

# Prostate Cancer Association Studies: Pitfalls and Solutions to Cancer Misclassification in the PSA Era

Elizabeth A. Platz,<sup>1,2,3\*</sup> Angelo M. De Marzo,<sup>2,3,4</sup> and Edward Giovannucci<sup>5,6</sup>

<sup>1</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

<sup>2</sup>Brady Urological Institute, Johns Hopkins Medical Institutions, Baltimore, Maryland

<sup>3</sup>Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins Medical Institutions, Baltimore, Maryland

<sup>4</sup>Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland

<sup>5</sup>Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, Massachusetts

<sup>6</sup>Channing Laboratory, Department of Medicine, Harvard Medical School and Brigham & Women's Hospital, Boston, Massachusetts

---

**Abstract** Widespread screening of American men for elevated PSA has changed the characteristics of prostate cancer cases in the U.S. The influence of the changed nature of prostate cancer cases in the PSA era and the need for careful consideration of who is a "case" and who is a "control" on the ability to detect associations of risk factors with prostate cancer in etiologic epidemiologic studies merits discussion. Issue 1: prostate cancer cases diagnosed in the PSA era are enriched with a pool of early lesions, which may differ in etiology, and are deficient in advanced lesions, which are the most likely to be the product of promotion and progression events. By admixing the two types of cases (i.e., imperfect specificity), the associations previously detected using epidemiologic designs when the majority of cases were clinically detected may no longer be apparent in the PSA era when the majority of cases are now detected in the pre-clinical phase. Researchers must now tailor hypotheses such that they are testable using early stage cases or specifically augment the number of advanced cases when testing hypotheses related to extraprostatic growth and progression. Issue 2: even when controls are screened for elevated PSA to rule out the presence of prostate cancer, some proportion of those controls currently will have one or more foci of prostate cancer. The imperfect sensitivity of the PSA test coupled with diagnostic work-up may in part result from (a) lack of PSA elevation in some men with prostate cancer or (b) failure of biopsy to sample the tumor focus in men with elevated PSA. Misclassification of men with undetected prostate cancer as controls usually produces a bias that tends to deflate associations. Given this type of disease misclassification, whether an association still can be statistically detected depends on the extent of misclassification, the magnitude of the true association, the prevalence of the exposure in the true controls, and the sample size, although in general moderate nondifferential misclassification does not lead to profound attenuation. However, under the same scenario attenuation does not occur in cohort or case-cohort studies in which the rate or risk ratio (RR) is calculated. That prostate cancer cases diagnosed in the PSA era are enriched with early stage, minimally invasive disease in our opinion is likely to pose a far more serious obstacle to epidemiologic research on the etiology of clinically important prostate cancer than the issue of inclusion as controls some men who have undiagnosed prostate cancer because of imperfect sensitivity of PSA screening and biopsy sampling error. *J. Cell. Biochem.* 91: 553–571, 2004. © 2003 Wiley-Liss, Inc.

**Key words:** prostate cancer; PSA; epidemiology; methods; misclassification

---

---

Grant sponsor: National Cancer Institute (Hopkins); Grant number: P50CA58236; Grant sponsor: National Cancer Institute (Harvard); Grant numbers: P50CA90381, CA55075, CA72036.

\*Correspondence to: Elizabeth A. Platz, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. E-mail: eplatz@jhsph.edu

Received 29 August 2003; Accepted 2 September 2003

DOI 10.1002/jcb.10700

© 2003 Wiley-Liss, Inc.

Our work as epidemiologists and pathologists has led us to pose the following question: What is the influence of widespread use of PSA screening for the early detection of prostate cancer on the ability to detect associations between exposures (includes genetics) and prostate cancer in epidemiologic studies? The influence of the changed nature of prostate cancer cases in the PSA era and the need for careful consideration of who is a "case" and who is a

“control” on the ability to detect associations with risk factors warrants attention. After briefly describing the use of PSA screening in the U.S., we discuss the following two issues that are relevant to epidemiologic research on the causes of prostate cancer.

- 1) Prostate cancer cases diagnosed in the PSA era are enriched with early lesions and are deficient in advanced lesions.
- 2) Even when controls are screened for elevated PSA to rule out the presence of prostate cancer, some proportion of those controls currently will have one or more foci of prostate cancer.

#### USE OF PSA SCREENING IN THE U.S.

PSA or prostate specific antigen is a protease secreted exclusively by epithelial cells lining the glandular acini of the prostate. Its levels are elevated in the serum of men with prostate cancer, but also in men with benign enlargement of the prostate (BPH) and prostatitis. Screening for elevated serum PSA to detect prostate cancer earlier in its natural history than with digital rectal examination alone began in 1988 in the U.S. and it rapidly became routine medical care, although the test was not approved for this use by the U.S. Food and Drug Administration until 1994 [Food and Drug Administration, 1994]. Due to widespread screening for PSA coupled with biopsy, in the early to mid-1990s, the prostate cancer incidence rate soared [Potosky et al., 1995] as it advanced in time the diagnosis of the pool of tumors still in the pre-clinical phase. In the late 1990s, as expected, the incidence rate resumed its pre-PSA era trajectory. By 2001, 75% of American men age 50 years old or older reported that they ever had a PSA test and 54% of men 50–69 years old reported that they recently had had a PSA test [Sirovich et al., 2003]. PSA testing is not routine elsewhere in the world.

Despite its common use in the U.S., the balance of its efficacy in reducing death from prostate cancer via early detection and treatment while the tumor is still histopathologically organ-confined against the over-diagnosis and treatment of as yet clinically insignificant tumors and associated physical and financial costs remains to be demonstrated. Although the death rate due to prostate cancer has been declining since the late 1990s in the U.S.

[American Cancer Society, 2003b], it is also declining in Europe [Quaglia et al., 2003]. This observation indicates that while PSA screening may be reducing death from prostate cancer, other factors such as improved treatment also are likely contributing to this decline. Despite the uncertainties, the American Cancer Society [2003a] and the American Urological Association [American Urological Association] recommend annual screening by digital rectal examination and the serum PSA test annually beginning at age 50 for men at normal risk of prostate cancer and earlier for men at higher than normal risk, including African-American men and men with a first degree family history of prostate cancer. Screening is not recommended for men whose lifespan is estimated to be less than 10 years. Because balance of risks and benefits remains unknown, the U.S. Preventive Services Task Force [2002] neither recommends for nor recommends against screening. To address the effect of screening with PSA and/or digital rectal examination on prostate cancer mortality, two large-scale randomized trials in the U.S. [Gohagan et al., 2000] and Europe [de Koning et al., 2002] are underway, with results expected in 2008.

The common use of screening for elevated PSA by American men, including those who have been or are being enrolled into epidemiologic studies of risk factors for prostate cancer, raises a new set of issues that if not considered in detail, may lead to false inferences drawn from the findings of these studies.

#### ISSUE 1: PROSTATE CANCER CASES DIAGNOSED IN THE PSA ERA ARE ENRICHED WITH EARLY LESIONS AND ARE DEFICIENT IN ADVANCED LESIONS

Widespread screening for elevated serum PSA coupled with broad recommendations for the age at which to start screening has changed the characteristics of prostate cancer cases at diagnosis. The median age at diagnosis is now slightly younger and the distribution of stage at diagnosis has shifted downward [Hankey et al., 1999]. In the mid-1980s prior to the introduction of PSA screening, 14.9% of prostate cancer cases had distant metastases at diagnosis whereas by 1995 only 6.6% of cases had distant metastases at diagnosis [Stanford et al., 1999]. PSA-detected tumors are those that were in a pre-clinical phase of their natural history, i.e.,

asymptomatic and not palpable, at the time of diagnosis (clinical stage T1c). PSA-detected prostate cancers tend to be pathologically organ-confined at prostatectomy and are of smaller volume than clinically-diagnosed cases. Even PSA-detected cases diagnosed when they are extraprostatic may have more favorable characteristics than advanced cases diagnosed in the pre-PSA era, again because of detection earlier in their natural history. As early as 1993 pathologists reported an increase in the percentage of prostate cancer cases with tumor volumes much less than 0.5 cc and Gleason score less than 7 on radical prostatectomy [DiGiuseppe et al., 1997], both indicators of better prognosis.

Investigators conducting epidemiologic studies in the PSA era have a new set of challenges: the case pool will be deficient in advanced disease and enriched with early disease. Prostate cancer cases that are advanced at diagnosis are those that have experienced not only initiating events, but also have been subject to promotion and progression. If the case pool is deficient in advanced disease because of early detection, exposures that are more important later in carcinogenesis may not be detectable as being risk factors for prostate cancer.

Essentially, the inability to detect associations in this context results from reduced disease specificity. Specificity is the probability that a person will test negative given that the person does not have the disease. For contrast, sensitivity is the probability that a person will test positive given that the person does have the disease. Men diagnosed with early prostate cancer are contaminating the case pool because these cases are either less strongly associated with the exposure than the advanced cases (e.g., exposure prevalence is intermediate between the advanced cases and the controls), not associated with the exposure at all (e.g., same exposure prevalence as in the controls), or even associated with the exposure possibly in the direction opposite from the advanced cases. Thus, when studying some late-acting exposures, the cases with early disease can be considered to be false-positive cases. The false-positive proportion is the complement of specificity. Note that the choice of the optimal control (or noncase) definition and the influence of imperfect specificity of disease classification have been widely discussed for case-control and cohort studies in [Copeland et al., 1977;

White, 1986; Brenner and Savitz, 1990] and in the many older articles referenced in those papers. We do not review the literature on these areas, but do illustrate the problem in the context of prostate cancer. Also, we do not address the problem of misclassification that differs in extent by exposure status (i.e., differential misclassification).

#### Quantifying the Influence of Classifying Early Prostate Cancer Cases Along With Advanced Cases in Case-Control Studies of Late-Acting Exposures

In quantifying the influence of imperfect specificity of the definition of "case" on the estimate of the strength of the association several assumptions are made: no confounding, selection, or observation bias, no effect modification, that the exposure is dichotomous and that exposure is perfectly classified, and that all men classified as controls are true controls (e.g., no false-negatives, perfect sensitivity). We also assume that the extent of misclassification is the same for the exposed and the unexposed (i.e., nondifferential misclassification). To illustrate the extent of error, we present the following study: a nested case-control study (a type of prospective study; details of this design are given later) of factors that are purported to be important late in carcinogenesis, in which all men in the cohort have an equal opportunity for detection of prostate cancer, the same number of controls as cases are sampled, the controls are sampled using the incidence density approach (described later), and at the same age all of the men have the same baseline risk.

