



Published in final edited form as:

*J Med Screen*. 2008 ; 15(3): 122–129. doi:10.1258/jms.2008.007058.

## Early reporting for cancer screening trials

**Stuart G Baker,**

*Stuart G Baker, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA*

**Barnett S Kramer,** and

*Barnett S Kramer, Office of Disease Prevention, National Institutes of Health, Bethesda, MD, USA*

**Philip C Prorok**

*Philip C Prorok, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA*

### Abstract

**Objective**—Many cancer screening trials involve a screening programme of one or more screenings with follow-up after the last screening. Usually a maximum follow-up time is selected in advance. However, during the follow-up period there is an opportunity to report the results of the trial sooner than planned. Early reporting of results from a randomized screening trial is important because obtaining a valid result sooner translates into health benefits reaching the general population sooner. The health benefits are reduction in cancer deaths if screening is found to be beneficial and more screening is recommended, or avoidance of unnecessary biopsies, work-ups and morbidity if screening is not found to be beneficial and the rate of screening drops.

**Methods**—Our proposed method for deciding if results from a cancer screening trial should be reported earlier in the follow-up period is based on considerations involving postscreening noise. Postscreening noise (sometimes called dilution) refers to cancer deaths in the follow-up period that could not have been prevented by screening: (1) cancer deaths in the screened group that occurred after the last screening in subjects whose cancers were not detected during the screening program and (2) cancer deaths in the control group that occurred after the time of the last screening and whose cancers would not have been detected during the screening programme had they been randomized to screening (the number of which is unobserved). Because postscreening noise increases with follow-up after the last screening, we propose early reporting at the time during the follow-up period when postscreening noise first starts to overwhelm the estimated effect of screening as measured by a  $z$ -statistic. This leads to a confidence interval, adjusted for postscreening noise, that would not change substantially with additional follow-up. Details of the early reporting rule were refined by simulation, which also accounts for multiple looks.

**Results**—For the re-analysis of the Health Insurance Plan trial for breast cancer screening and the Mayo Lung Project for lung cancer screening, estimates and confidence intervals for the effect of screening on cancer mortality were similar on early reporting and later.

**Conclusion**—The proposed early reporting rule for a cancer screening trial with post-screening follow-up is a promising method for making results from the trial available sooner, which translates into health benefits (reduction in cancer deaths or avoidance of unnecessary morbidity) reaching the population sooner.

## INTRODUCTION

The goal of cancer screening trials is to evaluate the effect of early detection of cancer coupled with early intervention on deaths from the cancer targeted by screening (which we call cancer deaths or cancer mortality). Cancer screening trials typically involve randomization of subjects to screening vs. no screening or usual care, with a long period of follow-up after the last screening. Particularly in recent years there has been interest in monitoring cancer screening trials for early reporting of results during the follow-up period.

If results from a randomized trial of cancer screening were reported early in the follow-up period and screening were found to have a benefit, a substantial number of lives could be saved prior to the originally scheduled end of follow-up. For example suppose the follow-up period were scheduled for 10 years and results were reported after only 6 years of follow-up when cancer screening was found to be beneficial. After 6 years of follow-up, persons in the general population could be influenced to start cancer screening. Otherwise an additional four years of follow-up would pass before persons in the general population could be influenced to start screening, and four years would be lost when screening could have had widespread benefits.

If results from a randomized trial of cancer screening were reported early and the upper bound of a 95% confidence interval for the difference in mortality rates indicated that the any possible benefit was probably small relative to harms (false-positives, unnecessary biopsies and overdiagnoses of non-life-threatening cancers), and the lower bound was less than zero, the current level of screening could be reduced before the originally scheduled end of follow-up, which would be a net benefit. For example, prostate-specific antigen (PSA) is widely used in some countries to screen for prostate cancer. Suppose that in reporting the results of a randomized trial of PSA screening the upper bound of the 95% confidence interval indicated a small reduction in prostate cancer mortality relative to harms, and the lower bound indicated an increase in prostate cancer mortality. Many persons would be then influenced to stop PSA screening or not to start PSA screening. The earlier this unfavourable confidence interval is reported, the sooner (1) persons currently undergoing PSA screening could stop receiving PSA screening and (2) persons contemplating PSA screening could avoid PSA screening. The key idea of early reporting during the follow-up period is that there is a time during the follow-up period after which the coverage of the confidence interval for the estimated difference in cancer mortality rates between randomization groups would not change substantially with further follow-up (discussed later). We developed an early reporting rule and refined it using a simulation that accounts for multiple looks. We then investigated the performance of the rule in randomized trials of breast and lung cancer screening in which there was no early reporting in the original designs and analyses.

