Analysis of Case-Control Studies of the Efficacy of Screening for Cancer:
How Should We Deal with Tests Done in Persons with Symptoms?

Noel S. Weiss

Many tests used to screen asymptomatic individuals for the presence of cancer also are used in symptomatic persons to help determine the presence and/or extent of that cancer. For example, mammography can be used both to screen for breast cancer in a woman coming for an annual checkup and to help in the diagnosis of a woman who has recently noticed a lump in her breast. Case-control studies that seek to evaluate the efficacy of a screening test of this type in reducing cancer mortality need to consider how to deal with tests done in persons with symptoms possibly due to cancer.

On the surface, the resolution to this issue seems obvious. If a test had been done in the absence of symptoms of the cancer, it should be categorized as a screening test; otherwise, it should be considered a diagnostic test (1). However, in the conduct of a specific study, circumstances may be present that make the choice among analytic options not at all obvious.

INADEQUATE INFORMATION REGARDING THE REASON(S) FOR TESTING

Often it is not possible to obtain reliable data on the reason that a test was performed, but only on whether it was performed. For example, some case-control studies of screening efficacy (2) have had access only to computerized data on persons enrolled in a given health care plan. These files generally contain information on diagnoses and on tests administered, but not on signs or symptoms present when those tests were done. Other studies have had access to medical records, but those records did not provide adequate detail regarding the reason that a particular test was done in a particular patient (3, 4). Case-control studies of the efficacy of screening for cancer that are based on interviews may have a similar problem, in that the interviews must be conducted months to years following the initial cancer diagnosis, and then often with quite ill patients or (in studies of fatal cancer) with their next of kin. These individuals could be expected to have difficulty in recalling any symptoms or signs that may have prompted a test to be done, or even whether the test was done at all.

Faced with data of uncertain quality and completeness regarding reasons underlying the performance of tests, some investigators have chosen to disregard altogether in their analyses tests done close to the time of diagnosis (and those done at a corresponding point in time among controls). These investigators have been concerned, appropriately so, that a far higher percent of cases than controls would have had recent diagnostic tests (i.e., tests performed in response to cancer symptoms), and that misclassification of some of these diagnostic tests as screening tests would lead to a falsely high proportion of cases believed to have received screening. This would lead to a falsely high odds ratio, and thereby a falsely low estimate of the efficacy of cancer screening. Unfortunately, this analytic approach produces a bias of its own. To illustrate this bias, let us consider the following hypothetical set of circumstances:
1. The cancer for which screening is being performed has a detectable, occult phase of 3 years duration prior to clinical diagnosis.
2. The screening test in question is 100 percent sensitive in identifying occult tumors.
3. The available treatment for preclinical lesions detected via screening is ineffective in altering the natural history of this disease, and so the test does not lead to any reduction in cancer mortality.

Assume that a case-control study of the ability of this screening test to reduce the occurrence of fatal cancer were done in a given population. What if that study were to consider only tests received during the period beginning 3 years prior to diagnosis and ending 6 months prior to diagnosis, in order to avoid the problems indicated above regarding the uncertainty of the reasons for testing around the time of diagnosis? If that were done, none of the cases in the study persons who died from the particular cancer would have received a test during that 2.5-year interval! Because the test is 100 percent sensitive, if a test had been performed in a person with cancer 6–36 months prior to the time it would otherwise have been diagnosed, the date of diagnosis would have occurred at the time of screening, not 6–36 months later. Whatever the frequency of screening present in the controls during that 6–36-month period, the odds ratio obtained in the study would be zero, and so the efficacy of the screening test in leading to a reduction in mortality would be estimated as 100 percent. However, because the therapy following diagnosis of the occult tumor in fact leads to no life being saved, the correct odds ratio is 1.0, and the efficacy of screening should have been estimated as zero. It is therefore clear that analyses of case-control studies of screening that confine their attention to tests done well before the diagnosis of cancer, most of which will be negative tests, can lead to a substantial downward bias in the odds ratio relating the conduct of the test to the likelihood of a fatal outcome.

The opposite approach has been put forward by Gill and Horwitz (5), who recommend that, in one analysis of a case-control study of screening efficacy, all tests performed be considered, irrespective of symptom status at the time of testing. They argue that this strategy would provide an upper bound for the estimate of the odds ratio associated with screening (and thus a lower bound to the estimated efficacy of the screening test). However, I believe the strategy would give highly misleading results in many instances. Take the example of the efficacy of screening mammography in reducing breast cancer mortality. Nearly all women who recently died of breast cancer in the United States would have undergone a mammogram around the time of diagnosis of their tumor. In some, it would have been the first step that led to the detection of the cancer (i.e., screening); in the rest, it would have been done to evaluate a breast mass or other symptom or sign of breast cancer that had been identified in some other way. Whatever the level of screening during that same period of time in the population from which the cases arose, an analysis which assumes that all tested persons actually are “screened” persons would obtain an odds ratio well in excess of 1.0. If every woman with breast cancer who had not been screen-detected had undergone a mammogram as part of her diagnostic evaluation, the observed odds ratio would be infinity. Gill and Horwitz (5) contend that counting all tests received, screening or otherwise, is analogous to the strategy in randomized controlled trials of maintaining all patients in the arm of the trial to which they have been assigned, irrespective of their compliance with the intervention they were supposed to have received. I believe 1) that infinity is not a very useful upper bound! And 2) that the analogy is not an apt one. In randomized trials, adherence to the policy of “intention-to-treat” will lead to a bias whose direction can be predicted, i.e., a bias toward the null. This policy avoids the possibility of bias in an unpredictable direction, a bias which often would be of substantial magnitude. In case-control studies of cancer screening, it is evident from the example above that an analysis which considers all tests, irrespective of the indication for those tests, can lead to a result that is biased quite a distance away from the null.

