% conservative management or active surveillance	By age, stage, and grade, on average 58% radical prostatectomy * 22% radiation therapy 20% conservative management or active surveillance
Life tables	MISCAN: by European country (Human Mortality Database) FHCRC: US life tables (Berkeley Mortality	MISCAN: by European country (Human Mortality Database)	US life tables (Berkeley Mortality Database)	US life tables (Berkeley Mortality Database)

	Database)	FHCRC: US life tables (Berkeley Mortality Database)		

* This category includes radical prostatectomy, radical prostatectomy with hormone therapy, and radiation therapy with hormone therapy, all assuming to have a hazard ratio of 0.62 on prostate cancer death.

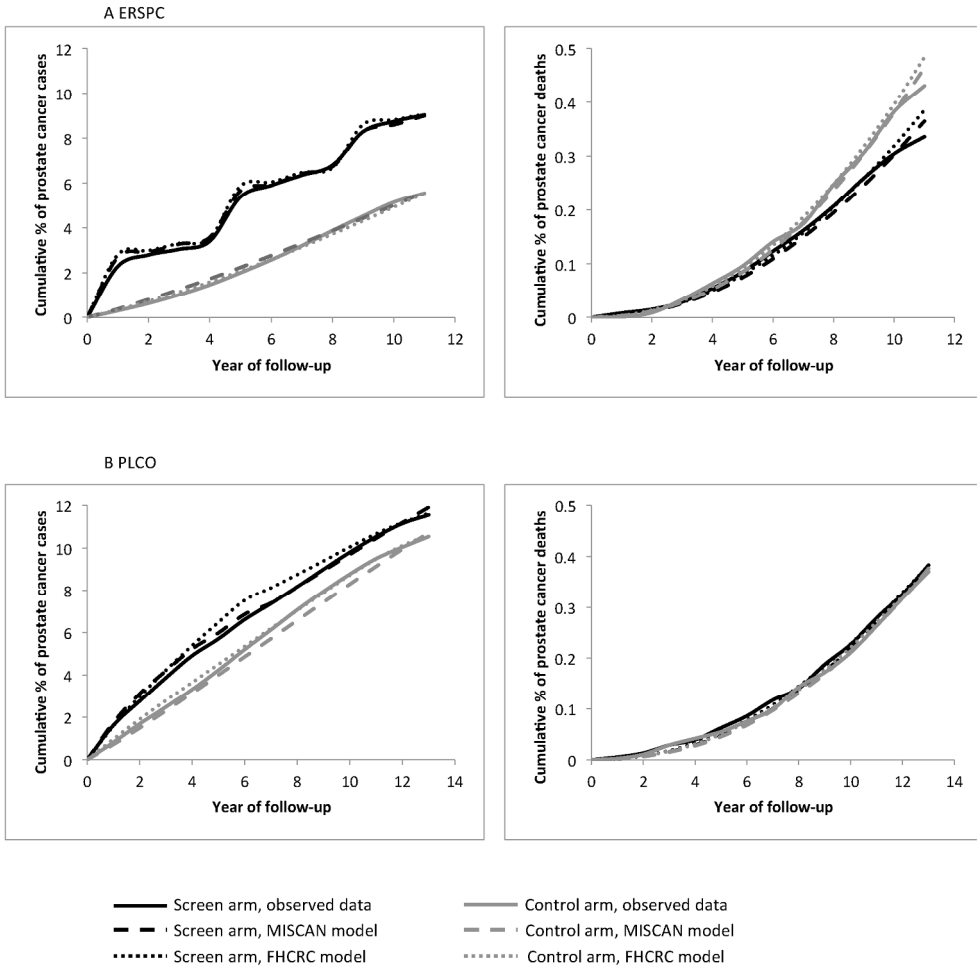
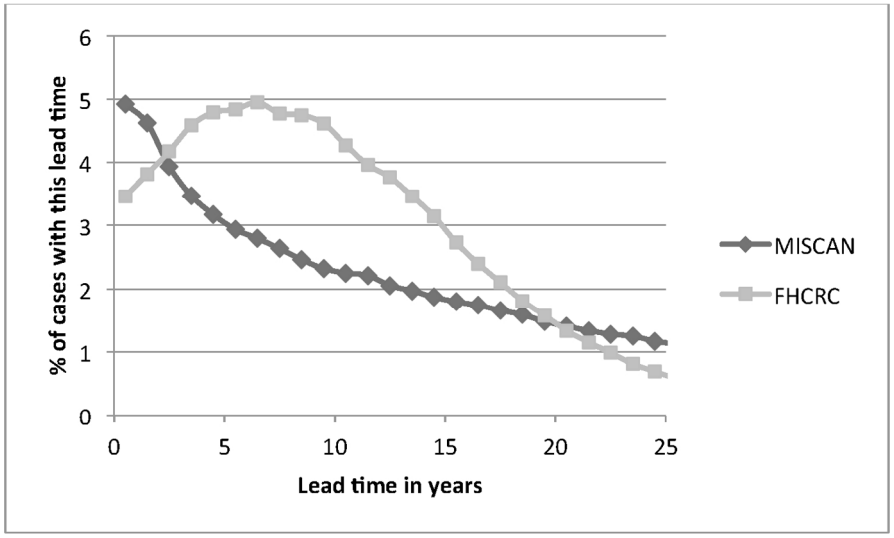


Figure 1: Observed and predicted cumulative percentage of prostate cancer incidence (left) and prostate cancer mortality (right) in the ERSPC (A) and PLCO (B) by year of follow-up.

213x212mm (300 x 300 DPI)



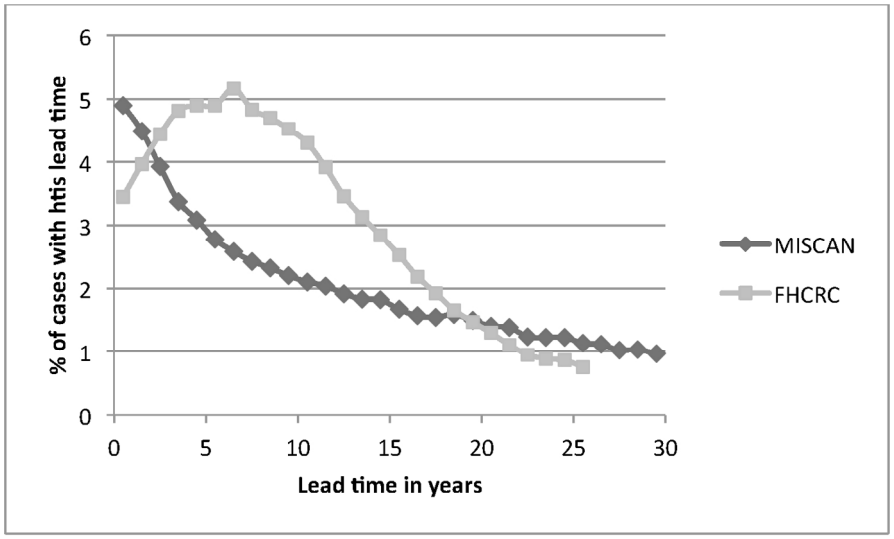


Figure 2: Lead time distribution of screen-detected cases in the models, using the base ERSPC model (A) or the PLCO model (B), for men aged 60-65 at prostate cancer diagnosis. This is defined as the time from detection (screen and interval) until clinical detection before age 100 in the absence of death from other causes. In the Erasmus-MISCAN and FHCRC models, 31% and 20%, respectively, of cases were clinically detected and therefore had a lead time of 0 (and corresponding cure probability of 0). Results for other ages at diagnosis were similar.

134x172mm (300 x 300 DPI)

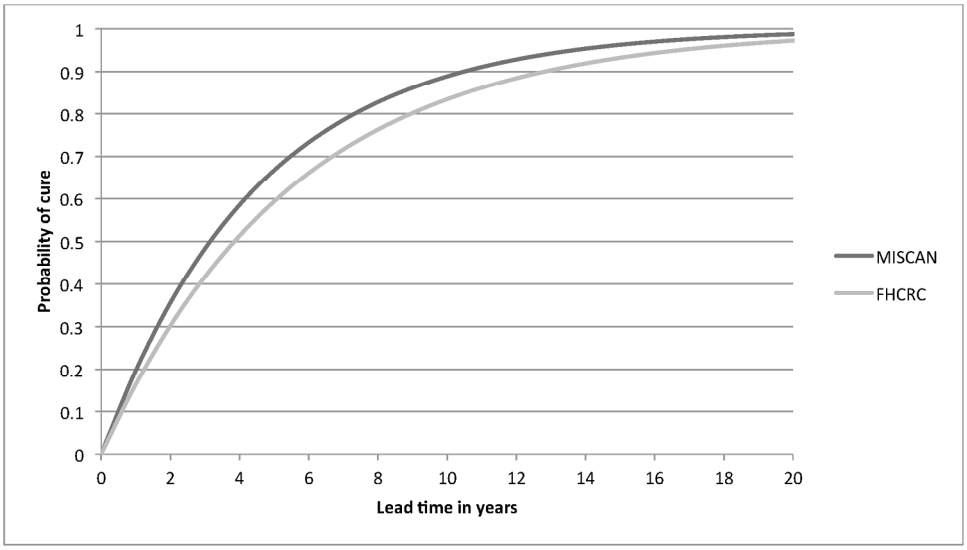


Figure 3: The cure probability for screen-detected cases by lead time in the ERSPC trial as estimated by the two models. In the models, cured men were assigned to die at their independently generated date of other-cause death. Men who were not cured died at the same time they would have died if they had not been screened. So, for example, 60% (FHCRC) to 70% (Erasmus-MISCAN) of men with a lead time of 5 years will not die from prostate cancer and the remaining 30% to 40% will die at the same time and from the same cause as if they had not been screened.