## METHODS

### Analysis without early reporting

As a prerequisite to the discussion of early reporting in cancer trials with follow-up after the last screening, we first discuss the analysis without early reporting. We follow the framework of Baker *et al.*<sup>1</sup> but with an extension to allow randomization in different years. Let  $t = 1, 2, \dots, T$  denote year since randomization, where  $T$  is the largest year corresponding to the end of follow-up. Let  $g$  denote randomization group, with  $g = 0$  = assigned to no screening and  $g = 1$  = assigned to screening. Let  $x_g(t; T)$  denote the number of deaths from the target cancer and let  $r_g(t; T)$  denote the number at risk for target cancer death in year  $t$  after randomization and group  $g$ . (The dependence on  $T$  is included in the notation for number of deaths and number at risk to be consistent with subsequent notation involving year of reporting.) If, as is typically the case, the analysis is done at a fixed point in calendar time, then the vast majority of censoring arises from staggered entry, so  $r_g(t; T)$  can be approximated from the number of subjects

randomized in successive years (Appendix). Let  $S(t)$  denote the probability of surviving all causes of mortality other than the target cancer from randomization to time  $t$ . The value of  $S(t)$  is treated as known and could be obtained from life table data for the target population adjusted for age and ideally underlying health. (In our examples, we set  $S(t) = 1$ , as a reasonable approximation for healthy subjects in the age range of the trial and over the time span of follow-up.) The estimated cumulative difference in target cancer mortality rates at time  $t$  is approximately

$$d(t) \approx \sum_{i=1}^t S(i)h_0(i) - \sum_{i=1}^t S(i)h_1(i), \quad h_g(t) = \frac{x_g(t;T)}{r_g(t;T)}. \quad (1)$$

A modification is needed to adjust for non-compliance (when some subjects randomized to screening refuse screening) and contamination (when some subjects randomized to no screening receive screening).<sup>1</sup> Based on the standard model for potential outcomes with all-or-none compliance,<sup>2-5</sup> if non-compliance and contamination occur soon after randomization, the estimated effect of screening among those subjects who would only receive screening if offered (sometimes called the causal effect) is

$$d_{\text{causal}}(t) = \frac{d(t)}{f_1 - f_0}, \quad (2)$$

where  $f_g$  is the fraction who receive screening in group  $g$  starting soon after randomization. The potential outcomes model underlying Equation (2) assumes three types of subjects: (1) those who would receive screening regardless of assigned randomization group; (2) those who not receive screening regardless of randomization group; and (3) those who would receive screening only if assigned to the screening group. The model underlying Equation (2) also assumes that the probability of cancer death among subject types (1) and (2) does not depend on randomization group. Equation (2) estimates the effect of screening among type (3) subjects. For example, see Baker *et al.*<sup>5</sup> for a simple derivation of Equation (2) in a general context. A nice feature of Equation (2) is that data are not needed on cancer deaths in each randomization group broken down by those who received screening and those who did not receive screening.

Another modification is needed to adjust for postscreening noise, which is sometimes called dilution. Postscreening noise refers to cancer deaths in the follow-up period that could not have been prevented by screening: (1) cancer deaths in the screened group that occurred after the last screening in subjects whose cancers were not detected during the screening program, and (2) cancer deaths in the control group that occurred after the time of the last screening and in subjects whose cancers would not have been detected during the screening program had they been randomized to screening. The number in (2) is not directly observed, but, in keeping with the theoretical underpinnings of randomization, the number in (1) would equal, on average, the number in (2). Failure to adjust for postscreening noise can yield incorrect confidence intervals for the estimated effect of screening.

Baker *et al.*<sup>1</sup> proposed estimating the effect of screening while adjusting for postscreening noise by analysing the data in the year since randomization when postscreening noise first starts to mask the effect of screening. The key is to note that after some point in the follow-up period, the difference between cancer mortality rates should remain fairly constant as there is little residual effect of screening, and the cancer mortality rates will be increasing due to postscreening noise (which is also related to age). The year when postscreening noise first starts to mask the effect of screening is identified using a  $z$ -statistic, the estimated difference in cumulative cancer mortality rates divided by its standard error,

$$z(t) = \frac{d(t)}{\sqrt{\text{vard}(t)}}, \quad \text{vard}(t) = \sum_{g=0}^1 \sum_{i=1}^t \frac{\{S(i)\}^2 h_g(i)}{r_g(i)}. \quad (3)$$

The estimated variance in the denominator of  $z(t)$  is computed under the assumption that cancer death rates follow a Poisson distribution. At some year after the last screening the  $z$ -statistic typically decreases because (1) postscreening noise increases the denominator of the  $z$ -statistic and (2) any effect of screening diminishes after the last screening, keeping the numerator relatively constant.

More formally, to mitigate the effect of postscreening noise, Baker *et al.*<sup>1</sup> proposed estimating the effect of screening (with slightly different notation here) by  $d_{\text{causal}}(t_0^*; X_{T0})$ , where  $t_0^*$  denotes the year after randomization when the  $z$ -statistic reaches a maximum and  $X_{T0}$  denotes the data  $x_g(t; T)$  and  $r_g(t; T)$  for  $g = 0, 1$  and  $t = 1, 2, \dots, T$ . The subscript '0' on  $t_0^*$  and  $X_{T0}$  denotes the original data, as opposed to random replications to be discussed later. The subscript 'T' on  $X_{T0}$  indicates the inclusion of all data until the end of follow-up at time  $T$  since first randomization is used. Because the data were used to select the year  $t_0^*$ , it is not appropriate to use standard methods to compute confidence intervals for  $d_{\text{causal}}(t_0^*; X_{T0})$  that pretend as if  $t_0^*$  was specified in advance, the fallacy of cutpoint optimization.<sup>6</sup>