Shown in Figure 1 is the usual display of data, called a two-by-two table, from a case-control study and the calculation of the odds ratio (OR) from the prevalences of exposure among the cases and controls. The OR is used as an estimate of the relative risk. The OR equals the number of exposed cases divided by the number of unexposed cases all divided by the number of exposed controls divided by the number of unexposed controls. The left panel shows the two-by-two table with perfect classification of the case/control status. If the prevalence of exposure is 50% in the true cases and 25% in the controls, then the OR is 3.0. The right panel shows the two-by-two table with 75% of the cases being false-positive cases (i.e., early cases). The proportion of the true-positive cases (i.e., advanced cases) that are exposed stays the

75% of cases are misclassified

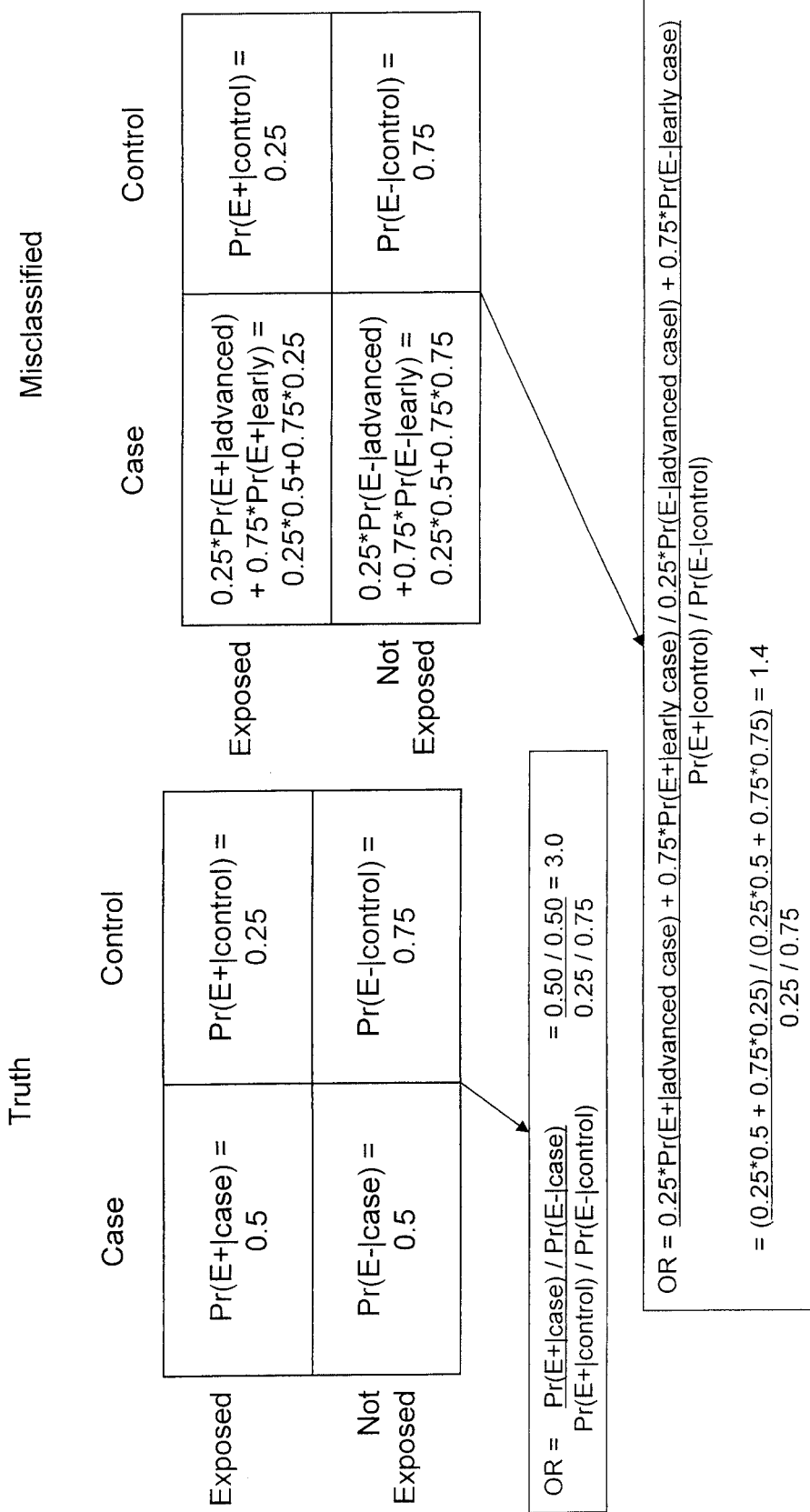


Fig. 1. What is the influence on the odds ratio (OR) in a case-control study of including as prostate cancer cases men with early disease when early disease is not associated with exposure? (Assumed that the prevalence of exposure in the early cases is the same as in the controls and that misclassification is nondifferential with respect to exposure.)

same, 50%. However, the proportion of the total cases (i.e., advanced cases plus early cases) that are exposed now equals the weighted average of the exposure proportion for the true-positive cases and the false-positive cases. If the early cases have the same exposure prevalence as the controls, then with 75% misclassification the OR is attenuated from 3.0 to 1.4.

If the prevalence of exposure of the early cases is intermediate between the advanced cases and the controls or equal to that of the controls, then this type of nondifferential misclassification results in an attenuation of the OR toward 1.0 (i.e., no association). The degree of the attenuation of the effect depends on the extent of false-positive cases and on the prevalence of exposure in the controls.

The effect of imperfect specificity on the statistical significance of the OR relative to the perfect classification scenario depends on how the perfect classification scenario is defined. We consider two possible ways of handling the false-positive cases (i.e., early cases) in an unmatched case-control study: (1) only advanced cases are selected for study along with an equal number of controls or (2) total cases (advanced plus early) are selected for study along with an equal number of controls, but for hypotheses related to extraprostatic growth and dissemination, the early cases are excluded from the analysis leaving advanced cases and all of the controls. In these two perfect classification scenarios the ORs calculated are identical; the difference is in the sample size. The width of the 95% confidence interval (CI) is narrower and the *P* value is smaller in the first scenario because the overall sample size is larger. However, the latter approach has been the more common approach in epidemiologic studies of prostate cancer, although exceptions can be found [Yoshizawa et al., 1998]. Comparing the imperfect specificity scenario to perfect classification scenario 1, the width of the confidence interval is comparable, but the power to detect the association as statistically significant is reduced because the OR is attenuated. Comparing the imperfect specificity scenario to the perfect classification scenario 2, the width of the confidence interval is narrower because of the larger number of cases, but the power to detect the association is again reduced because of the attenuated OR. A third correct classification scenario is given in Brenner and Savitz [1990] for hospital based case-control studies in which the total number

of true- and false-positive cases are fixed and the number of controls is selected in proportion to the number of true-positive cases (like scenario 2, but minus the extra controls that would have been selected for the false-positive cases). Comparing the imperfect specificity scenario to perfect classification scenario 3, the confidence interval is narrower, but the power to detect the association is reduced, again because of the attenuated OR.

Shown in Table I is the effect on the OR, the 95% CI, and the statistical significance of the OR for combinations of the true OR of 3.0 (strong association), 1.5 (moderate association typically expected in biomarkers studies), and 1.25 (modest association), prevalence of the exposure in the controls of 50% (e.g., dichotomize an exposure distribution at the median or a common allelic variant), 25% (e.g., highest quartile versus below or a moderately common allelic variant), and 10% (e.g., highest decile versus below or a less common allelic variant), the proportion of the total number cases that are misclassified cases of 25, 50, and 75%, and control sample sizes of 10,000 (very large size achievable possibly through pooling of data among studies), 1,000 (large size), 500 (moderate size), and 200 (small size). Note that an estimate of 75% of cases not being true cases in the context of hypotheses related to extraprostatic growth and dissemination may be a realistic guess given that the proportion of cases detected at a point where they already have distant metastases is now much less than 10%.

For the scenario in which 75% of the cases are false-positives, for typical sample sizes of 1,000 or 500 cases and controls and a typical sized true OR of 1.5, the associations are completely obscured. For strong associations and moderate to large sample sizes and prevalences of exposure in the controls, associations are substantially attenuated, but would still be detectable as statistically significant. Note that these statements are based on there being no association between the exposure and early disease. As the association between the exposure and early disease approaches that for advanced disease, the extent of attenuation would be reduced.

