What Do the Screening Trials Really Tell Us and Where Do We Go From Here?

Ruth D. Etzioni, PhD[a,*], Ian M. Thompson, MD[b]

KEYWORDS

• Prostate cancer • Mass screening • Clinical trials • Prostate-specific antigen • Simulation modeling • Public health policy

KEY POINTS

- Screening trials provide information that is critical for the development of screening policy, but cannot provide all the information needed for developing sound policies for population screening.
- Results from a modeling analysis of the Prostate, Lung, Colon, and Ovarian trial reveals that the empirical finding of no difference in prostate cancer mortality in this study could have easily occurred even if prostate cancer screening had a high degree of efficacy.
- The balance of screening harm with benefit will be materially affected by patient decisions following diagnosis, such as whether the patient selects aggressive curative treatment or active surveillance to reduce the chance of overtreatment.

INTRODUCTION

The prostate screening odyssey has captivated researchers, policymakers, and clinicians since the late 1980s when the prostate-specific antigen (PSA) test was approved by the Food and Drug Administration for monitoring the progression of prostate cancer. The test was rapidly adopted for screening in the United States[1] even as clinical trials to evaluate its efficacy in early detection were just beginning in the United States and Europe.

While the United States and European trials were ongoing, routine PSA screening became the standard of care in the United States, dramatically changing the profile of prostate cancer, and prompting concerns about overdetection and overtreatment of the disease. As rates of death from prostate cancer declined after the inception of screening, it became clear that policies for prostate cancer screening would have to carefully navigate the harm-benefit trade-offs of PSA testing. The results of the 2 large, randomized screening trials were eagerly awaited for what was hoped would be the final word regarding the lives saved and the price that would have to be paid for any screening benefit.

Five years after the publication of the primary trial results, there remains a vigorous debate about whether and how best to screen for prostate cancer. The randomized trial results were the basis for revised prostate screening recommendations from all of the major policy panels including the US Preventive Services Task Force (USPSTF),[2] the American Cancer Society,[3] and the American Urology Association.[4] Whereas the USPSTF has

Supported by U01 CA157224 (Cancer Intervention and Surveillance Modeling Network from the National Cancer Institute), P30 CA015704-36 (Cancer Center Support Grant from the National Cancer Institute) and U01 CA86402 (from the Early Detection Research Network of the National Cancer Institute).

[a] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M2-B230, PO Box 19024, Seattle, WA 98109-1024, USA; [b] The Cancer Therapy and Research Center, University of Texas Health Science Center at San Antonio, MC-7845, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, USA

* Corresponding author.
E-mail address: retzioni@fhcrc.org

Urol Clin N Am 41 (2014) 223–228
http://dx.doi.org/10.1016/j.ucl.2014.01.002

recommended against PSA screening at all ages, the other panels have generally recommended shared decision making except for men with a limited life expectancy.

This article reexamines the trials and their findings in light of what needs to be known to develop policies for population screening. The authors first review the empiric results from the trials and ask what they inform us about (1) screening benefit, (2) screening harms, particularly overdiagnosis, and (3) the harm-benefit trade-offs of screening. Statistical and modeling analyses that go beyond the trial results are considered, and how these results may modify perceptions of the aforementioned outcomes is discussed. All screening outcomes depend on the screening strategy used, including the screening ages, intervals, and cutoffs for biopsy referral. Varying these parameters can dramatically alter the balance of harm and benefit; unfortunately, the 2 randomized trials are inherently limited in their ability to compare alternative screening strategies. The authors conclude that screening trials in general, and the Prostate, Lung, Colon, and Ovarian (PLCO) trial and European Randomized Study of Screening for Prostate Cancer (ERSPC) in particular, provide information that is critical for the development of screening policy, but cannot provide all the information needed for developing sound policies for population screening.

THE LARGE RANDOMIZED PROSTATE CANCER SCREENING TRIALS

The 2 large screening trials, the United States–based PLCO cancer screening trial[5,6] and the ERSPC,[7,8] have been previously described in detail.

Several measures of screening benefit and harm are presented in the trial reports, and these are briefly reviewed here, as the manner by which harm and benefit are measured will significantly influence the perception of the value of screening. Several definitions are important in understanding screening outcomes. The relative screening benefit is expressed by the (prostate cancer) mortality rate ratio, which is the ratio of the risk of death from prostate cancer in the screened group relative to the control group over the follow-up period. The absolute screening benefit is expressed by the difference between the cumulative incidence of death from prostate cancer in the 2 groups, and may be thought of as an estimate of the lives saved by screening over the follow-up period. Both relative and absolute benefits are time-sensitive and generally increase with follow-up time.[9–11] Overdiagnosis is the detection by screening of cases that, in the absence of screening, would not have caused morbidity or mortality in the patient's lifetime. Overdiagnosis may be expressed as an absolute number of overdiagnosed cases, as a fraction of the number screened, or as a fraction of screen-detected cases. Depending on how overdiagnosis is estimated, the results may also be highly time-sensitive. Finally, a measure of harm-benefit trade-off that has become fairly standard is the (additional) number needed to detect (NND) to prevent 1 death from prostate cancer, defined as the estimated overdiagnoses divided by the estimated lives saved by screening. The NND has been referred to as the additional number needed to treat to prevent 1 death from prostate cancer but this is not, strictly speaking, accurate, because not all newly diagnosed prostate cancers receive immediate treatment. The concept of the NND, a harm-benefit trade-off measure pertaining specifically to screening that carries the possibility of overdiagnosis, should be distinguished from the similarly named number needed to treat or NNT, which is a concept of benefit most commonly used in analysis of treatment trials.

The PLCO Screening Trial

The PLCO trial randomized 76,693 men to screening or a control group managed according to community standards. Screening-arm participants were given annual PSA tests for 6 years with concomitant digital rectal examinations (DREs) for the first 4 years. Diagnostic follow-up for positive test results was left to participants who were referred to their doctors for PSA higher than 4.0 ng/mL or a suspicious finding on DRE. Approximately 40% of participants referred to biopsy for an abnormal screening test underwent prostate biopsy within 1 year.[12]

By the time the PLCO trial began randomizing participants, PSA screening was widespread.[1] This aspect had a critical impact on the trial and its outcomes. In brief, 45% of participants had had at least 1 PSA before enrollment[6]; moreover, over the course of the trial approximately half of the control-arm participants were screened every year, with 74% of the control group receiving at least 1 screening test during their participation in the trial.[13] By contrast, 95% of the screened group was screened at least once during the course of the trial. The average number of screening tests was 5 in the screened group and 2.7 in the control group.[13] Thus, screening in the control group was approximately half as intensive as that in the screened group.

The empirical results from the PLCO after 11 and 13 years of follow-up clearly show no relative