This is the accepted manuscript of the article, which has been published in European Urology, 2019, 75(6), 1015-1022. https://doi.org/10.1016/j.eururo.2019.03.010

Could differences in treatment between trial arms explain the reduction in prostate cancer mortality in the European Randomized Study of Screening for Prostate Cancer?

Sigrid V. Carlsson[1,2], MD, PhD, MPH, Marianne Månsson PhD[2], Sue Moss[3], PhD, Maciej Kwiatkowski[4,5] MD, Franz Recker[4] MD, Teuvo L.J. Tammela[6], MD, Chris Bangma[7], Monique J. Roobol, PhD[7], Anssi Auvinen[8,*], MD, PhD and Jonas Hugosson[2,*], MD, PhD

[1]*Departments of Surgery (Urology Service) and Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, USA*
[2]*Department of Urology, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Sweden*
[3]*Centre for Cancer Prevention, Queen Mary University of London, London UK*
[4]*Department of Urology, Kantonsspital Aarau, Aarau, Switzerland*
[5]*Department of Urology, Academic Hospital Braunschweig, Braunschweig, Germany*
[6]*Tampere University, Department of Urology, Tampere, Finland*
[7]*Erasmus Medical Center, Department of Urology, Rotterdam, The Netherlands,*
[8]*Tampere University, School of Health Sciences, Tampere, Finland*

** Both senior authors contributed equally*

Keywords: prostate cancer, screening, early detection, treatment

Corresponding author:
Prof. Jonas Hugosson
Department of Urology
Sahlgrenska University Hospital
Bruna Stråket 11B
Gothenburg, Sweden
E-mail: jonas@urol.se
Phone: +46-31-3429017

Abstract

Background: Differential treatment between trial arms has been suggested to bias prostate cancer (PC) mortality in the European Randomized Study of Screening for Prostate Cancer (ERSPC).

Objective: To assess the contribution of treatment differences on the observed PC mortality reduction between the screening (SA) and control arm (CA).

Design, Setting and Participants: 14,136 men with PC (SA: 7,310; CA: 6,826) in the core age group (55-69 years) at 16 years of follow-up.

Outcome Measurements: Observed versus estimated number of PC deaths by treatment allocation in the SA and CA, respectively.

Statistical Analysis: Primary treatment allocation was modelled using multinomial logistic regression adjusting for: center, age, year, prostate-specific antigen, Grade Group, and TNM stage. For each treatment, logistic regression models were fitted for risk of PC death, separately for the SA and CA, and using the same covariates as for the treatment allocation model. Treatment probabilities were multiplied with estimated PC death risks for each treatment based on one arm, then summed and then compared to the observed number of deaths.

Results: When comparing the estimated differences (after controlling for clinical and pathological variables) with the observed treatment distributions between the ERSPC trial arms, the residual discrepancies weremarginal: hormonal therapy [0.1% vs. 9.4%], radical prostatectomy [0.8% vs. 10.1%], radiotherapy [3.3% vs. 7.9%], and surveillance [2.6% vs. 7.2%].

We estimated a reduction in the between-group risk of PC mortality to be 0.01% [95% CI -0.3%, 0.2%] when applying the SA model to the CA and 0.05% (95% CI (-0.1%, 0.2%) when applying the CA model SA, had the two groups received identical primary treatment given clinical characteristics. Because the observed difference between groups was 4.2%, our findings suggest that differential treatment explains only a trivial proportion of the main findings of the ERSPC trial.

Limitation: Only data on primary treatment was available.

Conclusion: Use of PSA remains the predominant explanation for the reduction in PC mortality seen in the ERSPC trial and is not attributable to differential treatment between trial arms.

Patient Summary: This study shows that the prostate cancer deaths in the European screening trial (ERSPC) were prevented because men were diagnosed and treated earlier through PSA-screening and not because of different, or better, treatment in the screening arm compared to the control arm.

Take Home Message: The reduction in prostate cancer mortality in the European Randomized Study of Screening for Prostate Cancer (ERSPC) is a result of early detection and not differential treatment between trial arms.

Twitter Summary: Prostate cancer mortality reduction in ERSPC trial due to early detection and not differential treatment

Introduction

The European Randomized Study of Screening for Prostate Cancer (ERSPC) has shown that prostate-specific antigen (PSA) can reduce prostate cancer (PC) mortality by 21% at 13 years.(1) It has been questioned whether this reduction is an effect of screening and early detection, or whether differential treatments between trial arms could have resulted in superior outcomes in the screening arm (SA) vs. control arm (CA) of the trial(2, 3). Others have hypothesized that the net reduction in prostate cancer deaths in ERSPC was not a result of earlier detection of more curable cancers in the SA through PSA-screening, but rather a result of more prostate cancer deaths in the CA where "androgen deprivation therapy treatment contributed differentially to an increase in prostate cancer deaths in control patients".(4-6)

A difference in treatment distribution between the cases in the SA and CA of a screening trial overall is entirely expected and desirable, since the purpose of screening is to find cancer at a stage where curative treatment is still possible.(1) Conversely, because of later detection at an advanced stage, hormonal therapy is more commonly used in the CA with a larger fraction of advanced cancers.(1, 7)

Such treatment differences only cause bias if men with similar tumors are treated differently between arms. For instance, if a man with a low-risk PC would be equally likely to receive active surveillance or undergo surgery or radiotherapy, regardless of whether the cancer was detected through screening in the SA or through clinical detection in the CA, then a treatment selection bias is an unlikely explanation for the mortality differences between groups.

An earlier study from the ERSPC, comparing the treatment distribution between the trial arms by patient and tumor characteristics, showed that a man with non-metastasized high-risk PC in the CA was more likely to receive radiotherapy, hormonal treatment or surveillance, compared to a similar man

in the SA. However, trial arm had only a minor role in the choice of treatment as compared to clinical and pathological variables.(8) Nevertheless, the potential impact of any differential treatment on the PC mortality results of the ERSPC trial has not previously been investigated.

We undertook the present study to estimate the PC mortality that would have been observed in the SA, if similar primary treatment had been applied with identical outcomes as in the CA, by clinical risk.

Methods

The design of the ERSPC has been described previously.(9) The study population of the present analysis encompassed all 14,136 men diagnosed with PC (7,310 in the SA; 6,826 in the CA), in the core age group 55-69 years at randomization (analysis to core age group predefined at the outset(9) and age band that is shared by all centers), in four centers: Finland (n=8,046), Netherlands (n=3,700), Sweden (n=1,406), and Switzerland (n=984). These four centers were chosen as they are the largest centers that contribute the most data to the ERSPC trial, and that all had clinical prognostic and treatment variables available for comparability. Cases included those diagnosed from randomization date to December 31, 2013, truncated at 16 years of follow-up.

