American Journal of Epidemiology
Vol. 151, No. 10
Printed in U.S.A.

Investigation of Design and Bias Issues in Case-Control Studies of Cancer Screening Using Microsimulation

Robert J. Connor,[1] Rob Boer,[2] Philip C. Prorok,[1] and Douglas L. Weed[3]

Using a microsimulation approach, the authors examined design and bias issues in case-control studies of cancer screening. Specifically, they looked at the impact on the odds ratio of the way in which exposure to screening is defined, the type of age matching, the time scale used, and the criteria used to determine control eligibility. The results showed that defining exposure as "ever/never" screened produced, as expected, a serious bias in favor of screening. Defining exposure as being screened no later than the time the case's cancer is diagnosed has a serious bias against screening. An alternative exposure definition—screening can occur no later than the time the case *would have been clinically diagnosed*—eliminates the bias against screening. Further, the results showed that the type of age matching and the time scale used can produce a bias against screening and that this bias can be quite strong when case-control studies are performed in populations with a periodic screening program that is the only source of screening. Finally, control eligibility criteria had little effect. *Am J Epidemiol* 2000;151:991–8.

bias; cancer (epidemiology); case-control studies; neoplasms

Considerable interest has been generated regarding the methodology of cancer screening case-control studies, with the main focus on the design (1–10) and, more recently, on the relation between theory and practice (11). This interest is likely due to the increasing use of the matched case-control design as an alternative to a randomized controlled trial to evaluate cancer screening (12–22), whether the goal is to detect and treat early cancers (e.g., breast cancer screening) or precancerous lesions (e.g., cervical cancer screening).

Many questions remain regarding the appropriate design of such studies. In this paper, we use a microsimulation approach to examine how design choices affect mortality odds ratios from case-control studies of screening for breast cancer. We limit our considerations to a setting in which screening is available only through an organized periodic screening program. Microsimulation is a computer-based technique for creating hypothetical individuals whose life histories are created and maintained in the database and thus constitute a population. Incidence and survival of breast cancer and deaths from other causes are programmed into a complex stochastic model by using parameters from published data. The computer program ages individuals, updates their disease status according to a complex set of algorithms, and changes their survival time after detection in a hypothetical screening program offered to the hypothetical individuals. Microsimulation, with its focus on individuals, has been used in a number of areas. In the area of public health care, the uses include the study of the transmission of infectious diseases (23) and the study of the costs and benefits of cancer screening programs (24–26). Microsimulation differs from traditional statistical simulations and from Monte Carlo simulations. Statistical simulations are usually concerned with small-sample properties of equations representing analytic formula rather than properties of hypothetical populations. Monte Carlo simulations tend to focus on the behavior of subgroups within a hypothetical population rather than on the behavior of the individuals comprising the hypothetical population. Nevertheless, in each case, simulation studies involve repeated runs and considerable flexibility in changing parameter values and inputs to reveal how the underlying model performs under different conditions.

In particular, we used microsimulation to examine whether various design choices for matched case-control studies of screening result in biased odds ratios. These odds ratios estimate the ratio of the mortality rate of

Received for publication January 2, 1999, and accepted for publication June 4, 1999.

Abbreviation: DCIS, ductal carcinoma in situ.

[1] Biometry Branch, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.
[2] Department of Public Health and Social Medicine, Erasmus University, Rotterdam, the Netherlands.
[3] Preventive Oncology Branch, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.
Reprint requests to Dr. Philip C. Prorok, Division of Cancer Prevention, National Cancer Institute, EPN 344–MSC 7354, 9000 Rockville Pike, Bethesda MD 20892–7354.

those screened relative to what it would have been had they not been screened (i.e., the efficacy of screening). We examined the effect of 1) three different measures of exposure, 2) the way in which controls are age matched, 3) the choice of the time scale, and 4) the eligibility criteria used to determine the controls for each case. A microsimulation computer program, MISCAN (MIcrosimulation Screening ANalysis) (24), was used to generate individual life histories in the presence and the absence of a breast cancer screening program. Case-control studies were performed using the MISCAN-generated population with screening offered. Definitions of screening exposure, the type of age matching, the choice of time scale, and the control eligibility criteria were varied. Moreover, two screening scenarios were modeled, one with no benefit from screening and the other with a benefit. How well the odds ratio estimates the efficacy of screening under these varying conditions was assessed by comparing the expected odds ratio with the average efficacy observed.