189x109mm (300 x 300 DPI)

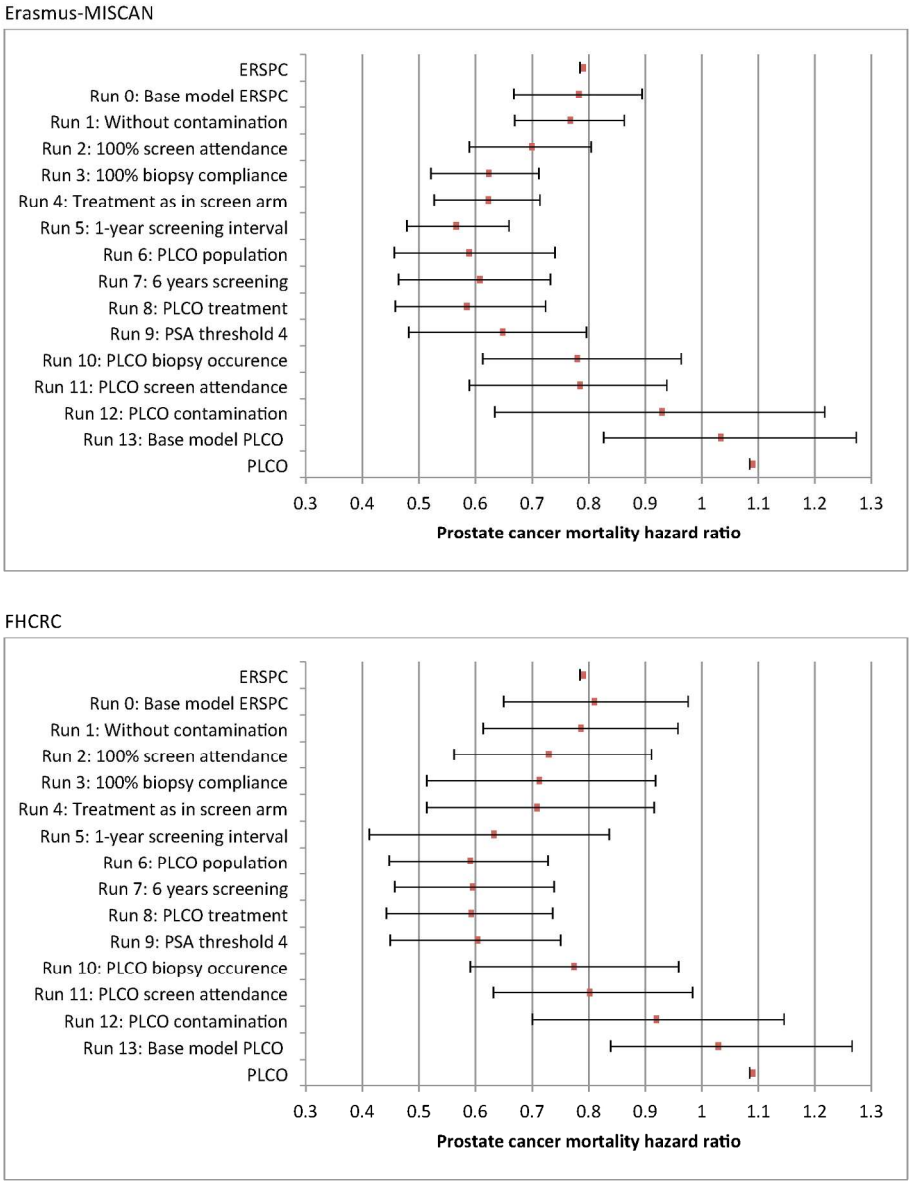


Figure 4: Step-by-step prostate cancer mortality rate ratios and simulation-based uncertainty ranges for the Erasmus-MISCAN and FHCRC models. The changes in the models are cumulative. In run 13 a cure parameter of 0 is used, in all other runs, the ERSPC-based cure parameter is used (0.22 for Erasmus-MISCAN, 0.18 for FHCRC). Appendix 9 provides intervals that incorporate variability in the estimated cure rate parameter (FHCRC model). For each run 0-13, 100 simulations of a single ERSPC or PLCO trial population were performed to generate sample mortality rate ratios; the bracketed line (uncertainty range) and dot represent, respectively, the range and mean of the sample mortality rate ratios observed over the 100 simulations.
In run 0-5 a follow-up of 11 years is used, in run 6-13 the follow-up is 13 years. In each step the listed implementation change is added to the previous step.

173x219mm (300 x 300 DPI)

Supplementary Material 1. Model descriptions

Erasmus-MISCAN model

The Erasmus-MISCAN prostate cancer model is a micro-simulation model that simulates individual disease histories stochastically.[1-3] The model is programmed in Delphi. Birth and life tables of a specific population are used to simulate a trial or country population with the observed age distribution. Simulated individuals face age-specific risks of onset of a preclinical prostate cancer. From the point of onset, the development of cancer in individuals is modeled as a sequence of tumor states (Figure 1). There are eighteen preclinical detectable states in the natural history of prostate cancer which are derived from combinations of three clinical T-stages (cT-stage 1, 2, and 3+), Gleason grade (<7, 7, and >7) and metastatic stage (local-regional or distant). Progression through the stages and grades are modeled by a semi-Markov process, and it is assumed that there is a stage- and grade-specific risk of transition from the local-regional to the distant stage. From each preclinical detectable state, cancer can progress to the clinical disease state (diagnosis). Screening (defined by year, age, attendance rate, and test sensitivity) is superimposed on the life histories generated in the absence of screening.

The parameters for the natural history of prostate cancer (background incidence, transition probabilities between the states, and durations in the states) are generally estimated using incidence and stage distribution data from the ERSPC trial or SEER.[1, 4, 5] This calibration process involves the maximum likelihood method. Parameters are estimated by minimizing the differences between observed and predicted counts, measured as a sum of the chi-square quantities using an adapted version of the simplex optimization method of Nelder and Mead.[6]

For this project, PSA screening frequency, screening ages, attendance rates, and thresholds for biopsy referral follow protocols in the PLCO and in each center of the ERSPC. Clinical diagnosis and test sensitivity parameters were estimated so that model-projected incidence counts match observed counts in each trial by age, year, stage, and grade at diagnosis.

After diagnosis, men receive treatment (radical prostatectomy, radiation therapy) or conservative management/active surveillance based on their age, stage, and grade at diagnosis and can die from prostate cancer or death from other causes (based on life tables). The parameters used for calculating survival are described in the main manuscript.

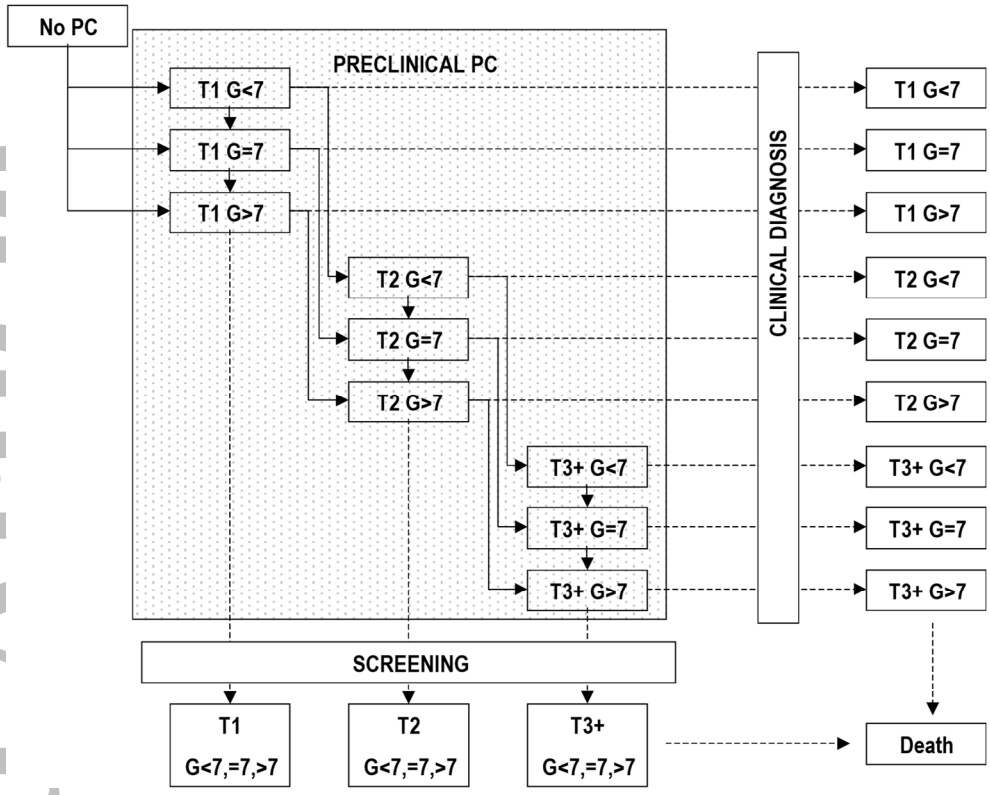


Figure 1. The Erasmus-MISCAN model. Prostate cancer develops from no prostate cancer via one or more screen-detectable preclinical states to a clinically diagnosed cancer or screen-detected cancer. The arrows indicate the possible transitions. Each state can be local or metastatic, but for simplicity this is not illustrated. Individuals in any state are at risk of death from other causes.

FHCRC model

The FHCRC model is also a microsimulation model. The model is programmed in C++. Its natural history model consists of two linked parts: PSA growth and disease progression.[7-9] Log PSA growth over time is modeled using a piecewise linear model, with a larger slope following onset of a low-grade (Gleason score ≤7) preclinical tumor and a still larger slope following onset of high-grade (Gleason score >7) preclinical tumor. To exploit more precise information about tumor grade in the trial data, we further partitioned low-grade tumors into Gleason score <7 vs 7. PSA slopes by grade category were estimated using random effects models fit to data from the control arm of the Prostate Cancer Prevention Trial (PCPT),[10] which screened 9,000 men on the control arm annually for up to 7 years with an exit biopsy regardless of PSA test results, and the finer grade partition was based on a Bayesian analysis of tumor grade given an individual's PSA slope. The estimated PSA growth model allows us to simulate individual PSA trajectories that reflect observed inter- and intra-individual variability before and after onset of a preclinical tumor in each grade category (Figure 2A).