To avoid the fallacy of cutpoint optimization, confidence intervals are computed by generating the data under a Poisson distribution  $j = 1, \dots, J$  times. For the  $j$ th random data generation let  $x_{gj}(t; T)$  denote the number of deaths for group  $g$  at time  $t$  randomly generated from a Poisson distribution with expected value  $x_g(t; T)$ . Let  $t_j^*$  denote the year after randomization when the  $z$ -statistic reaches a maximum for the  $j$ th random data generation. Also let  $X_{Tj}$  denote the data  $x_{gj}(t; T)$  and  $r_g(t; T)$  for  $g = 0, 1$  and  $t = 1, 2, \dots, T$  for the  $j$ th random data generation. Substituting  $x_{gj}(t; T)$  for  $x_g(t; T)$  into Equations (1) to (3) yields  $d_{\text{causal}}(t_j^*; X_{Tj})$ . The lower and upper bounds of a 95% confidence interval correspond to the lower and upper 2.5% quantiles of the distribution of  $d_{\text{causal}}(t_j^*; X_{Tj})$  for  $j = 1, 2, \dots, J$ .

### Analysis with early reporting

We extend these ideas to early reporting of results during the follow-up period. Suppose a cancer screening trial is monitored in successive years during the follow-up period after screening has ended. Let  $m$  denote the monitoring year since the start of the study corresponding to the initial enrollment. Importantly  $m$  is also the maximum follow-up time since randomization for the current year of monitoring. Therefore  $m \leq T$ . The value of  $m$  is directly related to chronological year of monitoring. For example if the first enrolment began in chronological year 1965 and the chronological year of monitoring year is 1970, which includes the data from 1965 to 1969, then  $m = 5$ , which also means the maximum follow-up is five years.

In this setting  $t_j^*$  is called the year of analysis. When we discussed the analysis without early reporting,  $t_j^*$  represented the year after randomization when the  $z$ -statistic reached a maximum (with ties going to the latest maximum). To be more conservative in accounting for chance fluctuations due to yearly monitoring, we also considered  $t_j^*$  as the year the  $z$ -statistic reached a maximum (with ties going to the latest maximum) plus one. This latter modification would likely increase the comfort level of a Data and Safety Monitoring Committee considering whether to recommend early reporting of results. As shown later in the simulation, this conservative choice of year of analysis gives approximately correct confidence intervals.

The key idea behind early reporting in the follow-up period is that once postscreening noise starts to mask the effect of screening on cancer mortality, the year of analysis should not change with increased follow-up and hence estimates and confidence intervals should not change substantially with further follow-up.

For each monitoring year  $m$ , random data are generated according to a Poisson distribution with means equal to the observed counts. Let  $X_{Tj}$  denote the data  $x_{g^j}(t; m)$  and  $r_{g^j}(t; m)$  for  $g = 0, 1$  and  $t = 1, 2, \dots, T$  for the  $j$ th random data generation for monitoring year  $m$ . For the  $j$ th random generation of the data for monitoring year  $m$ , we computed an indicator of whether or not the year of analysis,  $t_j^*$ , occurred prior to the current year of monitoring. Mathematically, this indicator is  $I(t_j^*, m; X_{mj}) = 1$  if  $t_j^* < m$ , and 0 otherwise. Then, we computed the percent of random data generations in which the year of analysis occurred prior to the current year of monitoring, namely

$$F(m) = 100 \sum_{j=1}^J \frac{I(t_j^*, m; X_{mj})}{J}. \quad (4)$$

Our early reporting rule is to report results if  $F(m) > F_{\text{target}}$ , where  $F_{\text{target}}$  is a target value determined by simulation (discussed later).

The following bootstrap approach is used to obtain estimates and 95% confidence intervals for the effect of screening on cancer mortality at the year of early reporting. Let  $d_{\text{causal}}(t_j^*, m; X_{mj})$  denote the estimated causal effect of screening in monitoring year  $m$  since the trial began for the  $j$ th random generation of data. If results are reported at monitoring year  $m$ , the effect of screening on the reduction in cancer mortality rate is estimated by

$$\text{dif}(m) = \sum_{j=1}^J \frac{d_{\text{causal}}(t_j^*, m; X_{mj})}{J}, \quad (5)$$

with estimated standard error,

$$\begin{aligned} \text{se dif}(m) &= \sqrt{\text{vardif}(m)}, \\ \text{vardif}(m) &= \sum_{j=1}^J \frac{(d_{\text{causal}}(t_j^*, m; X_{mj}) - \text{dif}(m))^2}{J}. \end{aligned} \quad (6)$$

A bootstrap 95% confidence interval is

$$(\text{dif}(m) + 1.96 \text{se dif}(m), \text{dif}(m) - 1.96 \text{se dif}(m)). \quad (7)$$

To minimize the impact of chance fluctuations in the observed data, the average over random data generations,  $\text{dif}(m)$ , is used in Equation (7) instead of the observed estimate,  $d_{\text{causal}}(t_0^*, m; X_{m0})$ . Note that the same random generations of data,  $X_{mj}$ , are used to compute  $F(m)$  and  $\text{dif}(m)$ . Another quantity of interest is the average year of analysis,

$$t_{\text{avg}}^* = \sum_{j=1}^J \frac{t_j^*}{J}, \quad (8)$$

which is the average over the random generations of data.

