What Is a Screening Test?

Misclassification Bias in Observational Studies of Screening for Cancer

John Concato, MD, MS, MPH

OBJECTIVE: To demonstrate the importance of accurately identifying clinical distinctions of subjects in observational studies of screening.

DESIGN: Simulated case-control studies.

SETTING: The West Haven Veterans Affairs Medical Center.

PATIENTS: Fifty-two men diagnosed with prostate cancer in 1988 or 1989 had 252 digital rectal examinations (DREs) in the preceding 5 years. A classification scheme used patient symptoms and the results of prior DREs to assign the last DRE before the diagnosis of cancer to one of the following categories: definite screening, likely screening, probable screening, not screening, or other and unknown. Sixty-five percent of the DREs were classified as definite or likely screening, and another 15% were classified as probable screening.

MAIN RESULTS: Changing the definition of a screening DRE from one including to one excluding probable DREs lowered the frequency of screening in case subjects more than it did in case controls, and thus lowered the odds ratio (OR), making screening appear to be more protective. Even when DRE was not protective, the ORs for the effectiveness of screening with the more restrictive definition ranged from 0.21 to 0.83 in 36 simulated case-control studies that differed according to the frequency of screening, the prevalence of cancer in case controls, and the extent of misclassification error.

CONCLUSIONS: If clinical distinctions in the performance of screening tests are not classified appropriately, observational studies will misrepresent the proportion of subjects exposed to screening interventions and produce biased results.

KEY WORDS: prostate cancer; screening; case-control study; bias; digital rectal examination.

J GEN INTERN MED 1997;12:607-612.

Observational research methods, including case-control studies, are increasingly used to evaluate the effectiveness of cancer screening.¹⁻¹⁰ In a typical case-control study of screening, case patients are persons who died of

cancer and control patients are persons who did not die of cancer. "Exposure" to screening is then measured in both groups. Such a study will conclude that screening was protective if it occurred more frequently in control than in case patients.

The Department of Veterans Affairs is funding a case-control study to determine whether screening for prostate cancer with prostate-specific antigen (PSA) and digital rectal examination (DRE) reduces mortality. In this study, case patients are men with prostate cancer who subsequently die, and control patients are age-matched men with or without prostate cancer who were alive when the corresponding case patient died. Exposure to PSA and DRE before the diagnosis of prostate cancer is being determined by a review of medical and laboratory records.

Accurate identification of PSA and DRE, and accurate classification of these tests as screening tests, are crucial to the validity of this approach. Misclassification is particularly likely for DRE, because DRE is used to screen for prostate cancer in men who have other prostatic disorders, and some of these disorders can mimic the symptoms of prostate cancer. For example, the analysis of screening DREs might exclude DREs in men with stable nocturia who subsequently were found to have cancer, by assuming that the purpose of the DRE was to diagnose cancer and nocturia was the reason to suspect cancer. In this approach, when a DRE is done in a case patient with stable nocturia and cancer is found, the DRE would not be classified as a screening DRE. In contrast, when a DRE is done in a control patient with identical symptoms from benign prostatic hyperplasia (BPH) and no cancer is found, the DRE would be classified as a screening DRE. This type of analysis would bias the results because all case patients, compared with very few control patients, have prostate cancer. Therefore, even if the reasons for the DREs were identical in the two situations, misclassification of screening DREs would be more common in case patients.

This type of misclassification also occurs if a study arbitrarily excludes some DREs rather than determining whether each DRE is a screening test or not. A report from one such study states that "all medical charts were reviewed to determine the date of diagnosis of prostate cancer and to conceal (for cases and controls) all material entered after three months before this date."⁶ This strategy apparently was intended to conceal the subjects' case-control status and thus prevent other types of bias, but it also ignored all DREs done less than 3 months before diagnosis, some of which could have been screening tests.

Received from the Medical Service, West Haven Veterans Affairs (VA) Medical Center, VA Connecticut Healthcare System; and the Clinical Epidemiology Unit, Department of Medicine, Yale University School of Medicine, New Haven, Conn.

Dr. Concato is supported by a Career Development Award from the VA Health Services Research and Development (HSR&D) Service; this project was funded in part by VA HSR&D Merit Review Award 133.

