Published in final edited form as:

Semin Oncol. 2010 June ; 37(3): 216–223. doi:10.1053/j.seminoncol.2010.05.009.

Cancer Screening Trials: Nuts and Bolts

Philip C. Prorok^{1,2} and Pamela M. Marcus¹

¹National Cancer Institute, Division of Cancer Prevention

Abstract

The most rigorous and valid approach to evaluating cancer screening modalities is the randomized controlled trial, or RCT. RCTs are major undertakings and the intricacies of trial design, operations, and management are generally under appreciated by the typical researcher. The purpose of this chapter is to inform the reader of the “nuts and bolts” of designing and conducting cancer screening RCTs. Following a brief introduction as to why RCTs are critical in evaluating screening modalities, we discuss design considerations, including the choice of design type and duration of follow-up. We next present an approach to sample-size calculations. We then discuss aspects of trial implementation, including recruitment, randomization, and data management. A discussion of commonly employed data analyses comes next, and includes methods for the primary analysis, comparison between the screened and control arms of cause-specific mortality rates for the cancer of interest, as well as for secondary endpoints such as sensitivity. We follow with a discussion of sequential monitoring and interim analysis techniques, which are used to examine the primary outcome while the trial is ongoing. We close with thoughts on lessons learned from past cancer screening RCTs and provide recommendations for future trials. Throughout the chapter we illustrate topics with examples from completed or on-going RCTs, including the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial and the National Lung Screening Trial (NLST).

Keywords

early detection; cancer screening; randomized trial; PLCO; NLST

Introduction

Screening for the early detection of cancer is regarded by many as an obvious intervention strategy, one that will help alleviate the burden of cancer in populations. What is not sufficiently recognized, however, is that cancer screening interventions are not automatically beneficial, that harms are associated with screening, and that it is important to carefully and rigorously evaluate a screening exam before it is introduced into a population, to ensure that benefits outweigh the harms. It is the purpose of this chapter to briefly discuss aspects of the most rigorous and valid approach to screening evaluation, the randomized controlled trial (RCT), by describing issues in the design, implementation, and analysis of such studies. Examples are provided.

²Corresponding Author: Philip C. Prorok, Ph.D., Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, 6130 Executive Blvd., EPN Room 3132, Bethesda, MD 20892, Telephone: 301-496-7709, pp2g@nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Design Issues

Several designs have been used for cancer screening RCTs. Most studies have employed the traditional two-arm design [1–24], which aims to determine whether the screening intervention results in benefit, that is, a reduction in cause-specific mortality. In the two-arm design, participants in one arm receive the screening exam of interest while those in the other arm serve as controls, receiving either no screening exam as part of the trial or an exam routinely used in the population as a screening modality. Similar designs have been used to address questions about screening frequency [25], age-specific effectiveness [26], and the effect of adding one screening modality to another [27–31]. One trial, the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, is testing multiple regimens by using two study arms: participants in intervention arm received screening exams for three cancers (all participants received lung (chest x-ray) and colorectal (flexible sigmoidoscopy) cancer screens, while women also received ovarian cancer screens (transvaginal ultrasound and CA-125 evaluation) and men also received prostate cancer screens (digital rectal exam and PSA level evaluation), while participants in the control arm received no trial exams [32]. PLCO provides an excellent example of many of the issues confronted when designing and conducting an RCT [33].

A major design issue for PLCO involved deciding whether to conduct one trial or four separate trials, one for each cancer site. A comparison of costs and logistics revealed that evaluation of screening modalities for the four cancers using one trial was less costly and more efficient, as one administrative structure and coordinating center could be used. Another design issue involved the evaluation of multiple tests for a single cancer site. Rather than evaluating each test for the same cancer site with a different arm, it was decided that combinations of screening tests would be evaluated. So, for example, DRE and PSA testing were administered together at each screening round, rather than conducting a separate trial for each. The primary reasons for combining modalities were cost constraints and the desire to evaluate the regimen of combined interventions. If a combined regimen does not work, testing the individual procedures would not be warranted. If the combined regimen does work, each test could be independently evaluated in separate RCTs.

