Bias in Case-Control Studies of Screening Effectiveness

Ronald S. Hosek, W. Dana Flanders, and Annie J. Sasco

Screening programs, such as annual mammography, are undertaken to reduce mortality and/or morbidity from chronic diseases such as cancer. Matched case-control studies have been used to assess the effectiveness of screening programs because of their relative simplicity and low cost. In such studies, the exposure history for controls consists of the number of screening examinations received prior to the date of diagnosis of the matched case. The authors know of no methodological evaluations that demonstrate the validity of such case-control studies. To examine the possible existence of bias due to design rules, the authors developed a simple deterministic model, which is used to calculate expected screening and disease patterns in a cohort. Cases and matched controls are selected from the cohort, and their screening histories are used to calculate an odds ratio, as is commonly done in practice. Results utilizing this simple model suggest that systematic inclusion of the examination from which diagnosis is made, which is the approach typically used in practice, leads to a positive bias (odds ratio > 1) in the absence of any real effect. Systematic exclusion of this examination appears to lead to a negative bias (odds ratio < 1). Although this simple approach has several limitations, the results suggest that a commonly used method of conducting case-control studies may yield biased odds ratios. Possible methods to reduce this bias may exist, such as defining exposure intervals differently.


bias (epidemiology); case-control studies; chronic disease; epidemiologic methods; mammography

Screening for chronic disease is performed with the goal of reducing mortality or morbidity. Often the target disease is chronic, such as cancer, which undergoes insidious development over a relatively long time period. The rationale for screening rests on the supposition that the targeted disease undergoes a detectable preclinical stage during which the disease has yet to manifest symptoms but may be detectable by means of a suitable screening test, such as mammography. If such early detection can be coupled with timely, efficacious, therapeutic intervention, morbidity or mortality from the disease should be reduced (1, 2). Because screening has associated costs as well as potential benefits, researchers must document that screening does, in fact, actually reduce morbidity or mortality.

To study the potential benefits of screening, many authors have used case-control studies. For example, some authors have used these studies to examine reduction in mortality from breast cancer after mammography (3, 4), from colon cancer after sigmoidoscopy (5, 6), and from cervical cancer after Papanicolaou smear (7). Such case-control studies have appeal, in part, because they can cost much less and consume much less time than clinical trials or follow-up studies. In addition, they can provide a direct measure of the effect on individuals rather than on populations and can avoid the ethical problems associated with clinical trials.

To study reduction of mortality due to screening, many researchers have selected cases from those who have died from the disease of interest and have chosen a matched control from all who remain alive at the time of death of the case. Having the disease does not necessarily exclude subjects from the control group. The researchers then compare the number and timing of screening examinations prior to the diagnosis of the case with the corresponding number of control examinations, perhaps using logistic regression to model a dose-response relation and determine an odds ratio. Epidemiologists interpret a result of (odds ratio (OR) < 1) from such studies as supporting the hypothesis that screening reduces mortality. (Please note that the definitions of all abbreviations used in the text may be found in table 1.)

Although case-control studies such as this have intuitive appeal and have received much attention (8-
nondetectable and detectable phases. We will confine the natural history of diseases such as cancer into two phases: preclinical and clinical (1, 2). The purpose of this paper is to evaluate possible bias in case-control studies of screening effectiveness because simple cohort studies of screening effectiveness may themselves yield biased results (1, 15). Valid analysis requires special techniques (1, 15, 16) because of biases such as the structural healthy screenee bias (1). Moreover, the authors know of no methodological evaluation that validly demonstrates that odds ratios from case-control studies estimate a concisely defined parameter, such as the mortality rate among those screened regularly, compared with the mortality rate among the unscreened.

The purpose of this paper is to evaluate possible bias in case-control studies of screening effectiveness. To do this a simple, deterministic model for disease occurrence and screening in a cohort is utilized to study the expected values of odds ratios that would arise in a case-control study. Screening examinations that can detect early disease but that impart no benefit in terms of reduction of mortality are assumed.