Address correspondence and reprint requests to Dr. Concato: Yale University School of Medicine, 333 Cedar St., SHM IE-61, New Haven, CT 06510.

The overall goal of the research reported here was to demonstrate the importance of accurately identifying clinical distinctions in patients and their diseases when conducting observational studies of screening. The example used is a case-control study of the effectiveness of screening DREs for reducing prostate-cancer mortality. To achieve that goal, a classification scheme of reasons for performing DREs was developed.

METHODS

Subjects

Eligible subjects were men diagnosed with prostate cancer at the West Haven Veterans Affairs Medical Center in 1988 or 1989. Seventy-two men with histologic evidence of prostate cancer were identified through Pathology Department reports. Medical records were unavailable for 14 subjects. Of the 58 subjects whose records were available, 6 subjects were excluded: 4 because their prostate cancer was diagnosed before 1988, 1 because the histologic diagnosis was squamous metaplasia, and 1 because the record was incomplete. The remaining 52 records were reviewed by an experienced research assistant using a form developed for this study. The author reviewed completed forms with the research assistant weekly, and problems were resolved by examining medical records together.

Data Collection

Information was collected on all DREs during the 5 years before the diagnosis of prostate cancer; this study focuses on the last DRE before the diagnosis of prostate cancer. Because a DRE often is done when prostate cancer is suspected but before it is diagnosed, the last DRE before the diagnosis may not be a screening test. For example, consider a man with a normal DRE 9 months ago who came to the emergency department with acute urinary-tract obstruction. A DRE in the emergency department that led to the diagnosis of prostate cancer was done to diagnose new urinary-tract symptoms suggestive of cancer, and thus was not a screening DRE. The DRE done 9 months earlier was the last screening DRE. Alternatively, consider a man who receives a routine DRE each year during a periodic health evaluation. After several years of normal DREs, the next DRE reveals a prostatic nodule, and a prostatic biopsy reveals cancer. Because the nodule was unexpected, the last DRE was a screening DRE.

To distinguish between screening and nonscreening DREs, information about three clinical factors was recorded for each DRE: the type of clinical setting where the DRE was done, for example, primary care clinic, urology clinic, emergency department, or hospital ward; the patient's symptoms, for example, no symptoms, urinary-tract symptoms, or gastrointestinal symptoms; and the results of prior DREs, for example, whether they were normal, consistent with BPH, or suggestive of cancer.

Although information about the type of clinical setting was collected, it could not be used to classify DREs as screening DREs. For example, many patients received annual DREs as screening tests in primary care clinics, but others received DREs in the same clinics as part of an evaluation for prostate cancer. Conversely, one patient presenting to the emergency department with symptoms of urinary-tract obstruction received a DRE to diagnose cancer, but another patient who presented after choking on his food received a routine screening DRE. In addition, although most DREs done in a urology clinic were done to diagnose cancer in patients referred for consultation, screening DREs also were done for patients followed for BPH.

The urinary-tract or gastrointestinal symptoms that prompted each DRE are listed in Table 1. Five symptom patterns were identified: no symptoms, only gastrointestinal symptoms, chronic and stable urinary-tract symptoms, chronic and worsening urinary-tract symptoms, and acute urinary-tract symptoms. (Patients with urinary-tract symptoms also may have had gastrointestinal symptoms.) These symptom patterns were used to determine whether the DRE was a screening test as follows:

- ◆ In the absence of symptoms, the DRE was done for screening or to follow-up another test, such as an abnormal PSA or a previously abnormal DRE.
- ◆ If only gastrointestinal symptoms were present, the DRE was done to diagnose gastrointestinal disorders. Such a DRE was a screening test for prostate cancer if specific results about the prostate were reported. If prostate-specific results were not mentioned, however, the DRE could not be further classified because it could not be determined whether prostate-specific results were appreciated but not documented or not appreciated.
- ◆ If there were chronic, stable urinary-tract symptoms, which usually occurred with a diagnosis of BPH, a DRE was a screening test unless a specific concern of malignancy was noted, for example, when "rule-out prostate cancer" was written.
- ◆ If there were chronic but worsening urinary-tract symptoms, the DRE was not a screening test, for example, when a patient with long-standing nocturia developed new urinary hesitancy.
- ◆ If there were new urinary-tract symptoms, the DRE was not a screening test.