Surveillance was defined as either active surveillance or watchful waiting. Only primary treatment was available for analysis and was categorised as radical prostatectomy, radiotherapy, surveillance, or hormonal therapy alone. If hormonal treatment was recorded as primary treatment, followed by radical prostatectomy or radiotherapy, then the treatment with curative intent was used. A total of 193 cases (108 in the SA; 85 in the CA) with other or unknown primary treatment were excluded from the main analysis. Deaths were ascertained by linkage with national registries, and causes of death were evaluated in a blinded manner by an independent cause of death committee following a standard

algorithm(10), except in Finland where a very high concordance (97%, kappa 0.95) between the committee assignments and official causes of death has been established and hence the official cause of death were used(11).

Statistical analysis

We used a potential outcomes approach applying a counterfactual method, estimating what the outcome would have been, if the (counterfactual) treatment allocation in one trial arm had been identical to that in the other arm, while the risk of dying from PC remained the same once the choice of treatment was fixed(12, 13). In short, this was done in the following steps (see **Figure 1**):

I. One model for treatment allocation was fitted for each arm separately.
II. For each arm and treatment, a model for PC death was fitted.
III. The expected number of men in SA dying from PC had they been allocated a treatment as in CA was calculated by combining I and II. The same procedure was carried out for the men in CA.
IV. The potential outcome in SA, that is the expected number of men in SA dying from PC, had they been allocated a treatment as in CA, was compared to the observed outcome, and correspondingly for CA.

Details of the different steps follow below:

I. *Treatment allocations models*

Treatment allocation was modeled by means of multinomial logistic regression, for the SA and CA separately. The following variables were used as covariates: center, age at diagnosis (years), year at diagnosis (≥year 2000 vs. <2000; this year was chosen to mark the start of the PSA era and noticeable uptick in the incidence of PC detected at a curable stage in Europe (14, 15); a sensitivity analysis was

performed split at year 2005, when combined treatment for localized disease became available), PSA level (to reflect clinical risk grouping: ≤10 ng/mL; >10 and ≤20; >20; missing or 0), Grade Group (1; 2-3; 4-5; missing), clinical T stage (T1abc/Tx/missing; T2; T3-4), N and M stage (both N0 and M0; at least one N1 or M1; otherwise missing).

II. *PC death models*

For each treatment, logistic regression models were fitted for risk of PC death. This was done separately for the screening and control arms. Log of follow-up time from randomization was used as offset(16, 17). The same variables were used as covariates as for the treatment allocation models.

III. *Potential outcomes in SA and CA, had the men been given treatments as in the opposite arm*

For the men in SA:

a. The treatment allocation model based on CA was used to estimate potential treatments based on the men´s individual values of the covariates.
b. For each treatment, the risk of PC death was estimated for each man based on his values of the covariates using the PC death models based on SA (the true arm).
c. The estimates in a. and b. were combined to achieve the risk of PC death for each man. The risk can be seen as a weighted average of the risk of PC deaths for the different treatments (using the total probability theorem).
d. Summing these numbers over all men in SA gives the expected number of PC deaths.

The same procedure was carried out for the men in CA.

IV. *Comparison of observed and potential number of PC deaths*

The potential outcome in SA, that is the expected number of men in SA dying from PC, had they been allocated a treatment as in CA, was compared to the observed outcome, and correspondingly for CA.

In order to estimate the uncertainty in the estimates, non-parametric bootstrapping was used; 2000 bootstrap samples were drawn, and the whole procedure was applied to each bootstrap sample. Confidence intervals for the difference in the number and proportion of the different treatments, and in PC deaths were then derived by means of the so-called basic method(18). Categories for missing data were used in the main analysis. As a sensitivity analysis, modeling was performed on complete cases only. Furthermore, three more sensitivity analyses were carried out: cut-off for year of diagnosis at 2005 instead of 2000; excluding one center at a time; men with "Other/unknown" treatment were given treatments in an extreme fashion (all men with missing in one arm got the same treatment and similarly for the other arm, for all combination of treatments) (Supplementary Appendix A).

Simulations of different scenarios were performed in order to show that the modeling strategy gave reliable results (Supplementary Appendix B). The actual trial data was used, but treatments and PC deaths were altered in a random fashion in two different scenarios. In the first scenario, treatment distributions and risks of death were chosen so that there was a theoretical difference between "observed" PC deaths in CA and predicted by the treatment model based on SA, and vice versa. In the second scenario, a random reordering of PC deaths and treatments were done, which should lead to that "observed" and predicted PC deaths were the same. Statistical analyses were performed using R version 3.3.1.

A step-by-step "toy" example of the methodology is illustrated in Supplementary Appendix C.

Results

The analysis comprised 14,136 men with a follow-up of 16 years. Supplementary Table 1 shows the number of PC cases by center and trial arm, with Finland contributing with the largest number of participants, followed by the Netherlands, Sweden and Switzerland.

Screening introduced a stage and grade shift, with a higher proportion of tumors of lower stage and grade in the SA compared to the CA and conversely, fewer men with metastasized disease in the SA (Table 1).

Supplementary Table 2 shows the distributions of treatment by trial center and trial arm. Surveillance and radical prostatectomy was used more frequently in the SA and, conversely, radiation therapy and hormonal therapy were more common in the CA. Surveillance was frequently utilized in both trial arms in Sweden (45% in the SA and 34% in the CA) whereas radiotherapy was frequently used in both trial arms in Finland and the Netherlands. Radical prostatectomy was common in Switzerland.

Figure 2 (data included in Table 2) shows the observed and estimated treatment distributions by trial arm for the four centers together. Adjusting for center, age at diagnosis, year at diagnosis, PSA-level, Grade Group, and TNM stage, the difference between the estimated (adjusted) and observed treatment distributions in the respective arms was marginal. When comparing the estimated differences (after controlling for clinical and pathological variables) with the observed treatment distributions between the ERSPC trial arms, the residual discrepancies weremarginal: hormonal therapy [0.1% vs. 9.4%], radical prostatectomy [0.8% vs. 10.1%], radiotherapy [3.3% vs. 7.9%], and surveillance [2.6% vs. 7.2%].

Similarly, as seen in Table 3, the difference in the estimated and observed number of PC deaths was very small. We estimated a reduction in the between-group risk of PC mortality to be 0.01% [95% CI -0.2%, 0.2%] when applying the SA model to the CA and 0.05% (95% CI (-0.1%, 0.2%) when applying the

CA model SA, had the two groups received identical primary treatment given clinical characteristics. Because the observed difference between groups was 4.2%, our findings suggest that differential treatment explains only a trivial proportion of the main findings of the ERSPC trial. Similar findings were seen in a complete case only analysis (Figure 3, data included in Table 2).