Several designs were considered for PLCO. The two primary competitors were the reciprocal control design and the all-versus-none design. The reciprocal control design would have had three arms: one devoted to screening for prostate or ovarian cancer, one to colorectal cancer screening, and one to lung cancer screening. Since screening would have been undertaken for only one cancer site (per gender in the case of the prostate/ovarian arm) in any given arm, participants in the other two arms would have served as controls. This design was ultimately deemed unfeasible because of the cost of screening all participants. Furthermore, the reciprocal control design was expected to result in substantial levels of contamination, as all participants, due to the fact that they were being screened, would be aware of the other trial-administered screening tests, which they might then seek out. The all-versus-none design, with participants randomized to one of two arms, thus was chosen. In the spirit of a multiphasic screening endeavor, one arm served as a control, while screens for all cancer sites of interest were administered in the other arm. Use of the all-versus-none design reasonably assumes that the screening tests for each cancer do not detect any of the other cancers of interest, and that the endpoints—death from each of the four cancers—are not related. It was further decided to employ the so-called “stop screen” approach, an approach in which screening is performed for a fixed number of years or screening rounds and then stopped, but follow-up continues for endpoint ascertainment [34]. This approach was chosen because it had been used successfully in breast and colorectal cancer screening trials, and because it is the only design that allows direct assessment of overdiagnosis. Overdiagnosis is the identification through screening of cancers that never would have

surfaced clinically in the absence of screening. This phenomenon has been observed repeatedly [35–40] and must now be considered the rule rather than the exception in cancer screening.

In addition to allowing for assessment of overdiagnosis, a stop-screen design is often necessary because resources typically are not available to screen throughout the life of the study. Therefore, the number of screening rounds also must be decided upon at the design stage in this type of trial. However, changes can be and often are made as the study progresses. In PLCO, the initial regimen of four annual screens for PSA and CA-125 was later expanded to six screens, and was a trade-off between the expected number of screens necessary to produce an effect (should one exist) versus available and anticipated resources. In the early years of PLCO, sigmoidoscopy was administered at the first and fourth annual screens, although the annual fourth screen eventually was replaced with a screen at the sixth annual visit to reflect emerging clinical practice. An annual interval between screens is typically chosen in RCTs, as it is the interval most likely to be used once mass screening programs are established in the community. Compared to less-frequent screening, an annual interval increases the likelihood of detection of a broad spectrum of preclinical conditions, thus providing a better representation of the natural history of the cancers under study. A longer interval is less desirable in most instances, as it might allow some rapidly growing lesions, lesions likely to be lethal but which could be cured if found early, to escape detection. An exception to the use of an annual interval is found in endoscopic screening for colorectal cancer: our current understanding of colorectal cancer progression resulted in a five-year screening interval in PLCO.

Duration of follow-up, the time from randomization to cessation of event ascertainment, is another important design parameter. In PLCO, a minimum of 10 years of follow-up was initially chosen to allow for sufficient time for mortality reductions, should any exist, to emerge. Although follow-up intervals of at least 7 years were typical in breast cancer screening trials [1,41], it was assumed that the longer natural history of prostate cancer, and perhaps other cancers under study, warranted a longer follow-up period. In the National Lung Screening Trial (NLST), modeling of the disease and screening processes resulted in the decision to capture endpoint events over an approximately 7 year period [19]. It must be recognized that these and other design parameters were chosen using the best information available at the time of design and may be subject to change as a result of data gathered during the trial and other information. In the Minnesota FOBT trial [25], the screening and follow-up periods chosen at the beginning of the trial were ultimately both extended, resulting in the opportunity for a mortality reduction to emerge.

Sample-Size Calculation

A crucial element of trial design is calculation of the necessary number of participants, or sample size. Calculation of sample size involves identification and estimation of key design parameters, including hypothesized mortality reduction, expected compliance and contamination, number of screening rounds and length of follow-up. The approach used for PLCO [33] provides an example of how sample size can be calculated for a cancer screening RCT. For a trial designed to detect a $(1-r) \times 100\%$ reduction ($0 < r < 1$) in the cumulative disease-specific death rate over its duration, the following are defined:

- N_C is the number of participants randomized to the control arm and N_S is the number randomized to the screened arm, with $N_S = f N_C$;
- P_C is the proportion of participants in the control arm who comply with the control arm protocol and P_S is the proportion of individuals in the screened group who comply with the screening arm protocol;

- Y is the duration of the trial from entry to end of follow-up, expressed in years;
- R_C is the average annual disease-specific mortality rate in the control arm, expressed in deaths per person per year;
- α is the type I error rate (the probability of finding an effect given that one truly does not exist);
- β is the type II error rate (the probability of not finding an effect given that one truly does exist);
- Q_C equals $r + (1-r)P_C$;
- Q_S equals $1 - (1-r)P_S$.