The risk of onset of a preclinical tumor increases with age and is modeled using a Weibull hazard, and the probability of high-grade cancer at onset increases quadratically with age.[9] Grade is set at onset and does not change over time. Given individual age and grade at onset, the risks of progression from a local-regional to a metastatic stage and from a preclinical to a symptomatic state increase proportionally to an individual's PSA level.[8, 9] Similar to grade categories, we further partitioned local-regional stage tumors at diagnosis into clinical T-stage ≤T2A vs > T2A given age, grade, and PSA at diagnosis using a logistic regression model fit to localized cases in the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) database.[11]

The risk of onset, the probability of high-grade cancer at onset, and risks of progression through states were estimated using simulated maximum likelihood based on a Poisson likelihood function for counts of cases by age, year, stage, and grade at diagnosis. More specifically, for particular candidate values of the natural and clinical history parameters, we simulated cancer incidence using trial data on screening ages, enrollment patterns, screen arm attendance rates, control arm screening rates, sensitivity of DRE[12] and receipt of biopsy in the PLCO and in each center of the ERSPC. The projected incidence counts were compared with observed incidence counts within patient and tumor strata in a likelihood function, and the candidate natural and clinical history parameters were systematically adjusted to replicate as closely as possible the observed incidence patterns.[8, 9] Additional details of the PSA growth model and natural history estimation methods are described in the Supplementary Materials in Gulati et al.[9] The estimated PSA growth and natural history model allows us to simulate PSA trajectories and detailed disease courses for all trial participants (Figure 2B).

As in the Erasmus-MISCAN model, after diagnosis, men receive primary treatment based on their age and the tumor stage and grade, and prostate cancer survival depends on these patient and tumor characteristics, the cure parameter associated with screening, and primary treatment. All individuals can die from other causes based on life tables. The parameters controlling the effects of screening and initial treatment on prostate cancer survival are described in the main manuscript.

Source code

The FHCRC model is programmed in C. Miscan is programmed in Delphi. The source code of the FHCRC model is available on request. The source code of the Miscan model is not available on request. Interested persons can contact the authors to do runs using the models.

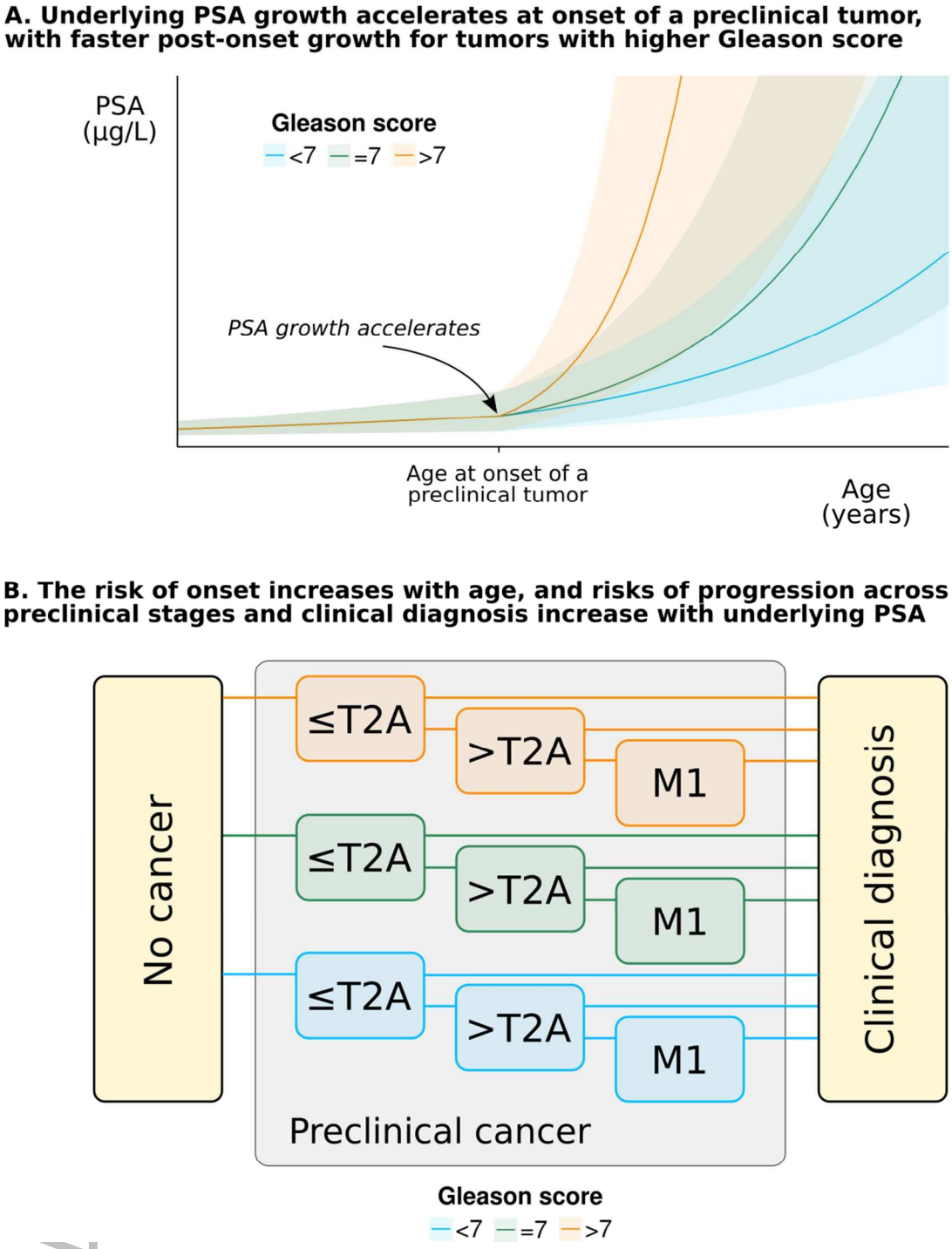


Figure 2. The FHCRC model. Individual PSA levels (A) are linked to the onset of a preclinical tumor, its Gleason score, and its transitions through preclinical states and clinical diagnosis (B). Individuals in any state are at risk of death from other causes.

Supplementary Material 2. Calibration of the models

Erasmus-MISCAN: For each set of parameter values, the model generates life histories and counts results (for example number of incident cases, number of prostate cancer deaths) by age and year. These counts are considered predictions from the model. Parameters are estimated by minimizing difference between observed and predicted counts, measured as the sum of the chi-square quantities. The minimization is accomplished using an adapted version of the simplex optimization method of Nelder and Mead.[6] Optimization is initiated with small sample sizes (i.e. 20,000) to reduce running time. Then it is repeated with larger sample sizes (i.e. 2,000,000) when optimization progress is no longer significant.

For this analysis, first the disease progression rates and the PSA test sensitivity were calibrated against the incidence by stage and grade of the clinically-detected cancers in the control arm and the screendetected and interval cancers in the screened arm. Then, with the disease progression rates and PSA sensitivity fixed at their calibrated values we jointly calibrated the baseline survival hazard ratio and the cure parameter against the published numbers of prostate cancer deaths after 11 years of follow-up in both arms of the ERSPC.[13] A similar approach was used to calibrate the baseline survival hazard ratio in both arms of the PLCO.

FHCRC: The FHCRC calibration targets for incidence and mortality were the same as those used by the Erasmus-MISCAN model. For incidence, the model parameters (risks of disease onset, grade category at onset, risks of progression to metastatic and clinical disease) were estimated using a simulated maximum likelihood algorithm. The likelihood on the observed counts of cases by age, year, grade, stage and trial arm was modelled as a Poisson distribution. The observed counts were those observed in the trial and the expected counts were those simulated by the model. The maximum likelihood estimate for the unknown parameters was obtained by optimizing over the simulated likelihood surface as a function of the unknown parameters.

For mortality, two approaches were used to estimate the survival parameters, i.e., the cure parameter and the hazard ratio to capture improvements in baseline survival (in the absence of screening and primary treatments) since the pre-PSA era (using SEER data for untreated cases diagnosed in 1983–1986). The first approach, with results shown in the main text, minimized the sum of squared errors between observed and simulated cumulative deaths for years 0–7 (ERSPC) and 0–10 (PLCO). The second approach, with results reported in Supplementary Material 9, used another simulated Poisson likelihood to estimate the survival parameters against published numbers of prostate cancer deaths in the ERSPC after 11 years of follow-up.[13] Additional details are given in Supplementary Material 9.

Supplementary Material 3. Treatment distributions

In this appendix the treatment model as used in both the Erasmus-MISCAN and FHCRC models is described. The ERSPC dataset included treatments defined as: prostatectomy (RP), radiation therapy (RT), and conservative management/active surveillance (CM), all with and without hormone treatment (ADT). For conservative management we did not take delayed treatment into account in the models. To account for the different effects of treatment on prostate cancer mortality in our models, the treatments are transformed into the following groups (Table 1):

Table 1: Treatment categories

Dataset	New category	Hazard ratio on mortality
RP, RP with ADT, RT with ADT	RP	0.62
RT	RT	0.7
CM, CM with ADT	CM	1

First we imputed the missing cases (about 6%) given patient and clinical characteristics at diagnosis. Patients in distant stage were excluded. Patient ages at diagnosis were divided in 5-year age groups (5559 until 80-84). Gleason score is presented as G6 (Gleason score 2-6), G7 (Gleason score 7) or G8 (Gleason score 8-10) and stage as early (≤ T2a) and advanced (> T2a).

We used the following multinomial logit model, in which the probability that an individual *i* with explanatory variable X_i will be treated in category *j* is given by

$$P\,Y[\,_i = j\,X|\,_i\,] = \frac{\exp(X_i\square_j)}{\sum_{l=1}\exp(X_i\square_l)}, \quad for\ j = 1,\ldots,J.$$

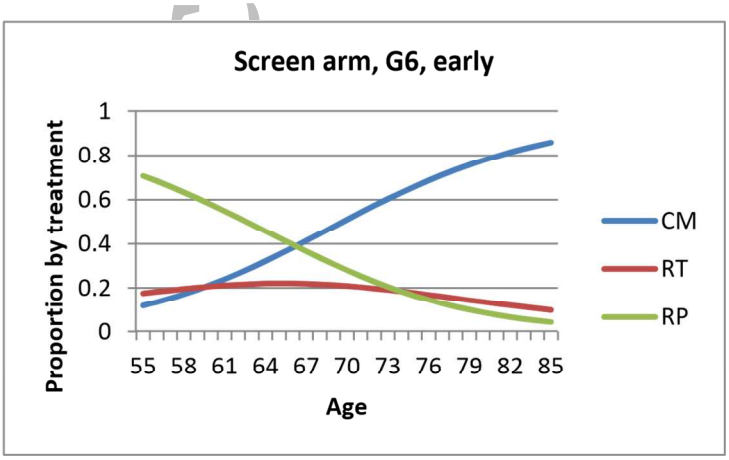
The treatment types are the categories. The treatment CM is taken as the baseline category, so that $\exp(X_i\square_j)$ equals 1 for *j* corresponding to the treatment CM.