In our simulation studies of our method, the estimates produced by our approach were similar to those entered as parameters in the simulations, hence supporting the reliability of the results presented here (Supplementary Appendix B). In the sensitivity analyses the absolute difference between observed and estimated PC deaths were less than 0.2%, except for the case where Finland was excluded. Then the SA model led to a somewhat lower number of estimated PC deaths in the control arm, and the difference was 0.5% (Supplementary Appendix A).

Discussion

We undertook the present study to assess the contribution of treatment differences on the observed PC mortality reduction between trial arms in ERSPC.

Despite slight differences in treatment distributions between trial arms adjusted for prognostic variables, these differences in treatment had little effect on PC mortality. The difference in the number of PC deaths between the estimated and observed models was small (0.01%-0.05%). Therefore, it is unlikely that the reduction in PC mortality in the ERSPC trial is attributable to differential treatment between the trial arms, rather, the PC mortality reduction mainly reflects the effect of PSA-screening leading to early detection, allowing for effective management.

The study has several strengths. Rather than using strata of tumor risk groups, prognostic variables were included in multivariable models predicting treatment and PC death, respectively. This approach minimizes subtle differences in tumor characteristics, which would have been missed if data had been categorized into risk groups. In particular the high-risk group, as generally defined, is very heterogeneous. For instance, a man with screen-detected small volume, Grade Group 4, clinically non-palpable tumor (clinical stage T1c) with a low PSA-value, would be classified as “high-risk”, as would a man with Grade Group 5, T3 tumor and a high PSA up to 100 ng/mL. Because of this heterogeneity in risk grouping, patients in the CA may still have more advanced tumor features compared to men in the SA, within similar risk groups. Using multivariable models rather than risk grouping thus aims to account for any slight difference in treatment distribution that can occur within risk groups, even if men with similar disease are treated similarly irrespective of trial arm.

Furthermore, men in the SA were slightly younger at diagnosis (median 67 years) compared to corresponding men in the CA (median 69 years), which might affect treatment choice. Even if the age

difference is small, it may result in more aggressive treatment for men in the SA. We addressed this by adjusting for age.

As men in the CA were treated later in time, because they were diagnosed more recently compared to men in the SA (due to lead-time gained by screening), new, effective treatments could have become available for men in the CA. For instance, during the study period, multimodal treatment of locally advanced PC has been shown to yield more favourable outcomes than surgery or radiation alone.(19) In the Dutch center, a prior study from ERSPC reported that treatment patterns differed slightly between the trial arms; for instance, RT was more often combined with hormonal therapy and radiation dosages was often higher than 69 Gy in the CA, relative the SA, likely reflecting later detection in the CA and development and adoption of novel treatment modalities over time.(20) Despite these differences favoring the CA, differences in disease-free survival between the arms were reported to be minimal.(20) We attempted to control for the effect of treatment development over the study period by adjusting for year of treatment, however lack of granular information regarding treatment received remains a limitation of this study.

Another strength is that the study is based on a very large patient material and the setting of a randomized screening trial, which is a major advantage, increasing comparability between the arms. Data from each center were sent in to an independent central database every 6 months from the beginning of this study. A quality control committee ensured standardized tumor classification within each center. A pathology committee co-ordinated biopsy grading. A PSA committee conducted inter-comparisons of all laboratories involved in the trial. The primary endpoint, i.e. death from PC, was evaluated by COD committees who analyzed deaths in men with PC in a blinded fashion i.e. unaware of trial allocation.(9, 10)

The study is not devoid of limitations. The study is limited by only being able to include four centres of the ERSPC in the analysis and still having some missing data on prognostic factors (slightly

more missing data in the CA) and lack of detailed treatment information, because only primary treatment data was collected in the trial. Any effect of secondary treatment upon disease recurrence on PC mortality would be smaller compared to primary treatment with curative intent and we have no reason to believe that men in the SA received more frequent or higher-quality secondary treatments compared to the CA. The current study did not aim to address whether treatment quality differed between the trial arms, or whether men received treatment at academic high-volume centers in one arm more frequently than in the other. It has been hypothesized that "men in the screening group received treatment at a few centers that specialized in treatment of prostate cancer [whereas] the men in the control group received standard care in their community [and] that is likely to account for some and possibly all of the survival benefit"(3). In the Netherlands' center, men diagnosed with PC were referred back to their primary care physicians for further management and referral. In the Swedish, Finnish and Swiss centers, and because the studies were concentrated to defined geographical areas, the same referral hospitals were largely responsible for treatments of men in both arms (21, 22).

Conclusion

Based on multivariable modeling data from the ERSPC trial, differences in the receipt of primary treatment between the SA and CA were minimal and the potential effect of these differences on PC mortality was extremely small. These findings suggest that the effectiveness of PSA screening in reducing PC mortality in the ERSPC trial was largely due to early detection, allowing for effective management, and was not attributable to differential treatment between trial arms.

Conflict of interest

S.C. has received a lecture honorarium and travel support from Astellas Pharma (unrelated to current study). M.K. reports travel support from Janssen and Astellas. No other author has any conflict of interest to disclose.

Acknowledgements

We thank Dr. Andrew Vickers at Memorial Sloan Kettering Cancer Center for advice on the statistical methods. Gathering detailed information on treatment within ERSPC Rotterdam was made feasible through unconditional support of the Rotterdam Prostate Cancer Foundation (SWOP).

Funding

The funding of the ERSPC trial was obtained from national cancer research funding agencies, European funding in the form of Framework programmes, private sponsors, and a grant from the former Beckman/Hybritech company. S.C.'s work on this study was supported by grants from AFA Insurance, the Swedish Cancer Society, the Sweden America Foundation, the Swedish Council for Working Life and Social Research, the Swedish Society for Medical Research and the Swedish Prostate Cancer Foundation. S.C. is further supported in part by a Cancer Center Support Grant from the National Cancer Institute made to Memorial Sloan Kettering Cancer Center (P30-CA008748), a SPORE grant from the National Cancer Institute to Dr. H. Scher (P50-CA092629), a grant from the National Cancer Institute as part of the Cancer Intervention and Surveillance Modelling Network (U01CA199338-02) and the David H. Koch prostate cancer research fund. This work was further supported by a Movember grant from the Cancer Society of Finland, and a grant from the Academy of Finland (#260 931).

Role of the funding source

None.

Author's contributions

Study concept and design: SC, SM, AA, JH.

Acquisition of data: All authors.

Analysis and interpretation of data: All authors.

Drafting of the manuscript: SC.

Critical revision of the manuscript for important intellectual content and final approval of the submitted version: All authors.