A stochastic microsimulation approach was used because a closed solution for an analytic model that reasonably reflects the complexity of population screening is not tractable. As noted above, the microsimulation approach generates populations of individual life histories within which case-control studies can be performed. The breast cancer MISCAN model was used in this study because it was well documented and had been validated for breast cancer screening (24–26).

Our general approach to the microsimulation and analysis is presented below. Details of the simulation model and the analysis are also provided.

MATERIALS AND METHODS

General approach

The output from one MISCAN run provides 1) a population of 50,000 life histories when screening is not available and is not offered and 2) the life histories of the same 50,000 individuals with a "hypothetical" organized screening program in which screening is offered through the program and is not otherwise available. A matched case-control study is performed, using the latter population. An odds ratio is calculated and provides an estimate of the efficacy of screening. With the two populations, the true impact of screening is directly assessed by comparing the outcomes with and without screening.

Odds ratio calculations use all eligible controls for each case. When an n-to-1 matched design is used, these calculations eliminate sampling variation (i.e., the variation in the odds ratio due to each case's n controls being randomly sampled from the case's potential set of controls). For investigation of the effect on the odds ratio of the definition of screening exposure, the type of age matching, the choice of time scale, and the control eligibility criteria, case-control studies were performed for various combinations of these factors. For estimation of the odds ratio expected for a specific combination of factors, the odds ratio was calculated in each of 100 populations of 50,000 life histories generated in 100 MISCAN runs. The average of these 100 odds ratios was used to estimate the odds ratio expected for a specific combination of factors. One hundred runs were used to ensure adequate precision in estimating the odds ratios. The standard error of the estimated odds ratio was calculated to assess precision.

The true efficacy of screening expected for a specific combination of factors was estimated by averaging the efficacy obtained in each of the 100 runs. The efficacy for a particular run was calculated directly by using the individual life histories with and without screening.

Two screening scenarios were used. The first reflects the situation in which there is no benefit from screening, i.e., women with breast cancer die at the same time that they would have if screening had not been offered. Put another way, there is no reduction in the number of breast cancer deaths among those detected by screening. For this "no benefit" scenario, the efficacy is 1.00. The second scenario reflects a situation in which screening confers a benefit. Here, it is assumed that there is a 50 percent reduction in the number of breast cancer deaths among those detected by screening. The resultant efficacy is less than 1.00, and it is estimated as indicated above.

Screening program simulated

The MISCAN model simulates individual life histories, taking into account three general areas of assumptions: demography, natural history, and screening effects (24, 26).

Demography. The population created by MISCAN has an age structure and a mortality from causes other than breast cancer, based on the Dutch female population.

Natural history. The natural history of breast cancer is modeled as a progression through several states. The first state is "no breast cancer." Women reside in this state until a transition occurs to one of the preclinical states when the tumor becomes detectable by screening. In the model used for this study, there are four preclinical disease states—ductal carcinoma in situ (DCIS) and three invasive states according to tumor size (<1, 1–2, and 2 cm)—and four corresponding clinical states. The duration in the different preclinical states follows an exponential distribution. The mean duration of the preclinical, screen-detectable

period increases from approximately 2.7 years at age 50 years to 6.2 years at age 70. Incidence and clinical stage distribution data govern the transition to the clinically diagnosed states. The incidence of breast cancer in the model follows that in the Dutch National Hospital Registry (SIG). After a diagnosis of breast cancer, the survival period depends on the disease state and the age of the individual at the time of diagnosis. Breast cancer mortality resulting from incidence and survival closely follows Dutch breast cancer mortality.

Screening. Within the MISCAN model, various assumptions about screening are made regarding the timing of screenings, the detection of cancer by screening, and in particular, the impact of screening. In this study, each woman without an earlier diagnosis of breast cancer is invited for screening at each of her birthdays from ages 50 through 70 years during the study period from January 1, 1990 through December 31, 2004. The population of the case-control studies consists of those women who are invited at least once. Screening attendance is modeled on Dutch experience (1, 2). The probability of attending a first screening is 0.75. When a woman is subsequently invited to attend a later round of screening, attendance depends on her behavior at the immediately preceding invitation: A woman who attended the previous screening has a probability of 0.85 of attending her next scheduled screening. For those who did not attend the previous screening, the probability is 0.20. A screening examination consists of two tests: mammography and clinical breast examination. Given the preclinical disease state at the time of screening, there is an assumed sensitivity of detecting the tumor in that state. Sensitivity of mammography for DCIS is 0.40; for invasive tumors of less than 1 cm, it is 0.70; and for larger tumors, it is 0.95. Sensitivity of clinical breast examination for DCIS and for invasive tumors of less than 1 cm is 0.00; for tumors of 1–2 cm, it is 0.50; and for tumors of 2 cm, it is 0.70 (26). The effect of detecting a cancer at screening has been simplified for easier interpretation of the outcomes of case-control studies: There is either no mortality benefit from the screening (i.e., all breast cancers follow their natural history as if screening had not been offered) or there is a benefit with a 50 percent reduction in the number of breast cancer deaths among the cancers detected by screening (i.e., 50 percent of those detected at screening with breast cancer who would die from breast cancer without screening do not die of breast cancer, but die at a later time from another cause, as governed by the demographics of the population model).