#### Possible Examples From Actual Studies

To illustrate Issue 1 using actual studies, we point to the results from an initial [Chan et al., 1998] and an expanded [Chan et al., 2002] case-control study nested in the prospective

**TABLE I. Estimate\* of the Odds Ratio (OR), 95% Confidence Interval (CI), and the Statistical Significance of the OR for Combinations of Extent of Nondifferential Misclassification of Early Cases as Advanced Cases, True ORs, Prevalence of Exposure in the Controls, and Sample Size for Controls**

	True OR = 3						True OR = 1.5						True OR = 1.25													
	Sample size for controls																									
	10,000			500			200			10,000			500			200			10,000			500			200	
Prevalence of exposure in controls	10,000	1,000	1,000	500	500	200	10,000	10,000	1,000	500	500	200	10,000	10,000	1,000	500	500	200	10,000	10,000	1,000	500	500	200		
Perfect classification if select only advanced cases (scenario 1)**																										
0.5	2.8-3.2	2.5-3.6	2.3-3.9	2.3-3.9	2.0-4.6	2.0-4.6	1.4-1.6	1.4-1.6	1.3-1.8	1.2-1.9	1.2-1.9	1.0-2.2	1.2-1.3	1.2-1.3	1.0-1.5	1.0-1.5	1.0-1.6	1.0-2.2	1.2-1.3	1.2-1.3	1.0-1.5	1.0-1.6	1.0-1.6	0.8-1.8		
0.25	2.8-3.2	2.5-3.6	2.3-3.9	2.3-3.9	2.0-4.6	2.0-4.6	1.4-1.6	1.4-1.6	1.2-1.8	1.1-2.0	1.1-2.0	1.0-2.3	1.2-1.3	1.2-1.3	1.0-1.5	1.0-1.5	0.9-1.7	0.8-1.9	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.7-2.3		
0.1	2.8-3.2	2.3-3.9	2.1-4.3	2.1-4.3	1.7-5.3	1.7-5.3	1.4-1.7	1.4-1.7	1.2-2.0	1.0-2.2	1.0-2.2	0.8-2.7	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.7-2.3	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.6-2.4		
Perfect classification for 25% misclassification scenario (scenario 2)***																										
0.5	2.8-3.2	2.4-3.7	2.2-4.0	2.2-4.0	1.9-4.7	1.9-4.7	1.4-1.6	1.4-1.6	1.2-1.8	1.1-2.0	1.1-2.0	1.0-2.2	1.2-1.3	1.2-1.3	1.0-1.5	1.0-1.5	1.0-1.6	1.0-2.2	1.2-1.3	1.2-1.3	1.0-1.5	1.0-1.6	1.0-1.6	0.8-1.9		
0.25	2.8-3.2	2.5-3.7	2.2-4.0	2.2-4.0	1.9-4.7	1.9-4.7	1.4-1.6	1.4-1.6	1.2-1.8	1.1-2.0	1.1-2.0	0.9-2.4	1.2-1.3	1.2-1.3	1.0-1.5	1.0-1.5	0.9-1.7	0.8-1.9	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.6-2.4		
0.1	2.8-3.3	2.3-3.9	2.0-4.3	2.0-4.3	1.6-5.3	1.6-5.3	1.4-1.7	1.4-1.7	1.1-2.0	1.0-2.2	1.0-2.2	0.8-2.8	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.7-2.3	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.6-2.4		
25% Misclassification																										
0.5	2.1-2.3	1.8-2.6	1.7-2.8	1.7-2.8	1.4-3.3	1.4-3.3	1.3-1.4	1.3-1.4	1.1-1.6	1.1-1.7	1.1-1.7	0.9-2.0	1.1-1.2	1.1-1.2	1.0-1.4	0.9-1.5	0.9-1.5	0.8-2.0	1.1-1.2	1.1-1.2	1.0-1.4	0.9-1.5	0.9-1.5	0.8-2.0		
0.25	2.2-2.5	1.9-2.8	1.8-3.0	1.8-3.0	1.5-3.5	1.5-3.5	1.3-1.5	1.3-1.5	1.1-1.7	1.0-1.8	1.0-1.8	0.9-2.1	1.1-1.3	1.1-1.3	1.0-1.4	0.9-1.6	0.9-1.6	0.7-1.8	1.1-1.3	1.1-1.3	1.0-1.4	0.9-1.6	0.9-1.6	0.7-1.8		
0.1	2.2-2.6	1.9-3.1	1.7-3.5	1.7-3.5	1.3-4.2	1.3-4.2	1.3-1.5	1.3-1.5	1.0-1.8	0.9-2.0	0.9-2.0	0.7-2.5	1.1-1.3	1.1-1.3	0.9-1.6	0.8-1.8	0.8-1.8	0.6-2.2	1.1-1.3	1.1-1.3	0.9-1.6	0.8-1.8	0.8-1.8	0.6-2.2		
Perfect classification for 50% misclassification scenario (scenario 2)***																										
0.5	2.8-3.2	2.4-3.8	2.1-4.2	2.1-4.2	1.8-5.1	1.8-5.1	1.4-1.6	1.4-1.6	1.2-1.9	1.1-2.0	1.1-2.0	0.9-2.4	1.2-1.3	1.2-1.3	1.0-1.5	0.9-1.7	0.9-1.7	0.8-2.0	1.2-1.3	1.2-1.3	1.0-1.5	0.9-1.7	0.9-1.7	0.8-2.0		
0.25	2.8-3.2	2.4-3.8	2.2-4.1	2.2-4.1	1.8-5.0	1.8-5.0	1.4-1.6	1.4-1.6	1.2-1.9	1.1-2.1	1.1-2.1	0.9-2.5	1.2-1.3	1.2-1.3	1.0-1.6	0.9-1.8	0.9-1.8	0.7-2.1	1.2-1.3	1.2-1.3	1.0-1.6	0.9-1.7	0.9-1.7	0.7-2.1		
0.1	2.7-3.3	2.2-4.0	2.0-4.5	2.0-4.5	1.6-5.7	1.6-5.7	1.4-1.7	1.4-1.7	1.1-2.1	1.0-2.4	1.0-2.4	0.7-3.0	1.1-1.4	1.1-1.4	0.9-1.8	0.8-2.0	0.8-2.0	0.6-2.6	1.1-1.4	1.1-1.4	0.9-1.8	0.8-2.0	0.8-2.0	0.6-2.6		
50% Misclassification																										
0.5	1.6-1.8	1.4-2.0	1.3-2.1	1.3-2.1	1.1-2.5	1.1-2.5	1.2-1.3	1.2-1.3	1.0-1.5	1.0-1.6	1.0-1.6	0.8-1.8	1.1-1.2	1.1-1.2	0.9-1.3	0.9-1.4	0.9-1.4	0.7-1.6	1.1-1.2	1.1-1.2	0.9-1.3	0.9-1.4	0.9-1.4	0.7-1.6		
0.25	1.7-1.9	1.5-2.2	1.4-2.4	1.4-2.4	1.2-2.8	1.2-2.8	1.2-1.3	1.2-1.3	1.0-1.5	0.9-1.6	0.9-1.6	0.8-1.9	1.1-1.2	1.1-1.2	0.9-1.4	0.8-1.5	0.8-1.5	0.7-1.7	1.1-1.2	1.1-1.2	0.9-1.4	0.8-1.5	0.8-1.5	0.7-1.7		
0.1	1.8-2.1	1.5-2.5	1.3-2.8	1.3-2.8	1.1-3.4	1.1-3.4	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.7-2.3	1.0-1.2	1.0-1.2	0.8-1.5	0.7-1.7	0.7-1.7	0.6-2.1	1.0-1.2	1.0-1.2	0.8-1.5	0.7-1.7	0.7-1.7	0.6-2.1		
Perfect classification for 75% misclassification scenario (scenario 2)***																										
0.5	2.7-3.3	2.2-4.1	1.9-4.5	1.9-4.5	1.4-5.7	1.4-5.7	1.4-1.6	1.4-1.6	1.1-2.0	1.0-2.2	1.0-2.2	0.8-2.8	1.1-1.4	1.1-1.4	0.9-1.6	0.8-1.8	0.8-1.8	0.6-2.2	1.1-1.4	1.1-1.4	0.9-1.6	0.8-1.8	0.8-1.8	0.6-2.2		
0.25	2.7-3.3	2.3-4.0	2.0-4.4	2.0-4.4	1.6-5.7	1.6-5.7	1.4-1.7	1.4-1.7	1.1-2.0	1.0-2.2	1.0-2.2	0.7-2.8	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.6-2.3	1.1-1.4	1.1-1.4	0.9-1.7	0.8-1.9	0.8-1.9	0.6-2.3		
0.1	2.7-3.4	2.1-4.2	1.8-4.9	1.8-4.9	1.3-6.3	1.3-6.3	1.3-1.7	1.3-1.7	1.0-2.3	0.8-2.7	0.8-2.7	0.6-3.7	1.0-1.4	1.0-1.4	0.8-1.9	0.7-2.3	0.7-2.3	0.5-3.2	1.0-1.4	1.0-1.4	0.8-1.9	0.7-2.3	0.7-2.3	0.5-3.2		
75% Misclassification																										
0.5	1.2-1.4	1.1-1.5	1.0-1.6	1.0-1.6	0.9-1.9	0.9-1.9	1.0-1.2	1.0-1.2	0.9-1.3	0.9-1.4	0.9-1.4	0.7-1.6	1.0-1.1	1.0-1.1	0.9-1.3	0.8-1.3	0.8-1.3	0.7-1.5	1.0-1.1	1.0-1.1	0.9-1.3	0.8-1.3	0.8-1.3	0.7-1.5		
0.25	1.3-1.5	1.1-1.7	1.0-1.8	1.0-1.8	0.9-2.1	0.9-2.1	1.0-1.2	1.0-1.2	0.9-1.4	0.8-1.5	0.8-1.5	0.7-1.7	1.0-1.1	1.0-1.1	0.9-1.3	0.8-1.4	0.8-1.4	0.7-1.7	1.0-1.1	1.0-1.1	0.9-1.3	0.8-1.4	0.8-1.4	0.7-1.7		
0.1	1.3-1.6	1.1-1.9	1.0-2.1	1.0-2.1	0.8-2.6	0.8-2.6	1.0-1.2	1.0-1.2	0.8-1.5	0.7-1.7	0.7-1.7	0.6-2.1	1.0-1.2	1.0-1.2	0.8-1.4	0.7-1.6	0.7-1.6	0.6-2.0	1.0-1.2	1.0-1.2	0.8-1.4	0.7-1.6	0.7-1.6	0.6-2.0		

\*ORs and 95% CIs estimated from a logistic regression model. Statistically significant associations are shown in bold.

\*\*Only advanced cases are selected for study along with an equal number of controls.

\*\*\*Total cases (advanced plus early) are selected for study along with an equal number of controls, but for hypotheses related to extraprostatic growth and dissemination, the early cases are excluded leaving advanced cases and all of the controls.