The classification of some DREs as screening tests also depended on the results of previous DREs. A DRE could not be a screening DRE if a previous DRE was done because cancer was suspected. For example, a DRE could not be a screening DRE when it was done in a urology clinic on a patient referred because of a suspicious prostate gland found during DRE in a primary care clinic. The types of results for previous DREs are listed in Table 2. Four patterns were identified: no DREs, normal results,

Table 1. Symptoms in Men Receiving Digital Rectal Examinations

Urinary-Tract Symptoms	Gastrointestinal Symptoms
Burning on urination	Abdominal pain
Decreased voiding	Diarrhea
Dribbling	Hemorrhoids
Flank pain	Rectal bleeding ("BRBPR")*
Hematuria ("blood")	Stomach pain
Incomplete emptying	Constipation
Infection	Hematemesis
Obstruction	Melena
Urgency	Rectal pain
Decreased force of stream	
Difficulty urinating	
Dysuria	
Frequency	
Hesitancy	
Incontinence	
Nocturia	
Obstructive voiding symptoms	
Urinary retention	

*BRBPR indicates bright red blood per rectum.

abnormal results that were consistent with BPH, or abnormal results that were suggestive of cancer. For example, a "negative" result was considered normal, a "2+ enlarged" or "35-g" prostate was classified as abnormal but consistent with BPH, and a "prominent left lobe" was abnormal and suggestive of cancer.

Any abnormal finding that was not consistent with BPH was classified as suggestive of cancer. For example, a "boggy" prostate gland was classified as suggestive of cancer because a patient with such a finding probably would not be followed for BPH, but rather would receive a urologic evaluation for suspected prostatitis that could reveal a cancer if it were present. Also, any abnormal finding that was consistent with both BPH and prostate cancer was classified as suggestive of cancer, for example, a "firm" prostate gland.

Classification Scheme

For each DRE, any symptoms present when the DRE was done and any results from previous DREs were combined to determine whether the DRE was a screening DRE. This classification scheme is described in Table 3 and classifies each DRE as a definite screening test, a likely screening test, a probable screening test, not a screening test, or other and unknown. The distinction between a likely and a probable screening test depends on the reasons for category assignment, not necessarily on the frequency of cancer.

Simulated Case-Control Studies

The impact of misclassification bias was assessed in a series of simulated case-control studies using the 52 case patients and hypothetical control patients (arbitrarily of the same sample size). The reference simulation included case and control subjects with the same frequency of screening, so that the odds ratio (OR) for the effectiveness of screening was 1.0. The effect of misclassification error was then measured by changing the frequency of screening, the prevalence of prostate cancer in control subjects, and the extent of misclassification error in case and control subjects with prostate cancer. The results are presented as ORs and 95% confidence intervals (CIs) for the effectiveness of screening.

RESULTS

The 52 patients had a median age of 69 years (range 56–92 years). Forty-five (87%) were white, 5 (10%) were African American, and 2 (4%) were Hispanic. The total number of DREs performed during the 5 years before the diagnosis of prostate cancer was 252, with a median of 4 per patient (range 0–16).

The last DRE before the diagnosis of cancer was a definite screening test in 14 patients (27%). It was a likely screening test in 20 patients (38%) without urinary-tract

Table 2. Results of Digital Rectal Examinations

Normal	Consistent with BPH*	Suggestive of Malignancy and Other Non-BPH Findings
Anodular	2+, etc., enlarged	Boggy
No masses	35 g, etc.	Firm
Normal	Consistent with post-TUR changes	Hard
Smooth	2 1/2 cm, etc.	Hard ridge
Benign	Consistent with BPH	Nodule
Nontender	Enlarged	Tender
Small		Change in consistency
Soft		Friable
		Hard lobe
		Induration
		Prominent lobe

*BPH indicates benign prostatic hyperplasia; TUR, transurethral resection.