Analysis program

Case-control study odds ratio. Each MISCAN-generated population of 50,000 individual life histories with screening being offered was converted into matched case-control files with the cases defined as those diagnosed with breast cancer and dying from breast cancer in the 15-year study period, January 1, 1990 through December 31, 2004. Using the cases and eligible controls, the odds ratio was calculated using the SAS conditional logistic model package, PROC PHREG (27). For each MISCAN-generated population, case-control studies were performed for all combinations of exposure measures, type of age matching, choice of time scale, and control eligibility criteria.

The measures for exposure to screening are:

A. One or more screens regardless of when they occur (note that it is recognized that this is not an appropriate measure of exposure since once a breast cancer is detected, no further screening can occur for that individual, and therefore, controls can have a greater opportunity to be screened. This measure is used for comparative purposes and to illustrate that our methodology does produce the expected bias.);
B. One or more screens but only those that occur before or at the time that the case's cancer is diagnosed; and
C. One or more screens but only those that occur before or at the time that the case would have been diagnosed in the absence of screening. (Note that, in a real population, the time of diagnosis in the absence of screening for cancers is not observable. However, it is available in the MISCAN life histories when screening is not offered.)

We presume for the cases that screening ceases at diagnosis and, thus, that the primary effect of the three exposure measures is to vary the latest time that a control is eligible to be screened.

Figure 1 presents the three exposure measures graphically when the case's cancer is diagnosed clinically. Figure 2 presents the case when the case's cancer is detected by screening.

The categories for type of age matching are 1) 1-year age window (i.e., the age of a control must be within plus or minus 1 year of the age of the case), and 2) birth year (i.e., a control must be born in the same year as the case).

The categories for time scale are 1) calendar time, where the time of an event is the date of the event (e.g., the time of a woman's screen is the date on which the screen was done), and 2) chronologic time, where the time of an event is the age of the woman when the event occurs (e.g., the time of a woman's screen is the woman's age when the screen was done).

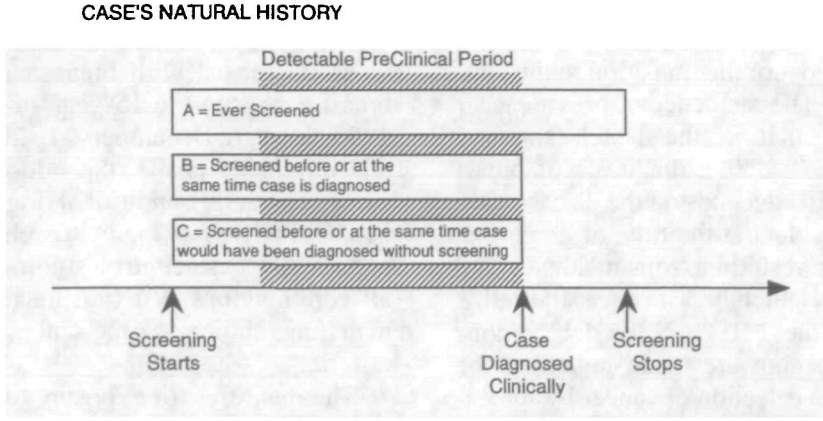

FIGURE 1. Screening exposure criteria for controls when the case's cancer is clinically detected as a result of symptoms.