Physicians' Health Study evaluating the association of plasma insulin-like growth factor-1 (IGF-1) with prostate cancer. IGF-1 is a peptide hormone that promotes growth in childhood and adolescence and at the cellular level it promotes proliferation and inhibits apoptosis, including in normal prostate and tumor cells *in vitro* [Cohen et al., 1991, 1994]. The initial study included 152 prostate cancer cases diagnosed after the date that a blood sample was collected in 1982 through 1992 and age-matched controls [Chan et al., 1998]. The majority of the cases were not PSA-detected and majority of the controls had not been screened for PSA. Risk of prostate cancer increased with increasing plasma IGF-1 concentration after taking into account concentration of IGFBP-3, the major binding protein of IGF-1. In the expanded study, the initial cases plus an additional 378 cases diagnosed through 1995 were included. After adjusting for IGFBP-3, IGF-1 was no longer associated with prostate cancer overall, but an association was observed for cases that were of advanced stage at diagnosis (extraprostatic and metastatic) or that were diagnosed in the pre-PSA era [Chan et al., 2002].

Exposures that are more important early in prostate carcinogenesis may still be detectable as risk factors for prostate cancer in the PSA era because these exposures likely would have influenced the development of both early and advanced cases. To illustrate this contention we point to a case-control study nested in the prospective Health Professionals Follow-up Study evaluating the association of plasma lycopene [Wu et al., *in press*], androgens and length of the androgen receptor gene CAG repeat with prostate cancer [Platz et al., 2003]. Lycopene, a carotenoid found in tomatoes, is an efficient antioxidant [Sies and Stahl, 1995] found in biologically active concentrations in the prostate [Clinton et al., 1996]. Lycopene might be expected to be more important early in prostate carcinogenesis by reducing the opportunity for oxidative DNA damage. By binding to the androgen receptor androgens might enhance the growth of prostate cancer cells. The androgen receptor contains a variable length CAG repeat and the fewer the number of CAG repeats the greater the transactivational activity of the receptor [Chamberlain et al., 1994; Kazemi-Esfarjani et al., 1995]. The study included 460 prostate cancer cases diagnosed

after the date that a blood sample was collected in 1993 through 1998. Controls were men from the cohort matched to cases on age, history of PSA prior to blood draw, and other factors, who did not have a diagnosis of prostate cancer by the date of the case's diagnosis, but who had had a PSA test since the date of blood draw. Requiring the controls to have had a PSA test in theory reduces the likelihood of undetected prostate cancer in the controls (the validity of this assumption is discussed as Issue 2 below), but more importantly reduces the opportunity for detection bias that would result from differential rates of undergoing screening by men with and without the exposure of interest. Many of the cases were PSA-detected and all of the controls had been screened for PSA. The investigators observed that plasma lycopene was inversely associated with subsequent prostate cancer [Wu et al., *in press*], but plasma testosterone and androstenediol glucuronide (a metabolite of the major intraprostatic androgen dihydrotestosterone) concentrations and length of the androgen receptor gene CAG repeat were not associated with subsequent prostate cancer. In the similarly conducted Physicians' Health Study with cases diagnosed from 1982 to 1992, plasma lycopene was inversely associated [Gann et al., 1999] and plasma testosterone and androstenediol glucuronide were positively associated [Gann et al., 1996] with subsequent prostate cancer. Also in the Physicians' Health Study, but with cases diagnosed from 1982 to 1995, risk of subsequent prostate cancer appeared to increase monotonically with decreasing CAG repeat number primarily for advanced cases [Giovannucci et al., 1997].

The early cases diagnosed following-PSA screening may be further divided into two types: (a) those that would have progressed to advanced disease had they not been detected early and (b) those that may never have progressed to become clinically apparent during the patient's lifetime. The latter may be divided further into those that would not have become clinically apparent because they had very little or no biological capacity to progress and those that would have progressed except the patient died from other causes sooner. Because at this time we cannot predict which of these early cases would have progressed to clinically significant disease, we do not know whether these two types of early disease have shared or distinct etiologies. If their etiologies differ,

grouping both types as one case group could obscure associations between exposures and prostate cancer that is clinically important.

### Comments and Solutions to Issue 1

The variability in the characteristics of prostate cancer cases between the pre-PSA era and the PSA era and between countries that currently do not and do routinely screen for elevated PSA has major implications for drawing inferences from epidemiologic studies individually and collectively. Researchers must now tailor hypotheses to fit the case mix in their studies. Alternatively they may design studies and analyzes to capture relevant case groups (which may require recruitment over many years or many centers and be very expensive) and with adequately large sample size, run analyzes separately for organ-confined disease versus extraprostatic/metastatic disease. When summarizing the literature on the etiology of prostate cancer, which has been frequently described as inconsistent, reviewers must consider differences in case mix among studies as a possible explanation for these apparent inconsistencies in findings among studies. We provided examples suggesting that this issue may in fact be relevant for at least some exposures of interest, such as the IGF axis and androgens.

### ISSUE 2: EVEN WHEN CONTROLS ARE SCREENED FOR ELEVATED PSA TO RULE OUT THE PRESENCE OF PROSTATE CANCER, SOME PROPORTION OF THOSE CONTROLS CURRENTLY WILL HAVE ONE OR MORE FOCI OF PROSTATE CANCER

In the pre-PSA era the contrast between the “cases” and the “controls” seemed to be substantial in epidemiologic studies. Cases had clinically overt disease that was palpable (large volume and/or extracapsular), that obstructed urinary flow, or that caused pain due to metastases to bone. Controls in retrospective case-control studies or in nested case-control studies were men without a diagnosis of prostate cancer or the time during which they did not have a diagnosis of prostate cancer, although the men were not necessarily evaluated to rule out disease. Some of the controls likely did have occult prostate cancer. Now cases are largely men with an elevated PSA who on biopsy are found to have one or more foci of prostate adenocarcinoma and most of these are organ-

confined. Controls are primarily men who had a PSA test, but who did not have an elevated PSA or who are men who had an elevated PSA, but on biopsy, tumor was not detected. However, in most studies some of these controls have undetected prostate cancer because of (a) imperfect sensitivity of screening for prostate cancer or (b) biopsy sampling error (Fig. 2). Related to the issue of undetected prostate cancer in the controls, is the influence on the findings of including as “controls” men who are known to be diagnosed with prostate cancer subsequently in a nested case-control study.

### Imperfect Sensitivity of Screening for Prostate Cancer

The usual trigger for diagnostic work up via prostate biopsy is a serum PSA concentration of  $>4$  ng/ml or a PSA velocity (slope) of  $>0.75$  ng/ml per year. The PSA cutpoint is decreased to 2.5 ng/ml in men who have a first-degree family history of prostate cancer. Because the reported sensitivity of the PSA test is roughly 67.5–80% [Carroll et al., 2001] a negative PSA test does not mean that a man is free of prostate cancer. For example, using samples from the Physicians' Health Study, Gann et al. [1995] showed that PSA concentrations of 2–3 ng/ml, lower than the typical cut-off, were associated with 5.5-times the risk of diagnosis of prostate cancer years later. Subsequently, in the Baltimore Longitudinal Study of Aging, Fang et al. [2001] showed that PSA concentrations above the median of 0.6 ng/ml in men 40–49.9 years old was associated with 3.75-times the risk of being diagnosed with prostate cancer later. These prospective studies suggest that some men with serum PSA in the “normal” range have prostate cancer already present prior to the time that their serum PSA exceeds 4 ng/ml. Nevertheless, these false-negative men would be eligible to be selected as “controls” in epidemiologic studies.

This hypothesis is now substantiated by the recently published results from the Prostate Cancer Prevention Trial [Thompson et al., 2003]. In that trial, 18,882 men aged 55+ years old (median = 63 years) who had serum PSA concentrations  $\leq 3$  ng/ml and a normal digital rectal examination were randomized to take finasteride, an inhibitor of 5  $\alpha$ -reductase type II the enzyme that catalyzes the conversion of testosterone to dihydrotestosterone, for 7 years or to placebo. The men underwent annual



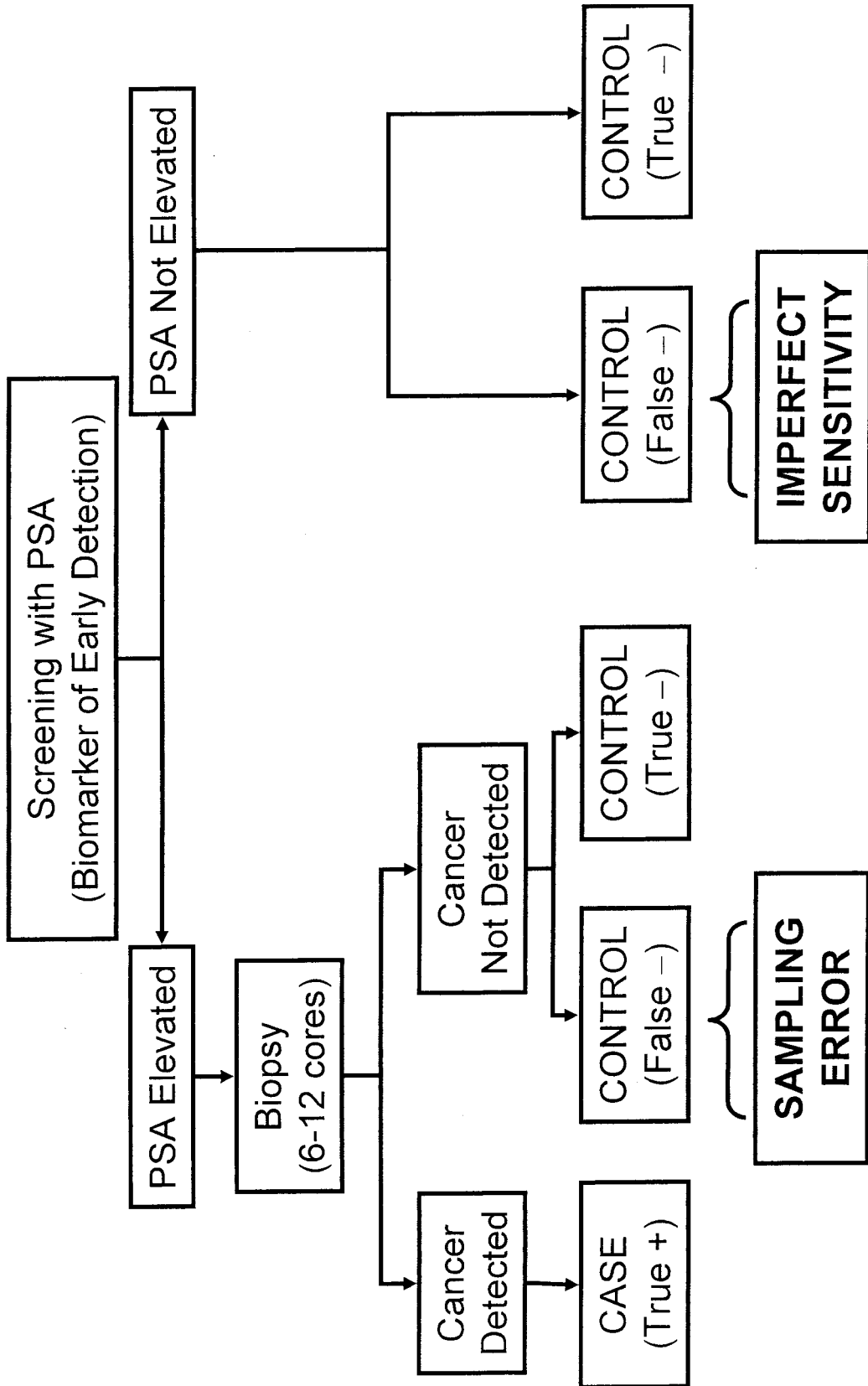


Fig. 2. Types of misclassification of prostate cancer cases in the PSA era.