Statistical analysis: MM, SM.

Obtaining funding: All authors.

Administrative, technical, or material support: N/A.

Table 1. Patient characteristics

	Screening arm	Control arm	Total
Number of men	7,310	6,826	14,136
Age at diagnosis, median (IQR) (years)	67 (64, 71)	69 (66, 73)	68 (65, 72)
PSA at diagnosis, median (IQR) (ng/mL)	6.0 [4.1, 10.5]	9.8 [6.3, 18.4]	7.7 [4.7, 14.0]
Missing, n (%)	159 (2%)	316 (5%)	475 (3%)
Time from diagnosis to treatment, median (IQR), days	54 (33, 89)	47 (26, 80)	50 (29, 84)
Missing, n (%)	330 (5%)	357 (5%)	687 (5%)
Follow-up from diagnosis, median (IQR), years	7.6 (4.2, 11.6)	5.5 (2.4, 8.8)	6.8 (3.2, 10.6)
Biopsy Grade Group, n (%)			
1	4680 (64%)	3,137 (46%)	7,817 (55%)
2-3	1671 (23%)	2,185 (32%)	3,856 (27%)
4	395 (5%)	574 (8%)	969 (7%)
5	339 (5%)	662 (10%)	1,001 (7%)
Missing	225 (3%)	268 (4%)	493 (4%)
Clinical T stage, n (%)			
T1a,b,c	4,585 (63%)	3,632 (53%)	8,217 (58%)
T2	1,818 (25%)	1,609 (24%)	3,427 (24%)
T3	708 (10%)	1,169 (17%)	1,877 (13%)
T4	128 (2%)	293 (4%)	421 (3%)
Missing	71 (1%)	120 (2%)	191 (1%)
Clinical M stage, n (%)			
M0	5,164 (71%)	4,371 (65%)	9,535 (68%)
M1	259 (4%)	559 (8%)	818 (6%)
Missing	1,848 (25%)	1,846 (27%)	3,694 (26%)
Clinical N stage, n (%)			
N0	2,942 (40%)	1,614 (24%)	4,556 (32%)
N1	85 (1%)	166 (2%)	251 (2%)
Missing	4,283 (59%)	5,046 (74%)	9,329 (66%)

Missing includes men with unknown or other treatment
IQR=Interquartile range, T=tumor, N=node, M=metastasis

Table 2. Observed and estimated treatments by trial arm among men with prostate cancer (data illustrated in Figure 2)

	Screening arm (SA)			**Difference (CA – SA)**	**Control arm (CA)**			**Difference (SA – CA)**
Treatment	Observed N (%)	Estimated CA model N (%)	Difference: Estimated CA model – observed in SA % (95% CI)	Observed (%)	Observed N (%)	Estimated SA model N (%)	Difference: Estimated SA model – observed in CA % (95% CI)	Observed (%)
Hormonal therapy	692 (9.6%)	709 (9.8%)	0.2% (-0.5, 0.9)	9.4%	1,274 (19.0%)	1,271 (18.9%)	-0.1% (-1.3, 1.2)	-9.4%
Radical Prostatectomy	2,315 (32.0%)	2,298 (31.8%)	-0.2% (-2.0, 1.5)	-10.1%	1,472 (21.9%)	1,523 (22.7%)	0.8% (-0.5, 2.1)	10.1%
Radiotherapy	2,202 (30.5%)	2,441 (33.8%)	3.3% (1.6, 5.0)	7.9%	2,575 (38.4%)	2,351 (35.0%)	-3.3% (-5.0, -1.6)	-7.9%
Surveillance	2,016 (27.9%)	1,777 (24.6%)	-3.3% (-4.9, -1.7)	-7.2%	1,387 (20.7%)	1,563 (23.3%)	2.6% (1.4, 3.8)	7.2%

Table 3. Observed and estimated numbers of prostate cancer deaths among men with prostate cancer (data illustrated in Figure 3)

	Screening Arm (SA)			Difference (CA – SA)	Control Arm (CA)			Difference (SA – CA)
	Observed N (%)	**Estimated CA model N (%)**	**Difference: Estimated CA model – observed in SA % (95% CI)**	**Observed (%)**	**Observed N (%)**	**Estimated SA model N (%)**	**Difference: Estimated SA model – observed in CA % (95% CI)**	**Observed (%)**
Main analysis: All data	442 (6.1%)	446.6 (6.2%)	0.05% (-0.1%, 0.2%)	4.2%	691 (10.3%)	691.8 (10.3%)	0.01% (-0.3%, 0.2%)	-4.2%
Sensitivity analysis: Complete cases	388 (5.7%)	390.4 (5.7%)	0.03% (-0.2%, 0.2%)	3.8%	580 (9.5%)	580.8 (9.5%)	0.01% (-0.3%, 0.3%)	-3.8%

Tumor nodal (N)- and metastasis (M)- stages are not used for death models in the sensitivity analysis

Supplementary Table 1. Number of men by trial center and trial arm

	Number of men with prostate cancer*			Number of men randomized (without prostate cancer at the time of randomization)		
Center	**Screening arm N (%)**	**Control arm N (%)**	**Total N (%)**	**Screening arm N (%)**	**Control arm N (%)**	**Total N (%)**
Finland	3,500 (48%)	4,546 (67%)	8,046 (57%)	31874 (53%)	48302 (63%)	80176 (59%)
Netherlands	2,376 (33%)	1,324 (19%)	3,700 (26%)	17422 (29%)	17379 (23%)	34801 (26%)
Sweden	814 (11%)	592 (9%)	1,406 (10%)	5901 (10%)	5950 (8%)	11851 (9%)
Switzerland	620 (9%)	364 (5%)	984 (7%)	4920 (8%)	4929 (6%)	9849 (7%)
Total	7,310 (100%)	6,826 (100%)	14,136 (100%)	60117 (100%)	76560 (100%)	136677 (100%)