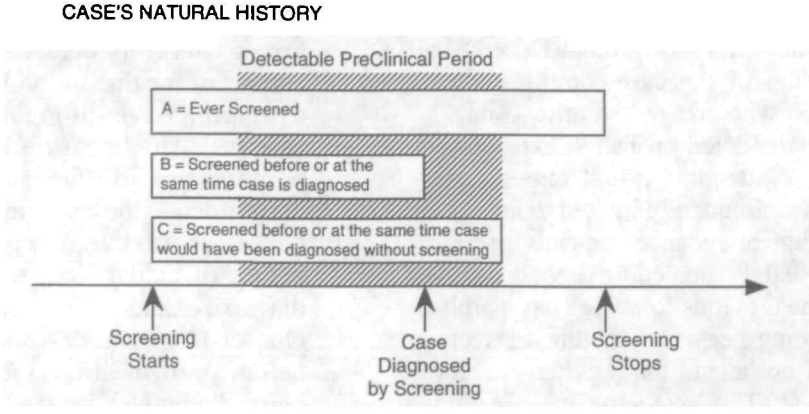

FIGURE 2. Screening exposure criteria for controls when the case's cancer is detected by screening.

The categories for eligibility criteria for age-matched controls are:

a. None (i.e., all the age-matched controls are eligible);
b. Must be alive when the case dies;
c. Must be "cancer-free" when the case is diagnosed (i.e., must not be diagnosed with breast cancer before the case's disease is diagnosed), and
d. Must be "cancer-free" when the case's disease is diagnosed and alive when the case dies.

There are 48 combinations of these factors for each MISCAN run. For both scenarios (the 100 runs with no benefit from screening and the 100 runs with benefit), the odds ratios obtained over the 100 runs for a particular combination are summarized by using the weighted average of the estimates, with the weights being the inverse of the variance of the estimates. The standard error of the weighted average is calculated to assess precision.

Efficacy. For a MISCAN run the efficacy of being screened is given by

$$\text{Efficacy} = \text{mortality rate (screened)}/\text{mortality rate (not screened)},$$

where the mortality rate (screened) is the screened-for-cancer mortality rate among those who are screened at least once, i.e., those who accepted the screening offer at least once, and the mortality rate (not screened) is the screened-for-cancer mortality rate among the same individuals had they not been screened, i.e., their screened-for-cancer mortality rate in the absence of screening.

This information is available from the MISCAN life histories with and without screening being offered and is used to calculate directly the efficacy obtained on each run. Note that this calculation is unnecessary for the scenario with no benefit from screening. Here, the true efficacy is 1.00 by definition. For the scenario with a benefit from screening, the efficacy of being screened must be calculated for each of the 100 runs with a benefit. The mean of the efficacy of being screened for the 100 runs with a benefit from screening is used to estimate the true efficacy of screening when there is a benefit from screening.

RESULTS

The results are given for exposure measures *A-C* by type of age matching (same birth year and 1-year age window), by time scale (calendar time and chronologic time), and by eligibility criteria *a-d*. Table 1 shows the summary odds ratio (and its standard error) for each combination of the factors for the 100 runs with no benefit from screening. Table 2 shows the summary odds ratio (and its standard error) for each combination of the factors for the 100 runs with a 50 percent reduction in the number of breast cancer deaths among those detected by screening. Note that in tables 1 and 2 the standard errors of the estimated odds ratios indicate that the estimates are precise.

In the title of each table, the estimate of the true efficacy for that scenario is given. For table 1, the true efficacy for the 100 runs with no benefit is 1.00, as noted earlier. For table 2, the estimate of the true efficacy of being screened for the 100 runs with a benefit from screening is 0.70.

Exposure measures

For exposure measure *A*, which allows screening anytime, we observe that 1) there is a substantial bias in favor of screening both when there is no benefit and when there is a benefit, and 2) the odds ratios are essentially the same for both types of age matching and for the two time scales and differ only slightly for the eligibility criteria.

For exposure measure *B*, an examination of tables 1 and 2 reveals a bias against screening. In table 1, all of the odds ratios for the no benefit model are, at best, 21 percent greater than 1.00 and may be as much as 176 percent greater. In table 2, all of the odds ratios for the benefit model, where the estimated efficacy is 0.70, are at least 12 percent greater than 0.70 and may be as much as 86 percent greater. The combination of factors with the smallest bias is birth year age matching with the chronologic time scale and with one of the eligibility criteria *b*, *c*, or *d*. For the no benefit scenario, the estimated odds ratio is 1.21 rather than 1.00, and for the benefit scenario, it is 0.79 when the estimated efficacy is 0.70.