The total number of disease-specific deaths needed for a one-sided α -level significance test with power $1-\beta$ is given by:

$$D = \frac{\{(Q_C + fQ_S) Z_{1-\alpha} - \sqrt{Q_C Q_S} (1+f) Z_\beta\}^2}{f (Q_C - Q_S)^2}.$$

The number of participants in the control arm is given by:

$$N_c = \frac{D}{(Q_C + fQ_S) R_C Y}.$$

It is important that those design parameters that need to be estimated are done so as accurately as possible to ensure the validity of the trial. Such estimation can emanate from previous studies, special investigations, modeling, or data from cancer registries [19,33,42]. In situations where the screening test is available in the community, it is crucial to estimate the likely compliance in both arms. In addition, the underlying event rates in the control arm must account for the likelihood of substantial healthy volunteer bias [43,44]. Adjustment for this bias when calculating expected event rates is particularly important as failing to do so will likely result in an undersized and therefore statistically underpowered trial [45].

Trial Implementation

Recruitment

Recruitment to cancer screening RCTs is often the most challenging task in such endeavors, as these trials need to enroll tens of thousands of participants during a relatively short time period. The process is simplified when population registers exist (such as in some European countries) as individuals can be recruited from these lists. Such lists generally do not exist in the US, so other approaches are needed. In recent large US trials, the most successful method has been direct mail. With direct mail, rosters of names and addresses from for-profit, not-for-profit, and government organizations are acquired and recruitment materials are mailed to individuals on the rosters. This method was used in the PLCO trial, with a 1% enrollment yield [46]. Other strategies are often necessary to recruit persons from minority populations. Community outreach, which involves the support of community-based organizations such as churches to lend credibility to a study, was used successfully in PLCO [47].

Randomization

Individuals who meet the eligibility criteria are randomized to a trial arm. A formal randomization scheme should be used, and may involve stratification by key risk factors. Random assignment is best accomplished using computerized systems as opposed to sealed envelopes, as a computerized system helps to ensure blinded allocation. In PLCO, randomization software and encrypted files were loaded on desktop microcomputers at each screening center. As each participant was successfully randomized into the trial, his or her name, gender, date of birth, and assigned arm were automatically stored in encrypted data tables so data could not be altered. At the same time, a second protected set of synchronized tables, stored on a backup device, was also updated.

Data Management

Cancer screening RCTs are typically large, long term projects that involve collection of vast amounts of data over many years. An automated data management system should be established at the outset, recognizing that hardware and software will need to be updated periodically. In PLCO, the data management system allows for the following: the ability for the trial leadership and the Coordinating Center (CC) to access screening center systems and data remotely, synchronization of databases on multiple platforms, preparation of high quality analysis datasets, secure backup and archiving of data, and robust configuration management [48]. Data are entered using a distributed data entry system and are transmitted regularly to the CC.

In PLCO, numerous forms were developed for collection of crucial information. These forms included eligibility and consent forms, baseline and mid-trial questionnaires that sought demographic and risk factor information, examination forms that documented the outcome of trial screening exams, diagnostic evaluation and treatment forms, and questionnaires for annual follow-up of participants. Additional forms were developed as data collection needs grew during the course of the trial. Scannable data collection forms were used to prevent typing errors, and quality control checks were built into systems as well to ensure data integrity [33,48]. With the increasing availability of electronic medical records systems in the US, future trials may have the opportunity to transmit this type of data directly into study systems.

Data elements that should be collected include:

- participant identification number;
- participant demographic and risk factor information, including date of birth;
- participant randomization assignment and date of trial entry;
- date and result of each screening test;
- diagnostic procedures performed as a result of a positive or suspicious screening test, to allow determination of whether a cancer was screen- detected;
- detailed physical findings and any complications or morbid events possibly associated with each screening test;
- date of diagnosis, histology, stage at diagnosis, and initial therapy for every cancer of interest diagnosed during the trial in every randomized arm;
- date and cause of death as noted on the death certificate, for every death which occurs during the trial in every randomized arm.

