Review Article

Prostate cancer screening: what can we learn from randomised trials?

Anssi Auvinen

Faculty of Social Sciences/Health Sciences, University of Tampere, Tampere, Finland
Correspondence to: Anssi Auvinen, MD, PhD. Faculty of Social Sciences/Health Sciences, University of Tampere, Tampere, Finland.
Email: Anssi.Auvinen@staff.uta.fi.

Abstract: In this article, the principle of randomised trials are first described and then prostate cancer screening trials published to date are evaluated based on these principles. A summary of the randomised prostate cancer screening is provided. The conclusion that can be made from the results of the screening trials, as well as limitations of the evidence and open questions are outlined in the end.

Keywords: Prostate cancer screening; screening trials; validity; effectiveness

Submitted Dec 11, 2017. Accepted for publication Dec 17, 2017.
doi: 10.21037/tau.2017.12.13
View this article at: http://dx.doi.org/10.21037/tau.2017.12.13

Introduction

Modern medicine requires that management decisions are based on solid science. The question that we seek to answer is whether an intervention can produce the defined goals. In the spirit of evidence-based medicine and effectiveness or outcomes research, claims of effectiveness must be derived from head-to-head comparisons of therapeutic alternatives, with direct observation in comparable patients and conditions with controlled design. A valid study design allows us to infer what would have happened to this group of patients subjected to an intervention, if they had received an alternative intervention (ceteris paribus).

Principles of randomised trials

A randomised trial is the gold standard for evaluating the effectiveness of any medical intervention (1). Its chief virtue is that it can eliminate bias from treatment assignments. By virtue of randomisation, they can create comparable groups at baseline and avoid any imbalance in the key determinants of the outcomes studied and hence minimise confounding and selection bias.

This requires, however, careful selection of subjects in terms of eligibility and exclusion criteria (2,3) (*Table 1*). Recruitment can already be a hurdle, if potential participants are not willing to join a trial. Population-based effectiveness trials that identify and randomise subjects prior to consent have at least an apparent advantage in this respect. However, low compliance or substantial contamination can abolish that benefit. Informed consent is crucial in a randomised trial, where participants are subjected to an intervention that usually also carries a risk of adverse effects. Also, information on the features defining eligibility should be readily available and up to date. For instance, in a screening trial the participants must be free from the target condition, and ideally recruitment attempts should exclude people with the disease at the outset, based on case lists from comprehensive cancer registries.

In addition, the randomisation procedure must be valid, i.e., based on genuinely unpredictable allocation to guarantee allocation concealment (4,5). If the assignment of a particular subject can be foreseen, allocation can become biased by certain subjects being preferably assigned to a given trial arm, which leads to selection bias (6). The allocation ratio, i.e., proportion of subjects assigned to the trial arms does not need to be equal, but it must be identical across the participants (overall in simple randomisation and within the group in stratified randomisation). The current requirement is that computer-generated random numbers are used, preferably assigned by a separate party not responsible for recruitment and only after registering

Table 1 Key features affecting the relevance and quality of randomised trials

Feature	Effect
Equipoise	Relevant study question, avoid excessive risks
Outcome	Clinically important end-point
Comparator	Relevant study question
Fair randomisation	Avoid selection bias and confounding
Blinding	Avoid information bias
Adequate sample size	Sufficient statistical power
Intention-to-treat analysis	Maintain balanced distribution

a participant to the trial so that potential for tampering the random allocation is minimised.

Randomisation can, however, generate comparability only at baseline, at the time of allocation. Selective loss during a study can lead to subsequent distortion of the initial balance (7,8). Use of passive data collection through e.g., registers can avoid this, but maintaining validity also requires that researchers are allowed to compile data on subjects who have withdrawn from the study. This is clearly an ethical issue, as there is a conflict between autonomy and common good that can be obtained from valid knowledge available only from unbiased research.