When exposure measure *C* is used, the bias against screening is greatly reduced. In particular, for the combination of birth-year age matching with the chronologic time scale and with one of the eligibility criteria *b*, *c*, or *d*, the bias is nearly eliminated. In each of these

TABLE 1. Estimated odds ratios (and standard errors) given by measure of exposure, by type of age matching, by time scale, and by eligibility criteria in which no benefit from screening is assumed; true efficacy = 1.00

Type of age match, time scale, and eligibility criteria	Exposure category		
	A	B	C
	OR* (SE)*	OR (SE)	OR (SE)
Birth year			
Calendar			
a	0.44 (0.008)	2.76 (0.055)	1.17 (0.020)
b	0.41 (0.007)	2.75 (0.055)	1.15 (0.020)
c	0.42 (0.007)	2.75 (0.055)	1.15 (0.020)
d	0.41 (0.007)	2.76 (0.055)	1.15 (0.020)
Chronologic			
a	0.44 (0.008)	1.22 (0.021)	1.03 (0.017)
b	0.41 (0.007)	1.21 (0.020)	1.01 (0.017)
c	0.42 (0.007)	1.21 (0.020)	1.01 (0.017)
d	0.41 (0.007)	1.21 (0.020)	1.01 (0.017)
Within 1 year			
Calendar			
a	0.44 (0.008)	2.69 (0.053)	1.19 (0.021)
b	0.41 (0.007)	2.67 (0.052)	1.17 (0.020)
c	0.42 (0.007)	2.68 (0.052)	1.17 (0.021)
d	0.41 (0.007)	2.68 (0.052)	1.17 (0.020)
Chronologic			
a	0.44 (0.008)	1.58 (0.028)	1.16 (0.020)
b	0.41 (0.007)	1.56 (0.027)	1.14 (0.020)
c	0.42 (0.007)	1.57 (0.027)	1.14 (0.020)
d	0.41 (0.007)	1.57 (0.027	1.14 (0.020)

* OR, odds ratio; SE, standard error.

TABLE 2. Estimated odds ratios (and standard errors) given by measure of exposure, by type of age matching, by time scale, and by eligibility criteria in which a benefit from screening is assumed (a 50% reduction in the number of breast cancer deaths among the cancers detected by screening); estimated true efficacy = 0.70

Type of age match, time scale, and eligibility criteria	Exposure category		
	A	B	C
	OR* (SE)*	OR (SE)	OR (SE)
Birth year			
Calendar			
a	0.32 (0.006)	1.30 (0.026)	0.79 (0.014)
b	0.29 (0.005)	1.28 (0.026)	0.77 (0.014)
c	0.29 (0.005)	0.28 (0.026)	0.77 (0.014)
d	0.29 (0.005)	1.28 (0.026)	0.77 (0.014)
Chronologic			
a	0.32 (0.006)	0.81 (0.014)	0.71 (0.013)
b	0.29 (0.005)	0.79 (0.014)	0.70 (0.012)
c	0.29 (0.005)	0.79 (0.014)	0.70 (0.012)
d	0.29 (0.005)	0.79 (0.014)	0.70 (0.012)
Within 1 year			
Calendar			
a	0.32 (0.006)	1.30 (0.026)	0.80 (0.015)
b	0.29 (0.005)	1.28 (0.025)	0.78 (0.014)
c	0.30 (0.005)	1.29 (0.025)	0.79 (0.014)
d	0.29 (0.005)	1.29 (0.025)	0.78 (0.014)
Chronologic			
a	0.32 (0.006)	0.97 (0.018)	0.79 (0.014)
b	0.29 (0.005)	0.96 (0.017)	0.77 (0.014)
c	0.30 (0.005)	0.96 (0.017)	0.77 (0.014)
d	0.29 (0.005)	0.96 (0.017)	0.77 (0.014)

* OR, odds ratio; SE, standard error.

circumstances, the odds ratio is close to the efficacy expected; with no benefit, the estimated odds ratio is 1.01 versus the efficacy of 1.00, and with a benefit, the estimated odds ratio is 0.70 when the estimated efficacy is 0.70.

Age matching, time scale, and eligibility criteria

As noted above, these factors have little effect on the odds ratio when exposure measure *A* is used. However, for exposure measures *B* and *C*, these factors do have an impact. For exposure measure *B*, the most obvious effect is the large bias against screening when the calendar time scale is used with either type of age matching. When chronologic time is used, the bias is reduced.

For exposure measure *C*, the results are similar to those for exposure measure *B*, but with the amount of bias substantially reduced. In particular, when birth-year age matching is chosen with chronologic time, the bias is essentially eliminated.