Pilot Phase

A pilot phase can be very informative when conducted prior to full implementation of large screening trials. A pilot phase provides the opportunity to test trial procedures and assess design assumptions. Pilot phases were conducted with great success in both the PLCO and NLST trials [19,33]. Long-term commitments to these trials were based on the success of these pilot phase studies, as they demonstrated that each trial was in fact feasible.

In PLCO, specific activities were planned for the pilot phase, which amounted to the first two years of the trial. Major activities involved refinement of protocol components including eligibility criteria, development of mechanisms for notifying participants of screening results and encouraging them to seek diagnostic evaluation for suspicious or positive results, collection of data on diagnostic work-up of participants with suspicious or positive screens, development of procedures for quality control of screening examinations, and development of follow-up procedures to determine cancer incidence and ascertain cause of death. All 10 PLCO screening centers participated in the pilot phase, allowing each screening center (SC) to gain invaluable experience. Furthermore, the pilot phase allowed study leaders to identify procedures that worked, and those that did not.

The major activities carried out by all PLCO SCs during the pilot phase were as follows:

- Test and evaluate recruitment and randomization procedures by enrolling participants. The ten SCs enrolled 12,000 participants during the pilot phase;
- Determine the logistics of performing screening exams and reporting exam results by performing the screening examinations on the pilot-phase participants randomized to the screened arm;
- Assess contamination (use of screening modalities outside the context of the trial) among randomized participants;
- Assess compliance for each screening modality;
- Test all data forms, collection procedures, and computer systems;
- Establish procedures to collect, ship, and analyze blood for PSA and CA125 assays;
- Establish procedures to collect and ship blood and tissue samples to the biorepository.

During PLCO's pilot phase, each SC identified sources from which to recruit participants and implemented recruitment strategies appropriate for its particular population. Trial investigators met to discuss how the trial was progressing and any need for procedural changes. The CC developed trial forms, established data entry and editing systems, and prepared a manual of operations and procedures. Efforts were made to standardize all screening examinations to the extent possible in a study that had multiple screening centers. In addition, performance at each SC was monitored and any deficiencies corrected so as to maintain adequate recruitment and ensure correct administration of activities during the main phase of the trial. Participants randomized during the pilot phase were treated as a vanguard group and included in the trial.

NLST's pilot phase was called the Lung Screening Study (LSS) Feasibility Phase [49,50]. It was conducted to address assertions by critics that it would be impossible to conduct an RCT of helical computed tomography (CT) screening for lung cancer because high risk individuals would not agree to be randomized. This study did in fact randomize an ample number of participants: 3318 within a two and a half month period at six PLCO screening centers. Participants were randomized to one of two screening modalities: helical CT or

single view chest X-ray. Screens were performed at baseline and one year later. The LSS Feasibility Phase demonstrated convincingly that an RCT of helical CT was feasible. In addition to meeting accrual targets, crossover contamination (in particular, use of screening CT by participants randomized to the chest X-ray arm) was measured and found to be low. The experience from the LSS Feasibility Phase yielded valuable input for the design and conduct of NLST. Unlike PLCO, participants from the LSS Feasibility Phase were not included in NLST.

Data Analysis

A data analysis plan should be prepared at the beginning of any screening trial. Of course, this plan is likely to be augmented as new secondary questions and analytic methods arise during the trial. Trial data are analyzed on an ongoing basis guided by the analysis plan, with more detailed analyses typically occurring for special meetings of trial investigators or the trial's Data and Safety Monitoring Board (DSMB). Commonly used analytical methods include standard descriptive statistics and techniques such as regression, analysis of rates and proportions, contingency table methods, and analysis of survival data. New methods of analysis or modeling are devised as needed.

Typical data analyses include description of characteristics of the enrolled population and comparison of those characteristics across trial arms; estimation of screening test operating characteristics including sensitivity, specificity, predictive value, and false positive rate; calculation of prevalence and incidence of the cancer of interest; description of cancer characteristics including stage, histology, treatment, survival, and method of detection (screen-detected versus interval); identification of prognostic variables; estimation and modeling of lead time; calculation of potential surrogate endpoints such as incidence rates of interval cancers and advanced stage disease; enumeration of complications resulting from screening exams or diagnostic evaluation; and calculation of cause-specific and all-cause mortality. Other so-called "ancillary studies" could investigate other topics of interest, such as quality of life or the cost-benefit ratio of the screening process.