**BACKGROUND**

For the purpose of analysis, Morrison and others divide the natural history of diseases such as cancer into two phases: preclinical and clinical (1, 2). The presence of either phase signals the presence of disease. During the preclinical phase, no symptoms are present; during the clinical phase symptoms occur.

The preclinical phase can be further subdivided into nondetectable and detectable phases. We will confine our attention to the detectable preclinical phase (DPCP), for only during this period can screening examinations detect disease.

In studies of several diseases, researchers have estimated the length of the preclinical phase by studying incidence rates and prevalence in screened populations (17-20). Such studies suggest that the preclinical phase, while variable by individual and disease, typically lasts from months to years. To reduce mortality effectively, the screening test must detect disease in the preclinical phase. To improve the chances of early detection, health maintenance organizations and other managed care organizations recommend screening at regular intervals. The appropriate design of such programs and the choice of intervals have been the subject of efficacy and cost-effectiveness analyses and other research (21-24).

As noted previously, matched-pair case-control studies are attractive for performing efficacy studies because of low relative cost, rapidity, and compatibility with registries and other clinical databases. The time course of a typical case-control study is shown in figure 1, where \( A_p \) marks the onset of the detectable preclinical phase, \( A_c \) marks the onset of the clinical phase, \( A_j \) is the age at the \( j \)th screen, between \( A_p \) and \( A_e \).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>Case-fatality</td>
</tr>
<tr>
<td>IR</td>
<td>Age-specific incidence rate in cohort</td>
</tr>
<tr>
<td>( M_A )</td>
<td>Number in cohort at age ( A )</td>
</tr>
<tr>
<td>DPCP</td>
<td>Detectable preclinical phase</td>
</tr>
<tr>
<td>CP</td>
<td>Length of clinical phase</td>
</tr>
<tr>
<td>PP</td>
<td>Duration of DPCP</td>
</tr>
<tr>
<td>SI</td>
<td>Length of interval between successive screens for those screened</td>
</tr>
<tr>
<td>( f )</td>
<td>Ratio of ( PP ) to ( SI ), assumed to be 1 or less</td>
</tr>
<tr>
<td>( A_p )</td>
<td>Age when we first count screening</td>
</tr>
<tr>
<td>( A_c )</td>
<td>Age at onset of clinically manifest disease</td>
</tr>
<tr>
<td>( A_j )</td>
<td>Age at death of cases in study</td>
</tr>
<tr>
<td>( A_p ) ( i = 1, 2, \ldots )</td>
<td>Age at ( i )th screen, between ( A_p ) and ( A_e )</td>
</tr>
<tr>
<td>( \delta, SI )</td>
<td>Time from ( A_p ) to first screen. ( \delta ) is uniformly distributed between 0 and 1.</td>
</tr>
<tr>
<td>( s )</td>
<td>Proportion of cohort who undergo screening</td>
</tr>
</tbody>
</table>

The screening history for cases is determined by counting the number of screening examinations that occur within a specified time interval to the date of the diagnosis of the case. If diagnosis is made by screening examination during the preclinical phase, the examination is typically counted in the screening history. A commonly used rule for determining the control screening history is to count only those screening examinations that occur within the exposure interval established by the matched case, i.e., from the start of some period of interest up to the date of the diagnosis of the case. We will show that, because the control’s screening history is dependent on that of the matched case, a positive bias can occur, resulting in artificially large values of the odds ratio.

**METHODS**

We begin with a simple model for disease occurrence and screening in a cohort.

We assume:

1. Preclinical disease occurs with age-specific incidence rate (IR) in a large cohort, with \( M_A \)
people of age $A$. The case-fatality is the proportion $CF$. Deaths from competing causes occur independently from death due to the disease of interest and can be ignored.

2. We wish to study the effect on mortality of screening that occurs after age $A_B$, the beginning of the period of interest for screening history, and before age $A_E$, the age of death.

3. The screening examinations have 100 percent sensitivity and 100 percent specificity, but have no impact on subsequent mortality. This scenario could hold, for example, if early detection and treatment did not reduce mortality.

4. For all subjects, those who develop disease have a DPCP of length $PP$; after a clinical phase of length $CP$, $CF$ percent of diseased subjects die, and the remainder are cured. We assume that competing risks can be ignored.