screening for prostate cancer by PSA test and digital rectal examination, and if either was abnormal, a biopsy was performed. At the end of the 7th year the men who were not previously diagnosed with prostate cancer underwent biopsy irrespective of indication. Unexpectedly, 24.4% of the participants in the placebo arm were diagnosed with prostate cancer. Of these, 15.1% were diagnosed on the exit biopsy, despite many having "normal" PSA and digital rectal examination over the 7 years [Thompson et al., 2003]. This relatively high proportion of otherwise missed cases might have been predicted based on the prevalence of prostate cancer in autopsy-based studies of middle-age and older men (reviewed in [Godley and Schell, 1999]). Given the relatively short interval between randomization and exit biopsy, most of these men with "normal" range PSA and digital rectal examination likely already had prostate cancer at randomization.

The influence of not recognizing men who have prostate cancer but who do not have an elevated serum PSA on the sensitivity of PSA as a screening test also has been addressed mathematically. The true sensitivity may be lower and the true specificity may be higher for the  $>4$  ng/ml definition of elevated PSA in particular in younger men [Punglia et al., 2003]. Punglia et al. estimated that in men less than 60 years old 82% of cases may be missed and in men 60 years old and older 65% of cancers may be missed. The explanation put forth is that the calculation of sensitivity and specificity is biased by not knowing the prostate cancer status of all of the screened men; prostate cancer status is only known among men who had elevated PSA and thus were biopsied [Punglia et al., 2003].

The observation that 15% of the men with "normal" range PSA and digital rectal examination had prostate cancer in the placebo arm of the prostate cancer prevention trial (PCPT) [Thompson et al., 2003] also indicates that some prostate cancer adenocarcinomas do not result in abnormal elevations in serum PSA. Ordinarily PSA secreted by prostate epithelial cells into the glandular lumen does not enter circulation. In theory the production of PSA should be proportional to the number of epithelial cells in tissue that is normal, hyperplastic, or malignant. The mechanisms by which prostate cancer as well as BPH and prostatitis result in elevated PSA are not known, but have been hypothesized

to include damage to prostate epithelial cells rendering them leaky coupled with enhanced vascular permeability that occurs during the inflammatory response to that damage. The volume of organ-confined tumors in the peripheral zone determined at prostatectomy accounts for only for 10% of the variability in pre-operative serum PSA concentration in the range of 2–22 ng/ml [Stamey et al., 2002]. Extrapolating from this result, in men with limited disease, prostate cancer may be serendipitously detected because of PSA elevations unrelated to the presence of tumor (e.g., BPH [Roehrborn et al., 1999; Stamey et al., 2002]), which then triggers biopsy. Men with limited prostate cancer, but without a concurrent condition that raises PSA will not have the opportunity to undergo biopsy to have their occult tumors detected. These false-negative cases also will be eligible to be selected as controls in epidemiologic studies.

#### Biopsy Sampling Error

Sampling error occurs because only a small portion of the total prostate volume is biopsied, even now that 12 cores, rather than 6, are often obtained. Issues about the interpretation of prostate biopsy have been discussed in detail [Bostwick, 1997]. For example, a study that included men with elevated PSA who underwent numerous rounds of biopsy found that of all the cancers ultimately detected, only 77% were detected on the initial biopsy [Roehl et al., 2002]. By the fourth biopsy 99% of the cases had been detected, suggesting that sampling error is at least 23%. Studies of men with persistently elevated PSA, but who had one or more negative biopsies, indicate that when the number of cores is increased towards saturation (the maximum tolerable) 13–34% of those men will have cancer detected [Stewart et al., 2001; Fleshner and Klotz, 2002]. The proportion of clinically insignificant tumors (e.g., low volume, low Gleason score) on saturation biopsy was greater in men who had had three previous negative biopsies (22%) compared to those with only one previous negative biopsy (11%) [Stewart et al., 2001]. The optimal number of cores that should be taken to detect important cancers remains controversial [Scheck, 2001]. If the goal is to maximize the detection of all foci to minimize the risk of missing a clinically important tumor then a large number of cores should be used. However, the greater the number of cores taken the greater the chance of detecting foci that would

never have become manifest coupled with the higher risk of adverse effects of saturation biopsy (e.g., bleeding, urinary problems) and subsequent treatment for prostate cancer (e.g., incontinence, impotence), psychological concern, and financial outlay. Whether an increase in the number of biopsy cores taken will improve the ability to predict prognosis or will result in better survival because of earlier treatment of a clinically significant tumor is under study.

From an epidemiologic perspective, the net effect of sampling error is that some men with elevated PSA, but one or more negative biopsies will be classified as “controls” even if one or more foci of prostate cancer is actually present. In fact, since up to 20% of men with PSA-detected cancers are found to have “insignificant” cancer at radical prostatectomy, it is likely that some “controls” will have greater tumor volumes and perhaps higher grade tumors than the “cases.” However, the contrast between who is a case and who is a control may not be as poor as at first glance. Prostate tumors that are of large volume or that are multifocal are statistically more likely to be sampled on biopsy than those that have a small volume or are unifocal [Egevad et al., 1998]. If so, then men with elevated PSA and who truly do have a focus or foci of tumor, but that tumor(s) is missed on biopsy will be called a “control.” However, on average that “control” will have less tumor involvement than the typical “case.”

#### Sampling of Controls in Nested Case-Control Studies

In studies of the association of genetic polymorphisms or other biomarkers of risk with prostate cancer, frequently the nested case-control design is used because it is efficient and it minimizes the total number of samples that must be assayed. This design has as its basis the prospective cohort study: a large group of men who do not have a diagnosis of prostate cancer (although they are not necessarily screened for the disease) is assembled from whom exposure information and biological samples are collected and stored. These men (the “cohort”) are then followed over time for the diagnosis of prostate cancer.

In prospective designs, person-time at risk, rather than persons, is the unit of observation. Person-time at risk is the time that an individual remains free of prostate cancer until the date of diagnosis with prostate cancer, the end

of follow-up, death, diagnosis of cancer of another site, or loss-to-follow-up. All of the men in the cohort who develop prostate cancer since the start of the cohort are the cases. Individual men are selected as controls so that biomarkers or genotypes can be measured in stored samples, but the controls really represent the pool of person-time at risk. To select the controls, all men in the cohort are aligned at the study start by date or age, including the men who later are diagnosed with the disease. At the date (or age) that the first case is diagnosed, one (or sometimes more) of the men still at risk for prostate cancer is selected as a control, even if that control is diagnosed with prostate cancer later. At the date (or age) that the next case is diagnosed, another man in the pool of men still at risk for prostate cancer is selected as a control and so on. This type of control sampling is called incidence density sampling. Only the biological samples from the cases and selected controls are assayed for the biomarker of risk.

The measure of the association between the biomarker of risk and prostate cancer that is calculated for a nested case-control study is the OR. The benefit of this design over a standard case-control study is that the biomarkers measured in these samples reflect levels that are not affected by the presence of overt disease. Further, because the samples were collected before overt disease occurred, the problem of observation bias due to differential survival by exposure status is avoided.

Note that the “person-time at risk” should equate to the person-time during which a man does not have prostate cancer, or at least a diagnosis of prostate cancer. In both the prospective cohort design and its variants the nested case-control and case-cohort designs, we know with certainty that some men ultimately have a diagnosis of prostate cancer by the end of follow-up, yet we include the time preceding their diagnosis as time at risk, despite the fact that this is a time when they likely do already of have existing foci of tumor. What is the influence of the inclusion of men with undiagnosed cancer in the “at risk” group? Consider the example in which each man in the cohort undergoes as part of his usual medical care an annual PSA test such that differential opportunity for detection of occult prostate cancer by exposure status is precluded. Why is one man diagnosed with prostate cancer at an earlier age, whereas another man at the same

age does not have the diagnosis, but instead is diagnosed years later? Assuming similar screening histories, no biopsy sampling error, and at the time of diagnosis the tumor volume was larger and more readily detectable in the man diagnosed at an earlier age compared to the tumor volume in at the time in the man diagnosed at an older age, a possible major explanation for the difference in the natural history of the two men's prostate cancers is a difference in some influential characteristic of exposure, such as type of exposure (includes genetic variation), timing of exposure, and dose of exposure. The man with the earlier diagnosis of prostate cancer may have reached a critical cumulative exposure of a risk factor that triggered tumor growth earlier than did the man who was diagnosed later. Thus, the fact that the first man was diagnosed earlier than the second man is etiologically relevant information.

#### Quantifying the Influence of Misclassification of Prostate Cancer Cases as Controls in Case-Control Studies

In the PSA-era, are occult prostate cancer cases admixed with controls a problem in standard case-control or nested case-control studies, or not? In some situations, admixture is probably not a problem in theory or in practice. First, the contrast between the "cases" and the "controls" may not be as poor as expected. The "cases" are true cases with minimal to advanced disease. The "controls" are true controls plus cases with minimal disease (note: this will be true most of the time, but not all of the time because big anterior tumors can be missed by biopsy). If the exposure of interest is more strongly associated with more fulminant disease than with less fulminant disease, then an association between the exposure and prostate cancer will not be fully obscured because the controls with undiagnosed (minimal) cancer will not differ in this exposure from true cancer-free controls.