The primary and most important analysis in any cancer screening trial is the comparison of cause-specific mortality rates for the cancer of interest between the screened and control arms on an intent-to-screen basis. In addition, the death rates from other causes and all-cause mortality should be scrutinized to assess the comparability of the randomized populations. Mortality rates are compared using Poisson methods or the log-rank test [51]. These methods assume that mortality rates in the screened and control arms are proportional over time; however, this may not always hold. In such an instance, other statistical methods are used [52].

Secondary variables worthy of special mention include sensitivity, therapy, and harms. Sensitivity is commonly calculated as the ratio of screen-detected cancers to all cancers found in the screened arm for some interval after a screen, typically one year. This approach, however, is biased and will yield an inflated estimate when overdiagnosed cancers are present among the screen-detected cancers. A correct method is calculated as one minus the interval cancer incidence divided by the control arm incidence, and can be adjusted for non-compliance in the screened arm if it is believed that screening attendance was correlated with the risk of disease [53]. With regard to therapy, data should be collected on treatment received for persons diagnosed with the cancer of interest in both the screened and control arms. These data then can be examined within each stage of disease to determine whether therapy is comparable between randomized groups, thereby assessing the possibility of a confounding effect of therapy on the impact of the screening.

Harms of screening include overdiagnosis, false positive screening exams, complications of the screening procedures administered to trial participants, and complications of diagnostic and treatment procedures administered to trial participants as a result of positive screens and screen-detected cancer, respectively. Complications include any adverse medical event and any deaths potentially related to or occurring as a downstream consequence of trial procedures. False positive exams are quite common, as most individuals with a positive screening test will not be diagnosed with cancer after diagnostic workup. False positive rates therefore can be substantial [54]. Cancer incidence is tracked after the cessation of screening to alert investigators to the possibility of substantial overdiagnosis, as evidenced by a persistent excess of cases in the screened arm relative to the control arm.

Sequential Monitoring and Interim Analysis

A process for regular, formal monitoring of safety issues and evaluation of accumulating data should be established early in the course of a cancer screening trial. This is best accomplished by the creation of a DSMB, a committee comprised of experts not associated with the trial who therefore can provide an independent assessment of trial progress. DSMBs were not employed in many early cancer screening trials, but more recent trials have included such a committee [19,33,55,56].

Sequential monitoring is an integral part of the PLCO and NLST trials [19,33]. In both trials, statistical monitoring guidelines were established for the DSMB to use when examining emerging data. The accruing mortality data and secondary endpoints are examined at least annually to determine if and when a protocol change is warranted, one that would result in an early decision about the efficacy of the screening intervention. The data are analyzed from two perspectives, one addressing the prospect of early termination due to a significantly large effect, and one addressing the prospect of early termination due to a negligible effect. For PLCO, this is done separately for the four cancers of interest.

Statistical group sequential procedures are used in PLCO [33]. To assess whether early termination due to a large effect should occur, the O'Brien-Fleming boundary [57] and Lan-DeMets sequential procedure are used [58]. The test statistic in PLCO is a weighted log-rank statistic with weights linear in and proportional to the cumulative mortality, where the weights are estimated from the empirical survival function based on the data from both arms [52]. Choice of this combination of boundary and weights was based on power computations conducted using simulation methods. For early termination due to a negligible effect, a stochastic curtailment procedure is used [59]. A similar approach is used for NLST [19].

A weighted rank statistic was chosen for these and other recent screening and prevention trials [24,60] to account for anticipated non-proportional cause-specific hazard rates between the trial arms. Non-proportionality can arise early in the trial due to a lag time before the screening benefit, should one exist, emerges. A weighted log-rank statistic down-weights early events, thus placing more emphasis on later events, but preserves the randomized, intent-to-screen comparison. Another potential source of non-proportionality is dilution of the screening effect due to inclusion in the mortality comparison of deaths among cancers that are diagnosed, in both arms, after screening stops in the intervention arm. Methods to account for dilution have been proposed [61,62] but further research is warranted.

