II-13

American Journal of Epidemiology
Copyright © 1992 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved

Vol. 135, No. 7
Printed in U.S.A.

Approaches to the Analysis of Case-Control Studies of the Efficacy of Screening for Cancer

Noel S. Weiss,[1] Barbara McKnight,[2] and Nancy G. Stevens[3]

To an increasing extent, case-control studies are being undertaken to determine if use of early detection procedures is associated with reduced mortality from cancer. The authors recommend that in such studies the analysis focus on screening activity in cases that occurs during an interval prior to diagnosis in which the cancer is believed to be detectable and still curable and to a corresponding time period in controls. This approach places a heavy burden on the investigator to estimate accurately the period during which the tumor ought to be detectable using the test in question and to sort out reliably tests done in response to signs or symptoms of the cancer from screening tests per se. Nonetheless, the authors feel that it offers the greatest ability to discern a true benefit of screening, while minimizing the numerous potential biases that can be present in this type of study. *Am J Epidemiol* 1992;135:817–23.

case-control studies; mass screening; multiphasic screening

As with other medical interventions, the effectiveness of measures designed to detect cancer at an early stage in reducing cancer mortality is best evaluated through randomized trials. However, since the mortality rate of most forms of cancer is relatively low, the application of such trials for this purpose requires a very large number of subjects and usually very high costs. Thus, to an increasing extent in recent years, nonexperimental studies have been used to investigate whether persons who undergo early detection procedures are at a decreased risk of dying from cancer. Because of their economy, most of the nonexperimental studies have been case-control rather than cohort in type.

Case-control studies of the early detection of cancer share a number of potential difficulties with case-control studies aimed at elucidating etiologic factors. For instance, the retrospective ascertainment of "exposure"—in this case, screening history—is often inaccurate, sometimes differentially so between cases and controls. In addition, there is the issue of the comparability of screened and unscreened individuals regarding their underlying risk of the occurrence of the cancer in question. However, there are several problems in the design and analysis of case-control studies of screening efficacy that have no counterpart in etiologic case-control studies. The purpose of this paper is to identify these additional problems and to suggest ways in which their potential for bias can be minimized.

THE PROBLEMS

In the discussion that follows, it will be assumed that "cases" in these studies are

Received for publication March 4, 1991, and in final form October 4, 1991.

Abbreviations: DCPP, detectable and still curable preclinical phase; DPP, detectable preclinical phase.

[1] Department of Epidemiology, University of Washington, and the Fred Hutchinson Cancer Research Center, Seattle, WA.

[2] Department of Biostatistics, University of Washington, and the Fred Hutchinson Cancer Research Center, Seattle, WA.

[3] Departments of Family Medicine and Epidemiology, University of Washington, Seattle, WA.

Reprint requests to Dr. Noel S. Weiss, Department of Epidemiology SC-36, University of Washington, Seattle, WA 98195.

This investigation was supported by PHS grants R35 CA39779 and RO1 CA40644 from the National Cancer Institute.

persons who sustained an outcome of cancer that the screening test sought to prevent (1). Usually, this outcome is death. However, for studies that require information from interviews with the subjects themselves, it seems reasonable to choose as cases persons with "late-stage" disease, i.e., persons with clinically evident cancer who are highly likely to die from it. Controls will be assumed to have been selected in such a way as to reflect the level of screening activity in the general population (at risk of developing and dying from cancer) from which the cases arose (2) and, specifically, to be similar to the cases in terms of age and other relevant respects. Screening histories are sought for the cases up to the time of the diagnosis of their cancer and for the controls prior to a corresponding calendar date. If the test leads to actions that enhance the chances of survival, beyond the influence of those same or different actions administered when the cancer would have otherwise become evident, a history of screening during the relevant period (as discussed below) should be less common among cases than among controls (table 1).

The definition of exposure (i.e., screening status) in a randomized controlled trial of screening efficacy, in the absence of major noncompliance, is unambiguous: Persons assigned to be screened are compared with those not so assigned, irrespective of the actual screening activity engaged in by members of the two groups of subjects. In a case-control study, of course, no such random assignment to screening is made. Nearly always, an individual's screening status is determined by the screening that actually took place. Thus, we must shift our attention from a subject's "assigned" screening program to that screening which actually has the potential to achieve some benefit, i.e., screening that takes place during the detectable and still curable preclinical phase (DCPP) of the cancer for which screening is being performed. (Cole and Morrison (3) have described a related, but not identical, concept with a similar, but not identical, abbreviation—the detectable preclinical phase (DPCP). When the ability of a screening test to initiate a series of events that can reduce cancer mortality is being considered, we believe it is necessary to include the additional notion of "still curable" when describing preclinical lesions.) Persons who received negative screens prior to the DCPP of *their* cancer clearly were not benefited by those screens; ideally, a history of such screens should enter the analysis only as a means of controlling for possible bias (as discussed below).