Even if the above contentions are not true, i.e., assuming that the cases misclassified as controls had the same extent of disease, same etiology, and the same prevalence of exposure as the correctly classified cases, what would be the extent of the error in estimating the association between an exposure and prostate cancer if some prostate cancer cases are misclassified as controls? The effect of disease misclassification

on the estimates of effect in case-control studies has been described in the epidemiologic literature in detail [Copeland et al., 1977] and specifically in the context of prostate cancer [Godley and Schell, 1999] previously. We provide additional comment and show the influence of this type of misclassification on both the magnitude of the association and the statistical significance of that association using recently published estimates of the extent of undiagnosed disease in men who would otherwise be classified as controls. We do not discuss the problem of misclassification that differs in extent by exposure status (i.e., differential misclassification) or the combined effect of imperfect sensitivity and specificity.

We made the same assumptions here for Issue 2 as we did for Issue 1: no confounding, selection, or observation bias, no effect modification, that the exposure is dichotomous and that exposure is perfectly classified, and that all men classified as cases are true cases (e.g., no false-positives, perfect specificity). We further assume that the extent of misclassification is the same in the exposed and the unexposed (i.e., nondifferential). We also illustrate the error for Issue 2 using the same type of study: a nested case-control study in which all men in the cohort have an equal opportunity for detection of prostate cancer, controls are selected using incidence density sampling, the same number of controls as cases are sampled, and at the same age all of the men have the same baseline risk.

Shown in Figure 3, are two-by-two tables and the calculation of the OR from the prevalences of exposure among the cases and controls. Using the same exposure prevalences as before, with perfect classification of the case/control status the OR equals 3.0 (left panel). The right panel shows the two-by-two table with 10% of the controls being false-negative cases. The proportion of the correctly classified cases that are exposed stays the same, 50%. However, the proportion of the "controls" that are exposed now equals the weighted average of the exposure proportion for the true controls and the false-negative cases. With 10% misclassification the OR is attenuated from 3.0 to 2.6.

In general, the type of nondifferential misclassification described here results in an attenuation of the OR toward the null value for the OR of 1 and a reduction in the statistical significance of the OR. The degree of the attenuation of the effect depends on the prevalence

10% of controls are cases

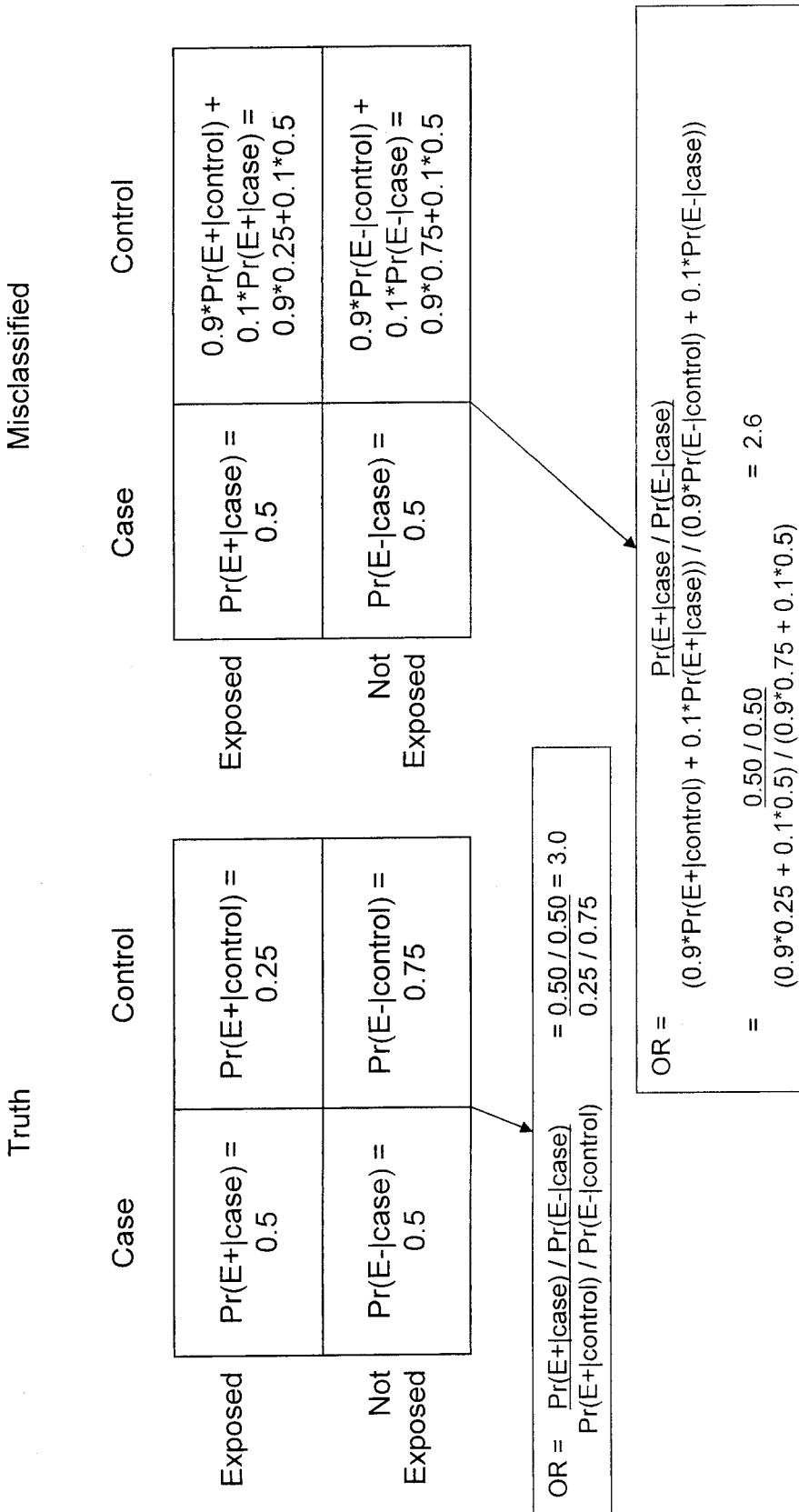


Fig. 3. What is the influence of misclassification of cases as controls on the OR and its statistical significance in a case-control study? (Assumed that the prevalence of exposure in the misclassified cases is the same as the correctly classified cases and that misclassification is nondifferential with respect to exposure.)

of misclassification of the cases as controls and on the prevalence of exposure in the controls. The degree of the reduction in the statistical significance of the OR depends on the size of the attenuated effect and the sample size. Shown in Table II is the effect on the OR, the 95% CI, and the statistical significance of the OR for when 10, 20, 30, 40, and 50% of controls are truly cases. Based on imperfect sensitivity estimated from the Prostate Cancer Prevention Trial and sampling error in men with persistent elevated PSA estimated from saturation biopsy studies, as many as 30% of controls may be false-negatives. This proportion might be even higher, if imperfect sensitivity is underestimated because of sampling error for the end-of-study-biopsies in the Prostate Cancer Prevention Trial. Thus, as in some autopsy studies, perhaps up to 50% of men in the age range at risk for a prostate cancer diagnosis have occult prostate cancer. In Table II, misclassification is shown in conjunction with varying true ORs of 3.0 (strong association), 1.5 (moderate association), and 1.25 (modest association); varying sample sizes of 10,000 cases and 10,000 controls (very large size), 1,000 cases and 1,000 controls (large size), 500 cases and 500 controls (moderate size), and 200 cases and 200 controls (small size); and varying prevalences of exposure among the controls of 50, 25, and 10%.

For very large sample sizes, common exposures, and strong associations, although the effect estimate is attenuated relative to the truth, the inferences drawn about the associations are unchanged; the associations are strong and are statistically significant, even with 50% misclassification. Even for moderate associations (OR = 1.5) if 30% of the controls are false-negatives, then a modest elevation in the OR is still seen and is statistically significant for a typically sized case-control study of 500 cases and 500 controls and common exposures. For modest associations, small sample sizes, or low prevalences of exposure in the controls, admixture of cases in the controls may nullify the association or preclude the detection of the association as being statistically significant.

#### **Cohort Studies—Exception to the Rule**

The preceding quantification of the influence of the admixture of cases among the controls applies to standard case-control and nested case-control studies, but not to prospective cohort studies in which the risk or rate ratio

(RR) is calculated as the measure of association. Among many of the strengths of the prospective cohort design is that no attenuation of the RR occurs when cases are admixed with the controls in a scenario where misclassification is nondifferential with respect to exposure, and the specificity of control identification is perfect (i.e., no false-positives). This fact is well established [White, 1986], but not often remembered.