The choice of eligibility criterion has only a small impact on the odds ratio. Criterion *a* yields the most distinct odds ratio. Its odds ratio stands apart from those for the three sets of results that require a specific vital status condition in addition to the age matching.

Thus, bias is essentially eliminated for the combination of exposure measure *C* with same birth year age matching with chronologic time and any one of the eligibility criteria *b*, *c*, or *d*. That is, if exposure measure *C* is used, it appears that, in the absence of other biases such as self- selection bias, the case-control study can give an unbiased estimate of the efficacy of being screened.

DISCUSSION

Exposure measures

A substantial bias favoring screening when exposure measure *A* is used was not unexpected. Those who do not have breast cancer are expected to have a higher rate of exposure to screening than are those who develop breast cancer. This follows because those who do not develop breast cancer will continue to be invited for screening, whereas those who develop breast cancer will not be invited for screening after their cancer is diagnosed. Hence, exposure measure *A*, which considers screening at any time as exposure to screening, will yield a greater exposure to screening for the women without breast cancer. Thus, the probability of being exposed to screening is greater for controls, most of whom will not have breast cancer, than it is for the cases, all of whom were diagnosed with and died of breast cancer in the 15-year period. Exposure measure *B* avoids this bias by not considering as exposure to screening those examinations that take place after the case's breast cancer was diagnosed. This definition of exposure has been used in cancer screening case-control studies (12–22). In our results, exposure measure *B* removes the bias of exposure measure *A* in favor of screening but at the price of introducing a bias against screening. That this would occur had been suggested (4, 10) (F. Berrino, National Cancer Institute of Milan, personal communication, 1993). In particular, Berrino postulated that exposure measure *B* would be biased against screening because a case is eligible to be screened until cancer is detected either clinically or by screening, while controls matched to a case with a cancer detected by screening are eligible to be screened only until the cancer of their matched case was detected by screening. Thus, the time interval in which controls matched to a case whose cancer was detected by screening are eligible to be screened is shorter than the time interval to the end of the preclinical period of the case. To eliminate this bias, Berrino suggested that the definition of exposure to screening be modified to include any screen up to and including the time the case's cancer would have been clinically diagnosed in the absence of screening. Figure 1 illustrates that the situation when the case's cancer is clinically detected does not change when following this suggestion, while figure 2 illustrates the change for a situation when the case's cancer is detected by screening. In practice, because the end of the preclinical period is unknown, the suggested exposure measure cannot be used exactly. At best, it might be approximated using estimates of preclinical duration. However, in our microsimulation, the detailed life histories for each individual both with and without screening are known. As a result, exposure could be evaluated when using exposure measure *C*. The results support Berrino's hypothesis. Exposure measure *B* results in a bias against screening, and exposure measure *C* eliminates this bias. At this point, our most important finding is that there is a bias against screening with exposure measure *B* that exposure measure *C* eliminates.

Age matching, time scale, and eligibility criteria

At first, it is surprising that the type of age matching (a 1-year window vs. same birth year) and the choice of time scale (calendar vs. chronologic) can have a large effect on the odds ratio. For example, for age matching, compare in table 1 the odds ratios for birth year age matching with those for age matching within 1-year when chronologic time and exposure measure *B* or *C* are used, and for time scale, compare in table 1 the odds ratios for calendar time with those for chronologic time when birth year age matching and exposure measure *B*

or *C* are used. However, on reflection, it can be seen that at least some of the bias associated with these factors is due to the characteristics of the screening program that was simulated. Recall that the screening program started on January 1, 1990, that it offered annual breast cancer screening to women from ages 50 through 70 years, and that each year's screening examinations were scheduled to be on the women's birthdays, which are uniformly distributed over the calendar year.

First, consider the strong bias against screening that occurs when calendar time is used. This is seen for both types of age matching. Because, in the simulation, birth dates are uniformly distributed over the calendar year and the birthdays are the scheduled dates for the screening, it follows that controls are equally likely to be screened before or after their case's screening date. Thus, in any screening cycle, one half of the case's controls will be screened after the case, and thus, controls will tend to have a lower probability than the cases of being screened. In particular, in the cycle in which a case is detected, about one half of the controls who are screened in that cycle would be considered as not exposed to screening if exposure measure *B* is used. If exposure measure *C* is used, the reduction in the exposure to screening is smaller because the "cutoff" time for screening is not as closely tied to the date of detection. An extreme example of this calendar time bias is a situation in which exposure measure *B* is being used and in which the case is detected at her first year of screening. Here, about one half of her controls who were screened in the first year would not be considered as exposed to screening (they are those whose screening date was later than the date on which the case was screened and breast cancer detected), and thus, there would be a bias against screening. In contrast, the chronologic time scale, which uses the individual's age at the time of the event, removes this source of bias since all of the controls of the same age would be considered exposed, even if their screens occurred after the date of diagnosis of the case.