5. All subjects in the cohort are screened at intervals of fixed length $SI$. We assume $SI > PP$; otherwise the 100 percent sensitive examination would detect all cases.

6. The first screen for subject $i$ after age $A_B$ occurs at age $A_B + \delta_{SI}$, where $\delta$ is uniformly distributed between 0 and 1.

Next, consider a case-control study in this cohort.

7. Cases consist of all people in the cohort who die at age $A_E$ during a given year. With this assumption, we expect $N = (M_A)(CF)$ cases in one year, where disease onset for these cases occurred at age $A = A_E - PP - CP$.

8. For each case, we select a matched control from all in the cohort who remain alive at age $A_E$. We assume rare disease so that essentially no controls have disease.

9. The number of screening examinations for each control is the number of examinations prior to the date of diagnosis of the matched case.

10. For cases, we count all examinations from age $A_B$ up to $A_C$, including the screening examination for screen-detected cases.

   By assumption, the first screening examination for subject $i$ after age $A_B$ occurs at age $A_B + \delta_{SI}$, and the $(n + 1)$th such examination occurs at age $A_B + (n + \delta_{SI})$. For cases with $\delta = 0$ (figure 2a), the number of examinations prior to the DPCP is therefore given by $k$, where:

   $$k = 1 + \text{int}[(A_P - A_B)/SI]$$

   and where $\text{int}[ ]$ is the integer part of the argument.

   First, consider cases for whom $A_P \leq A_B + kSI \leq A_C$ so that the $(k + 1)$st examination for cases with $\delta = 0$, falls within the DPCP. We depict this situation in figure 2a, with $k = 3$. Call this “case type I” (we cover the situation when $A_B + kSI > A_C$ so that the $(k + 1)$st examination falls after the DPCP in appendix 1.)

   We now derive the expected number of examinations for cases and matched controls.

   For the $ith$ case, $Ca_i$, for whom $\delta_{Ca_i}$ is less than $(A_C - (A_B + kSI))/SI$, the $(k + 1)$st examination will fall in the DPCP, and the case will be screen detected and will have $k + 1$ examinations. To visualize this result, imagine that the line in figure 2a slides to the right from $A_B$ an amount equal to $\delta_{Ca_i}$, as shown in figure 2b. For $\delta_{Ca_i} = 0$, the fourth examination in figure 2a falls prior to the interval between $A_P$ and $A_C$, and the examination is negative. As $\delta_{Ca}$ increases further, $A_4$ falls closer to $A_P$, until finally $\delta_{Ca} > (A_C - (A_B + kSI))/SI$ and the fourth examination falls beyond the DPCP; in this case, disease is symptomatic and detected clinically. For convenience, label this breakpoint $x$. Thus, we expect a proportion $x$ of type I cases to be screen detected and to have $k + 1$ examinations because this will occur if $\delta_{Ca}$ is less than $x$ and $\delta$ is uniformly distributed between 0 and 1.
FIGURE 2. Effect of timing of initial screening examination on exposure history. a. $A_{1-4}$ are the ages of screening examinations with the first occurring at the beginning of the study period ($\delta = 0$), so that three examinations occur before $A_P$, implying $k = 3$. b. For $0 < \delta < x$, case will be screen detected with an exposure history of $k + 1$ examinations. The equi-interval examination presentation may be considered a sampling array that slides to the right as $\delta$ increases. c. For $\delta = x$, there is a transition from screen to symptom detection. d. For $y < \delta < y$, cases are symptom detected with an exposure history of $k + 1$ examinations. e. For $\delta \leq y$, cases are screen detected with an exposure history of $k$ examinations.

For the screening history of the matched control, the “cutoff” date for counting the examinations of the control is the diagnosis date of the case. Thus, the control will have $k + 1$ examinations prior to age $A_C$ if the time from age $A_B$ until the first examination ($\delta_{CA}$) is between 0 and $x$, and $k$ examinations otherwise. However, if the control has $k + 1$ examinations prior to age $A_C$, there is a 50-50 chance that this $(k + 1)$st examination will follow that of the case and hence not be counted, since we only count control examinations prior to the date of diagnosis of the case. Thus, we expect $k + 1$ examinations for the matched control with probability $x/2$ and $k$ examinations with probability $1 - x/2$. Overall, then, we expect $N x^2/2$ case-type I pairs in which both the case and the control have $k + 1$ examinations, and $N(1 - x/2)$ pairs in which the case has $k$ examinations and the control has $k$ examinations.