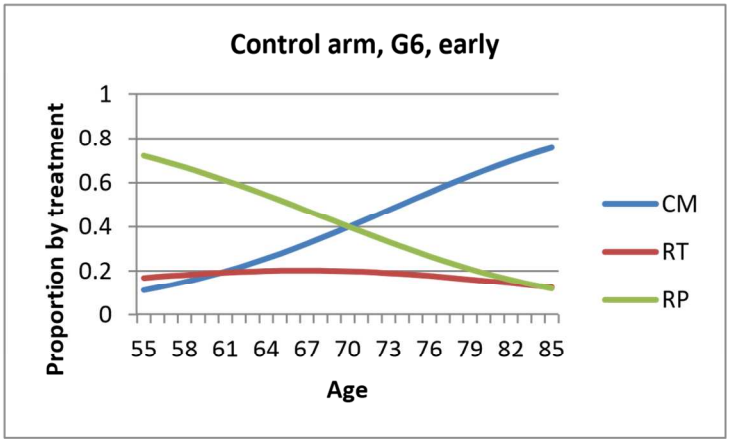
The explanatory variables in the model are: the age of the individual (5-year age categories), the Gleason score and stage (treatment as dependent variable, Gleason score and stage as factor and age as covariant). The parameter estimates for the multinomial logit model were obtained by fitting a multinomial logit model using SPSS (version 21). Two different models were fitted:

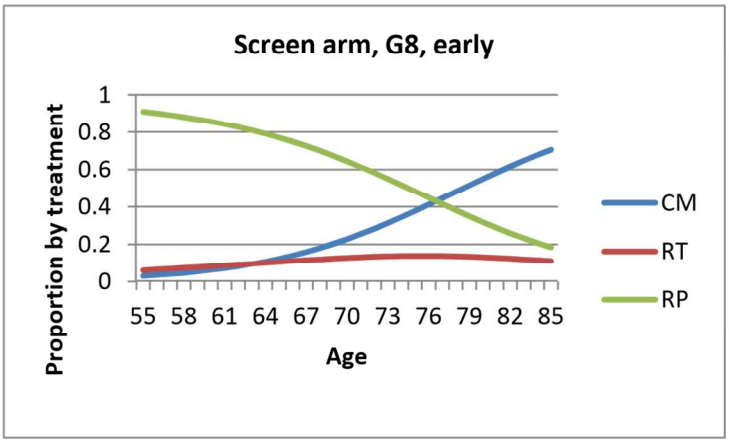
1. Treatment in the screen arm and in the control arm.
2. Treatment in the screen arm, for the "true screen detected" cases and for the "true interval cases" only.

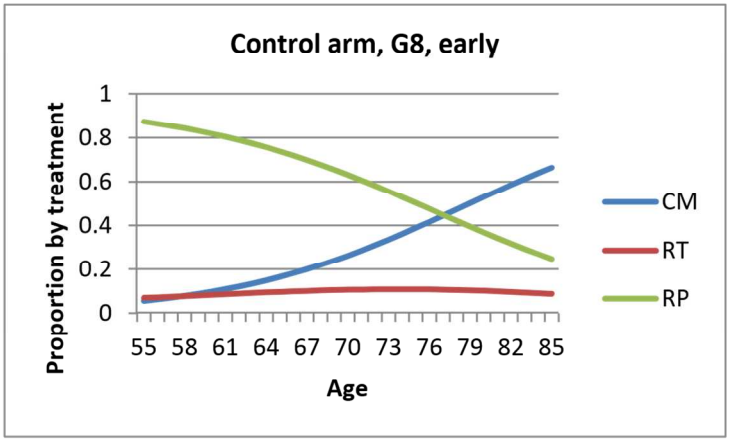
Model 1 Treatment distribution in screen and control arm.

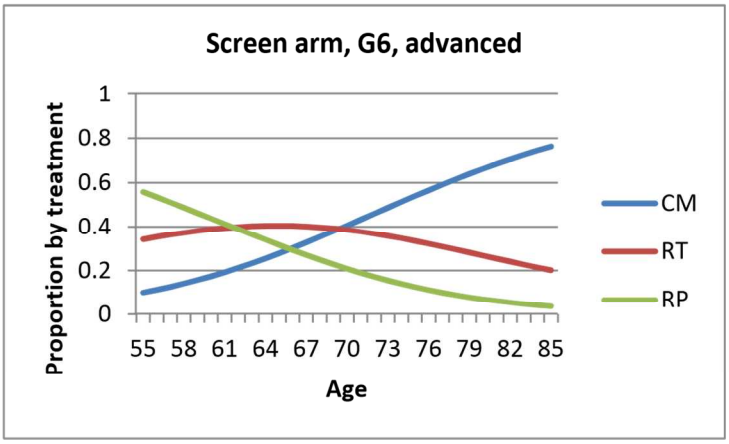
There are differences in treatment between the screen arm and the control arm, especially in the advanced stages, where there is more RT in the screen arm (Figure 1). Increasing Gleason score leads to more RP and CM and less RT. The parameters of the model are shown in Table 2.

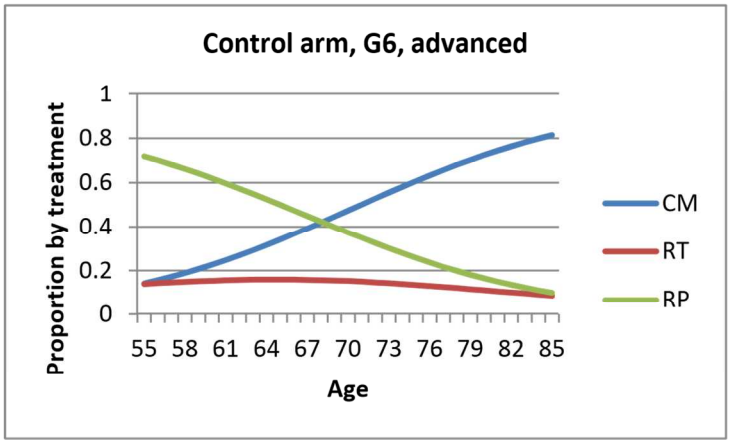












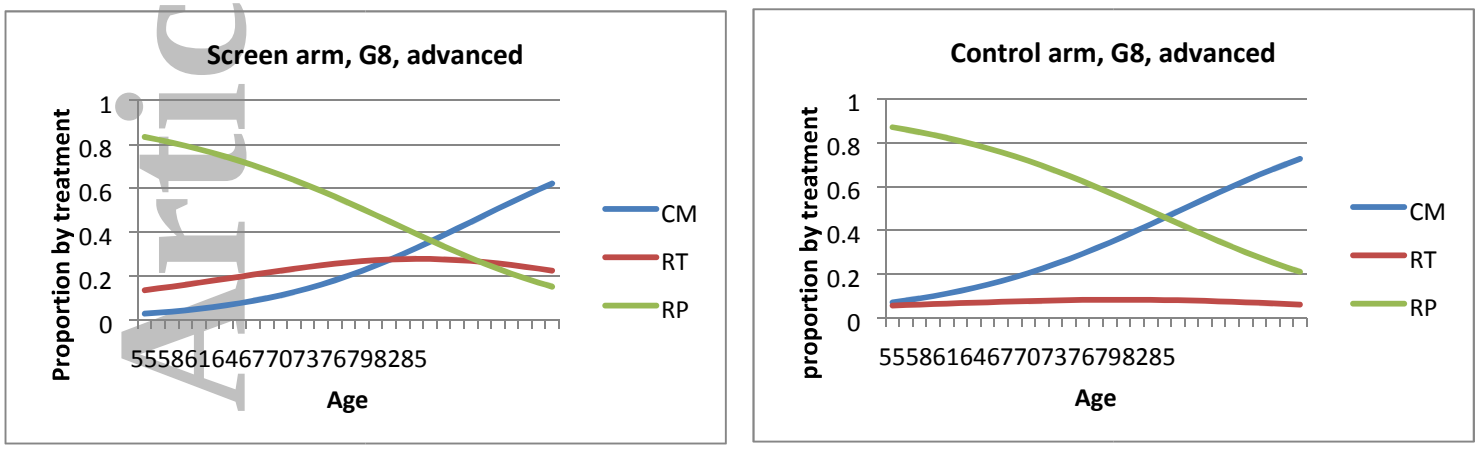


Figure 1: Model of the treatment distribution by arm.

Table 2: Parameter estimates of model 1. Using these parameters, for each age, Gleason score and state the proportion of men following RP, RT and CM can be calculated.

	Screen arm		Control arm	
	RP vs. CM	RT vs. CM	RP vs. CM	RT vs. CM
Intercept	12.221	6.472	10.019	4.464
Age	-0.159	-0.088	-0.132	-0.081
Gleason 6 vs 8	-1.606	-0.310	-0.878	0.203
Gleason 7 vs 8	-0.294	0.337	-0.188	0.272
Early vs Advanced	0.011	-0.852	0.229	0.390

Model 2 Treatment distribution in the interval detected cases and screen detected cases in the screen arm

In this model we only included interval and screen detected cases. In all categories, the interval cases received more CM and less RT (see examples in Figure 2). The parameters are shown in Table 3.

Table 3: Parameter estimates of model 2. Using these parameters, for each age, Gleason score and state the proportion of men following RP, RT and CM can be calculated.