We considered a wide range of scenarios depicted in Figure 1 in which screening ended at year 3 after randomization. (The number of years of screening is not relevant if the total number of cancer deaths during the screening period is specified). In all the scenarios, the yearly number of cancer deaths reached a constant to reflect postscreening noise. The large-effect scenario corresponded to a roughly 33% reduction in cancer deaths during the follow-up period before the period of constant postscreening noise (which started at different times). The moderate-effect scenario corresponded to a roughly 22% reduction in cancer mortality in same period. We also considered scenarios in which the number of deaths in the follow-up period was doubled for moderate and large-effect scenarios.

From each set of possible ‘true’ counts (namely the counts of cancer deaths in the two groups over the planned follow-up period) plotted in Figure 1, data for 1000 simulated trials were randomly generated according to a Poisson distribution with means equal to the true counts. For each simulated trial, we computed  $J = 20$  random generations of data, from which we computed  $F(m)$  using Equation (4). If  $F(m)$  was greater than or equal to  $F_{\text{target}}$ , we selected  $m$  as the year to report results and computed the bootstrap 95% confidence interval using Equation (7). Otherwise, if  $F(m)$  was less than  $F_{\text{target}}$ , we incremented the year of monitoring by one and repeated the same type of calculation but with an additional year’s data, thus mimicking the process of multiple looks.

We computed the coverage, namely the fraction of the simulated trials that the 95% confidence interval for each trial enclosed the true difference, where the true difference was computed as  $d_{\text{causal}}(t_0^*, T; X_{T_0})$ , where  $X_{T_0}$  are the ‘true’ counts plotted in Figure 1. Ideally, the coverage should be close to 95%. We ran the simulation for  $F_{\text{target}} = 90\%$ , 60% and 30%. With 1000 simulated data-sets, differences in estimated coverage of 3–4% (equals 4 standard deviations) are not statistically significant, even without adjusting for multiple comparisons associated with the many sets of ‘true’ counts.

For year of analysis equal to the maximum  $z$ -statistic we found that the coverages of 95% confidence intervals ranged from 85–93%, 72–92%, and 53–94% for  $F_{\text{target}} = 90\%$ , 60%, 30%, respectively. These coverages were too low to recommend this statistic.

In contrast, for year of analysis equal to the maximum  $z$ -statistic plus one, we found that the coverage of 95% confidence intervals ranged 91–96%, 90–94% and 84–95% for  $F_{\text{target}} = 90\%$ , 60% and 30%, respectively. Among these values, the smallest  $F_{\text{target}}$  that yielded a reasonable coverage was 60%. Thus we concluded that a reasonable rule for early reporting has  $F_{\text{target}} = 60\%$  when year of analysis equals the maximum  $z$ -statistic plus one

### Extension to relative risk

Sometimes there is interest in the estimating the relative risk (RR) in addition to the risk difference. The causal estimate for  $\text{RR}^5$  requires additional data. Let  $c$  denote the compliance status soon after randomization, with  $c = 1$  denoting receipt of screening and  $c = 0$  denoting refusal of screening. Let  $w_{gc}(t)$  denote the number of deaths from the target cancer among those in randomization group  $g$  with compliance status  $c$ . The causal estimate of RR (the estimate for subjects who would only accept screening if offered) is

$$\text{RR}_{\text{causal}}(t) = \frac{p_{11}(t) - p_{01}(t)}{p_{00}(t) - p_{10}(t)}, \quad p_{gc}(t) = \sum_{i=1}^t S(i) \frac{w_{gc}(t)}{r_g(t)}. \quad (9)$$

Let  $W_{mj}$  denote the data  $w_{gc}(t; m)$  and  $r_g(t; m)$  for  $g = 0, 1, c = 0, 1$  and  $t = 1, 2, \dots, m$ , corresponding to year of monitoring  $m$ . Similar to Equation (5), let  $\text{RR}_{\text{causal}}(t_j^*, m; W_{mj})$  denote the estimated causal RR of screening in monitoring year  $m$  since the trial began for the  $j$ th

random generation of data. Analogous to Equation (5), we computed

$RR(m) = \sum_{j=1}^J RR_{\text{causal}}(t_j^*, m; W_{mj}) / J$ . Estimated variances and bootstrap confidence intervals are computed analogously to Equations (6) and (7).

## RESULTS

We applied the early reporting rule ( $F_{\text{target}} = 60\%$  and year of analysis equal to the maximum of  $z$ -statistic plus one) to data from two randomized cancer screening trials in which there was no early reporting and one arm was offered screening while the control arm received no planned screening per protocol. Our goal was to compare two estimates of the causal effect of screening and 95% confidence interval, one based on early reporting and one based on the last year of follow-up.

For all the calculations, we used  $J = 20$  random data generations to compute bootstrap confidence intervals and approximated the probability of overall survival during the course of the trial by  $S(t) = 1$ . For reporting the results we omit the  $m$  in writing  $F(m)$  and  $dif(m)$ , with the understanding that  $m$  corresponds to the difference between the chronological year and year of first enrolment.