For cases for whom $x \leq \delta_{CA} < y$, where $y = A_P - (A_B + (k - 1) \text{SI})/\text{SI}$, the $k$th screen will occur before $A_P$ and the $(k + 1)$st examination will occur after $A_C$ (figure 2d). These cases are clinically detected, comprise a fraction $(y - x)$ of type I cases, and have $k$ screens. Note that $y \geq x$, since $(y - x) = 1 - f \geq 0$, where $f = \text{PP}/\text{SI}$ and $\text{PP} < \text{SI}$. The cutoff date for counting control examinations is $A_C$, so the matched control will have $k + 1$ examinations with probability $x$ and $k$ examinations with probability $1 - x$. Overall, then, we expect $N(y - x)(1 - x)$ case-type I pairs in which both the case and the control have $k$ examinations, and $N(y - x)x$ pairs in which the case has $k$ examinations and the control $k + 1$ examinations.

For cases from whom $\delta_{CA}$ lies between $y$ and 1, the $k$th examination falls in the DPCP (figure 2e). These cases are screen detected, have $k$ screens, and comprise a fraction $1 - y$ of cases because of the uniform distribution for $\delta$.

We now consider examinations of the matched control. Since the cutoff date for counting control examinations is the diagnosis date of the matched case, the control will have $k$ examinations if his or her $k$th examination falls before $A_P$ or if it falls between $A_P$ and $A_C$ and before the $k$th examination of the case. This timing will occur with probability $y + (1 - y)/2$. 

TABLE 2. Expected counts in table 2.

<table>
<thead>
<tr>
<th>No. of examinations (cases)</th>
<th>No. of examinations (controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k + 1$</td>
</tr>
<tr>
<td>$k + 1$</td>
<td>$N(x - y)/2$</td>
</tr>
<tr>
<td>$k$</td>
<td>$N(x' - y)(1 - x' + y/2)^2/(x'(1 - x'))$</td>
</tr>
</tbody>
</table>

* Cell entries represent the expected number of the $N$ matched pairs with the indicated screening history.

† Assumes $(k + 1)$st examination falls in the detectable preclinical phase.
exposure for the controls, one might use fixed time bias differs from the well-known "structural healthy cases (2)" and by morbidity (Type B cases (2)). The undergo screening. This restricts artificially the chances of the controls having undergone a screening examination. The defining exposure is determined by the time of screening. This restricts artificially the chances of the controls having undergone a screening examination. The date of diagnosis of the matched case. For screening-detecting cases, the upper end of the time window for defining exposure is determined by the time of screening. This restricts artificially the chances of the controls having undergone a screening examination. The bias holds both for cases defined by mortality (Type A cases (2)) and by morbidity (Type B cases (2)). The bias differs from the well-known "structural healthy screenee bias," whereby a subject's risk of disease is linked to his or her eligibility and propensity to undergo screening.

As an alternative for defining the period of possible exposure for the controls, one might use fixed time periods that could be defined by age and/or secular period, perhaps using the considerations of Weiss (14) in defining these periods. Since we have not modeled this approach, our comments are speculative. One might also use the approach proposed by Morrison (1) for matched studies in which the exposed subject is the member of the matched pair screened earlier, and the unexposed subject is the one screened later or never. However, these approaches may well introduce limitations on the evaluation of potential effects because the inherent comparisons may not be sharp.