Table 3. Classification Scheme for Screening Digital Rectal Examinations

Purpose of DRE	Criteria*
Definite screening	No urinary-tract symptoms; no or normal previous DRE Gastrointestinal symptoms only (prostate mentioned)
Likely screening	No urinary-tract symptoms; previous exam consistent with BPH
Probable screening	Stable urinary-tract symptoms; previous exam none, normal, or consistent with BPH
Not screening	Worsening or new urinary-tract symptoms
Other or unknown	Unable to classify, including gastrointestinal symptoms only (no mention of prostate)

*BPH indicates benign prostatic hyperplasia.

symptoms but with a previous DRE consistent with BPH; a probable screening test in 8 patients (15%) with stable BPH symptoms whose previous DREs were normal, not available, or consistent with BPH; and not a screening test in 8 patients (15%) with worsening or new urinary-tract symptoms. The last DRE was classified as other for one patient (2%) whose prostate cancer was diagnosed during surgery to treat bladder cancer. The last DRE was classified as unknown for one patient (2%) without any medical record entry before the diagnosis of prostate cancer.

Impact of Misclassification Bias

The potential impact of misclassification bias is illustrated in Table 4, in which all of the case subjects but none of the control subjects has cancer, and only the last DRE before the diagnosis of cancer is included in the analyses. If definite, likely, and probable DREs are included as screening tests, the frequency of screening in case subjects is 81% (42/52) (see top of Table 4). When the same screening frequency is assigned to control subjects, the OR is 1.0. In this example, screening is not associated with a reduction in mortality from prostate cancer.

When probable DREs are excluded and only definite and likely DREs are included as screening tests, the frequency of screening in case subjects decreases from

about 80% to 65% (34/52) (see bottom of Table 4); 15% of DREs are excluded as screening tests in case patients with stable urinary-tract symptoms and previous DREs consistent with BPH because cancer was diagnosed. Because none of the controls has prostate cancer, all of their probable DREs are counted as screening tests. In this example, the OR is 0.45, and screening appears to be associated with a reduction in mortality from prostate cancer.

Therefore, an apparent protective effect of screening can be created merely by changing the definition of a screening DRE. Restrictive definitions for a screening DRE lower the frequency of screening in case subjects more than they do in control subjects, and thus lower the OR, making screening appear to be more protective.

The extent of the effect on the OR that could be created by misclassification bias was measured in a series of simulated case-control studies. Three factors were varied across a range of clinically pertinent values. The frequency of screening in case and control subjects was 20%, 40%, 60%, or 80%. The prevalence of prostate cancer in control subjects was 0%, 5%, or 10%. The size of misclassification error in case and control subjects with prostate cancer was 5%, 10%, or 15%. This strategy produced 36 ($4 \times 3 \times 3$) hypothetical case-control studies. The ORs for the effectiveness of screening ranged from 0.21 to 0.83 (data not shown). The 95% CI for 3 of the 36 simulations excluded 1.0 (data not shown), despite the small sample sizes in these simulations.

Table 4. The Effect on Odds Ratio of Changing the Definition of a Screening Digital Rectal Examination

Analysis Including Definite, Likely, and Probable DREs as Screening Tests		
Screening	Cases	Controls
Yes	42 (81%)	42 (81%)
No	10 (19%)	10 (19%)
Total	52 (100%)	52 (100%)
OR = (42/10)/(42/10) = 1.0; 95% CI 0.38, 2.7.		
Analysis Including Only Definite and Likely DREs as Screening Tests		
Screening	Cases	Controls
Yes	34 (65%)	42 (81%)
No	18 (35%)	10 (19%)
Total	52 (100%)	52 (100%)

OR = (34/18)/(42/10) = 0.45; 95% CI 0.18, 1.1.

DISCUSSION

Observational studies that do not consider clinical distinctions in the performance of screening tests may misrepresent the proportion of subjects classified as having screening interventions. This study demonstrates that screening DREs in patients with prostate cancer can be misclassified as nonscreening tests in the presence of BPH. The results suggest that a too-strict definition of screening can falsely lower the observed frequency of screening tests in case subjects unless suitable precautions are taken.