*Includes men with unknown or other treatment

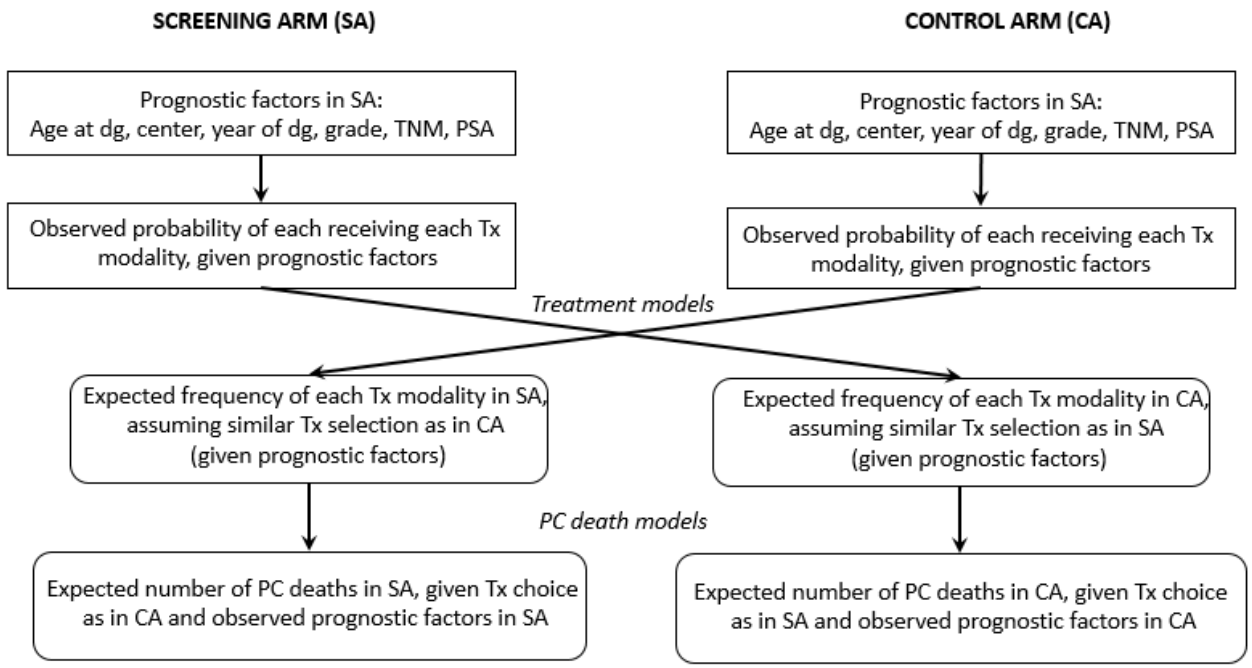
Supplementary Table 2. Primary treatment by trial center and trial arm

Center	Hormonal Therapy		Radical Prostatectomy		Radiotherapy		Surveillance		Other / Unknown		Total	
	SA	CA	SA	CA	SA	CA	SA	CA	SA	CA	SA	CA
Finland	493 (14%)	898 (20%)	951 (27%)	910 (20%)	1,230 (35%)	1,855 (41%)	775 (22%)	787 (17%)	51 (2%)	96 (2%)	3,500 (100%)	4,546 (100%)
Netherlands	90 (4%)	184 (14%)	774 (33%)	229 (17%)	833 (35%)	598 (45%)	667 (28%)	308 (23%)	12 (0.5%)	5 (0.4%)	2,376 (100%)	1,324 (100%)
Sweden	98 (12%)	167 (28%)	271 (33%)	150 (25%)	72 (9%)	60 (10%)	364 (45%)	200 (34%)	9 (1%)	15 (3%)	814 (100%)	592 (100%)
Switzerland	11 (2%)	25 (7%)	319 (52%)	183 (50%)	67 (11%)	62 (17%)	210 (34%)	92 (25%)	13 (2%)	2 (0.5%)	620 (100%)	364 (100%)
Total	692 (10%)	1,274 (19%)	2,315 (32%)	1,472 (22%)	2,202 (30%)	2,575 (38%)	2,016 (28%)	1387 (20%)	85 (1%)	118 (2%)	7,310 (100%)	6,826 (100%)

SA=screening arm, CA=control arm

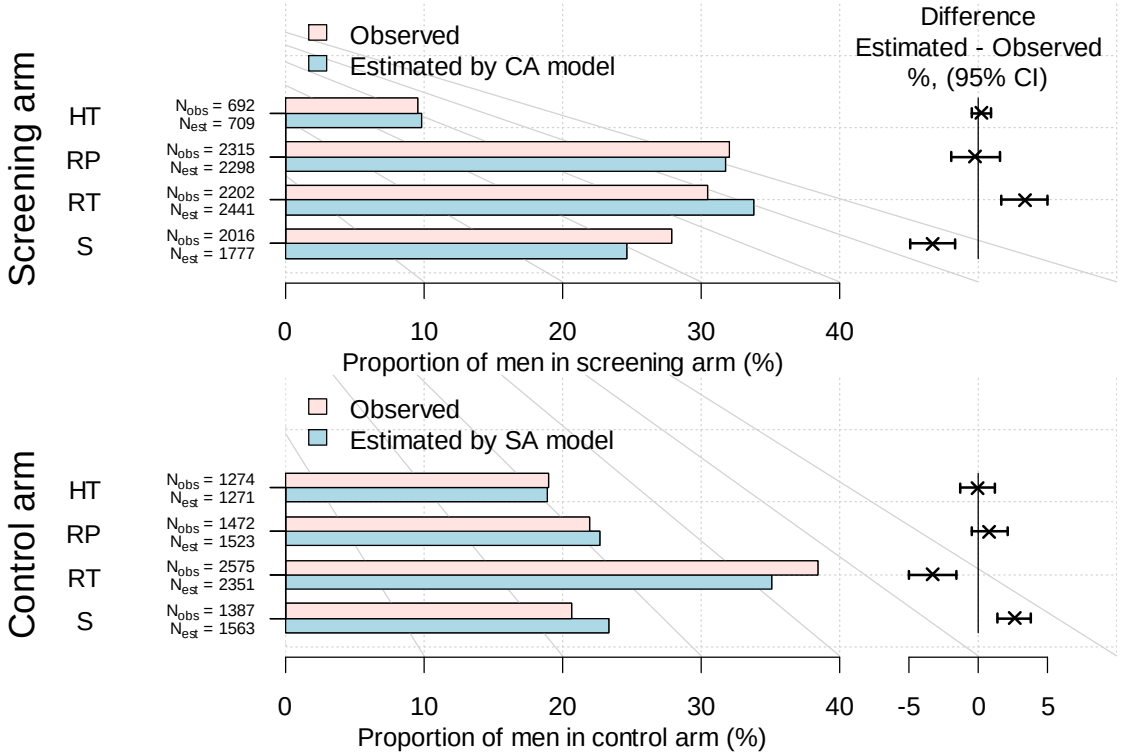
Includes men with unknown or other treatment