When evaluating the impact screening may have on mortality (Type A screening), the relevant screening history for the case consists of all screening tests performed prior to the diagnosis of disease, but no test performed between diagnosis and death. The treatment of the last screening examination should depend on the circumstances of diagnosis. For cases for whom the diagnosis was discovered during screening, i.e., fortuitous discovery of a cancer during a routine screening examination and without symptoms, that screening test should be included in the exposure history. By contrast, for cases who underwent the test because of the presence of suspicious symptoms, the test is part of the diagnostic workup and does not count as a screening test. It should, therefore, be excluded from the screening history. In practice, this approach requires an adequate and valid exposure history to obtain information on the reasons leading to the last test performed around the time of diagnosis. Although the definition is conceptually clear, the information may not always be available on clinical records.

Our work has several limitations. First, we only considered the null case wherein there is no benefit of screening despite 100 percent sensitivity; to evaluate bias in a more general situation would require a more complex model. Second, we only utilized the uniform distribution for all probabilistic considerations. Other distributions might be utilized to extend the model, although it is unclear which would be most appropriate for analyzing the effect in question. Third, we did not allow for variability in several key parameters, such as

<table>
<thead>
<tr>
<th>No of screening examinations</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$N_s^2(1/2 - f^2/6)$</td>
</tr>
<tr>
<td>2</td>
<td>$N_s^2(1/2 - f^2/2 - (1 - f)^2/3)$</td>
</tr>
<tr>
<td>1</td>
<td>$N_s^2(1 - s) f^2/2$</td>
</tr>
<tr>
<td>0</td>
<td>$N_s^2(1 - s) f^2/2$</td>
</tr>
</tbody>
</table>

* Cell entries represent the expected number of the N matched pairs with the indicated screening history.
† s, proportion of cohort screened at regular screening intervals (SI); 1 - s, proportion never screened.
‡ Assumes $k = 2$: all subjects who develop disease and in the screened cohort have two screening examinations at A.∞
the magnitudes of SI and DPCP, both of which seem to have an important effect on the magnitude of bias. For example, we assumed that SI > PP for all subjects, a situation that presumably does not hold in practice. Again, extension of the model, perhaps with randomization techniques, is needed to overcome this limitation.

Despite these limitations, our results suggest that case-control studies of screening in which the exposure history of the control is linked to that of the case as described here will likely yield biased results. If the final examination is counted for screen-detected cases, the odds ratio may be too large and could underestimate the degree of protective effect. If the final examination is not counted for screen-detected cases, the odds ratio may be too small and thus overestimate the degree of protective effect.

REFERENCES


APPENDIX 1

Here we consider the situation when the (k + 1)st examination occurs after $A_C$ for $\delta = 0$, so that $A_B + k SI > A_C$: this represents type II cases.

For type II cases for whom $0 \leq \delta_i < y$, where $y = (A_B - (A_B + (k - 1) SI)/SI$, the kth screen will occur before $A_P$ and the (k + 1)st examination would occur after $A_C$. These cases are clinically detected, have k countable screens, and comprise a fraction $y$ of all cases of type II. The "cutoff" date for counting the examinations of the matched control is $A_C$. Therefore, the matched control will have $k$ examinations if the kth examination falls before $A_C$, an event which arises with probability $1 - |x|$, where $|x|$ is the absolute value of $x$. Similarly, the control will have $k - 1$ examinations if the kth examination falls after $A_C$ with probability $|x|$.

For type II cases for whom $y \leq \delta_i \leq 1 - |x|$, the k examination falls in the DPCP. These cases are screen detected, comprise a fraction $1 - y - |x|$ of cases, and have k screens if we count the screening examination. (Note $1 - y - |x| = PP/SL$.) The cutoff date for counting the examinations of the matched controls is the date of the diagnosis of the case. Thus, the control will have $k$ examinations with probability $y + (1/2)PP/SL = y + (1/2)(1 - y - |x|)$ and $k - 1$ examinations with probability $|x| + (1/2)PP/SL = |x| + (1/2)(1 - y - |x|)$.
Finally, for $1 - |x| \leq \delta_i \leq 1$, the $k$th examination fall after the DPCP. These cases are clinically detected, comprise a fraction $|x|$ of cases, and have $k - 1$ screens. The cutoff date for counting control examinations is $A_C$, so the matched control will have $k$ examinations with probability $1 - |x|$, and $k - 1$ examinations, if the $k$th examination falls after $A_C$ with probability $|x|$. Based on a logistic model as described in the text, the matched odds ratio is:

$$\text{OR} = \frac{\left(y|x| + (1 - y - |x|)(|x| + (1/2)(1 - y - |x|))\right)/(|x|(1 - |x|))}{\left((|x|(1 - |x|) + (1 - y - |x|)^2/2\right)/(|x|(1 - |x|)) \geq 1.}$$

This odds ratio is clearly greater than or equal to 1, a bias for all valid values of $x$ and $y$.