	Screen detected cases		Interval cases	
	RP vs. CM	RT vs. CM	RP vs. CM	RT vs. CM
Intercept	12.951	5.930	11.637	5.271
Age	-0.158	-0.060	-0.157	-0.099
Gleason 6 vs 8	-2.099	-1.036	-1.219	0.574
Gleason 7 vs 8	-0.680	-0. 129	-0.050	0.649
Early vs Advanced	-0.351	-1.378	0.255	0.100

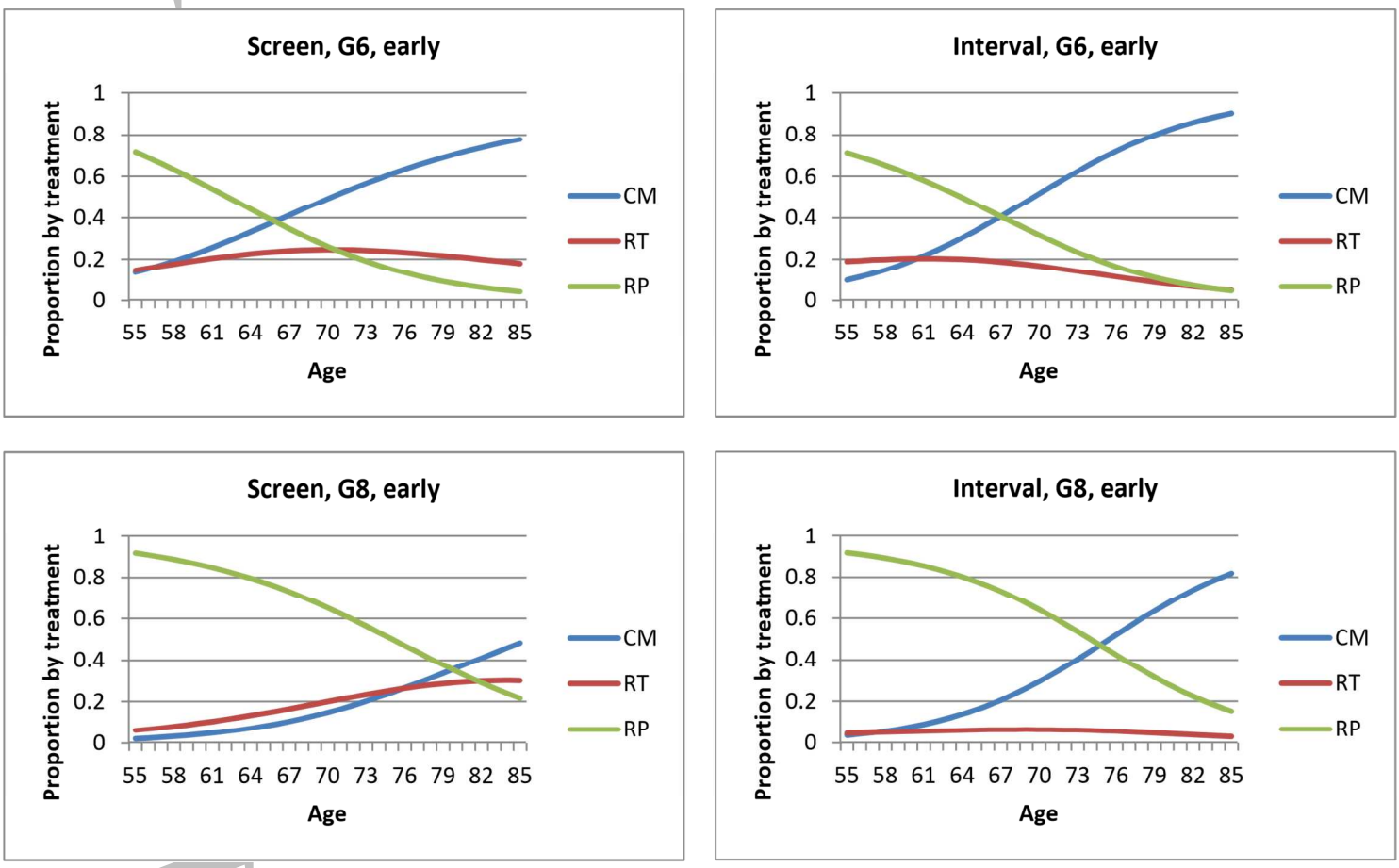


Figure 2: Examples of the model of the treatment distribution by Gleason scores and stage for the screen detected and interval cases.

We followed the same procedure for the PLCO trial.

Model 1 Treatment distribution in screen and control arm.

There is no clear difference in treatment between the screen arm and the control arm. There are small differences in treatment distributions between the early and advanced stages. Increasing Gleason score leads to more RP and CM and less RT. The parameters of the model are shown in Table 4.

Table 4: Parameter estimates of model 1. Using these parameters, for each age, Gleason score and state the proportion of men following RP, RT and CM can be calculated.

	Screen arm		Control arm	
	RP vs. CM	RT vs. CM	RP vs. CM	RT vs. CM
Intercept	13.549	5.362	14.025	5.148
Age	-0.164	-0.080	-0.173	-0.083
Gleason 6 vs 8	-0.970	0.562	-0.908	0.844
Gleason 7 vs 8	-0.211	0.823	0.097	1.280
Early vs Advanced	-0.243	-0.147	-0.174	0.023

Model 2 Treatment distribution in the true interval detected cases and true screen detected cases in the screen arm

In all categories, the interval cases received more CM and less RT. The parameters are shown in Table 5.

Table 5: Parameter estimates of model 2. Using these parameters, for each age, Gleason score and state the proportion of men following RP, RT and CM can be calculated.

	Screen detected cases		Interval cases	
	RP vs. CM	RT vs. CM	RP vs. CM	RT vs. CM
Intercept	11.313	3.281	11.373	6.718
Age	-0.138	-0.045	-0.138	-0.101
Gleason 6 vs 8	-0.514	0.356	-0.829	0.326
Gleason 7 vs 8	0.182	0.570	-0.113	0.639
Early vs Advanced	-0.165	-0.211	0.170	0.077

Both model 1 and 2 were used to assign the treatments by Gleason score, stage, age (single year ages based on interpolation between the 5-year age groups) and arm and/or mode of detection in the microsimulation models.

Supplementary Material 4. Consecutive runs starting at ERSPC input and ending with PLCO input

Run 0 starts with all parameters of the ERSPC trial. Each next step, the changes are cumulative.

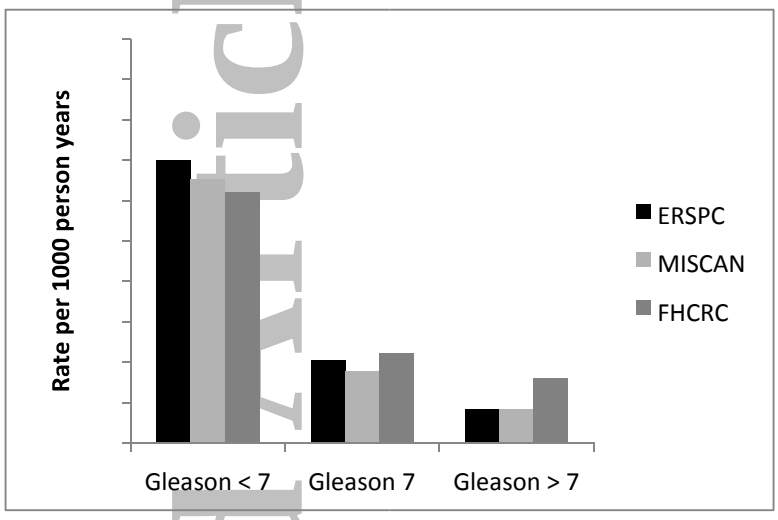
Run	Title	Change in arm*	Description
0	Base model ERSPC	S & C	Base run of calibrated ERSPC model, 11 years of follow-up
1	Without contamination	S & C	Screening before the trial (in both arms) and during the trial (in control arm) removed
2	100% screen attendance	S	100% screen attendance in screen arm for each round added
3	100% biopsy compliance	S	100% biopsy compliance after a positive test in screen arm added
4	Treatment as in screen arm	C	Treatment in control arm replaced with treatment as observed in screen arm
5	1-year screening interval	S	Annual screening for all centers until the maximum stop age by center
6	PLCO population	S & C	Age distribution and size of the PLCO population, US clinical background (probability of clinical detection and baseline survival hazard ratio), 13 years of follow-up
7	6 years screening	S	6 years of annual screening followed by no screening
8	PLCO treatment	S & C	Treatment in screen and control arm replaced with treatment as observed in PLCO
9	PSA threshold 4	S	Threshold of PSA > 4 ng/ml for biopsy referral (and DRE for the FHCRC model)
10	PLCO biopsy occurrence	S	PLCO biopsy occurrence rate added
11	PLCO screen attendance	S	PLCO attendance added
12	PLCO contamination	S & C	Screening before and after the trial (in both arms) and during the trial (in control arm) added
13	Base model PLCO	S & C	Cure parameter of 0, base run of calibrated PLCO model, 13 years of follow-up

*S means screen arm, C control arm

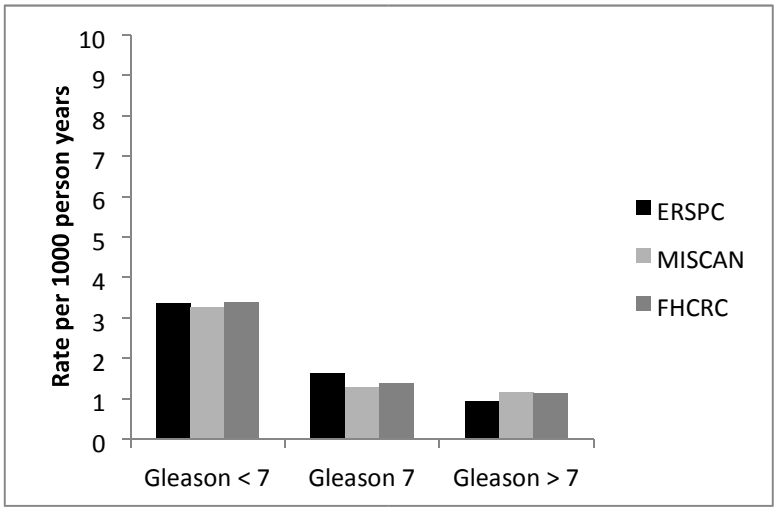
Supplementary Material 5. Observed and predicted grade distributions in both trials

In the figures, the observed data of the ERSPC and PLCO are compared by the model predictions of the MISCAN and the FHCRC models.