In the Health Insurance Plan (HIP) study of breast cancer screening, about 60,000 women were assigned on alternate days to either an offer of four annual breast cancer screenings (with 2/3 accepting the offer) or to a control group that received no screening.<sup>7</sup> In the years from 1964 to 1966 the numbers entering the study were 22036, 27742 and 10918, respectively, which were approximated equally allocated to the two randomization groups. We analysed the data in successive years of follow-up from 1969 to 1976 (Table 1, FFigure 2). Early reporting was recommended in 1971 (= 70%) and the estimated reduction in breast cancer mortality due to screening was  $dif = 19$  per 10,000 with 95% confidence interval of (9, 29) per 10,000. In comparison, in our last year of follow-up data, 1976, the estimated reduction in cancer mortality due to screening was similar,  $dif = 22$  per 10,000 with 95% confidence interval of (9, 34) per 10,000. Based on a separate set of randomly generated data (because additional random generations were needed), in 1971 ( $F = 100\%$ ), the estimated RR was 0.71 with 95% confidence interval of (0.57, 0.85) and in the last year 1976, the estimated RR was 0.72 with 95% confidence interval of (0.56, 0.89). The reason for the similar results in 1971 and 1976 is that the average year of analysis was similar, 6.3 for 1971 and 7.0 for 1976. (A naïve analysis in 1976 that did not adjust for postscreening noise would have given different and incorrect results).

In the Mayo Lung Project about 9200 male heavy smokers who tested negative on an initial screening were randomized to either radiological and sputum cytology screening examinations every four months for six years (with only 7% not screened) or a control group that received only an initial recommendation for annual chest X-rays.<sup>8</sup> In the years from 1972 to 1976 the numbers entering the study were 1603, 1586, 2733, 2154 and 1135, respectively, which were approximated equally allocated to the two randomization groups. We analysed the data for successive years of follow-up from 1979 to 1984 (Table 2, FFigure 3). Early reporting was recommended in 1982 (= 85%) with an estimated reduction in lung cancer mortality due to screening of  $dif = -39$  per 10,000 with 95% confidence interval of (-110, 32) per 10,000. In comparison, in our last year of follow-up data, 1984, the estimated reduction in lung cancer mortality due to screening was  $dif = -35$  per 10,000 with 95% confidence interval of (-136, 67) per 10,000, which also indicates no effect of screening on cancer mortality. (A negative number indicates more lung cancer deaths in the screened arm than the control arm.) The fact that  $F$  was smaller for 1983 and 1984 than 1982 probably reflects the variability in the maximum of the  $z$ -statistic when it is near zero. Based on a separate set of randomly generated

data (because additional random generations were needed), in 1982 ( $F = 70\%$ ), the estimated RR was 1.03 with 95% confidence interval of (0.99, 1.07) and in the last year 1984, the estimated relative risk was 1.03 with 95% confidence interval of (0.97, 1.09). The reason for the similar results in 1982 and 1984 is that the average year of analysis was similar, 9.1 for 1982 and 10.0 for 1984.

## DISCUSSION

Because the variability associated with postscreening noise can mask the estimated effect of screening, an early reporting rule based on postscreening noise is sensible. Postscreening noise was measured via a  $z$ -statistic. To relate the  $z$ -statistic to early reporting, a simulation was conducted that also incorporated multiple looks at the data. The simulation let us select the early reporting rules that yielded appropriate coverages of the confidence intervals for the estimated effect of screening on cancer mortality. Because of the time needed for the computer to do the calculations, it was not possible to perform a large number of simulations or random generations of data within each simulation.

Risk difference is a useful statistic for reporting results of cancer screening studies because it easily allows weighing of costs and benefits. However, RR is often reported as well. Both the risk difference and RR have the desirable property (not shared by the odds ratio) that when applying the estimates to a different population there is no bias from an unobserved covariate that does not interact with treatment in its effect on outcome.<sup>9</sup> We used simple formulas for the estimated risk difference or RR for subjects who would only receive screening if offered. These formulas can be viewed as approximations to formulas that incorporate estimated mortality rates from competing risks.<sup>10</sup>

In summary, the proposed early reporting rule is designed to give correct estimates and approximately correct confidence intervals for the effect of screening on cancer mortality once postscreening noise starts to have a large impact on estimation. The method is advantageous from a public policy standpoint because it can shorten the time until reporting of results, which translates into health benefits reaching the population sooner. It has been validated using data from the HIP study of breast cancer screening and the Mayo Lung Project.

## Acknowledgements

We thank Diane Erwin for pre-processing the data.

## References

1. Baker SG, Kramer BS, Prorok PC. Statistical issues in randomized trials of cancer screening. *BMC Med Res Methodol* 2002;2:11. [PubMed: 12238954]
2. Baker SG, Lindeman KS. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Stat Med* 1994;13:2269–78. [PubMed: 7846425]
3. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996;92:444–55.
4. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med* 1997;16:1017–29. [PubMed: 9160496]
5. Baker SG, Kramer BS. Simple maximum likelihood estimates of efficacy in randomized trials and before-and-after studies, with implications for meta-analysis. *Stat Methods Med Res* 2005;14:1–19. Correction (vol. 14, pg. 349, 2005)
6. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using ‘optimal’ cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994;86:829–35. [PubMed: 8182763]
7. Shapiro, S.; Venet, W.; Strax, P.; Venet, L. *Periodic Screening for Breast Cancer, The Health Insurance Plan Project and Its Sequelae, 1963–1986*. Baltimore: Johns Hopkins University Press; 1988.