An adequate sample size is a prerequisite for achieving a balanced distribution of the characteristics between the trial arms through randomisation. Further, realistic power calculations with reasonable expected event rates and allowance for non-compliance are needed to define the target sample size (9).

The research question of a trial should fulfil the requirement for equipoise, i.e. there should be a balance of uncertainty regarding the study hypothesis (10). This means that there is sufficient preliminary evidence to support a hypothesis (proof of principle) and safety data (often from phase 1–2) studies, but also genuine uncertainty about the effects of the experimental intervention. For this purpose, a systematic literature review should be incorporated as part of the proposal. Too early randomized controlled trial (RCT) ('a long shot') can lead to low chance of achieving a benefit and a poorly characterised risk profile. Besides the intervention studied, the comparator must also be chosen equally carefully. As a general principle, it should represent normal health care, i.e., the approach that would be provided in the absence a trial. The outcome should be chosen so that it quantifies the real goal of the intervention to the patient, i.e., benefit to the subject and it should not be only an indirect indicator such radiological or biochemical marker of an early stage of the disease process (11).

Blinding (efforts at keeping either participants or investigators unaware of the allocation) can improve comparability of end-point data and hence reduce information bias (11). It is, however, not possible for interventions that cannot be mimicked by sham procedures. Further, blinding is important primarily for outcomes involving a degree of judgment. Subjective outcome measures can be easily influenced by beliefs about interventions and their effects (placebo effect). Therefore, it is important to blind the evaluators of any 'soft' end-points. Objective outcomes, on the other hand, are not prone to placebo effects.

In order to maintain the benefits of randomisation, the analysis should follow the intention-to-treat (or screen) principle (12). This means that the groups compared are fully in compliance with random allocation. Only predefined exclusions can be justified and even they only in case that it had later turned out that subjects had been randomised even if they were not eligible. It is crucial that similar criteria are applied consistently across the arms and information is obtained in a similar fashion for the trial arms (to avoid asymmetrical exclusions of ineligible subjects) (13,14). 'Treatment received' analysis compares the participants with and without the intervention (either just excluding non-compliers in the intervention arm, or also combining them with 'contaminators' in the control arm), but such an approach loses entirely the advantage of randomisation and treats the data as a cohort study. Sometimes even the data analysis is performed in a blinded fashion, before opening the code revealing which arm is which. This can avoid subtle biases in the analysis phase of the study.

A randomised trial usually requires monitoring, frequently conducted by external, independent experts.

Table 2 Prostate cancer mortality in prostate cancer screening trials

Trial	Size of study population (screening + control)	Target age group (years)	Screening tests	Participation (%)	Follow-up (years)	Number of prostate cancer deaths (screening + control)	RR for prostate cancer mortality
Norrköping	1,494+7,532	50–69	DRE, PSA	70–78	20	30+130	1.2 (0.8–1.7)
Stockholm	2,400+24,772	55–70	DRE, TRUS, PSA	74	13	53+506	1.1 (0.8–1.5)
Quebec	31,133+15,353	45–80	PSA, DRE	24	11	153+75	1.0 (0.8–1.3)
PLCO	38,340+38,343	55–74	DRE, PSA	85–89	15	255+244	1.0 (0.9–1.2)
ERSPC	72,891+89,352	55–69	PSA	83	13	355+545	0.8 (0.7–0.9)

RR, risk ratio.

Trials can be terminated early, if the results show convincing evidence of either benefit or futility of the intervention, or unacceptable risks (15,16). This required for ethical conduct and good research practice, but it can also lead to premature discontinuation of a study, with inconclusive results, if careful consideration is not applied. A pre-specified monitoring plan should be employed and alpha spending considered to avoid type one error. For instance, a Finnish population-based colorectal cancer screening trial was discontinued when the 5-year results showed no benefit (17). The follow-up may have been too short, especially given evidence of benefit from previous trials. An Estonian trial of menopausal hormone treatment was also terminated early, mainly based on results from other trials that were far from identical (18).