The question "What is a screening exam?" requires reexamining traditional concepts about classifying diagnostic tests for the early detection of cancer. Various terms, including "screening," "mass screening," "case finding," and "differential diagnosis," are used in the litera-

ture, often without standard definitions or uniform applications. A commonly cited classification scheme for potential screening tests considers the variables "initiator of encounter" (provider or patient) and "symptom status" (none, unrelated to cancer, possibly related to cancer), and defines terms as follows:

Screening: testing of healthy volunteers from the community, for example, provider-initiated encounter and no symptoms present.

Case finding: testing of patients who have sought health care for disorders unrelated to their chief complaint(s), for example a patient-initiated encounter with symptoms unrelated to cancer.

Diagnosis: testing in patients who have actively sought health services to identify a cause for specific complaint(s), for example, a patient-initiated encounter with symptoms possibly related to cancer.¹¹

The categories of this classification scheme are not exhaustive, however, because other combinations exist for the variables "initiator of encounter" and "symptom status." For example, patients often request and receive tests without symptoms, such as an asymptomatic woman who seeks a mammogram after learning of a diagnosis of breast cancer in a friend or relative (patient-initiated encounter and no symptoms). In addition, providers regularly obtain tests with the intention of screening in patients with chronic, benign symptoms that may also be considered "possibly related to cancer." An example is a patient with documented hemorrhoids and intermittent rectal bleeding who receives periodic sigmoidoscopy from his primary care physician as a screen for colorectal cancer (provider-initiated encounter and symptoms possibly related to cancer).

Another problem occurs when subjects recruited from the community have symptoms that may be manifestations of cancer. In one example,¹² half the volunteers recruited for the American Cancer Society Prostate Cancer Detection Project had symptoms that may have prompted their participation (provider-initiated encounter and symptoms possibly related to cancer). Nonetheless, the report describes all exams as screening tests. Thus, this classification scheme¹¹ does not adequately classify all types of exams that would be encountered in an observational study of screening.

In contrast, the classification scheme in this study classifies DREs as either screening tests or nonscreening tests. The term screening applies to tests done with a baseline suspicion of cancer, including encounters that are provider- or patient-initiated, and patients who are asymptomatic, have unrelated symptoms, or have possibly related symptoms that were previously ascribed to a benign cause. The term case finding is not used because the provider, as a clinician, is screening for a disease regardless of whether a patient is recruited from the community,

receives an annual health maintenance exam, or presents for an unrelated problem. (In addition, the clinician usually does not have the perspective of finding a series of cases of cancer.) Finally, the term nonscreening tests describes tests done when a provider for a given patient has an increased suspicion of cancer above the baseline.

Besides selecting terms to describe exposure status, the current research also develops criteria for accurately classifying the purpose of tests encountered in the conduct of an observational study. The nonexhaustive categories of previous classification schemes¹¹ are replaced by a mutually exclusive and exhaustive classification scheme of reasons why a test was done. The new classification scheme also indicates the investigator's degree of certainty that an exam was done as a screening test.

This study focuses on the last DRE before the diagnosis of prostate cancer because it is the most difficult one to classify. The most inclusive criterion, which includes definite, likely, and probable DREs, is the most clinically relevant and methodologically conservative. The definition is relevant to clinicians because screening for prostate cancer is often conducted in the presence of stable, benign conditions such as BPH. The definition is also methodologically conservative because the OR for the effectiveness of screening will be closer to 1.0 if probable screening exams are included in the analysis (Table 4).

The current study combines urinary-tract symptoms and DRE findings in cogent categories to characterize each subject's status regarding BPH. The results, therefore, are consistent with evolving concepts of BPH that define the condition by symptoms, DRE findings, or other criteria.¹³ In addition, a recent editorial states that "no single definition [of BPH] has gained universal acceptance," but notes that either an enlarged prostate gland or urinary-tract symptoms are cardinal features.¹⁴

This study has several limitations. First, the role of DREs other than the last DRE before the diagnosis of prostate cancer was not examined. These DREs are important, because a case-control study can examine any versus no screening DREs; can require a threshold frequency of screening DREs, such as annual DREs, to define exposure; or can measure the effect of multiple DREs. In future studies, these DREs can be evaluated with the same classification scheme used in this study for the last DRE before the diagnosis of prostate cancer (Table 3). Second, PSA testing was not done before the diagnosis of prostate cancer in any of the men in this study. The approach used in this study to classify DREs as screening tests, however, could be used to classify PSAs as screening tests, but that approach must be developed and validated. Third, in this study DREs done at non-VA sites were not identified unless the results were noted in the medical record. Fourth, simulated rather than actual case and control subjects were used to illustrate the impact of the differential identification of screening. Finally, the new classification scheme (Table 3) must be evaluated for interobserver variability. Despite these limitations, the re-

sults demonstrate that misclassification bias is a serious threat when conducting observational studies of screening.