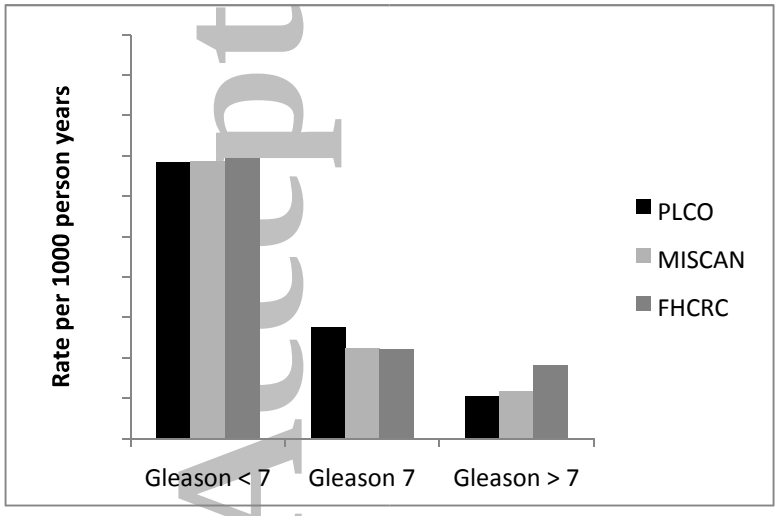
ERSPC, screen arm, 11 years of follow-up



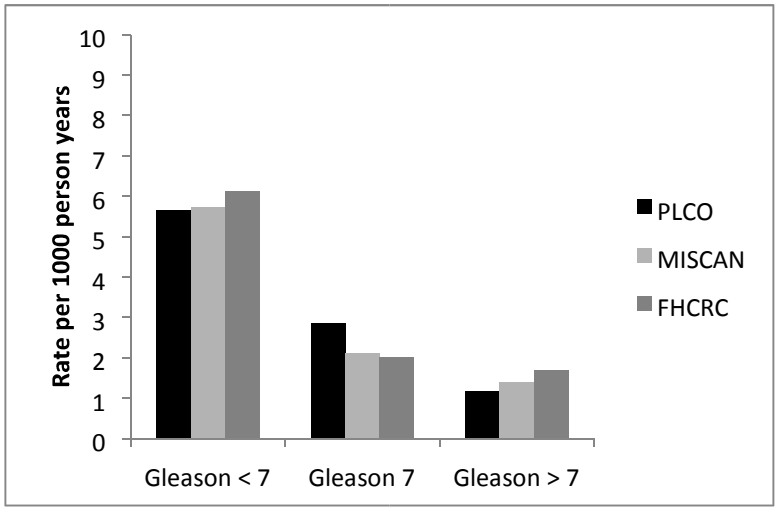
ERSPC, control arm, 11 years of follow-up



PLCO, screen arm, 13 years of follow-up



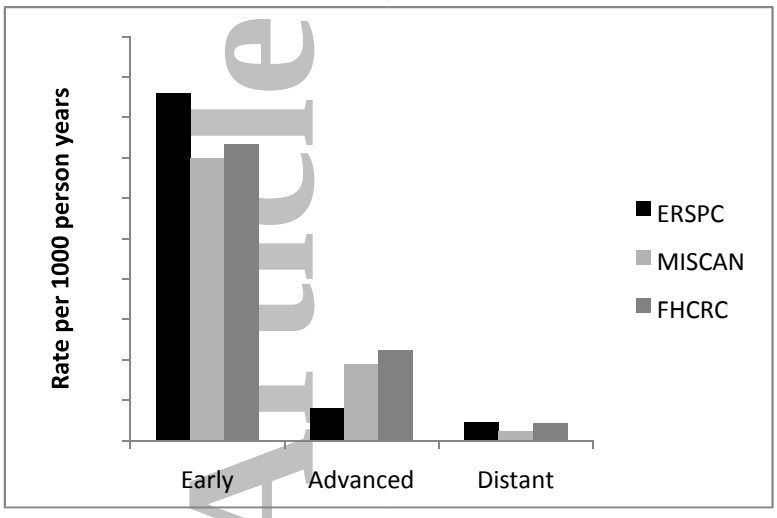
PLCO, control arm, 13 years of follow-up



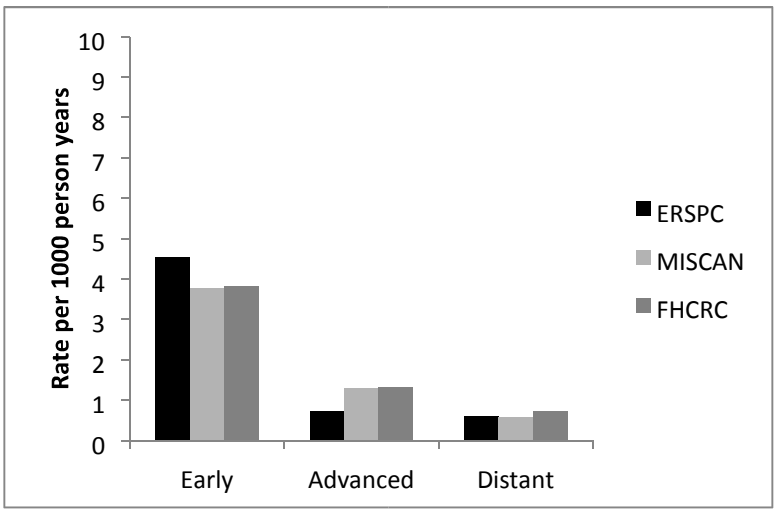
Supplementary Material 6. Observed and predicted stage distributions in both trials

In the figures, the observed data of the ERSPC and PLCO are compared by the model predictions of the MISCAN and the FHCRC models.

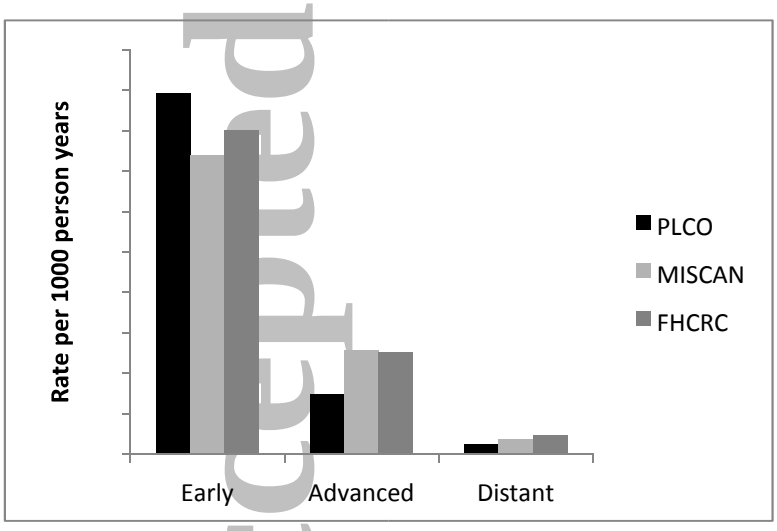
ERSPC, screen arm, 11 years of follow-up



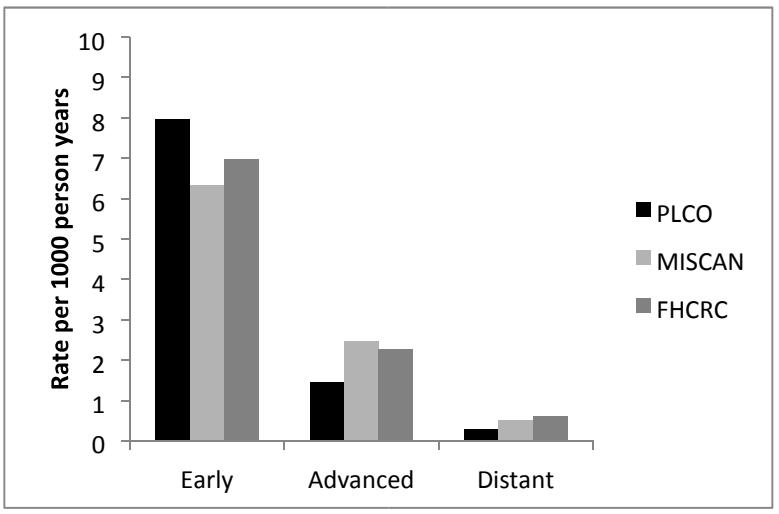
ERSPC, control arm, 11 years of follow-up



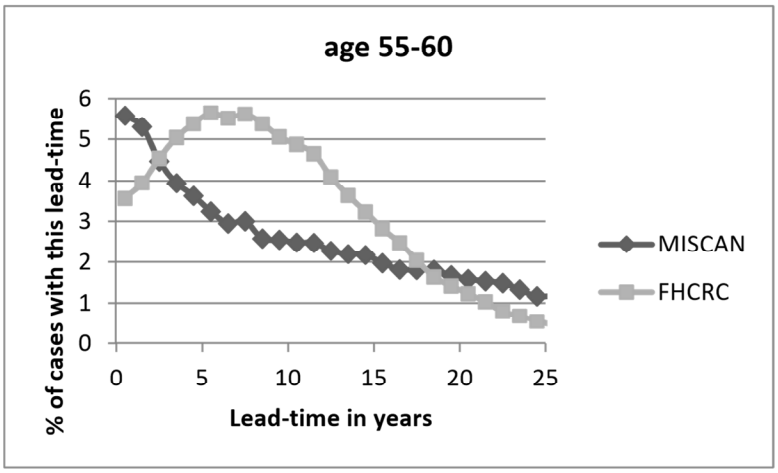
PLCO, screen arm, 13 years of follow-up

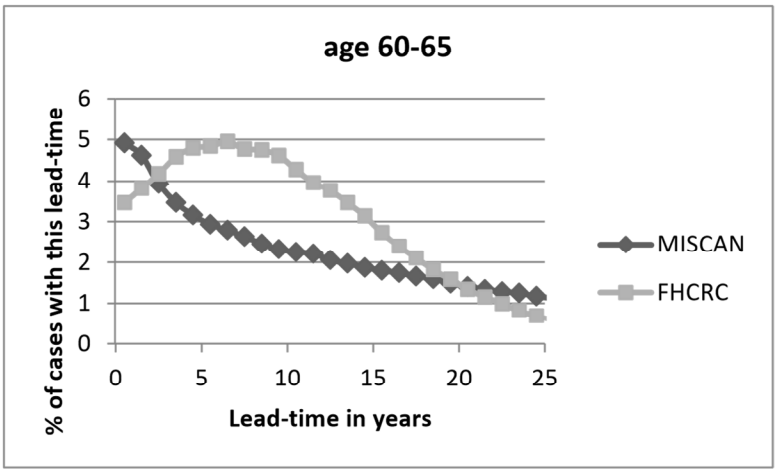


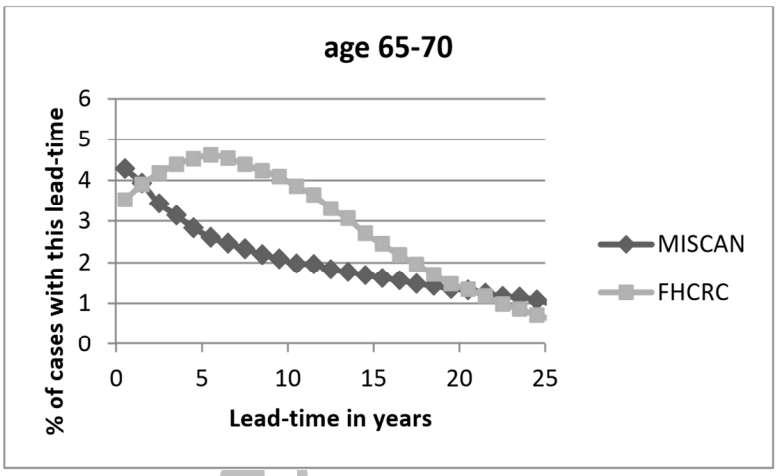
PLCO, control arm, 13 years of follow-up