As noted above, the randomness in birth dates and the use of the calendar time scale is of particular concern for exposure measure *B*, in which the date of the case's diagnosis ends the control's "ability to be screened," but is less important for exposure measure *C* since with this exposure measure the cutoff date for counting exposure for controls matched to cases detected by screening is not fixed by the date that the case's cancer was detected.

Next, consider the bias against screening that is seen using the 1-year window age matching with chronologic time. In the simulated screening program, when controls are age window matched, there will be matched controls whose birth year is the year before the case's birth year. In such a situation, the controls' first invitation to be screened will be either 1) for the screening cycle or round that follows the cycle in which the case was first invited, or 2) for the same cycle as the case but at an older age as well as at a later date than the case's screening would be. In the former situation, cases will have one more round in which they can be screened than these controls will have. In the latter situation, if the case is detected by screening, the control's screening in the cycle in which the case was detected would be after the case was detected; therefore, that screening would not be considered as exposure to screening under exposure measure *B* (and it might not be considered as exposure under exposure measure *C*). The result is that if age window age matching is used, cases will have a higher probability of being screened than will controls. As an extreme example, consider a case born between 1920 and 1940 whose breast cancer was detected by screening at the first screening cycle in 1990. With 1-year window age matching, there will be controls matched to the case who were born in the calendar year preceding the case's year of birth. Thus, the earliest these controls could have been screened is on their birthday in 1990. At this time, the controls would be invited for screening at both a later date and an older age than the case. Thus, under exposure measure *B*, these controls will be considered as not exposed to screening; using exposure measure *C*, fewer of these controls would be considered as not exposed. Clearly, this results in a bias against screening. For a case detected at a later round of screening, it follows that the controls will also have a lower probability of being screened than will the case, but not as low as for the situation in which the case is detected in the first round of screening.

When same birth year age matching is used, this type of bias is greatly reduced. This follows since in the simulated screening program, yearly screening is offered, and the screening cycle is a calendar year (January 1 through December 31 each year). Hence, the cases and their same-birth-year controls are eligible for screening over the same screening rounds. Screening programs to be evaluated by a case-control study may experience poorer adherence to the scheduled periodic screenings and have different lengths of screening cycles than the program simulated in this study, but similar, large biases can occur if the chronology of the particular program is not considered well enough.

The choice of control eligibility criterion *a*, *b*, *c*, or *d* has little effect on the odds ratio. This is expected because in a normal-risk population—one simulated by MISCAN—there will be relatively few individuals designated as ineligible because they died before the case died or were diagnosed with cancer before the case was diagnosed, since few individuals experience

these events. Although the differences are small, criterion *a*, which matches only on age, yields the most distinct result with, at most, a very small bias against screening. The results are essentially the same for the last three criteria. These microsimulation results reveal a substantial bias against screening when the measure of exposure to screening used is exposure measure *B* (i.e., with one or more screens but only those that occur before or at the time that the case's cancer is diagnosed). A number of studies have used this exposure measure and have shown a substantial benefit from screening. This suggests either a strong benefit from the screening or a large self-selection bias in favor of screening, or both. It is difficult to determine which explanation holds or whether both do.

The biases related to type of age matching and choice of time scale seem to be strongly related to the particular setting of screening within an organized screening program. These results may not strictly apply to settings other than the screening program scenarios that were simulated. This study did not consider a setting where the screening is "opportunistic," i.e., one in which screening is recommended but invitations to be screened are not offered, so that the time of screening is determined by the individuals themselves and is likely to be somewhat haphazard. In such a setting, similar biases can therefore be expected, but may be of a smaller magnitude.

Finally, although simulations are not truly reflective of reality and have strengths and weaknesses, these results make it advisable, when undertaking a screening case-control study, to consider the potential for bias from the measure of exposure, the type of age matching, and the time scale.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Charles Brown, Dr. Dana Friedman, Dr. David Levin, Dr. Gerrit van Oortmarssen, and Tom Riley for many helpful discussions and preparation of the manuscript.