Lessons Learned and Recommendations for Future Trials

Much has been learned from completed and ongoing cancer screening trials. Hopefully this knowledge will lead to improved trial design and analysis. The following is a list of thoughts and recommendations for future endeavors:

- New screening tests can rapidly become widely used, especially in the U.S., often with neither valid scientific evidence of benefit nor proper assessment of harm. It is therefore important, when a new test becomes available, to undertake rigorous trials as soon as possible before widespread community use precludes establishment of a proper control arm;
- It is possible to evaluate several screening tests and cancers in the same trial to save resources, as in PLCO;
- Sample size calculations should account for non-compliance, contamination, and the healthy volunteer effect. Models, including simulations, can be a valuable adjunct;
- In trials, adverse events and harms associated with screening can be identified and quantitatively measured just as can benefits;
- Overdiagnosis has been observed repeatedly in cancer screening. It should be expected and accounted for in study design, analysis, and interpretation;
- A pilot phase prior to or at the beginning of a trial can be extremely valuable for testing operational components and evaluating study centers;
- Treatment data for all diagnosed cancers of interest should be carefully collected and reported by stage for each trial arm to examine possible impacts of differential treatment on trial outcomes;
- Quality assurance of all trial operations is crucial;
- A DSMB should be established early in the trial. In addition to using sequential stopping rules to determine if early reporting of results is warranted, the DSMB can monitor compliance, contamination, and other design parameters to determine if protocol changes are necessary;
- Analysis methods such as weighted log-rank tests can be used to account for varying patterns of screening effect over time. Models of screening can be improved and validated using good quality trial data, including mortality endpoints.

A cancer screening RCT is a major endeavor, requiring a long-term commitment by participants, investigators and funding organizations. If a decision is made to conduct a trial, necessary resources must be committed for the duration of the study. To accomplish this in the today's climate of intense resource competition is difficult at best. One strategy is full commitment sequentially to the primary phases – pilot, recruitment, screening, and follow-up - with funding for each successive phase contingent on successful completion of the previous phase. Regardless of these difficulties, researchers should strive to establish such trials when a new screening modality appears promising, as the alternative, community acceptance of an unproven modality, could result in wasted resources and an overall detriment to public health.

REFERENCES

1. Shapiro, S.; Venet, W.; Strax, P., et al. Periodic Screening for Breast Cancer- the Health Insurance Plan Project and its Sequelae, 1963–1986. Baltimore, MD: Johns Hopkins University Press; 1988.
2. Shapiro S, Venet W, Strax P, et al. Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 1982;69:349–355. [PubMed: 6955542]
3. Tabar L, Fagerberg CJ, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography: randomized trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1985;1:829–832. [PubMed: 2858707]

4. Tabar L, Fagerberg G, Duffy SW, et al. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* 1992;30:187–210. [PubMed: 1732926]
5. Tabar L, Faberberg G, Day NE, et al. What is the optimum interval between mammographic screening examinations? – an analysis based on the latest results of the Swedish two-county breast cancer screening trial. *Br J Cancer* 1987;55:547–551. [PubMed: 3606947]
6. Nystrom L, Rutqvist LE, Wall S, et al. Breast cancer screening with mammography: overview of Swedish randomized trials. *Lancet* 1993;341:973–978. [PubMed: 8096941]
7. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *Br Med J* 1988;297:943–948. [PubMed: 3142562]
8. Friskell J, Eklund G, Hellstrom L, et al. Randomized study of mammography screening – preliminary report on mortality in the Stockholm trial. *Breast Cancer Res Treat* 1991;18:49–56. [PubMed: 1854979]
9. Miller AB, Howe GR, Wall C. The national study of breast cancer screening: protocol for a Canadian randomized controlled trial of screening for breast cancer in women. *Clin Invest Med* 1981;4:227–258. [PubMed: 6802546]
10. Miller AB, Baines CJ, To T, et al. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Can Med Assoc J* 1992;147:1477–1488. [PubMed: 1423088]
11. Kewenter J, Bjork S, Haglund E, et al. Screening and re-screening for colorectal cancer: a controlled trial of fecal occult blood testing in 27,700 subjects. *Cancer* 1988;62:645–651. [PubMed: 3292038]
12. Hardcastle JD, Thomas WM, Chamberlain J, et al. Randomized, controlled trial of faecal occult blood screening for colorectal cancer: results for first 107,349 subjects. *Lancet* 1989;1:1160–1164. [PubMed: 2566735]
13. Kronborg O, Fenger C, Snergaard O, et al. Initial mass screening for colorectal cancer with fecal occult blood test. *Scand J Gastroenterol* 1987;22:677–686. [PubMed: 3659830]
14. Kronborg O, Fenger C, Olsen J, et al. Repeated screening for colorectal cancer with fecal occult blood test: a prospective randomized study at Funen, Denmark. *Scand J Gastroenterol* 1989;24:599–606. [PubMed: 2762760]
15. UK Flexible Sigmoidoscopy Screening Trial Investigators. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomized trial. *Lancet* 2002;359:1291–1300. [PubMed: 11965274]
16. Hoff G, Grotmol T, Skovlund, Bretthauer M. for the Norwegian Colorectal Cancer Prevention Study Group. *Br Med J* 2009;338:1363–1366.
17. Thiis-Evensen E, Hoff GS, Sauar J, Langmark F, Majak BM, Vatn MH. Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer. *Scand J Gastroenterol* 1999;34:414–420. [PubMed: 10365903]
18. Fontana RS, Sanderson DR, Woolner LB, et al. Screening for lung cancer: a critique of the Mayo Lung Project. *Cancer* 1991;67 Suppl(4):1155–1164. [PubMed: 1991274]
19. The National Lung Screening Trial Research Team. The National Lung Screening Trial: overview and study design. *Radiology*. submitted.
20. Van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection for the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomized lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007;120:868–874. [PubMed: 17131307]
21. Infante M, Cavuto S, Lutman FR, et al. A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial. *Am J Respir Crit Care Med* 2009;180:445–453. [PubMed: 19520905]
22. Pedersen JH, AShrif H, Dirksen A, et al. The Danish randomized lung cancer CT screening trial – overall design and results of the prevalence round. *J Thorac Oncol* 2009;4:608–614. [PubMed: 19357536]

23. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer* 2009;64:34–40. [PubMed: 18723240]
24. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009;360:1320–1328. [PubMed: 19297566]
25. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* 1993;328:1365–1371. [PubMed: 8474513]
26. Roberts MM, Alexander FE, Anderson TJ, et al. al. Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet* 1990;335:241–246. [PubMed: 1967717]
27. Miller AB, Baines CJ, To T, et al. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Can Med Assoc J* 1992;147:1477–1488. [PubMed: 1423088]
28. Doria-Rose VP, Marcus PM, Szabo E, Tockman MS, Melamed MR, Prorok PC. Randomized controlled trials of the efficacy of lung cancer screening by sputum cytology revisited: a combined mortality analysis from the Johns Hopkins Lung Project and the Memorial Sloan-Kettering Lung Study. *Cancer* 2009;115:5007–5017. [PubMed: 19637354]
29. Berry DP, Clarke P, Hardcastle JD, Vellacott KD. Randomized trial of the addition of flexible sigmoidoscopy to faecal occult blood testing for colorectal neoplasia population screening. *Br J Surg* 1997;84:1274–1276. [PubMed: 9313712]
30. Rasmussen M, Kronborg O, Fenger C, Jorgensen OD. Possible advantages and drawbacks of adding flexible sigmoidoscopy to hemoccult-II in screening for colorectal cancer. A randomized study. *Scand J Gastroenterol* 1999;34:73–78. [PubMed: 10048736]
31. Verne JE, Aubrey R, Love SB, Talbot IC, Northover JM. Population based randomized study of uptake and yield of screening by flexible sigmoidoscopy compared with screening by faecal occult blood testing. *BMJ* 1998;34:182–185. [PubMed: 9665902]
32. Gohagan JK, Prorok PC, Hayes RB, Kramer BS. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: History, Organization, and Status. *Control Clin Trials* 2000;21:251S–272S. [PubMed: 11189683]
33. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 2000;21:273S–309S. [PubMed: 11189684]
34. Etzioni RD, Connor RJ, Prorok PC, Self SG. Design and analysis of cancer screening trials. *Stat Methods Med Res* 1995;4:3–17. [PubMed: 7613636]
35. Jorgensen KJ, Gotzsche PC. Overdiagnosis in publicly organized mammography screening programmes: systematic review of incidence trials. *BMJ* 2009;339:206–209.
36. Esteve J, Parker L, Roy P, et al. Is neuroblastoma screening evaluation needed and feasible? *Brit J Cancer* 1995;71:1125–1131. [PubMed: 7779698]
37. Marcus PM, Bergstralh EJ, Fagerstrom RM, Williams DE, Fontana R, Taylor WF, Prorok PC. Lung cancer mortality in the Mayo Lung Project: Impact of extended follow-up. *J Natl Cancer Inst* 2000;92:1308–1316. [PubMed: 10944552]
38. Melnikow J, Nuovo J, Willan AR, Chan BKS, Howell LP. Natural history of cervical squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol* 1998;92:727–735. [PubMed: 9764690]
39. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986–2005. *J Natl Cancer Inst* 2009;101:1325–1329. [PubMed: 19720969]
40. Miller AB. Conundrums in screening for cancer. *Int J Cancer* 2009;126:1039–1046. [PubMed: 19960430]
41. Tabar L, Fagerberg G, Duffy SW, et al. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiologic Clin North Am* 1992;30:187–210.
42. De Koning HJ, Liem MK, Baan CA, Boer R, Schroder FH, Alexander FE. on behalf of ECRSPC. Prostate cancer mortality reduction by screening: power and time frame with complete enrollment in the European Randomised Screening for Prostate Cancer (ERSPC) Trial. *Int J Cancer* 2002;98:268–273. [PubMed: 11857418]