The effectiveness of screening for prostate cancer is currently being evaluated in a randomized clinical trial of men who receive screening versus usual care.¹⁵ The primary outcome is mortality, and the results will be available in approximately 10 years. Meanwhile, observational studies, especially case-control studies, can help guide patient-care and health-policy decisions regarding screening. These studies can produce misleading results unless suitable attention is given to the clinical characteristics of patients and their diseases.

The author thanks the Editor and two anonymous reviewers for helpful comments; Peter Peduzzi, Ralph I. Horwitz, and Alvan R. Feinstein for reviewing an earlier version of this manuscript; and Karen Anderson and Ayumi Kamina for technical assistance.

REFERENCES

1. Weiss NS. Application of the case-control method in the evaluation of screening. *Epidemiol Rev.* 1994;16:102-8.
2. Gill TM, Horwitz RI. Evaluating the efficacy of cancer screening: clinical distinctions and case-control studies. *J Clin Epidemiol.* 1995;48:281-92.
3. Clarke EA, Anderson TW. Does screening of "Pap" smears help prevent cervical cancer? A case-control study. *Lancet.* 1979;2:1-4.
4. Verbeek AL, Hendricks JH, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography: first results of the Nijmegen project, 1975-1981. *Lancet.* 1984;1:1222-4.
5. Macgregor E, Moss SM, Parkin DM, Day NE. A case-control study of cervical cancer screening in north east Scotland. *BMJ.* 1985; 290:1543-6.
6. Friedman GD, Hiatt RA, Quesenberry CP, Selby JV. Case-control study of screening for prostatic cancer by digital rectal examination. *Lancet.* 1991;337:1526-9.
7. Selby JV, Friedman GD, Quesenberry CP, Weiss NS. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med.* 1992;326:653-7.
8. Thompson RS, Barlow WE, Taplin SH, et al. A population-based case-cohort evaluation of the efficacy of mammographic screening for breast cancer. *Am J Epidemiol.* 1994;140:889-901.
9. Herrinton LJ, Selby JV, Friedman GD, Quesenberry CP, Weiss NS. Case-control study of digital-rectal screening in relation to mortality from cancer of the distal rectum. *Am J Epidemiol.* 1995;142: 961-4.
10. Fukao A, Tsubono Y, Hisamichi S, Sugahara N, Takano A. The evaluation of screening for gastric cancer in Miyagi Prefecture, Japan: a population-based case-control study. *Int J Cancer.* 1995; 60:45-8.
11. Sackett DL, Holland WW. Controversy in the detection of disease. *Lancet.* 1975;2:357-9.
12. Mettlin C, Murphy GP, Babaian RS, et al. The results of a five-year early prostate cancer detection intervention. *Cancer.* 1996;77:150-9.
13. Barry MJ, Boyle P, Garraway M, et al. Epidemiology and natural history of BPH. In: Cockett ATK, Khoury S, Aso Y, et al., eds. *The 2nd International Consultation on Benign Prostatic Hyperplasia (BPH).* Jersey, Channel Islands: Scientific Communication International Ltd; 1993.
14. Walsh P. Treatment of benign prostatic hyperplasia. *N Engl J Med.* 1996;335:586-7.
15. Gohagan JK, Prorok PC, Kramer BS, Cornett JE. Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute. *J Urol.* 1994;152: 1905-9.

SGIM Headquarters Change of Address

The office of the Society of General Internal Medicine has moved. Please note our new address and number:

Society of General Internal Medicine
2501 M Street, NW, Suite 575
Washington, DC 20037

(800) 822-3060
(202) 887-5150
Fax: (202) 887-5405