In analyses of data from randomized trials of *transient* screening interventions, the concept of a DCPP proves useful as well. In the numerator of cancer mortality rates of groups assigned to be screened or not screened, only deaths in cases who plausibly had a tumor present during the time that

TABLE 1. Hypothetical data from a study of cancer screening efficacy

Screening during "relevant" period	No. of cancer deaths	No. of controls	Odds ratio
Screening status correctly classified			
Yes	35	91	0.18
No	105	49	1.00
Total	140	140	
*Screening status incorrectly classified in 40 cancer patients whose "screening" resulted from symptoms related to cancer**			
Yes	75	91	0.62
No	65	49	1.00
Total	140	140	

* Screening correctly classified in all controls.

persons who sustained an outcome of cancer that the screening test sought to prevent (1). Usually, this outcome is death. However, for studies that require information from interviews with the subjects themselves, it seems reasonable to choose as cases persons with "late-stage" disease, i.e., persons with clinically evident cancer who are highly likely to die from it. Controls will be assumed to have been selected in such a way as to reflect the level of screening activity in the general population (at risk of developing and dying from cancer) from which the cases arose (2) and, specifically, to be similar to the cases in terms of age and other relevant respects. Screening histories are sought for the cases up to the time of the diagnosis of their cancer and for the controls prior to a corresponding calendar date. If the test leads to actions that enhance the chances of survival, beyond the influence of those same or different actions administered when the cancer would have otherwise become evident, a history of screening during the relevant period (as discussed below) should be less common among cases than among controls (table 1).

The definition of exposure (i.e., screening status) in a randomized controlled trial of screening efficacy, in the absence of major noncompliance, is unambiguous: Persons assigned to be screened are compared with those not so assigned, irrespective of the actual screening activity engaged in by members of the two groups of subjects. In a case-control study, of course, no such random assignment to screening is made. Nearly always, an individual's screening status is determined by the screening that actually took place. Thus, we must shift our attention from a subject's "assigned" screening program to that screening which actually has the potential to achieve some benefit, i.e., screening that takes place during the detectable and still curable preclinical phase (DCPP) of the cancer for which screening is being performed. (Cole and Morrison (3) have described a related, but not identical, concept with a similar, but not identical, abbreviation—the detectable preclinical phase (DPCP). When the ability of a screening test to initiate a series of events that can reduce cancer mortality is being considered, we believe it is necessary to include the additional notion of "still curable" when describing preclinical lesions.) Persons who received negative screens prior to the DCPP of *their* cancer clearly were not benefited by those screens; ideally, a history of such screens should enter the analysis only as a means of controlling for possible bias (as discussed below).

In analyses of data from randomized trials of *transient* screening interventions, the concept of a DCPP proves useful as well. In the numerator of cancer mortality rates of groups assigned to be screened or not screened, only deaths in cases who plausibly had a tumor present during the time that

TABLE 1. Hypothetical data from a study of cancer screening efficacy

Screening during "relevant" period	No. of cancer deaths	No. of controls	Odds ratio
Screening status correctly classified			
Yes	35	91	0.18
No	105	49	1.00
Total	140	140	
*Screening status incorrectly classified in 40 cancer patients whose "screening" resulted from symptoms related to cancer**			
Yes	75	91	0.62
No	65	49	1.00
Total	140	140	

* Screening correctly classified in all controls.

the intervention was available ought to be considered. The inclusion in either group of cases who were diagnosed long after screening had ceased would tend to dilute any beneficial influence on mortality that screening might have had.

This attention in the analysis to screening that occurred during the DCPP poses some potential difficulties, as discussed below.

Often it is true that the very same test is administered whether the person is asymptomatic and is being screened for the presence of cancer or instead is symptomatic and is being tested for it.

Mammography, cervical smears, or fecal occult blood testing may all be used in the evaluation of both asymptomatic and symptomatic patients. It will be necessary to make a judgment by whatever means are available—usually a review of medical records—as to whether symptoms plausibly related to the cancer were present at the time(s) the test was performed. In both cases and controls, only tests done in the absence of such symptoms are tabulated as screens. Failure to make this judgment in an accurate way could lead to a substantial underestimate of the benefit (overestimate of the odds ratio) associated with screening, since it may falsely claim the presence of screening in a sizable number of individuals who had eventually succumbed to their cancer. Table 1 indicates, in a hypothetical example, how this bias might operate.