REFERENCES

1. Verbeek ALM, Hendriks JHCL, Holland R, et al. Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen Project, 1975–1981. Lancet 1984;1:1222–4.
2. Collette HJA, Day NE, Rombach JJ, et al. Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. Lancet 1984;1:1224–6.
3. Palli D, Del Turco MRD, Buiatti E, et al. A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). Int J Cancer 1986;38:501–4.
4. Clarke EA, Anderson TW. Does screening by "Pap" smear help prevent cervical cancer? A case-control study. Lancet 1979;2:1–4.
5. Aristizabal N, Cuello C, Correa P, et al. The impact of vaginal cytology on cervical cancer risks in Cali, Columbia. Int J Cancer 1984;34:5–9.
6. MacGregor JE, Moss SM, Parkin DM, et al. A case-control study of cervical cancer screening in north east Scotland. Br Med J (Clin Res Ed) 1985;290:1543–6.
7. Celentano DD, Klassen AC, Weisman CS, et al. Cervical cancer screening practices among older women: results from the Maryland Cervical Cancer Case-Control Study. J Clin Epidemiol 1988;41:531–41.
8. Oshima A, Hirata N, Ubukata T, et al. Evaluation of a mass screening program for stomach cancer with a case-control study design. Int J Cancer 1986;38:829–33.
9. Ebeling K, Nischan P. Screening for lung cancer—results from a case-control study. Int J Cancer 1987;40:141–4.
10. Selby JV, Friedman GD, Quesenberry CP Jr, et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med 1992;326:653–7.
11. Cronin KA, Weed DL, Connor RJ, et al. Case-control studies of cancer screening: theory and practice. J Natl Cancer Inst 1998;90:498–504.
12. Berrino F, Gatta G, D'Alto M, et al. Use of case-control studies in evaluation of screening programmes. UICC Tech Rep Ser 1984;78:24–43.
13. Etzioni RD, Weiss NS. Analysis of case-control studies of screening: impact of misspecifying the duration of detectable preclinical pathologic changes. Am J Epidemiol 1998;148:292–7.
14. Friedman GD, Hiatt RA, Quesenberry CP Jr, et al. Problems in assessing screening experience in observational studies of screening efficacy: example of urinalysis screening for bladder cancer. J Med Screen 1995;2:219–23.
15. Gill TM, Horwitz RI. Evaluating the efficacy of cancer screening: clinical distinctions and case-control studies. J Clin Epidemiol 1995;48:281–92.
16. Hosek RS, Flanders WD, Sasco AJ. Bias in case-control studies of screening effectiveness. Am J Epidemiol 1996;143:193–201.
17. Morrison AS. Case definition in case-control studies of the efficacy of screening. Am J Epidemiol 1982;115:6–8.
18. Moss SM. Case-control studies of screening. Int J Epidemiol 1991;20:1–6.
19. Sasco AJ, Day NE, Walter SD. Case-control studies for the evaluation of screening. J Chronic Dis 1986;39:399–405.
20. Weiss NS. Control definition in case-control studies of the efficacy of screening and diagnostic testing. Am J Epidemiol 1983;118:457–60.
21. Weiss NS, Lazovich D. Case-control studies of screening efficacy: the use of persons newly diagnosed with cancer who later sustain an unfavorable outcome. Am J Epidemiol 1996;143:319–22.
22. Weiss NS, McKnight B, Stevens NG. Approaches to the analysis of case-control studies of the efficacy of screening for cancer. Am J Epidemiol 1992;135:817–23.
23. Habbema JDF, DeVlas SJ, Plaisier AP, et al. The microsimulation approach to epidemiologic modeling of helminthic infections, with special reference to schistosomiasis. Am J Trop Med Hyg 1996;55:165–9.
24. Habbema JDF, van Oortmarssen GJ, Lubbe JTN, et al. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79–93.
25. Habbema JDF, Lubbe JTN, van Oortmarssen GJ, et al. A simulation approach to cost effectiveness and cost benefit calculations of screening for early detection of disease. Eur J Operations Res 1987;29:159–66.
26. van Oortmarssen GJ, Habbema JDF, van der Maas PJ, et al. A model for breast cancer screening. Cancer 1990;66:1601–12.
27. SAS Institute Inc. SAS technical report P-217. SAS/STAT software: the PHREG procedure. Version 6. Cary, NC: SAS Institute, Inc., 1991:63.