Besides quantifying the impact on the primary goals of interventions (providing the benefit to the patient, avoiding adverse effect of disease), they can also provide best estimates of the cost-effectiveness and quality of life impacts.

Randomised trials of prostate cancer screening

Several small prostate cancer screening trials have been carried out, while only the PLCO and ERSPC have sufficient sample size for assessing the effect of PSA screening on prostate cancer mortality (*Table 2*). Some other studies have also been conducted such as Stockhom-3 study (19) and Tyrol study (20), which are not trials, as they involved no randomisation. Also, several case-control studies have been conducted (21-23), but they are not covered here.

The Norrköping trial (24) was started in 1987. It did not involve real randomisation, but every sixth man was allocated to screening from a list of birth dates. Four screening rounds were used with a 3-year interval. Hence, the control group was 5 times larger than the intervention group (*Table 2*). Initially only DRE was used, but in the two last rounds also PSA (in the last round men older than 69 years, 46% of the screening group, were no longer invited). Participation ranged 70–78% by round. There were 43 screen-detected and 42 interval cases in the screening groups and 292 prostate cancers in the control group (cumulative incidence 5.7% *vs.* 3.9%). Cumulative prostate cancer mortality at 20 years was 2% (30/1,494) in the screening group and 1.7% (130/7,532) in the control group, RR =1.2, 95% CI: 0.8–1.7 (though the researchers report deaths among prostate cancer cases as their main results). The study was underpowered for assessing a mortality effect.

The Stockholm trial (25) had a very small screening group of 2,400 men. It had only a single screening round in 1988 using DRE, TRUS and PSA. The PSA cut-off for biopsy was 10 µg/L. Of the invited men, 74% participated and 65 cancers were detected (3.6%). Overall, prostate cancer incidence was 4.0 per 100 in the screening group and 5.2 per 1,000 in the control group (*Table 2*). There were 53 deaths from prostate cancer in the screening arm and 506 in the control arm, corresponding to cumulative mortality of 2.2% *vs.* 2.0%, with RR =1.1, 95% CI: 0.8–1.5.

The Quebec trial was started in 1988 and used PSA as the screening test (cut-off 3 µg/L), at the initial screen also DRE was used (26). Unlike the Swedish studies, it allocated more men to the screening than control arm (2:1). The target age group was very broad, 40–79 years. Screening was offered on annually, but compliance was low (24%) and 7% of the men in the control group sought screening on their own at Laval University. The results demonstrated

no mortality reduction between the trial arms with RR =1.0, 95% CI: 0.8–1.3 (*Table 2*) (though the investigators presented an analysis of screened *vs.* non-screened men and claimed that the results showed a screening benefit).

The PLCO trial recruited volunteers aged 55–74 years from 10 centers. Both PSA and DRE were used in the three first annual screens and PSA only (cut-off 4 µg/L) in the last two rounds. The incidence of prostate cancer at 13 years was 108 per 10,000 person-years in the screening arm and 97 in the control arm (RR =1.12, 95% CI: 1.02–1.17) (27). There were 255 deaths from prostate cancer in the screening and 244 in the control arm by 15 years, with an RR =1.04, 95% CI: 0.87–1.24 (28). The study population was large, but prostate cancer mortality was lower than in the US population at large. It had weaknesses in diagnostic evaluation, as only a third of the screen-positive men underwent a prostate biopsy. Contamination was substantial, as 45% of the men in the control arm had been screened within 3 years prior to baseline, and extensive PSA testing continued in the control arm during the intervention phase (29). A recent modelling study suggested that if the conduct of the PLCO trial had been similar to the ERSPC, it would have shown a mortality benefit (and vice versa for ERSPC) (30). Similar conclusions have also been reached by other investigators (31-33).