Figure 1. Description of study methodology



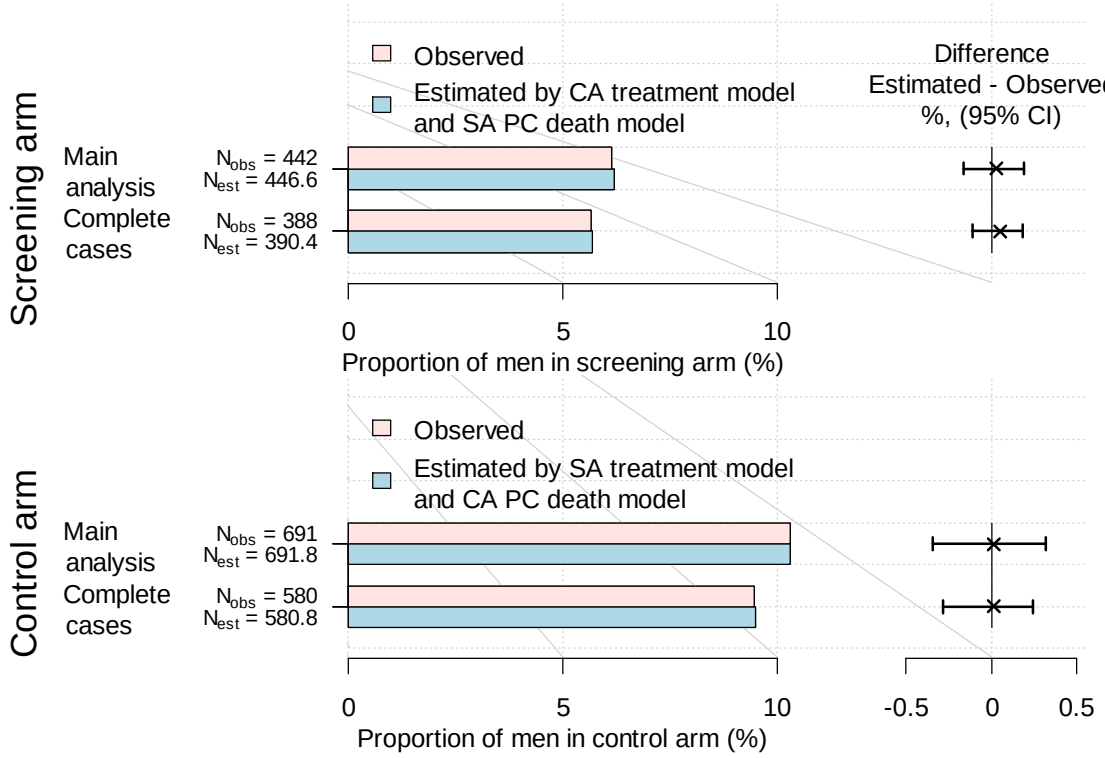
Abbreviations: SA=screening arm; CA=control arm, dg=diagnosis, TNM=tumor, nodes, metastasis stage, PSA=prostate specific antigen, Tx=treatment; PC=prostate cancer

Figure 2. Observed and estimated treatments by trial arm (data included in Table 2)



SA=Screening Arm, CA=Control Arm, N=Number, CI=Confidence Interval, obs=observed, est=estimated, HT=Hormone Therapy, RP=Radical Prostatectomy, RT=Radiotherapy, S=Surveillance

Figure 3. Observed and estimated numbers of prostate cancer deaths (data included in Table 3)



SA=Screening Arm, CA=Control Arm, N=Number, CI=Confidence Interval, obs=observed, est=estimated, HT=Hormone Therapy, RP=Radical Prostatectomy, RT=Radiotherapy, S=Surveillance

Supplementary Appendix A. Sensitivity analyses

One center at a time excluded

Excluded center	Obs SA	Est CA model, SA	Diff Est_Obs_SA	Obs CA	Est SA model, CA	Diff Est_Obs_CA
Finland	192 (5.1%)	197.5 (5.2%)	5.5 (0.2%)	277 (12.3%)	265.4 (11.8%)	-11.6 (-0.5%)
Netherlands	324 (6.7%)	329.4 (6.8%)	5.4 (0.1%)	516 (9.6%)	520 (9.7%)	4 (0.07%)
Sweden	389 (6.1%)	392.4 (6.1%)	3.4 (0.05%)	613 (10%)	613.1 (10%)	0.1 (0%)
Switzerland	421 (6.4%)	425.2 (6.4%)	4.2 (0.06%)	667 (10.5%)	668 (10.5%)	1 (0.02%)

Cut-off for year of diagnosis

CutOff year	Obs SA	Est CA model, SA	Diff Est_Obs_SA	Obs CA	Est SA model, CA	Diff Est_Obs_CA
2005	442 (6.1%)	447.9 (6.2%)	5.9 (0.08%)	691 (10.3%)	688.8 (10.3%)	-2.2 (-0.03%)

An additional PSA category (0-5 ng/ml)

Obs SA	Est CA model, SA	Diff Est_Obs_SA	Obs CA	Est SAmodel, CA	Diff Est_Obs_CA
442 (6.12%)	446.5 (6.2%)	4.5 (0.06%)	691 (10.3%)	692 (10.3%)	1 (0.01%)

Unknown / Other treatment replaced and all men included in analysis

SA Replacement for other/unknown treatment	CA Replacement for other/unknown treatment	Obs SA	Est CAmodel, SA	Diff Est_Obs_SA	Obs CA	Est SAmodel, CA	Diff Est_Obs_CA
Hormonal Therapy	Hormonal Therapy	448 (6.1%)	452.3 (6.2%)	4.3 (0.06%)	712 (10.4%)	718.4 (10.5%)	6.4 (0.09%)
Hormonal Therapy	Radical Prostatectomy	448 (6.1%)	446.7 (6.1%)	-1.3 (-0.02%)	712 (10.4%)	725.5 (10.6%)	13.5 (0.2%)
Hormonal Therapy	Radiotherapy	448 (6.1%)	446.5 (6.1%)	-1.5 (-0.02%)	712 (10.4%)	722.3 (10.6%)	10.3 (0.2%)
Hormonal Therapy	Surveillance	448 (6.1%)	446.4 (6.1%)	-1.6 (-0.02%)	712 (10.4%)	723.7 (10.6%)	11.7 (0.2%)
Radical Prostatectomy	Hormonal Therapy	448 (6.1%)	455.6 (6.2%)	7.6 (0.1%)	712 (10.4%)	699.4 (10.3%)	-12.6 (-0.2%)
Radical Prostatectomy	Radical Prostatectomy	448 (6.1%)	451.3 (6.2%)	3.3 (0.05%)	712 (10.4%)	712.7 (10.4%)	0.7 (0.01%)
Radical Prostatectomy	Radiotherapy	448 (6.1%)	448.4 (6.1%)	0.4 (0%)	712 (10.4%)	704.5 (10.3%)	-7.5 (-0.1%)
Radical Prostatectomy	Surveillance	448 (6.1%)	448.3 (6.1%)	0.3 (0%)	712 (10.4%)	705.7 (10.3%)	-6.3 (-0.09%)
Radiotherapy	Hormonal Therapy	448 (6.1%)	458.4 (6.3%)	10.4 (0.14%)	712 (10.4%)	704.5 (10.3%)	-7.5 (-0.1%)
Radiotherapy	Radical Prostatectomy	448 (6.1%)	451.1 (6.2%)	3.1 (0.04%)	712 (10.4%)	712.5 (10.4%)	0.5 (0.01%)