43. Pinsky PF, Miller A, Kramer BS, Church T, Reding D, Prorok P, Gelmann E, Schoen RE, Buys S, Hayes RB, Berg CD. Evidence of a healthy volunteer effect in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Am J Epidemiol* 2007;165:874–881. [PubMed: 17244633]
44. Church TR, Ederer F, Mandel JS, Watt GD, Geisser MS. Estimating the duration of ongoing prevention trials. *Am J Epidemiol* 1993;137:797–810. [PubMed: 8484371]
45. Ederer F, Church TR, Mandel JS. Sample sizes for prevention trials have been too small. *Am J Epidemiol* 1993;137:787–796. [PubMed: 8484370]
46. Gren L, Broski K, Childs J, et al. Recruitment methods in the PLCO Cancer Screening Trial. *Clin Trials* 2009;6:52–59. [PubMed: 19254935]
47. Pinsky PF, Ford M, Gamito E, et al. Enrollment of racial and ethnic minorities in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *J Natl Med Assoc* 3:291–298. 100.
48. Hasson MA, Fagerstrom RM, Kahane DC, et al. Design and evolution of the data management systems in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 2000;21:329S–348S. [PubMed: 11189686]
49. Gohagan J, Marcus P, Fagerstrom R, Pinsky P, Kramer B, Prorok P. Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs. chest radiography. *Chest* 2004;126:114–121. [PubMed: 15249451]
50. Gohagan JK, Marcus PM, Fagerstrom RM, et al. Final results of the Lung Screening study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. *Lung Cancer* 2005;47:9–15. [PubMed: 15603850]
51. Lawless, JF. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons; 1982.
52. Fleming, TR.; Harrington, DP. *Counting Processes and Survival Analysis*. New York: John Wiley & Sons; 1991.
53. Hakama M, Auvinen A, Day NE, Miller AB. Sensitivity in cancer screening. *J Med Screen* 2007;14:174–177. [PubMed: 18078561]
54. Croswell JM, Kramer BS, Kreimer AR, et al. Cumulative incidence of false-positive results in repeated, multimodal cancer screening. *Ann Fam Med* 2009;7:212–222. [PubMed: 19433838]
55. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med* 1993;328:1365–1371. [PubMed: 8474513]
56. Smith PH. The Data Monitoring Committee – bridging the gap between urology and public health epidemiology. *Br J Urol* 2003;92 suppl2:55–56.
57. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979;35:549–556. [PubMed: 497341]
58. Lan KKG, DeMets DL. Group sequential procedures: calendar versus information time. *Stat Med* 1989;8:1191–1198. [PubMed: 2814068]
59. Lin DY, Yao Q, Ying Z. A general theory on stochastic curtailment for censored survival data. *J Am Stat Assoc* 1999;94:510–521.
60. The Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Control Clin Trials* 1998;19:51–109.
61. Baker SG, Kramer BS, Prorok PC. Early reporting for cancer screening trials. *J Med Screen* 2008;15:122–129. [PubMed: 18927094]
62. Baker SG, Kramer BS, Prorok PC. Statistical issues in randomized trials of cancer screening. *BMC Medical Research Methodology* 2002;2:11. doi:10.1186/1471-2288-2-11. [PubMed: 12238954]