To sum up, if the reason for testing cannot be determined, a case-control study cannot give a valid estimate of efficacy either by ignoring tests done close to the time of diagnosis or by including all such tests in the analysis. Only those case-control studies that can obtain accurate information on the reason(s) for tests done on study subjects will be able to contribute useful results.

UNCERTAIN RELATION OF CANCER TO THE SYMPTOM THAT LED TO TESTING

It may be that a variety of signs or symptoms prompt the performance of a given test which, if positive, can be indicative of more than one form of pathology. For example, a fecal occult blood test may be obtained in a patient who complains of indigestion or heartburn as part of the assessment of possible pathology in the upper gastrointestinal tract. In a case-control study of the efficacy of fecal occult blood screening in reducing mortality from colorectal cancer, such a patient would be included along with asymptomatic persons in the group deemed to have
undergone a “screening” examination if there were no reasons to believe that indigestion and heartburn are a manifestation of a colorectal tumor. Undoubtedly, in some situations there will be considerable uncertainty regarding the relation of a given symptom or sign to the presence of a particular cancer. In these situations, it might be wise to consider a modified version of the suggestion of Gill and Horwitz (5) and to conduct two analyses. In each analysis, persons whose test was performed in response to likely cancer symptoms would be labeled as “not screened.” However, for persons with symptoms of uncertain relation to the cancer, separate analyses could be done that would include or exclude them from the analysis.

DEALING WITH SYMPTOMATIC PERSONS WHO WERE PREVIOUSLY EVALUATED FOR THE PRESENCE OF CANCER

Another potential ambiguity in the analysis and interpretation of case-control studies of screening can arise when some subjects have symptoms that bear a complex relation to the presence of a particular cancer (6). For example, while most breast lumps do not indicate the presence of a cancer per se, women with these lumps (depending on their histology) are at an increased risk of developing breast cancer in the future. Most women with a breast lump would receive a mammogram once that lump comes to the attention of their physician and, even if no cancer were present at that time, subsequently they would tend to be screened more frequently than other women. To help decide what to do about persons who previously have had a symptom or sign and who, while free of cancer at that time are later at an increased risk, let us consider the hypothetical data set in table 1. In this example, it is assumed that there is a one-year period prior to clinical diagnosis during which the tumor would be detectable with the use of the test. The data in table 1 describe two groups of persons:

1. those with no history of the symptom or sign; and
2. those who developed the symptom or sign more than one year earlier, and who were tested and found to be negative at that time. (Tests done in persons who developed the symptom or sign during the most recent year prior to diagnosis would be entered into the “no history” stratum and counted as not having been screened, unless the test had been done prior to the presence of the symptom or sign.)

In both group 1 and group 2, a history of “recent” screening is associated with a 50 percent lowering of mortality (see table 1). However, due to 1) the over-representation of persons with a prior indication for testing among cases, and 2) the relatively greater frequency of recent screening in persons with that prior indication, the crude odds ratio is not 0.50, but 0.57.

In order to obtain a result free of this confounding, the strata must be kept separate to allow for adjustment. If the estimated efficacy differs between persons with or without a prior indication for testing, as it may, for example, for mammography in women who previously did or did not have a prior lump identified, then the stratum-specific results can be presented.

It should be noted that the information required for this type of stratified analysis, i.e., the presence or absence of a given symptom or sign prior to the time the tumor was potentially detectable, both in persons who were and were not tested during that earlier period, may not be easy or even possible to collect in many settings.

CONCLUSIONS

Based on the foregoing, I recommend that in the analysis of case-control studies of the efficacy of screening for cancer:

1. Among persons with symptoms judged to be the result of the cancer itself, we should disregard tests that are first performed in response to those symptoms within the period of time that the

<table>
<thead>
<tr>
<th>Screening during year prior to diagnosis</th>
<th>Sign or symptom first evaluated &gt;1 year before diagnosis*</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No†</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>No</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Total</td>
<td>90</td>
<td>120</td>
</tr>
</tbody>
</table>

Odds ratio 0.50 0.50 0.57

* Among controls, one year prior to a corresponding date.
† Either no sign or symptom was ever present, or it was present only within the year prior to diagnosis.

Am J Epidemiol  Vol. 147, No. 12, 1998
cancer is believed to be detectable prior to diagnosis. The person should be entered into the analysis as “not screened” during that period of time.

2. We should consider tests performed in response to symptoms or signs that likely are indicative of some condition other than the cancer of interest as screening tests for that cancer. If there is uncertainty about the relation of the cancer to a particular symptom, dual analyses can be done that either include or exclude persons with tests done for this reason.

3. Once a symptomatic person has been tested as negative, and thereby believed to be free of the cancer at that time, we should tabulate subsequent tests performed in that person as screens (until or unless the symptoms change in such a way as to suggest the presence of a cancer). However, if the presence of the symptom or sign is found to be associated with an increased risk of the cancer, such persons (both cases and controls) are to be kept in a separate stratum when comparing cases and controls for the receipt of screening during the presumed period of detectability of the case’s cancer.

ACKNOWLEDGMENTS
This work was supported in part by grant no. R35 CA 39779 from the National Cancer Institute.
The author thanks Drs. DeAnn Lazovich and Mary Anne Rossing for their helpful suggestions.

REFERENCES