ERSPC trial is the largest RCT on prostate cancer screening, with eight centres, albeit the French data have not been included in the mortality analyses. A 4-year screening interval was used, with a PSA cut-off of 3 µg/L. Of the men assigned to screening, 83% were screened at least once and there were on average 2.3 tests per man. Nearly 5,000 prostate cancers were detected through 140,000 screening test (detection 3.5%). Cumulative incidence of prostate cancer was 10.2% in the screening arm and 6.2% in the control arm. A 20% reduction in prostate cancer mortality was demonstrated already at nine years, and it has remained similar at 11 and 13 years of follow-up (34-36) (*Table 2*). A recent analysis of cause of death attribution demonstrated that there is practically no impact of any bias in the adjudication (37). The ERSPC trial has also been criticised for treatment imbalance between the arms. Some of the criticisms are however misguided, as the requirement for comparability is that similar men with similar disease should be treated equally across the arms. A shift in disease characteristics should be reflected in treatment distributions and it does not bias the results. Yet, some imbalance has been shown even after stratification for major prognostic factors (stage or risk group) (38). It appears, however, that the difference is small that it could account for only a small fraction of the screening effect. A full analysis of the issue is being conducted within the ERSPC. Substantial differences between the trial centres have emerged, with large mortality reductions in Sweden and the Netherlands, but very little screening impact in Finland (36,39). These remain to be explained in full but may be influenced by a diluting effect of contamination (40,41), as well as larger effect of continued screening past there round or 8 years.

Conclusions

Currently, randomised screening trials, ERSPC above others, have shown that it is possible to reduce prostate cancer mortality.

A randomised trial is a blunt instrument, very much like an epistemic sledgehammer. This means that it can provide a very reliable answer to a single research question (notwithstanding a parallel design with several interventions, or with adequate sample size and pre-specified protocol for a sub-group analysis). Hence it is crucial to design the aim carefully.

A screening trial needs a well-designed intervention and cannot assess the impact of other interventions. Hence, the optimal screening test can be defined by a randomised trial only if one comparing several alternative approaches side by side could be conducted, but it does not appear practicable.

Are there high-risk groups where balance of benefits and harms of PSA-based screening is more favourable? It is often assumed that targeting a high-risk population will increase the benefits of screening. This requires, however, empirical assessment and no material differences have been shown between men with a positive family history and those without it (42,43). The target group with the largest screening benefit can be evaluated within a trial population. Statistical power is, however, often inadequate for small subgroups, such as high-risk groups defined by family history or specific genetic alterations. A polygenic risk score is potential method for stratifying men by probability of aggressive prostate cancer and prostate cancer death. However, empirical evidence remains scarce (44,45).

Any cancer screening trial must have a long follow-up to show a mortality reduction, as prostate cancer deaths occur mainly after 75 years of age, long after the window of opportunity for early detection with curative treatment. Hence, novel approaches will have always been developed by the time the final results are available. In this respect,

evidence from the trials is always outdated. This can also apply to the potential effect of introduction of new treatment approaches on the screening effect.

Even if prostate cancer is among the most frequent causes of cancer death, it accounts only for a small percentage of all deaths. This means that it will not be possible to show an effect in overall mortality.

The ERSPC trial has shown that PSA-based screening can reduce prostate cancer mortality. However, the substantial excess detection weighs heavily against the benefit. How could overdiagnosis be avoided? PSA alone does not appear to have sufficient specificity for clinically important cancers and those screening trials that have not shown clearly higher risks of prostate cancer in the screening arm have not shown mortality benefit either. Also within the ERSPC, the centers with substantial mortality reduction have also markedly elevated prostate cancer risk in the screening arm (39). Kallikrein panels such as Prostate Health Index (PHI) and 4Kscore and magnetic resonance imaging (MRI) hold promise for improving this (46,47), but their impact remains to be evaluated in randomised trials. One is on-going in Sweden and another being launched in Finland (48).