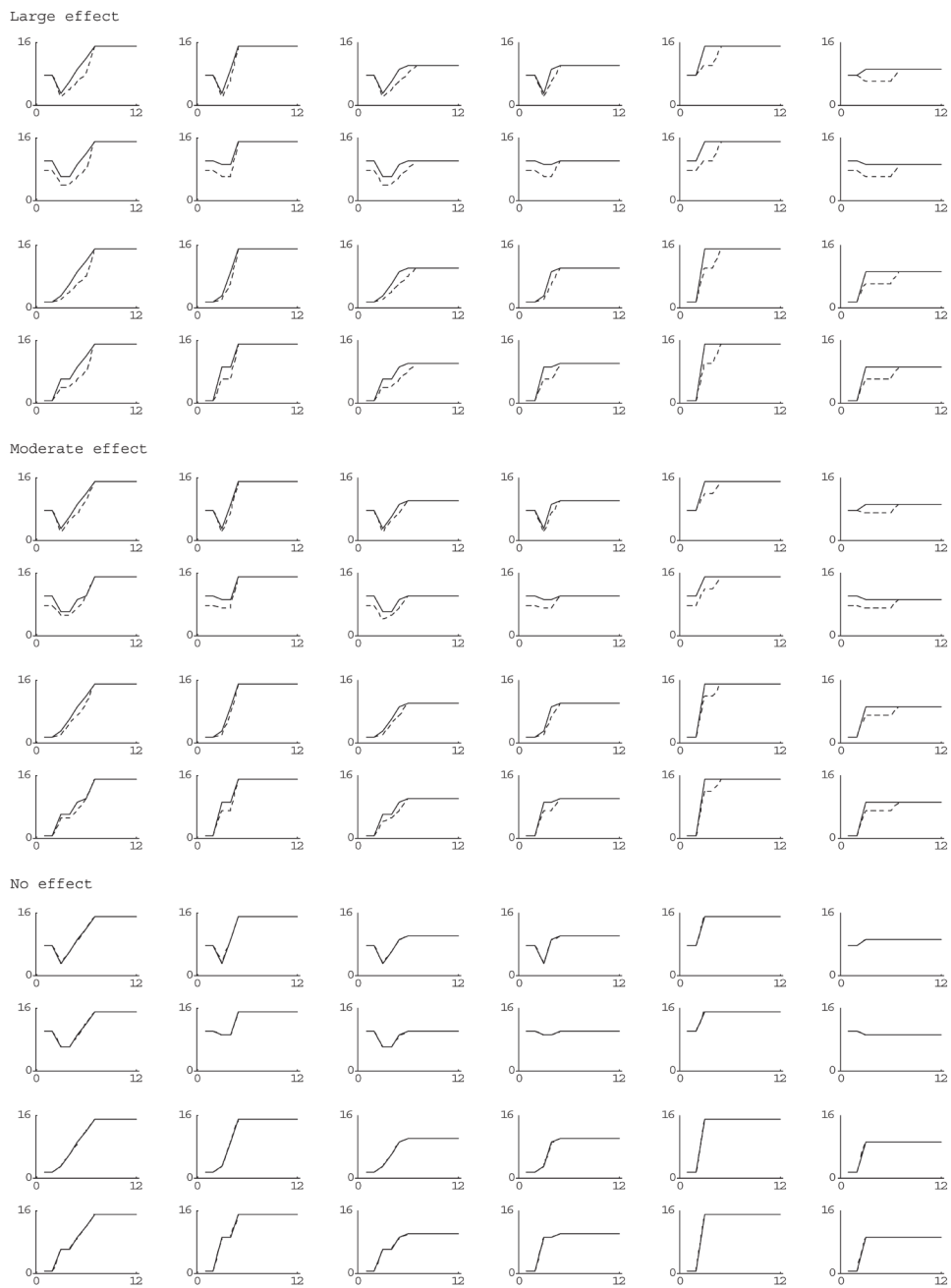
8. Marcus PM, Bergstralh EJ, Fagerstrom RM, Williams DE, Fontana R, Taylor WF, Prorok PC. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst* 2000;92:1308–16. [PubMed: 10944552]
9. Baker SG, Kramer BS. Randomized trials, generalizability, and meta-analysis: graphical insights for binary outcomes. *BMC Med Res Method* 2003;3:10.
10. Baker SG. Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. *J Am Stat Assoc* 1998;93:929–34.