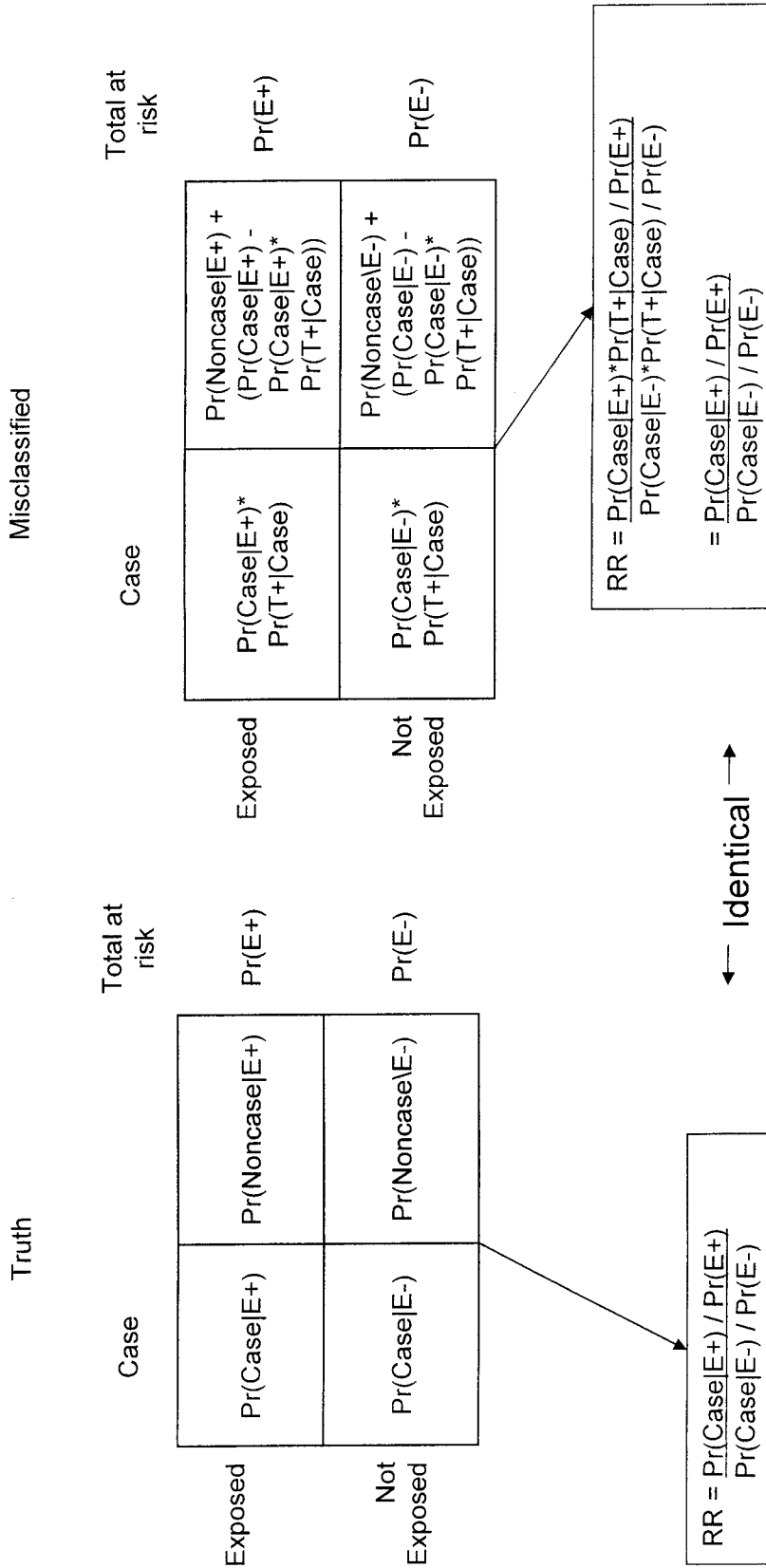
In the prospective cohort design the men at risk or the number of person-years at risk for each man is summed separately across exposed and unexposed men. The risk ratio is the number of exposed men who develop prostate cancer divided by the number of men at risk at baseline who were exposed all divided by the number of unexposed men who develop prostate cancer divided the number of men at risk at baseline who were not exposed. The risk ratio is calculated similarly by substituting person-years at risk for the number of men at risk at baseline.

Shown in Figure 4 is the calculation of the RR when some cases are admixed with the controls. Note that in a standard cohort study the total number of exposed and total number of unexposed is fixed at the start of the study, so the denominator of the risk of disease in the exposed and in the unexposed does not change between true scenario and the scenario when some of the cases are misclassified as controls. What does change is the numerator of the risk of disease in the exposed and in the unexposed. In the misclassified scenario, the risk of disease in the exposed is reduced by the proportion of cases misclassified as controls (i.e.,  $1 - \text{sensitivity}$ ). If this misclassification is nondifferential with respect to exposure then the risk of disease in the unexposed is reduced by the same proportion. Although the risks (or rates) of disease in the exposed and in the unexposed in the misclassified scenario are incorrect in calculating the RR, the error in the classification of the cases among the exposed is cancelled out by the equal error in the classification of the cases among the unexposed. Although the RR calculated in the perfect classification scenario and the RR calculated in the misclassified scenario are identical, the statistical power to detect an association will be reduced in the latter because the variance of this estimate depends on the number of observed exposed cases and unexposed cases, which is reduced with this type of case misclassification. However, the larger the

**TABLE II. Estimate\* of the OR, 95% CI, and the Statistical Significance of the OR for Combinations of Extent of Nondifferential Misclassification of Cases as Controls, True ORs, Prevalence of Exposure in the Controls, and per Group Sample Size**

Prevalence of exposure in true controls	True OR = 3					True OR = 1.5					True OR = 1.25					
	Per group sample size															
	10,000	1,000	500	200	10,000	1,000	500	200	10,000	1,000	500	200	10,000	1,000	500	200
Perfect classification																
0.5	2.8-3.2	2.5-3.6	2.3-3.9	2.0-4.6	1.4-1.6	1.3-1.8	1.2-1.9	1.0-2.2	1.2-1.3	1.0-1.5	1.0-1.6	1.0-2.2	1.2-1.3	1.0-1.5	1.0-1.6	0.8-1.8
0.25	2.8-3.2	2.5-3.6	2.3-3.9	2.0-4.6	1.4-1.6	1.2-1.8	1.1-2.0	1.0-2.3	1.2-1.3	1.0-1.5	0.9-1.7	0.9-1.7	1.2-1.3	1.0-1.5	0.9-1.7	0.8-1.9
0.1	2.8-3.2	2.3-3.9	2.1-4.3	1.7-5.3	1.4-1.7	1.2-2.0	1.0-2.2	0.8-2.7	1.1-1.4	0.9-1.7	0.8-1.9	0.8-2.7	1.1-1.4	0.9-1.7	0.8-1.9	0.7-2.3
10% Misclassification																
0.5	2.6-2.9	2.2-3.3	2.1-3.6	1.8-4.1	1.4-1.5	1.2-1.7	1.1-1.9	1.0-2.1	1.2-1.3	1.0-1.5	1.0-1.6	1.0-2.1	1.2-1.3	1.0-1.5	1.0-1.6	0.8-1.8
0.25	2.5-2.8	2.2-3.2	2.0-3.4	1.7-4.0	1.4-1.5	1.2-1.8	1.1-1.9	0.9-2.2	1.1-1.3	1.0-1.5	0.9-1.6	0.9-2.2	1.1-1.3	1.0-1.5	0.9-1.6	0.8-1.9
0.1	2.4-2.8	2.0-3.3	1.8-3.6	1.5-4.4	1.3-1.6	1.1-1.9	1.0-2.1	0.8-2.7	1.1-1.3	0.9-1.6	0.8-1.8	0.8-2.7	1.1-1.3	0.9-1.6	0.8-1.8	0.7-2.3
20% Misclassification																
0.5	2.3-2.6	2.0-3.0	1.9-3.2	1.6-3.8	1.3-1.5	1.2-1.7	1.1-1.8	0.9-2.1	1.1-1.3	1.0-1.4	0.9-1.5	0.9-2.1	1.1-1.3	1.0-1.4	0.9-1.5	0.8-1.8
0.25	2.2-2.5	1.9-2.8	1.8-3.0	1.5-3.5	1.3-1.5	1.1-1.7	1.1-1.8	0.9-2.1	1.1-1.3	1.0-1.5	0.9-1.6	0.9-2.1	1.1-1.3	1.0-1.5	0.9-1.6	0.8-1.9
0.1	2.1-2.4	1.8-2.8	1.6-3.1	1.3-3.8	1.3-1.5	1.1-1.8	1.0-2.0	0.8-2.5	1.1-1.3	0.9-1.6	0.8-1.8	0.8-2.5	1.1-1.3	0.9-1.6	0.8-1.8	0.7-2.3
30% Misclassification																
0.5	2.1-2.4	1.8-2.7	1.7-2.9	1.5-3.5	1.3-1.4	1.1-1.6	1.0-1.7	0.9-2.0	1.1-1.2	1.0-1.4	0.9-1.5	0.9-2.0	1.1-1.2	1.0-1.4	0.9-1.5	0.8-1.7
0.25	2.0-2.2	1.7-2.5	1.6-2.7	1.4-3.2	1.2-1.4	1.1-1.6	1.0-1.7	0.8-2.0	1.1-1.2	1.0-1.4	0.9-1.5	0.8-2.0	1.1-1.2	1.0-1.4	0.9-1.5	0.7-1.8
0.1	1.8-2.1	1.6-2.5	1.4-2.7	1.2-3.4	1.2-1.4	1.0-1.7	0.9-1.9	0.7-2.4	1.1-1.3	0.9-1.5	0.8-1.7	0.7-2.4	1.1-1.3	0.9-1.5	0.8-1.7	0.6-2.2
40% Misclassification																
0.5	1.9-2.1	1.7-2.4	1.5-2.6	1.3-3.1	1.2-1.4	1.1-1.5	1.0-1.6	0.9-1.9	1.1-1.2	1.0-1.4	0.9-1.5	0.9-1.9	1.1-1.2	1.0-1.4	0.9-1.5	0.8-1.7
0.25	1.8-2.0	1.6-2.2	1.4-2.4	1.2-2.8	1.2-1.3	1.1-1.5	1.0-1.7	0.8-1.9	1.1-1.2	0.9-1.4	0.9-1.5	0.8-1.9	1.1-1.2	0.9-1.4	0.9-1.5	0.7-1.8
0.1	1.6-1.9	1.4-2.2	1.3-2.4	1.1-2.9	1.2-1.4	1.0-1.6	0.9-1.9	0.7-2.3	1.0-1.2	0.9-1.5	0.8-1.7	0.7-2.3	1.0-1.2	0.9-1.5	0.8-1.7	0.6-2.2
50% Misclassification																
0.5	1.7-1.9	1.5-2.2	1.4-2.4	1.2-2.8	1.2-1.3	1.0-1.5	1.0-1.6	0.8-1.8	1.1-1.2	0.9-1.3	0.9-1.4	0.8-1.8	1.1-1.2	0.9-1.3	0.9-1.4	0.8-1.7
0.25	1.6-1.8	1.4-2.0	1.3-2.2	1.1-2.5	1.1-1.3	1.0-1.5	0.9-1.6	0.8-1.8	1.0-1.2	0.9-1.4	0.8-1.5	0.8-1.8	1.0-1.2	0.9-1.4	0.8-1.5	0.7-1.7
0.1	1.5-1.7	1.3-2.0	1.2-2.2	1.0-2.6	1.1-1.3	0.9-1.6	0.8-1.7	0.7-2.1	1.0-1.2	0.8-1.5	0.8-1.7	0.7-2.1	1.0-1.2	0.8-1.5	0.8-1.7	0.6-2.0

\*ORs and 95% CIs estimated from a logistic regression model. Statistically significant associations are shown in bold.