More general clinical issues not limited to screening but with substantial impact on screening outcomes include improved prognostic stratification to guide which men should be curatively treated (or how to identify patients at high risk of disease progression), and tailoring therapeutic approaches to maximise benefits and minimise harms by e.g., early endocrine treatment or chemotherapy.

Acknowledgements

None.

Footnote

Conflicts of Interest: The author has no conflicts of interest to declare.

References

1. Chalmers I. Development of fair tests of treatment. Lancet 2014;383:1713-4.
2. Van Spall HG, Toren A, Kiss A, et al. Eligibility criteria of randomized controlled trials published in high-impact general medical journals. JAMA 2007;297:1233-40.
3. Blümle A, Meerpohl JJ, Rücker G, et al. Reporting of eligibility criteria of randomised trials. BMJ 2011;342:d1828.
4. Savović J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomized controlled trials. Ann Intern Med 2012;157:429-38.
5. Page MJ, Higgins JO, Clayton G, et al. Empirical evidence of study design biases in randomised trials. PLoS One 2016;11:e0159267.
6. Kahan BC, Rehal S, Cro S. Risk of selectin bias in randomized trials. Trials 2015;16:405.
7. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials. Control Clin Trials 1995;16:62-73.
8. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601-5.
9. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 updated guidelines for reporting parallel groups randomised trials. J Clin Epidemiol 2010;63:e1-37.
10. London AJ. Equipoise in research. JAMA 2017;317:525-6.
11. Glasser SP, Howard G. Clinical trial design issues. J Clin Pharmacol 2006;46:1106.
12. Hopewell S, Dutton S, Yu LM, et al. The quality of reports of randomised trials in 2000 and 2006. BMJ 2010;340:c723.
13. Mills EJ, Wu P, Gangnier J, et al. The quality of reporting randomized trials in leading medical journals. Contemp Clin Trials 2005;26:480-7.
14. Kane RL, Wang J, Garrard J. Reporting of randomized clinical trials improved after adoption of the CONSORT statement. J Clin Epidemiol 2007;60:241-9.
15. Grant AM, Altdman DG, Babiker AB, et al. Issues in data monitoring and interim analysis of trials. Health Technol Assess 2005;9:1-238, iii-iv.
16. Harman NL, Conroy EJ, Lewis SC, et al. Exploring the role and function of trial steering committees. Trials 2015;16:597.
17. Pitkäniemi J, Seppä K, Hakama M, et al. Effectiveness of screening for colorectal cancer with a faecal occult-blood test in Finland. BMJ Open Gastroenterol 2015;2:e000034.
18. Veerus P, Hovi SL, Fischer K, et al. Results from the Estonian postmenopausal hormone therapy trial. Maturitas 2006;55:162-73.
19. Grönberg H, Adolfsson J, Aly M, et al. Prostate cancer screening in men aged 50-69 years (STHLM3). Lancet Oncol 2015;16:1667-76.
20. Oberaigner W, Siebert U, Horninger W, et al. Prostate-