## APPENDIX

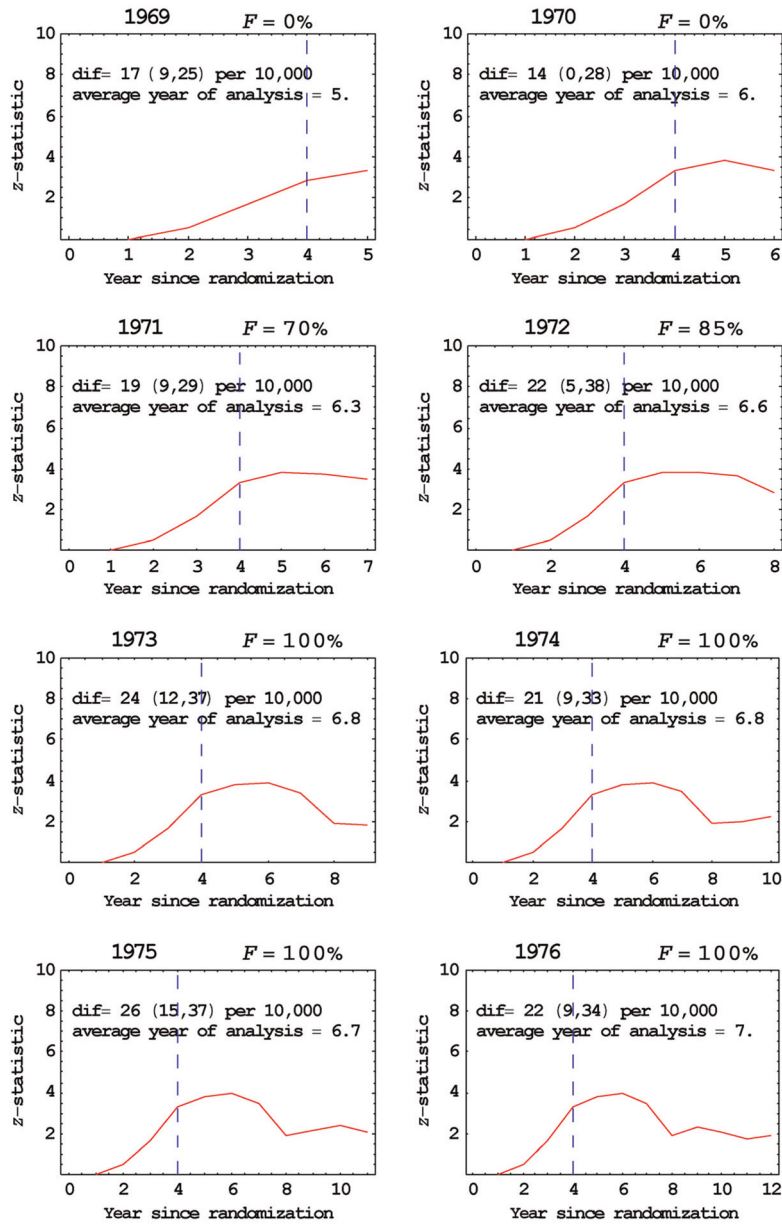
Recall that  $m$  is the monitoring year since the start of the study corresponding to the initial enrolment. Let  $n_{gi}$  denote the number of subjects in randomized group  $g$  who are enrolled in year  $i$  and let  $k$  denote the last year of enrolment. The number at risk in each group for year  $t$  after randomization is

$$r_g(t;m) = \begin{cases} n_{g1} + n_{g2} + n_{g3} + \dots + n_{gk}, & \text{if } t \leq m - k + 1, \\ \vdots & \\ n_{g1} + n_{g2} + n_{g3}, & \text{if } t = m - 2, \\ n_{g1} + n_{g2}, & \text{if } t = m - 1, \\ n_{g1}, & \text{if } t = m. \end{cases}$$

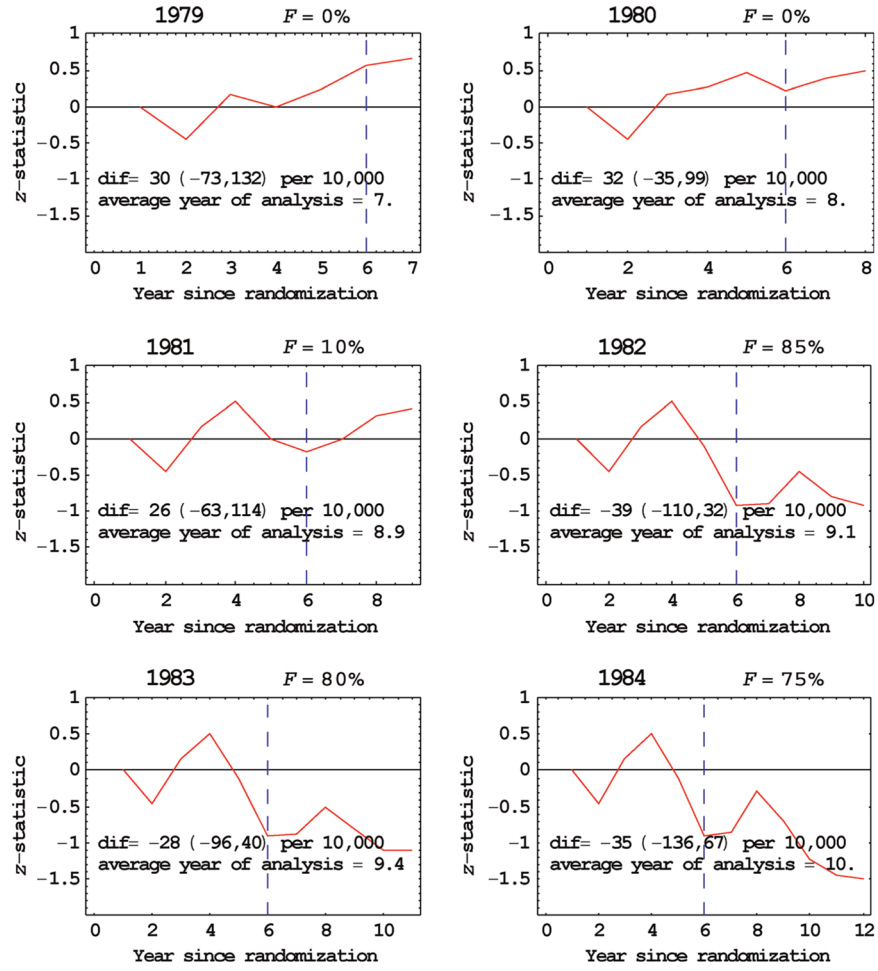
For example in the HIP study, in the years from 1964 to 1966 the numbers entering the study were 22,036, 27,742 and 10,918, approximately equally allocated to each group. Therefore at monitoring year  $m = 5$ , the number at risk in each group was  $(22,036 + 27,742 + 10,918)/2$  for  $t = 1, 2, 3$ ;  $(22,036 + 27,742)/2$  for  $t = 4$ ;  $22,036/2$  for  $t = 5$ .