APPENDIX 2

In this appendix, we first allow for variability in the timing of $A_C$ relative to $kSI$ and combine results for the two types of cases, type I and type II. We then consider the situation when a fraction, $s$, of the cohort is offered (and accepts) screening at interval $SI$ and the remainder is never screened.

We allow for variability in the timing of $A_C$ relative to $kSI$ and combine results by assuming that $(A_C - (A_B + kSI))/SI = x$ is uniformly distributed between $f - 1$ and $f$. For fixed $k$, (say $k = 2$), these limits correspond to the minimum and the maximum value of $A_C$, respectively (by the definition of $k$). This uniform distribution should arise, for example, if we include in the study all cases who die in the age range from $(k - 1)SI + PP + CP$ to $kSI + PP + CP$ and if cases die uniformly over this interval. In particular, this distribution implies that the proportion of cases of type I is $f = PP/SI$.

These two assumptions allow us to calculate expected values of the matched case-control pairs by averaging over cases who die in the age range of interest. We need only consider discordant matched pairs, since concordant matched pairs do not contribute. Using the relation $y = (A_E - PP - CP - A_B)/SI - (k - 1)$, we see that $y = x = 1 - f$ for $x > 0$ and that $y + |x| = 1 - f$ for $x < 0$. For fixed $PP$ and $CP$, we can treat $y$ as a function of $x$. Thus, we obtain the expected value of the numerator of the odds ratio by substituting $y = 1 + x - f$ to yield:

$$\text{NUM} = \int_0^f \left[x(1 - x/2) + (f - x)^2/2\right] dx + \int_{f-1}^0 \left[|x'|(1 - |x'|) + f^2/2\right] dx'.$$

Similarly, for the denominator, we have:

$$\text{DEN} = \int_0^f \left[x(1 - f)\right] dx + \int_{f-1}^0 \left[|x'|(1 - |x'|)\right] dx'.$$

Combining results gives:

$$\text{OR} = \frac{\left[3f^2 + 3(1 - f)^2 - 2(1 - f)^3 + 3(1 - f)f^2\right]}{\left[3(1 - f)^2 - 2(1 - f)^3 + 3f^2(1 - f)\right]}.$$

This odds ratio is clearly greater than or equal to one, representing a positive bias for cases of type II.

We now extend these results to the situation in which a proportion $s$ of the cohort is screened at regular intervals $SI$, and the remaining proportion, $1 - s$, is never screened. We assume that all cases have $k = 2$. The situation and study design is otherwise as described above. To illustrate the extension, consider the expected number of matched pairs in which the case has 0 screens and the control has one examination. Since screening has no effect on mortality, a proportion $1 - s$ of the cases will have had no examinations by $A_C$. The "cutoff date" for controls is, therefore, $A_C$. The matched control will have had one examination if he or she is in the screened part of the cohort of size $Ns$ and the time of his/her second scheduled examination, $A_2 = A_B + (1 + \delta) * SI$, falls after $A_C$. If the matched case is of type II ($A_C \leq A_B + 2SI$), then this would occur with probability $s|x|$. If the matched case is of type I ($A_P < A_B + 2SI \leq A_C$), then this situation can not occur since $\delta < 1$. The
projected number matched pairs, is then $N s (1 - s) |x|$. We now allow for variability in the timing of $A_C$ relative to $kSI$, by assuming as before a uniform distribution of $x$. Taking expectations with respect to $x$ gives $N s(1 - s) (1 - f)^{2/2}$ for the expected number of pairs for which the case is unscreened and the control is screened once. Similar considerations for other types of pairs lead to the results in table 3.