The length of the DCPP, and its proximity to the clinical onset of disease, probably varies substantially among different individuals with a particular type of tumor.

It would be only an educated guess as to what was even the average length of the DCPP. Thus, it is necessary to be somewhat arbitrary in choosing an interval prior to diagnosis (corresponding dates for controls) during which the presence of screening is to be tabulated. Unfortunately, the wrong arbitrary choice will bias the results.

Misspecification of the location of the DCPP within the detectable preclinical phase (DPP). Consider, hypothetically, a cancer that for a particular test has a DPP of 3 years in all persons. Assume that the test is 100 percent sensitive, i.e., there are no false negatives, and that detection of this cancer during its first 2 preclinical years leads uniformly to curative treatment, whereas detection in the last preclinical year or after symptoms develop is of no benefit (and also assume that such tumors uniformly prove fatal). A case-control study of deaths from this cancer, assuming completely accurate ascertainment of screening histories during the DPP, ought to obtain data of the sort shown in figure 1. If we knew—although, in reality, we never can—how long those cases detected by screening were identified prior to what would have been the date of clinical onset of their disease, we could exclude those in the last year of their DPP. Our analysis would show that during years 0–2 of the DPP, i.e., the DCPP, the relative risk of death associated with having been screened is zero (0/120 cases vs. 40/120 controls).

Now let us return to reality, in which we have no knowledge of when cancers first become detectable in any individual or when they would have come to clinical attention in the absence of screening. In reality, it is also true that the duration of neither the DPP nor the DCPP is the same from person to person. For example, some persons diagnosed with symptomatic colon cancer will be cured of it; in them, the DPP and DCPP completely coincide. On the other hand, persons with colon cancer detected by screening who eventually died of their disease had a DPP that exceeded the length of the DCPP. For these reasons, we must include in our comparison of cases and controls any effort at early detection that occurs between the start of the presumed DPP and the development of symptoms. If, in truth, detection relatively late in the DPP does not benefit some individuals, we would obtain an overall estimated benefit from screening that is smaller than the one we would have obtained had it been possible to focus atten-

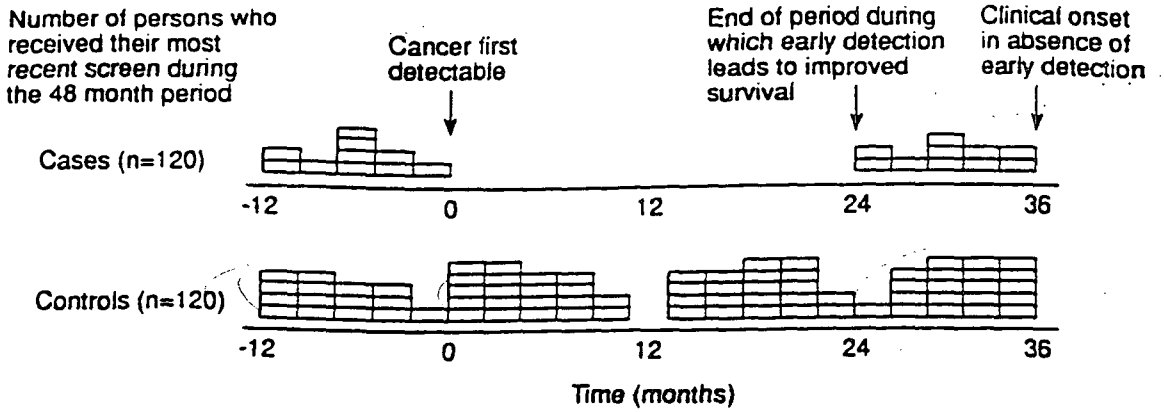

FIGURE 1. Screening during the detectable preclinical phase of persons who died of their cancer (cases) and during the corresponding period among controls. Figure 1 assumes that there is a detectable preclinical phase of 3 years for all cases, that screening during the first 2 years of the detectable preclinical phase inevitably leads to the cancer being cured (i.e., odds ratio = 0) associated with screening during that interval, and that screening after the time is of no benefit to survival.