**Figure 1.** Plots of yearly deaths since year of randomization (screening ends at year 3) for various simulation scenarios (solid line is controls; dashed line is screened group)



**Figure 2.** Early reporting the Health Insurance Plan breast cancer screening trial. Vertical dashed line indicates year of last screening; 'dif' refers to the average estimated difference in cancer mortality between groups at year of analysis with 95% confidence intervals in parentheses



**Figure 3.** Early reporting for the Mayo Lung Project trial involving lung cancel screening. Vertical dashed line indicates year of last screening ‘dif’ refers to the average estimated difference in cancer mortality between groups at year of analysis with 95% confidence interval in parentheses

**Table 1**

Data from Health Insurance Plan breast cancer screening trial

Year of monitoring	Group	Number of breast cancer each year deaths since randomization, $x_g(t, m)$ for $t = 1, 2, \dots, m$
1969 $m = 5$	Control	2, 6, 11, 10, 6
	Screened	2, 4, 4, 1, 1
1970 $m = 6$	Control	2, 6, 11, 19, 16, 5
	Screened	2, 4, 4, 4, 7, 7
1971 $m = 7$	Control	2, 6, 11, 19, 25, 15, 5
	Screened	2, 4, 4, 4, 13, 11, 6
1972 $m = 8$	Control	2, 6, 11, 19, 25, 31, 19, 0
	Screened	2, 4, 4, 4, 13, 21, 16, 10
1973 $m = 9$	Control	2, 6, 11, 19, 25, 32, 28, 8, 4
	Screened	2, 4, 4, 4, 13, 21, 27, 27, 4
1974 $m = 10$	Control	2, 6, 11, 19, 25, 32, 29, 15, 16, 4
	Screened	2, 4, 4, 4, 13, 21, 27, 34, 12, 0
1975 $m = 11$	Control	2, 6, 11, 19, 25, 32, 29, 17, 29, 15, 3
	Screened	2, 4, 4, 4, 13, 21, 27, 36, 21, 9, 9
1976 $m = 12$	Control	2, 6, 11, 19, 25, 32, 29, 17, 31, 20, 17, 5
	Screened	2, 4, 4, 4, 13, 21, 27, 36, 21, 22, 21, 2

Initial enrolment in 1964; year of monitoring includes data only through previous year

**Table 2**

Data from Mayo Lung Project for lung cancer screening

Year of monitoring	Group	Number of lung cancer deaths each year since randomization, $x_g(t, m)$ for $t = 1, 2, \dots, m$
1979 $m = 7$	Control	2, 7, 10, 8, 7, 6, 3
	Screened	2, 9, 7, 9, 5, 3, 2
1980 $m = 8$	Control	2, 7, 10, 10, 9, 8, 6, 2
	Screened	2, 9, 7, 9, 7, 10, 4, 1
1981 $m = 9$	Control	2, 7, 10, 13, 9, 13, 13, 10, 3
	Screened	2, 9, 7, 10, 13, 15, 11, 6, 2
1982 $m = 10$	Control	2, 7, 10, 13, 9, 13, 16, 15, 7, 3
	Screened	2, 9, 7, 10, 14, 22, 17, 10, 12, 5
1983 $m = 11$	Control	2, 7, 10, 13, 9, 14, 19, 20, 11, 5, 2
	Screened	2, 9, 7, 10, 14, 23, 20, 16, 16, 10, 2
1984 $m = 12$	Control	2, 7, 10, 13, 9, 14, 21, 23, 14, 9, 5, 2
	Screened	2, 9, 7, 10, 14, 23, 22, 16, 21, 18, 9, 3

Initial enrolment in 1972; year of monitoring includes data only through previous year