Adjusting for Screening History in Epidemiologic Studies of Cancer: Why, When, and How to Do It

Noel S. Weiss

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

Received for publication April 17, 2002; accepted for publication October 17, 2002.

In epidemiologic studies of cancer, differences between exposed and nonexposed persons with regard to a history of cancer screening during the time the malignancy (or an antecedent lesion) typically is present prior to diagnosis can be a source of confounding if one of the following conditions is present: 1) the screening modality identifies premalignant changes whose treatment has the potential to prevent the cancer from developing; or 2) the number of cases included in the study would have been smaller but for the presence of screening. These situations occur commonly, arguing that consideration be given to screening history in the design (with attention to distinguishing true screening tests from those administered to persons with signs or symptoms of cancer) and analysis of epidemiologic studies of those cancers for which screening modalities are in use in the study population.

confounding factors (epidemiology); effect modifiers (epidemiology); neoplasms

Abbreviation: NSAID, nonsteroidal antiinflammatory drug.

Editor’s note: An invited commentary on this commentary appears on page 962, and the author’s response appears on page 965.

In at least some parts of the world, it is now relatively common for screening to be done for cancers of the mouth, esophagus, stomach, colon and rectum, lung, breast, prostate, and cervix and for melanoma. The presence of such screening can influence studies of the etiologies of these cancers in at least two ways. First, to the extent that the introduction of screening in a population (or its cessation) affects the actual or reported incidence of the cancer, it can interfere with the interpretation of comparisons of incidence by place and time. Second, in some circumstances a history of screening can act as a confounder in cohort and case-control studies of that cancer. The purpose of this commentary is to characterize those circumstances and to describe the aspects of the subjects’ screening histories for which control of confounding is desirable.

HOW CAN FAILURE TO CONSIDER SCREENING HISTORY DISTORT OUR ASSESSMENT OF AN EXPOSURE’S POSSIBLE INFLUENCE ON CANCER INCIDENCE?

At a minimum, in order for a variable to be a confounder in an epidemiologic study, it must bear a relation to both

Reprint requests to Dr. Noel S. Weiss, Box 357236, Department of Epidemiology, University of Washington, Seattle, WA 98195-7236 (e-mail: nweiss@u.washington.edu).
exposure and disease status. Often, the likelihood or frequency of screening is associated with a given exposure, since a history of that exposure may be the basis for selectively offering and accepting screening. For example, some providers of care might recommend that a woman initiate mammography when she is between 40 and 49 years of age only if she has a positive family history of breast cancer, a late age at first birth, or some other characteristic (or combination of them) that places her at relatively high risk of developing breast cancer. If receipt of screening mammography is related as well to the likelihood that women with breast cancer will be detected as having it during the period of study, the association between any of these characteristics or their correlates and risk of breast cancer would be confounded. Conversely, if the exposure under study is unrelated to receipt of screening (e.g., a woman’s serum level of insulin-like growth factor 1, a potential etiologic factor in breast cancer, is unlikely to be related to her having had a recent mammogram), ignoring screening histories in the analysis would not give rise to confounding.

With regard to a possible association between screening history and cancer incidence, at least one of the following conditions needs to be fulfilled before it would be appropriate to treat screening history as a confounder.

**Condition 1**

The screening modality has the potential to identify not only the cancer per se but also the treatable antecedents of the cancer. Although the available screening modalities for cervical or colon cancer can reveal the presence of invasive cancer, most of these also have the potential to detect premalignant conditions (intraepithelial neoplasia and adenomatous polyps, respectively) whose treatment can lead to a reduction in cancer incidence (1, 2). So, for example, in a study of use of nonsteroidal antiinflammatory drugs (NSAIDs) in relation to the incidence of colon cancer, adjustment for a history of colon cancer screening would be needed to yield a valid result if NSAID use were found to be associated with screening. If a relatively higher proportion of users of NSAIDs had received screening during the period of time in which the premalignant condition is detectable prior to its malignant degeneration, a negative association would be observed between NSAID use and colon cancer incidence even in the absence of any true antineoplastic action of these drugs. In contrast to this example, consider a hypothetical study of NSAID use and prostate cancer. Screening for prostate cancer, by means of a digital rectal examination or the measurement of prostate-specific antigen levels in serum, can detect the presence of prostate cancer itself, but not any precursor lesions. Since screening is unable to lead to the prevention of prostate cancer occurrence, in order for it to act as a confounder the next condition would have to be met.

**Condition 2**

The number of cases included in the study would have been smaller but for the presence of screening. Often, the pathologic changes in a tissue that can later give rise to malignancy will be detectable only by means of screening. This is true for cervical preneoplasia, which would not be identified in the absence of a Papanicolaou smear or the application of some other cervical screening modality. Most case-control studies of cervical preneoplasia have sought to control for potential confounding by restricting controls to women who have been screened as negative (3, 4). During investigation of exposures that themselves are related to the likelihood or frequency of being screened (use and type of contraception, perhaps), this approach is sound: If a study of cervical preneoplasia fails to select cases and controls who are comparable with regard to a history of screening during the period of time the preneoplastic condition is presumed to be detectable, the study will not be able to separate potential risk factors for receipt of screening from those for the presence of the condition itself.