TABLE 2. Hypothetical data from a case-control study of the efficacy of screening, using the correct estimate of the length of the detectable preclinical phase prior to the diagnosis of a case (reference data for controls), but not considering the presence of the detectable and still curable preclinical phase

Screen during detectable preclinical phase*	No. of cases	No. of controls	Odds ratio
Yes	10	60	0.09
No	110	60	1.00
Total	120	120	

* See figure 1.

tion on an interval earlier during the DPP. With the scenario portrayed in table 2, the estimated relative risk associated with screening is now increased to 10/110 ÷ 60/60 = 0.09.

Misspecification of the duration of the DPP. If the interval chosen for analysis does not extend back far enough in time to encompass most patients' DPP, then only a portion of the controls' screening during the length of the true average DPP will be counted. The degree of this undercounting would be far smaller among cases, in whom screening (if it occurred at all) would have taken place relatively close to the time of diagnosis and after the preclinical portion of the disease during which detection would have the greatest likelihood of leading to effective treatment (figure 1). The effect of falsely underestimating the length of th DPP would be to minimize falsely the est mated benefit associated with screening (t ble 3).

If the length of the DPP is overestimate then screens prior to the true DPP will l included for both controls and cases. Th type of misclassification generally would " expected to reduce falsely the size of t benefit associated with screening (table 4)

Persons who undergo screening are likely be the ones who had been screened earlier their lives.

While this earlier screening did not ben these persons because they did not h cancer at the time of the screening, it

TABLE 3. Hypothetical data from a case-contr study of the efficacy of screening, using an underestimate of the detectable preclinical phase—2 years, instead of 3 years—and not considering the presence of the detectable and curable preclinical phase (20 fewer controls, bu fewer cases, would be categorized as having b screened)

Screen during detectable preclinical phase*	No. of cases	No. of controls	Odd ratio
Yes	10	60 − 20 = 40	0.18
No	110	60 + 20 = 80	1.0
Total	120	120	

* See figure 1.

BLE 4. **Hypothetical data from a case-control study of the efficacy of screening, using an overestimate of the detectable preclinical phase—4 years, instead of 3 years—and not considering the presence of the detectable and still curable preclinical phase (which assumes that in the year prior to the cancer being detectable, 15 additional controls and 10 additional cases would be categorized as having been screened)**

Screen during detectable preclinical phase*	No. of cases	No. of controls	Odds ratio
Yes	10 + 10 = 20	60 + 15 = 75	0.12
No	110 − 10 = 100	60 − 15 = 45	1.00
Total	120	120	

* See figure 1.

serve to identify other persons who *did* have cancer, and so the remainder who had been screened as negative have a period (equal to the maximum length of the DPP) during which they are at lower risk than an otherwise similar, unscreened group. This is a form of "healthy screenee" bias (4). As long there is heterogeneity in the length of the clinical course of disease, negative screens performed before the presumed start of the DPP will be associated with a lower likelihood of dying of cancer and, consequently, of being a "case" in a case-control study. Thus, screening prior to the average duration of the DPP can be thought of as a confounding factor, one related both to an increased likelihood of a person's being screened during the DPP and to a decreased likelihood of their dying of cancer.

POSSIBLE SOLUTIONS

Determining whether tests that were performed were screening in character

When a patient's medical record is examined for the presence of symptoms or signs of a particular cancer, it is necessary to restrict attention to only those notations that appeared prior to the test results being known to the health care provider. A standardized method for recording and judging ymptoms/signs should be used. If feasible, the person(s) recording and judging symptom relatedness should do so without knowledge of the case/control status of the subject.

Because costs for screening tests are not reimbursed by some health insurance plans, there may exist an incentive for providers to label a screening test as diagnostic in some patients. For example, it is likely that some women for whom mammograms were ordered solely because of "nodularity" have no palpable abnormalities of their breasts. For this reason, and because of the uncertain relation of nodularity to the occurrence of breast cancer, such mammograms probably should be classified as "screening" in any event. Care must be taken to identify such situations, perhaps by means of reviews of a sample of representative records together with discussions with local physicians.

Dealing with uncertainties as to the length of the DPP

While some educated guesses can be made regarding this characteristic, it will not be known to us in any study. Thus, in the analysis of a case-control study of screening, it is appropriate to generate a series of odds ratios for each of a series of guesses about the length of the DPP. Since the biases resulting from misspecification of this parameter appear falsely to raise the odds ratio (toward 1.0, assuming the screening test truly does have some efficacy), one could argue that the lowest odds ratio obtained would be the best estimate of the benefit associated with screening. This procedure could well yield biased estimates of the effect of screening, however, since even in the absence of any true effect, odds ratios that result from some choices of the DPP would be less than one because of sampling variability.