Radiotherapy	Radiotherapy	448 (6.1%)	453.7 (6.2%)	5.7 (0.08%)	712 (10.4%)	712 (10.4%)	0 (0%)
Radiotherapy	Surveillance	448 (6.1%)	451 (6.2%)	3 (0.04%)	712 (10.4%)	710.9 (10.4%)	-1.1 (-0.02%)
Surveillance	Hormonal Therapy	448 (6.1%)	456.6 (6.3%)	8.6 (0.12%)	712 (10.4%)	702.8 (10.3%)	-9.2 (-0.1%)
Surveillance	Radical Prostatectomy	448 (6.1%)	449.7 (6.2%)	1.7 (0.02%)	712 (10.4%)	711 (10.4%)	-1 (-0.01%)
Surveillance	Radiotherapy	448 (6.1%)	449.5 (6.2%)	1.5 (0.02%)	712 (10.4%)	708 (10.4%)	-4 (-0.06%)
Surveillance	Surveillance	448 (6.1%)	453.5 (6.2%)	5.5 (0.08%)	712 (10.4%)	713.5 (10.5%)	1.5 (0.02%)

Supplementary Appendix B. Validation of modelling procedure by simulations

Does modelling procedure give unbiased results regarding number of prostate cancer (PC) deaths between observed and modelled in case:

1. There is a true difference between observed and expected
2. There is no true difference between observed and expected

In both cases the covariate values were kept for each individual. The same scripts as for the results in the manuscript are used.

<u>There is a true difference</u>

In this part treatments and risk of death are chosen so that there is a theoretical difference between "observed" and what is expected by the treatment model based on the other group. This is done by assuming that there is a higher risk of death if hormonal treatment is given, and that the chance of getting hormonal treatment is different in the two arms.

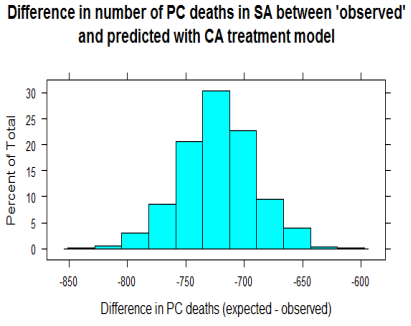
Assumptions ("truth"):

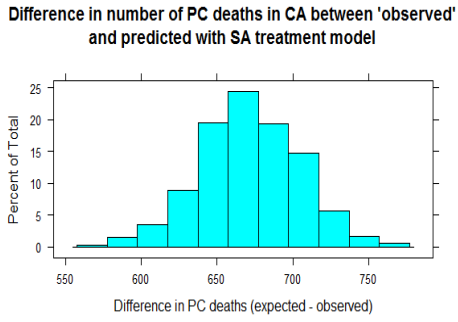
- Control arm (CA):
  - ➔ Treatments: p = 1/4 for all treatments
  - ➔ PC death: All treatments except hormonal: p = 1/10; Hormonal treatment: p = ½
  - ➔ Truth:
    - Exp[# of deaths] = 1341.6
    - Exp[# of deaths w treatments as in SA] = 2012.4
    - → Δ = 670.8

- Screening arm (SA):
  - ➔ Treatments: All treatments except hormonal: p = 1/6; Hormonal treatment: p = ½
  - ➔ PC death: Same as for CA
  - ➔ Truth:
    - Exp[# of deaths] = 2167.5
    - Exp[# of deaths w treatments as in CA] = 1445.0
  - → Δ = -722.5

Modelling (1000 simulations):

- Control arm: Mean difference in PC deaths with SA treatment model and "truth" in CA: 671.3
- Screening arm: Mean difference in PC deaths with CA treatment model and "truth" in SA: -722.7





There is no true difference

This is illustrated by means assuming that PC deaths are distributed at random between men (the same number as in the original data)

Assumptions ("truth"):

- Control arm (CA):
  - Treatments: Kept as in original data
  - PC deaths: Distributed at random between men (the same number as in the original data) whichever treatment → treatment distribution does not matter

  ➔ Truth:
  - Exp[# of deaths] = Observed = 691
  - Exp[# of deaths w treatments as in SA] = 691

  → Δ = 0

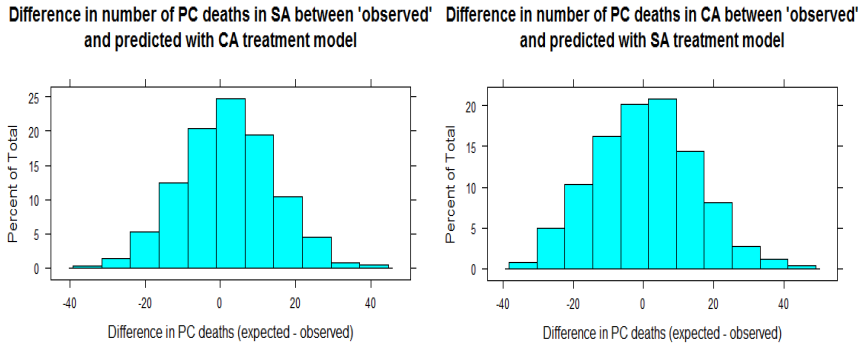
- Screening arm (SA): As for CA

  ➔ Truth:
  - Exp[# of deaths] = Observed = 442
  - Exp[# of deaths w treatments as in CA] = 442

  → Δ = 0

Modelling (1000 simulations):

- Control arm: Mean difference in PC deaths with SA treatment model and "truth" in CA: 0.4
- Screening arm: Mean difference in PC deaths with CA treatment model and "truth" in SA: 2