**Etiologic studies of cancer.** Apart from those instances in which screening identifies premalignant changes whose successful treatment averts the development of some cancers, screening does not have the capacity to cause or prevent cancer. Nonetheless, the observed incidence of cancer in a population over a given period of time can be substantially influenced by the level of early detection activity. This can occur for two reasons.

First, a test identifies some lesions that are labeled as cancer but that truly do not have malignant potential. Among endometrial hyperplasias that develop in postmenopausal women who take unopposed estrogens, it is likely that some are misdiagnosed as cancer when an endometrial biopsy is performed in response to estrogen-induced uterine bleeding (5). The distinction between endometrial hyperplasia and cancer can be subtle, and pathologists differ regarding the criteria they use for distinguishing the two conditions. In studies of the relation of estrogen use to risk of endometrial cancer, this nondifferential misclassification of disease status (false positives particularly occurring among hormone users) likely is responsible for a portion of the observed excess risk. As another example, there is reason to believe that screening for what is typically a very aggressive malignancy, lung cancer, can uncover some lesions that though labeled as cancer, in actuality do not have the potential to cause death. The existence of these cases of “pseudodisease” (6) has been deduced from data gathered in randomized trials of intensive screening for lung cancer by means of chest radiographs and sputum cytology (7). These trials have not observed a reduction in the mortality rate from lung cancer compared with a regimen of less intensive screening, even though the case fatality after more than 20 years of follow-up among persons diagnosed with lung cancer is relatively smaller in intensively screened persons (8).

Even if a screen-detected lesion truly does have malignant potential, if the lesion fails to manifest that potential for an extended period of time, the person harboring it may die from other causes before it is detected. This is likely the situation for a number of screen-detected prostate cancers in older men: Some of these lesions are indolent enough that death from coronary disease, stroke, and so on would take place before the cancer could produce symptoms or signs sufficient to lead to its diagnosis in the absence of screening.

Second, the use of the screening test in the population under study is not constant over time. During the follow-up
interval of a cohort study or the intake period of a case-control study, screening will give rise to the recognition of cancer in some persons who would otherwise not have been diagnosed until later, after that period had ended. If the level of screening in the base population for the study were higher during the period of case accrual than it had been previously, there would be more screen-detected cases “gained” from the subsequent period than would be “lost” to the preceding one. Overall, there would be more cases in the study than if screening had not been on the rise, and each of the “excess” cases would be positive for a history of screening.

FOR WHICH ASPECT OF A STUDY SUBJECT’S SCREENING HISTORY SHOULD WE CONTROL?

The only screening that can influence the recognition of cancerous or precancerous lesions is that which takes place while the pathology is present and detectable by means of the particular screening modality. Screening prior to that time cannot lead to recognition of the cancer or, in the case of premalignant pathology, its prevention. We cannot know in a given person how far back the relevant period of time extends. (For premalignant lesions, we also cannot know when it ends, i.e., when the malignant degeneration occurred.) However, using 1) the proportion of previously unscreened persons who are found to have a particular cancer once they are screened, 2) incidence rates of that cancer, and 3) the knowledge that the duration of preclinical, detectable disease (if it inevitably leads to cancer) is approximated by the ratio of “1” and “2” just described (9), it is possible to estimate on average how long screen-detectable lesions have been present. In a case-control study of breast cancer, for example, the relevant period during which to consider screening mammography as a potential confounder is the most recent 1–2 years prior to a woman’s first symptoms or signs of breast cancer (and a corresponding duration in controls), the estimated duration of preclinical, detectable breast cancer by means of mammography (10). In a study of cervical cancer, that interval would be considerably longer, corresponding to the period of time that a screen-detectable cervical neoplasm or treatable precursor lesion typically is present (11). Since most cancer-screening tests have relatively high sensitivity, potential confounding can be dealt with adequately simply by controlling for a history of any screening examinations during the period corresponding to the preclinical detectable phase of the particular tumor. Had the tumor been present, even one examination would have had a high likelihood of identifying it. For those tests of lower sensitivity, there is little potential for confounding in the first place.

THROUGH WHAT MEANS SHOULD POTENTIAL CONFOUNDING BY SCREENING HISTORY BE CONTROLLED?

Most often, confounding by screening during the period of time corresponding to the preclinical phase of the tumor will be dealt with in the same way as confounding by any other variable, that is, by means of a stratified or multivariate analysis. It is also possible to confine the analysis to that portion of the study population in which there is little or no confounding on the basis of screening. For example, in a cohort study examining the association between use of postmenopausal hormone therapy and the incidence of breast cancer, Joffe et al. (12) restricted one analysis to those women who had a mammogram in the 2 years prior to a follow-up period in which incident cases were to accrue. Only in this subgroup of the study population did hormone users not go on to have an appreciably higher utilization of mammography than did hormone nonusers.