Perhaps a better approach is to estimate the length of the DPP and the associated odds ratio simultaneously by the method of maximum likelihood. The Appendix describes how to do this using standard logistic regression software. If the length of the DPP varies from individual to individual, however, even this method may provide conservative estimates of the true benefit of screening during the DPP, since the length of the DPP will be over- or underestimated for many subjects (as discussed above).

Minimizing healthy screenee bias

For estimation of the length of the DPP and the associated odds ratio for a particular analysis, it will be necessary to control for a history of screening examinations prior to the start of the DPP. (Screening among controls that takes place after the diagnosis in the corresponding case will, of course, be excluded from consideration in the analysis (5)). The particular aspects of the prior history that need to be taken into account may vary from test to test, but recency and, possibly, frequency usually will be most relevant. This approach assumes that screening prior to and during the candidate DPP are not perfectly correlated, an assumption that will be true in nearly all instances. There may be the rare situation in which an individual's past screening activity entirely predicts more recent activity. Perhaps, for example, all women who performed regular breast self-examinations in the past, prior to a hypothesized 2-year DPP ending just before diagnosis, continued to perform breast self-examinations, whereas no woman who had failed to do so in the past initiated it later on. Statistical adjustment for past breast self-examinations would vitiate any true benefit associated with breast self-examinations during the DPP. The only possible way out of this dilemma would be to examine the ages at which breast self-examinations had been initiated. If they were begun early enough in life, before an appreciable risk of breast cancer had been present, a valid comparison of "lifetime" screenees and nonscreenees could be made.

CONCLUSION

At the present time, it is clear that case-control studies are needed to help evaluate the ability of screening tests to reduce mortality from various forms of cancer. Unfortunately, it is equally clear that there are substantial threats to the validity of these studies, above and beyond those encountered in etiologic case-control studies. The suggestions made in this paper are intended to be beginning steps in helping us to understand these threats. As experience grows with the conduct and analysis of case-cont studies of screening, we hope that furt steps can be taken toward this end, even in the process some of ours must be retrac

REFERENCES

1. Morrison AS. Case definition in case-control st ies of the efficacy of screening. Am J Epider 1982;115:6–8.
2. Weiss NS. Control definition in case-control stuc of screening and diagnostic testing. Am J Epider 1983;118:457–60.
3. Cole P, Morrison AS. Basic issues in populat screening for cancer. J Natl Cancer Inst 1980: 1263–72.
4. Morrison AS. Screening in chronic disease. N York: Oxford University Press, 1985.
5. Friedman DR, Dubin N. Case-control evaluat of breast cancer screening efficacy. Am J Epider 1991;133:974–84.

APPENDIX

Let p be the probability that a subject the sample is a case, and let $x_1, x_2, \ldots$ represent variables such as age and indic tors of screening prior to the DPP (see te for which adjustment will be made. For DPP of length l let

$$x_E(l) = \begin{cases} 1 & \text{subject was screened during the DPP} \\ 0 & \text{subject was not screened duri the DPP.} \end{cases}$$

Using standard statistical software to fit t logistic model,

$$\text{logit}p = \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_p x_p + \beta_{lx_E}(l$$

by maximum likelihood will yield the ma imum likelihood estimate $\hat{\psi}_l = e^{\hat{\beta}_l}$ of the a justed odds ratio associated with screeni during a DPP of length l. Repeated fits the above model can be used to make sir ulations estimates of l and the adjusted od ratio ψ associated with screening during t DPP, whatever its length.

1) Choose a minimum and maximu credible value for the length l of the DI

and make a list of values of l ranging from the minimum to the maximum. The steps separating different values of l in the list should be the smallest size that would make a difference as estimates of the length of the DPP (e.g., 1 month).

2) Fit the logistic model above for each value of l in the list. This will require computing a different exposure variable $x_E(l)$ and a different indicator for prior screening exposure for each l, since as l increases the number of subjects who are exposed to screening during the DPP will increase and the number exposed prior to the DPP will decrease.

3) Record the value of the maximized log-likelihood function for each model in the list. The values of l and $\hat{\psi}_l = e^{\hat{\beta}_l}$ for the model with the largest value of the maximized log-likelihood will be the joint maximum likelihood estimates $\hat{l}$ and $\hat{\psi}$ of the length of the DPP and the adjusted odds ratio associated with screening during the DPP.

Confidence intervals for ψ that take into account the estimation of l might be computed by inverting the likelihood ratio test of $H_0: \psi = \psi_0$, although the appropriate reference distribution for this test would need to be determined.