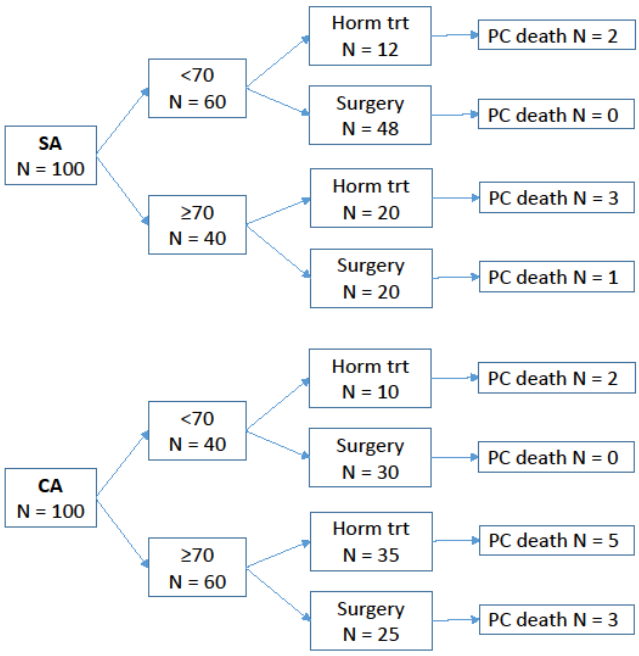


Supplementary Appendix C. "Toy" example to illustrate the methodology

We want to compare observed number of PC deaths in the CA with the expected if the men had received treatments as in the SA, given their clinical characteristics. For simplicity we assume

- In each of the SA and the CA, 100 men were diagnosed with PC
- The only covariate we consider is age at diagnosis, which is divided into <70 and ≥ 70 years
- Only two treatments are available: hormonal (H) and surgery (S)

The number of men in the different age categories, on the different treatments, and whether they died from PC or not, were distributed as follows in the two arms:



Since there is only one covariate (age at diagnosis) with only two levels, the modelling can be seen simply as calculating a number of proportions. Below, the notation P(A) should be read as "the probability of A)".

To be able to calculate the expected number of deaths in CA if they had received treatments as in SA, we need a treatment model based on SA and a death model based on CA:

I. A model for treatment allocation was fitted for the **SA** with age as covariate:
   - If age < 70 then P(hormonal) = 12/60 and P(surgery) = 48/60
   - If age ≥ 70 then P(hormonal) = 20/40 and P(surgery) = 20/40

II. For each treatment, a model for PC death was fitted for the **CA** with age as covariate:
- Hormonal treatment:
  - If age < 70 then P(PC death) = 2/10
  - If age ≥ 70 then P(PC death) = 5/35
- Surgery:
  - If age < 70 then P(PC death) = 0
  - If age ≥ 70 then P(PC death) = 3/25

The expected number of men in CA dying from PC had they been allocated a treatment as in SA is calculated by combining the treatment model based on SA (from I) and the PC death model based on CA (from II):

III. The probability of PC death for each man in CA is a weighted average of the risk of PC deaths for the different treatments:
- If age < 70 years:
  P(PC death) = P(PC death if trt=H) * P(trt = H) + P(PC death if trt=S) * P(trt = S) = 2/10 * 12/60 + 0 * 48/60 = 0.04
- If age ≥ 70 years:
  P(PC death) = P(PC death if trt=H) * P(trt = H) + P(PC death if trt=S) * P(trt = S) = 5/35 * 20/40 + 3/25 * 20/40 = 0.13

Summing over the 100 men in CA, 40 of which were less than 70 years at diagnosis, leads to 40*0.04 + 60*0.13= 9.5 expected deaths.

IV. This number is then compared to the observed number of PC deaths in CA which was 10 deaths.

References

1. Schroder 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(9959):2027-35.
2. Barry MJ. Screening for prostate cancer--the controversy that refuses to die. N Engl J Med. 2009;360(13):1351-4.
3. Brawley O. American Cancer Society press room. https://acspressroom.wordpress.com/2010/07/01/digging-deeper-into-prostate-screening-study/ Date accessed: 2014-12-10. 2014.
4. Haines IE, Gabor Miklos GL. Prostate-specific antigen screening trials and prostate cancer deaths: the androgen deprivation connection. J Natl Cancer Inst. 2013;105(20):1534-9.
5. Carlsson S, Roobol MJ, Schroder FH, Hugosson J, Auvinen A. RE: Prostate-specific antigen screening trials and prostate cancer deaths: the androgen deprivation connection. J Natl Cancer Inst. 2014;106(5).
6. Carter HB. Reducing prostate cancer deaths: unsupported speculation about the androgen deprivation hypothesis. Eur Urol. 2014;66(6):973-4.
7. Schroder FH, Hugosson J, Carlsson S, et al. Screening for Prostate Cancer Decreases the Risk of Developing Metastatic Disease: Findings from the European Randomized Study of Screening for Prostate Cancer (ERSPC). Eur Urol. 2012;62(5):745-52.
8. 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(10):2387-93.
9. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360(13):1320-8.
10. De Koning HJ, Blom J, Merkelbach JW, et al. Determining the cause of death in randomized screening trial(s) for prostate cancer. BJU Int. 2003;92 Suppl 2:71-8.
11. Makinen T, Karhunen P, Aro J, Lahtela J, Maattanen L, Auvinen A. Assessment of causes of death in a prostate cancer screening trial. Int J Cancer. 2008;122(2):413-7.
12. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. Stat Med. 1997;16(9):1017-29.
13. Kilpelainen TP, Tammela TL, Malila N, et al. The Finnish prostate cancer screening trial: analyses on the screening failures. Int J Cancer. 2015;136(10):2437-43.
14. Jonsson H, Holmstrom B, Duffy SW, Stattin P. Uptake of prostate-specific antigen testing for early prostate cancer detection in Sweden. Int J Cancer. 2011;129(8):1881-8.
15. Bray F, Lortet-Tieulent J, Ferlay J, Forman D, Auvinen A. Prostate cancer incidence and mortality trends in 37 European countries: an overview. Eur J Cancer. 2010;46(17):3040-52.
16. Anderson D, Feldblum C, Modlin D, Schirmacher D, Schirmacher E, Thandi N. A Practitioner's Guide to Generalized Linear Models. Casualty Actuarial Society Discussion Paper Program. 2004:1-116.
17. Hilbe J. Generalized Linear Models and Extensions. College Station, TX: Stata Press, 2001.
18. Davison AC, Hinkley DV. Bootstrap Methods and Their Application. Chapter 5.2. Cambridge University Press. 1997.

19. Widmark A, Klepp O, Solberg A, et al. Endocrine treatment, with or without radiotherapy, in locally advanced prostate cancer (SPCG-7/SFUO-3): an open randomised phase III trial. Lancet. 2009;373(9660):301-8.
20. Bokhorst LP, Kranse R, Venderbos LD, et al. Differences in Treatment and Outcome After Treatment with Curative Intent in the Screening and Control Arms of the ERSPC Rotterdam. Eur Urol. 2015;68(2):179-82.
21. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. Lancet Oncol. 2010;11(8):725-32.
22. Kilpelainen TP, Tammela TL, Malila N, et al. Prostate cancer mortality in the Finnish randomized screening trial. J Natl Cancer Inst. 2013;105(10):719-25.