MISCLASSIFICATION OF SCREENING STATUS

In epidemiologic studies, there can be errors in ascertaining both the fact and timing of receipt of cancer screening. Subjects’ memories may be faulty, and if the test is one that is performed on a blood specimen (e.g., a prostate-specific antigen test), the person from whom the specimen was taken may not have been told the test was done. Outpatient medical records generally can be expected to be a more accurate means of documenting the receipt of a particular test, but often these records are not (or not readily) available. When there is misclassification of screening status, even to a similar degree between persons with and without cancer, efforts to control for this variable will not fully remove the confounding it introduces.

Unfortunately, in these studies there is considerable opportunity for differential misclassification of screening status as well. Many of the same tests used to screen for the presence of cancer are also administered as part of the diagnostic evaluation of persons with signs or symptoms of cancer, and medical records may not capture the distinction. For example, consider a positive prostate-specific antigen test done in an asymptomatic man who has just had a prostatic nodule detected on a digital rectal examination. Even if the presence of the nodule had been noted in the record, the status of the prostate-specific antigen test would depend on whether or not it would have been done irrespective of the result of the digital examination: “diagnostic” if done only in response to the presence of the nodule or “screening” if it would have been performed in any event. However, the intent of the provider of care is unknown, and such tests cannot be unambiguously classified as “screening” or not.

Since a far greater proportion of persons with cancer than persons without cancer will have tests done in response to signs or symptoms of that condition, the potential for misclassification of these as screening tests particularly affects the former group. An example of the impact of differential misclassification of this sort on the results of an epidemiologic study of cancer is illustrated in table 1. The upper portion contains hypothetical data from a case-control study in which screening status has been ascertained without error. A history of one or more screens during the presumed preclinical detectable period was more common among cases than controls and also more common among the exposed than the nonexposed. Thus, even though the odds ratio within each of the two strata based on screening history was 2.0, the crude odds ratio relating exposure and disease was 2.33.

In the lower portion of table 1, it is assumed that half of the cases who truly had not been screened had undergone diag-
nostic tests that were incorrectly categorized as being screening in nature. It is also assumed that no such misclassification took place among controls, since perhaps only a few of them at most would have had signs or symptoms suggestive of the presence of the cancer. Although the odds ratio in the unscreened stratum remains correct (odds ratio = 2.0), that in the screened stratum is now biased (odds ratio = 1.56). The odds ratio adjusted for (misclassified) screening status is now 1.77, lower than both the true adjusted value of 2.0 and the crude odds ratio of 2.33. Thus, unlike the typical result of nondifferential confounder mismeasurement, in this instance it is biased toward the null and away from the crude.

Even when the screening histories of study participants are obtained without error, residual confounding can be present when we restrict or adjust for screening status if we misestimate the interval during which screening history is relevant for the potential detection of early cancers or precancers. In practice, it would be desirable to see if the size of the association between exposure and cancer incidence is sensitive to changes in the width of that interval. Residual confounding also can be present if the exposure has a bearing on the sensitivity of screening. For example, use of hormone therapy by postmenopausal women often leads to increased mammographic density, diminishing the ability of the screening modality to identify breast tumors (13). In this instance, adjustment for the fact of screening during the 2 years prior to diagnosis of breast cancer, for example, will not remove all possible confounding by differences in the receipt of effective screening between hormone users and nonusers.

IS IT DESIRABLE TO CONTROL FOR THE RECEIPT OF TESTS PERFORMED IN RESPONSE TO SIGNS OR SYMPTOMS OF THE CANCER?

As discussed earlier, studies of the potential influence of hormonal medications on the incidence of endometrial cancer potentially can be biased by the combination of 1) differential receipt of endometrial biopsy between users and nonusers of hormones and 2) ambiguities in distinguishing benign from malignant endometrial changes. Some investigators (14) have sought to overcome this problem by selecting as controls women who received an endometrial biopsy and were not found to have cancer. If the conditions that led to the endometrial biopsy in controls were unrelated to hormone use, this approach would have provided a valid result.

Unfortunately, in this instance, this was probably not the case: Women who undergo endometrial biopsy generally do so for uterine bleeding, which is a much more common result of hormone therapy than the development of cancer. In this case-control study, a substantial amount of selection bias likely was present, in that the proportion of hormone users in the controls likely was considerably in excess of the proportion among women in the population at risk of endometrial cancer. This example serves to warn us that, while control for receipt of screening is often desirable to control confounding in epidemiologic studies of cancer, control for receipt of diagnostic testing can produce misleading results and should be undertaken only with great caution. Bias due to misclassification of disease generally can be dealt with in other ways, such as by conducting analyses in which cases are restricted to persons whose tumor is unequivocally malignant.

ACKNOWLEDGMENTS

The author is grateful to Mary Anne Rossing and Lynda Voigt for helpful comments on an earlier draft of this manuscript.

REFERENCES