specific antigen testing in Tyrol, Austria. Int J Public Health 2012;57:57-62.
21. Bergstralh EJ, Roberts RO, Farmer SA, et al. Population-based case-control study of PSA and DRE screening on prostate cancer mortality. Urology 2007;70:936-41.
22. Concato J, Wells CK, Horwitz RI, et al. The effectiveness of screening for prostate cancer: A nested case-control study. Arch Intern Med 2006;166:38-43.
23. Weinmann S, Richert-Boe K, Glass AG, Weiss NS, Prostate cancer screening and mortality: a case-control study. Cancer Causes Control 2004;15:133-8.
24. Sandblom G, Varenhorst E, Rosell J, et al. Randomised prostate cancer screening trial: 20-year follow-up. BMJ 2011;342:d1539.
25. Kjellman A, Akre O, Norming U, et al. 15-year follow-up of a population-based prostate cancer screening study. J Urol 2009;181:1615-21.
26. Labrie F, Candas B, Cusan L, et al. Screening decreases prostate cancer mortality. Prostate 2004;59:311-8.
27. Andriole GL, Crawford ED, Grubb RL, et al. Prostate cancer screening in the randomized PLCO cancer screening trial. J Natl Cancer Inst 2012;104:125-32.
28. Pinsky PF, Prorok PC, Yu K, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. Cancer 2017;123:592-9.
29. Pinsky PF, Blacka A, Kramer BS, et al. Assessing contamination and compliance in the prostate component of the PLCO cancer screening trial. Clin Trials 2010;7:303-11.
30. Tsodikov A, Gulati R, Heijnsdik EA, et al. Reconciling the effects of screening on prostate cancer mortality. Ann Intern Med 2017;167:449-55.
31. 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.
32. Palma A, Lounsbury DW, Schlecht NF, et al. A systems dynamic model of serum prostate specific antigen screening for prostate cancer. Am J Epidemiol 2016;183:227-36.
33. de Koning HJ, Gulati R, Mosss SM, et al. The efficacy of prostate-specific antigen screening: Impact of key components in the ERSPC and PLCO trials. Cancer 2017. [Epub ahead of print].
34. Schröder FH, Hugosson J, Roobol MJ, et al. Screeing and prostate cancer mortality in a randomized European study. N Engl J Med 2009;360:1320-8.
35. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate cancer mortality 11 years of follow-up. N Engl J Med 2012;366:981-90.
36. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality. Lancet 2014;384:2027-35.
37. Walter SD, de Koning HJ, Hugosson J, et al. Impact of cause of death adjudication on the results of the European prostate cancer screening trial. Br J Cancer 2017;116:141-8.
38. Wolters T, Roobol MJ, Steyerberg EW, et al. The effect of study arm on prostate cancer treatment in the larger screening trial ERSPC. Int J Cancer 2010;126:2387-93.
39. Auvinen A, Moss SM, Tammela TL, et al. Absolute effect of prostate cancer screening: Balance of benefits and harms by center within the ERSPC. Clin Cancer Res 2016;22:243-9.
40. Nevalainen J, Stenman UH, Tammela TL, et al. What explains the differences between centres in the ERSPC? Cancer Epidemiol 2017;46:14-9.
41. Kilpeläinen TP, Pogodin-Hannolainen D, Kemppainen K, et al. Estimate of opportunistic PSA testing in the Finnish randomised study of prostate cancer screening. J Urol 2017;198:50-7.
42. Saarimäki L, Tammela TL, Määttänen L, et al. Family history in the Finnish prostate cancer screening trial. Int J Cancer 2015;136:2172-7.
43. Randazzo M, Müller A, Carlsson S, et al. A positive family history is a risk factor for prostate cancer in a population-based study with organised PSA screening. BJU Int 2016;117:576-83.
44. Pashayan N, Pharoah PD, Schleutker J, et al. Reducing overdiagnosis by polygenic risk-stratified screening. Br J Cancer 2015;113:1086-93.
45. Castro E, Mikropoulos C, Bancroft EK, et al. The PROFILE feasibility study. Oncologist 2016;21:716-22.
46. Bryant RJ, Sjöberg DD, Vickers AJ, et al. Predicting high-grade cancer at ten-core biopsy using a four-kallikrein panel markers. J Natl Cancer Inst 2015;107:djv095.
47. Parekh DJ, Punnen S, Sjöberg DD, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. Eur Urol 2015;68:464-70.
48. Auvinen A, Rannikko A, Taari K, et al. A randomised trial of early detection of clinically significant prostate cancer (ProScreen). Eur J Epidemiol 2017;32:521-7.

Cite this article as: Auvinen A. Prostate cancer screening: what can we learn from randomised trials? Transl Androl Urol 2018;7(1):12-17. doi: 10.21037/tau.2017.12.13