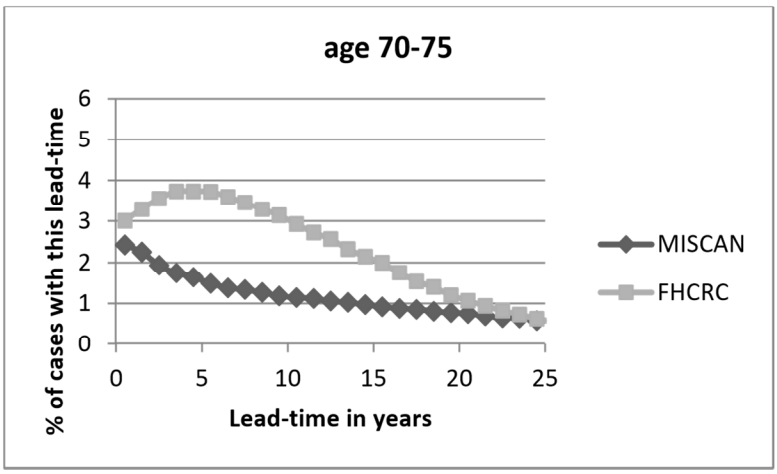


Supplementary Material 7. Lead-time distribution of screen detected cancers ERSPC

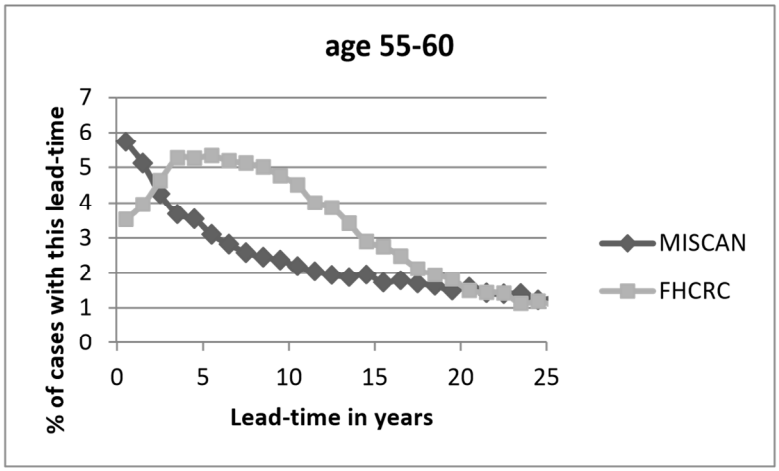


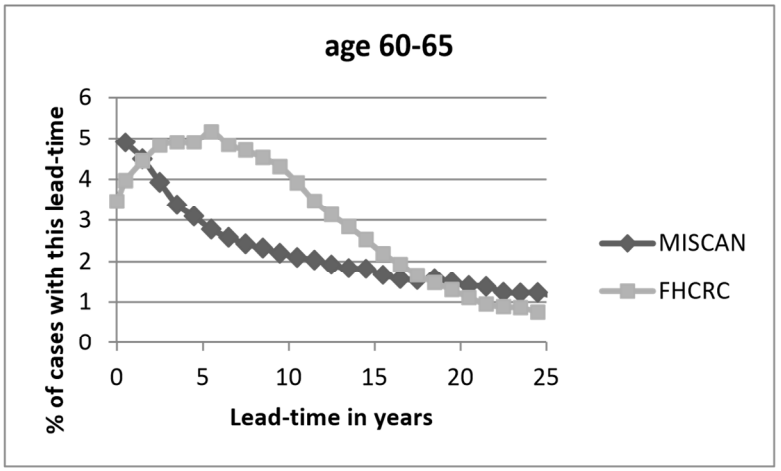


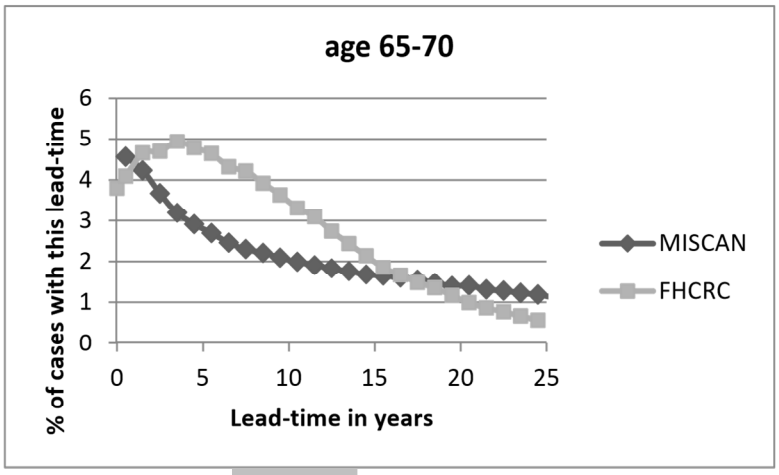


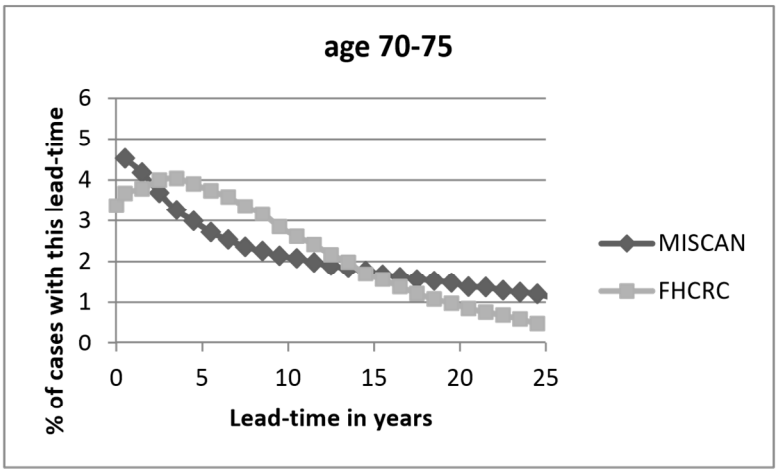


PLCO









Supplementary Material 8. Sensitivity analysis

In a sensitivity analysis, the effect of varying the cure parameter was explored. The cure parameter was calibrated to the limits of the 95% confidence interval of the prostate cancer mortality hazard ratio of 0.68-0.91 after 11 years in the ERSPC trial.[14] For the MISCAN model, the cure parameter for a prostate cancer mortality hazard ratio of 0.68 and 0.91 in the ERSPC was 0.50 and 0.08 respectively.

The MISCAN results are shown in Figure 1. Assuming no contamination, 100% attendance, 100% receipt of biopsies after a positive screen, and a 1-yearscreening interval (run 5), the effect of screening on prostate cancer mortality would be between 20% and 64%.

Using these two cure rates, the prediction of the prostate cancer mortality reduction in the PLCO would be 3%-15% (run 12), which is almost fully in the confidence interval of the PLCO trial. Therefore, the sensitivity analysis supports the primary analysis.

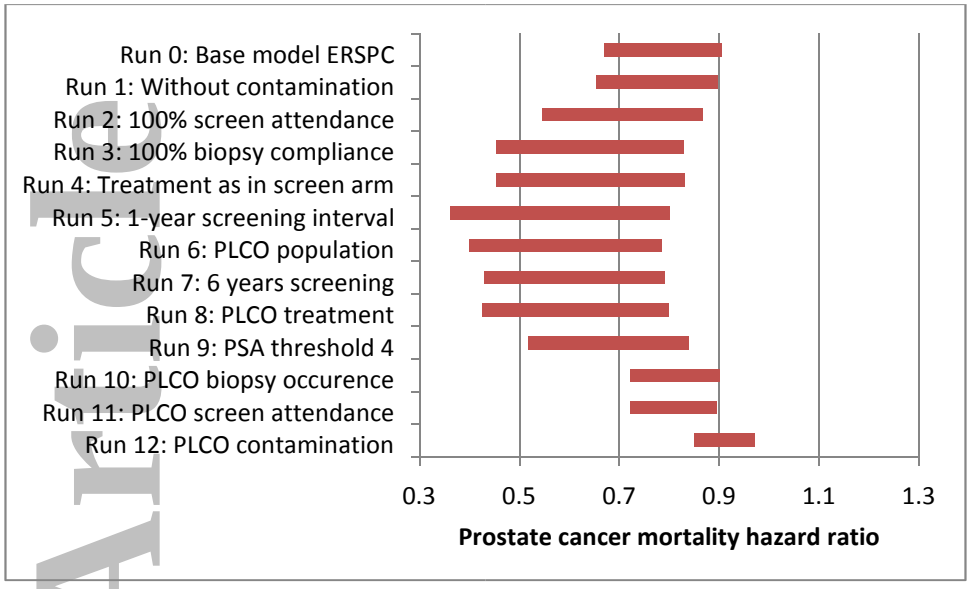


Figure 1. The prostate cancer mortality hazard reduction predicted by the MISCAN model using the cure parameters fitted to the limits of the 95% confidence interval of the published end results of the ERSPC. The left side of the bar represents the prostate cancer mortality hazard ratio corresponding to the prostate cancer mortality hazard ratio of 0.68 in the ERSPC and the right side is corresponding to the prostate cancer mortality hazard ratio of 0.91 in the ERSPC.

Supplementary Material 9. Uncertainty in the survival parameters

We used the FHCRC model to explore parameter uncertainty in the estimated cure parameter and the baseline survival hazard ratio, which we denote here as the parameter vector □. To do so, we evaluated a Poisson likelihood for each pair of candidate values of □ over a regular grid on the unit square. As for the results in the main manuscript, we used published numbers of deaths after 11 years of follow-up[13] as the observed data in the likelihood. Similar to the estimates obtained using simulated maximum likelihood, we found that a cure parameter of 0.20 and a baseline survival hazard ratio of 0.74 maximized the likelihood, which we denote $\hat{\square} = (0.74, 0.2)$. Furthermore, values of □ that satisfied the condition:

$$2\left(l(\square\square\square\hat{\square}) - l(\square)\right) \leq \square^2_{2,1-\square},$$

where $l(\cdot)$ is the log-likelihood function and $\square^2_{2,1-\square}$ denotes the $1-\square$ quantile from chi-square distribution with 2 degrees of freedom, determine a profile likelihood $100(1-\square)\%$ confidence region for □. Setting α= 5%, we obtained the profile 95% confidence region shown in Figure 1. This confidence region excludes values of □ with a cure parameter of 0. In particular, after accounting for improvements in baseline prostate cancer survival in the ERSPC trial relative to that observed among untreated cases in

SEER before screening began, we estimate a profile 95% confidence interval for the cure parameter to range from 0.07 to 0.40.

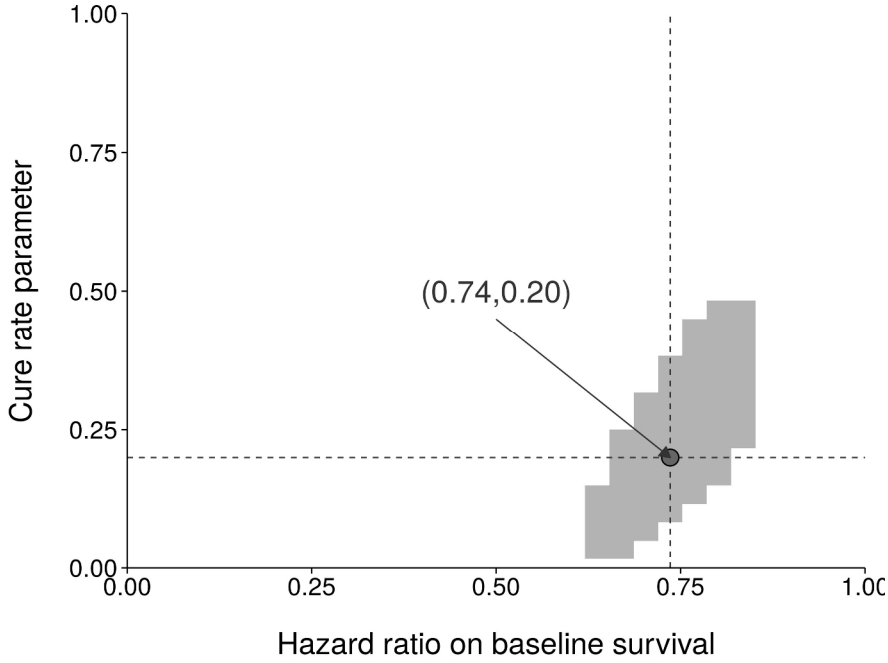


Figure 1. Maximum likelihood estimate and corresponding profile likelihood 95% confidence region for the survival parameters based on a Poisson likelihood for the number of prostate cancer deaths after 11 years of follow-up in the ERSPC using the FHCRC model.

Next, we drew random samples from the profile 95% confidence region and, for each value, projected the corresponding mortality rate ratio between trial arms for each run involving the benefit of screening estimated for the ERSPC in Figure 4 (i.e., runs 0–12). The results, averaged across 10 simulations to minimize error due to the microsimulation modeling framework, are shown in Figure 2. For each run, the dots show the projected mortality rate ratio corresponding to the maximum likelihood estimates, and the intervals are true confidence intervals reflecting uncertainty around the predicted rate ratio due to uncertainty in the estimated survival parameters. We observe the same effects of varying individual characteristics of the trials as we observed in Figure 4. In particular, we find comparable estimates under the run with the greatest mortality reduction (run 5), namely a 47% mortality reduction with profile 95% confidence interval 16% to 65%. We also find a prostate cancer mortality rate ratio in the PLCO setting that is much closer to 1 due to its circumstances of implementation even if screening had the same effect as in the ERSPC.

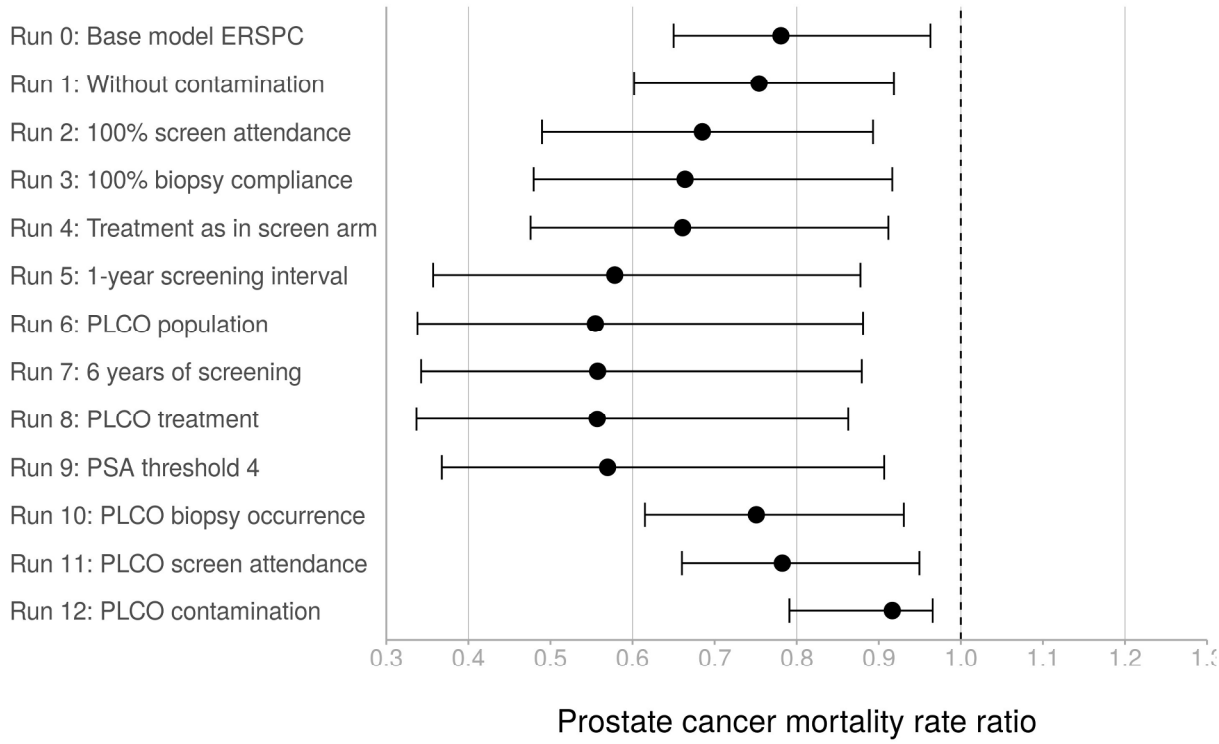


Figure 2. Step-by-step variation in prostate cancer mortality rate ratios due to varying trial populations, protocols, and practices and corresponding profile 95% confidence intervals using the FHCRC model.

Supplementary Material 10. Model predictions for Europe

To assess what this PSA efficacy would mean to the screen-naïve European population, we predicted the absolute number of prostate cancer deaths that would be averted annually if screening were implemented following different protocols.

We used the following assumptions: We modeled the European population from the year 2015 to 2040. An 80% attendance rate to screening and a threshold for biopsy referral PSA > 3 ng/ml were assumed. The biopsy compliance was 90% and the biopsy sensitivity 90%. The distribution and efficacy of treatment by age, stage, and Gleason score was the same as in the ERSPC trial. The cure parameter as a result of screening as fitted to the ERSPC data was used (0.22).

The following screening protocols were modeled using the MISCAN model:

- One screen at age 55
- Screens at ages 55, 57, and 59 (cost-effective protocol from Heijnsdijk et al.[15])
- Biennial screening from age 55 to 69 (AUA guideline)

The model output was the number of prostate cancer deaths per 100,000 men in 5-year age groups in the year 2040 (when a steady-state will be reached) in both a situation without screening and a situation

with screening. These rates were converted to absolute numbers of prostate cancer deaths by using the predictions of the male population size of the 28 EU countries made by EuroStat for the year 2040.

The results show that, depending on the screening protocol, 6,657 to 62,529 prostate cancer deaths could be prevented annually in the 28 EU countries compared to a situation without screening (Table 1). A cost-effective protocol would prevent 20,258 prostate cancer deaths annually, which is a reduction of 14%.

Table 1. The absolute number of prostate cancer deaths prevented annually, predicted by the MISCAN model.

Protocol	Absolute number of prostate cancer deaths averted annually (prostate cancer mortality reduction)
1 screen at age 55	6,657 (-5%)
Screens at ages 55, 57, and 59	20,258 (-14%)
Biennial screening from age 55 to 69	62,529 (-44%)

References

1. 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;95:868-78.
2. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med 2012;367:595-605.
3. Wever EM, Draisma G, Heijnsdijk EA, et al. Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. J Natl Cancer Inst 2010;102:352-5.
4. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst 2009;101:374-83.
5. Draisma G, Postma R, Schröder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. Int J Cancer 2006;119:2366-71.
6. Barton RR, Ivey JS. Nelder-Mead simplex modifications for simulation optimization. Manage Sci 1996;42:954-73.
7. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative PSA-based prostate cancer screening strategies. Ann Intern Med 2013;158:145-53.
8. Gulati R, Inoue L, Katcher J, Hazelton W, Etzioni R. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. Biostatistics 2010;11:70719.

9. Gulati R, Tsodikov A, Etzioni R, et al. Expected population impacts of discontinued prostatespecific antigen screening. Cancer 2014;120:3519-26.
10. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med 2003;349:215-24.
11. Lubeck DP, Litwin MS, Henning JM, et al. The CaPSURE database: a methodology for clinical practice and research in prostate cancer. CaPSURE Research Panel. Cancer of the Prostate Strategic Urologic Research Endeavor. Urology 1996;48:773-7.
12. Schröder FH, van der Maas P, Beemsterboer P, et al. Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 1998;90:1817-23.
13. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014;384:2027-35.
14. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012;366:981-90.
15. Heijnsdijk EA, de Carvalho TM, Auvinen A, et al. Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. J Natl Cancer Inst 2015;107:366.