**Fig. 4.** What is the influence of misclassification of cases as controls on the RR and its statistical significance in a cohort study? (Assumed that the prevalence of exposure in the misclassified cases is the same as the correctly classified cases and that misclassification is nondifferential with respect to exposure.)



study and the more common the disease under study, the greater the likelihood that the association will still be detectable as statistically significant.

### Comments and Solutions to Issue 2

Even if 10–30% and even possibly 50% of the controls are actually men with undetected prostate cancer, this type of nondifferential misclassification of undetected cases as controls is unlikely to fully obscure moderate to strong associations between an exposure and prostate cancer in case-control and nested case-controls studies of moderate to large size especially if the misclassified cases have an etiology that is somewhat distinct from the correctly classified cases. Godley and Schell [1999] suggest that the etiology of occult and detected prostate cancers are likely similar because some tumors found in autopsy studies were as large as those that were clinically detected. However, this argument pertains mainly to the pre-PSA era when even “clinical” cases (e.g., palpable or occult metastases detected) may have gone undetected, but pertains so to the PSA era when the opportunity for detection of even small volume, subclinical cases is high in the population. Ultimately, detailed evaluation and extended clinical follow-up of the prostate cancer cases that were detected on end-of-study biopsy compared with those that were detected on biopsy for indication in the Prostate Cancer Prevention Trial will provide empirical data on whether the former cases are biologically similar to the latter cases. Given the fact that many cases detected by PSA screening would be currently considered clinically insignificant based on tumor volume and Gleason score at radical prostatectomy, there is little reason to believe currently that cases detected by random biopsy in men without elevated PSA would be markedly different biologically than those detected with an elevated, but low PSA.

Nevertheless, investigators who are concerned about misclassification of cases as controls as an explanation for their null findings on risk factors for prostate cancer can perform secondary analyses to estimate what the true OR would have been for a range of misclassified proportions. Approaches to estimating the extent of misclassification using PSA concentration [Whittemore et al., 1995] and methods to adjust for misclassification in statistical analysis have been described [Copeland et al., 1977;

Godley and Schell, 1999]. Alternatively, where feasible investigators may conduct prospective cohort studies or the case-cohort variant and perform analyses to avoid bias to the null if sensitivity is imperfect (but specificity is perfect).

What should be avoided when attempting to limit the extent of cases being admixed with controls is defining the controls as men with very low PSA (e.g., <2 ng/ml). The net effect of this restriction is to eliminate from the control not just men with prostate cancer, but those with BPH and prostatitis. The case group on the other hand would still include men with BPH and prostatitis. In fact the elevation in PSA that triggered the cases' biopsy and prostate cancer diagnosis actually may have been due to their BPH and prostatitis. The lack of comparability between the cases and controls on the presence of BPH and prostatitis might result in the inadvertent study of risk factors for benign prostate conditions, not prostate cancer.

### CONCLUSIONS

Widespread use of PSA testing for early detection of prostate cancer has necessitated discussion of the influence of the changed case mix and the nature of prostate cancer “cases” and “controls” when drawing inferences about the etiology of prostate cancer from epidemiologic studies conducted in the PSA era. That prostate cancer cases diagnosed in the PSA era are enriched with early stage, minimally invasive disease in our opinion is likely to pose a far more serious obstacle to epidemiologic research on the etiology of clinically important prostate cancer than the issue of inclusion as controls some men who have undiagnosed prostate cancer because of imperfect sensitivity of PSA screening and biopsy sampling error. Although the imperfect sensitivity of prostate cancer detection is widely perceived as a problem, it is unlikely to fully obscure associations in reasonably sized case-control studies of exposures with modest to moderate prevalences. As the sensitivity of screening for prostate cancer increases even further, the effect of including clinically significant cases in the control group on the OR will further diminish, whereas the effect of including clinically insignificant cases in the case group on the OR will further increase. In fact, emerging evidence indicates that the broadening of the scope of prostate cancer

diagnosis in the PSA era may obscure associations with potentially important factors, such as IGF-1 and androgens, which may influence the progression of prostate cancer. Paradoxically, aggressive attempts of including all cases may actually hinder attempts at elucidating risk factors for the most clinically relevant cancers. Also, in general the risk factors that are of most interest for study and intervention may be those for which the admixture of early cases with the late cases (Issue 1) is more important to avoid than admixture of cases with controls (Issue 2). For example, if 50% of middle aged and older men do actually have prostate cancer, then the causal exposures must be nearly ubiquitous and may occur early in life making them less feasible to be targeted for prevention. In contrast, exposures that cause progression are those that are more important from the perspective of preventing metastasis and death and are those that are more likely to be influenced by the admixture of early cases with the late cases (Issue 1). With attention to the choice of case criteria, study populations, designs, sample size, and hypotheses, research on the etiology of prostate cancer will continue to move forward now that we are conducting etiologic research in the PSA era.

#### ACKNOWLEDGMENTS

We thank Sabine Rohrmann, PhD for her helpful comments on this paper. This work was presented at the 11th Annual SPORE Investigators' Workshop, July 8–10, Baltimore, MD.

#### REFERENCES

- American Cancer Society. 2003a. Can Prostate Cancer Be Found Early? [http://www.cancer.org/docroot/CRI/content/CRI\\_2\\_4\\_3X\\_Can\\_prostate\\_cancer\\_be\\_found\\_early\\_36.asp?sitearea=](http://www.cancer.org/docroot/CRI/content/CRI_2_4_3X_Can_prostate_cancer_be_found_early_36.asp?sitearea=)
- American Cancer Society. 2003b. Cancer Facts & Figures, 2003. <http://www.cancer.org/downloads/STT/CAFF2003PWSecured.pdf>
- American Urological Association Adult conditions. Prostate—Prostate cancer screening. <http://www.urologyhealth.org/adult/index.cfm?cat=09&topic=250>
- Bostwick DG. 1997. Evaluating prostate needle biopsy: Therapeutic and prognostic importance. *CA Cancer J Clin* 47:297–319.
- Brenner H, Savitz DA. 1990. The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control studies. *Am J Epidemiol* 132:181–192.
- Carroll P, Coley C, McLeod D, Schellhammer P, Sweat G, Wasson J, Zietman A, Thompson I. 2001. Prostate-specific antigen best practice policy—Part I: Early detection and diagnosis of prostate cancer. *Urology* 57:217–224.
- Chamberlain NL, Driver ED, Miesfeld RL. 1994. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res* 22:3181–3186.
- Chan JM, Stampfer MJ, Giovannucci E, Gann PH, Ma J, Wilkinson P, Hennekens CH, Pollak M. 1998. Plasma insulin-like growth factor-I and prostate cancer risk: A prospective study. *Science* 279:563–566.
- Chan JM, Stampfer MJ, Ma J, Gann P, Gaziano JM, Pollak M, Giovannucci E. 2002. Insulin-like growth factor-I (IGF-1) and IGF binding protein-3 as predictors of advanced-stage prostate cancer. *J Natl Cancer Inst* 94:1099–1109.
- Clinton SK, Emenhiser C, Schwartz SJ, Bostwick DG, Williams AW, Moore BJ, Erdman JW, Jr. 1996. Cis-trans lycopene isomers, carotenoids, and retinol in the human prostate. *Cancer Epidemiol Biomarkers Prev* 5:823–833.
- Cohen P, Peehl DM, Lamson G, Rosenfeld RG. 1991. Insulin-like growth factors (IGFs), IGF receptors, and IGF-binding proteins in primary cultures of prostate epithelial cells. *J Clin Endocrinol Metab* 73:401–407.
- Cohen P, Peehl DM, Rosenfeld RG. 1994. The IGF axis in the prostate. *Hormone Metab Res* 26:81–84.
- Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. 1977. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 105:488–495.
- de Koning HJ, Liem MK, Baan CA, Boer R, Schroder FH, Alexander FE. 2002. Prostate cancer mortality reduction by screening: Power and time frame with complete enrollment in the European Randomised Screening for Prostate Cancer (ERSPC) trial. *Int J Cancer* 98:268–273.
- DiGiuseppe JA, Sauvageot J, Epstein JI. 1997. Increasing incidence of minimal residual cancer in radical prostatectomy specimens. *Am J Surg Pathol* 21:174–178.
- Egevad L, Norberg M, Mattson S, Norlen BJ, Busch C. 1998. Estimation of prostate cancer volume by multiple core biopsies before radical prostatectomy. *Urology* 52:653–658.
- Fang J, Metter EJ, Landis P, Chan DW, Morrell CH, Carter HB. 2001. Low levels of prostate-specific antigen predict long-term risk of prostate cancer: Results from the Baltimore Longitudinal Study of Aging. *Urology* 58:411–416.
- Fleshner N, Klotz L. 2002. Role of “saturation biopsy” in the detection of prostate cancer among difficult diagnostic cases. *Urology* 60:93–97.
- Food and Drug Administration. 1994. FDA approves test for prostate cancer. <http://www.fda.gov/bbs/topics/ANSWERS/ANS00598.html>
- Gann PH, Hennekens CH, Stampfer MJ. 1995. A prospective evaluation of plasma prostate-specific antigen for detection of prostate cancer. *JAMA* 273:289–294.
- Gann PH, Hennekens CH, Ma J, Longcope C, Stampfer MJ. 1996. Prospective study of sex hormone levels and risk of prostate cancer. *J Natl Cancer Inst* 88:1118–1126.
- Gann PH, Ma J, Giovannucci E, Willett W, Sacks F, Hennekens CH, Stampfer MJ. 1999. Lower prostate cancer risk in men with elevated plasma lycopene levels: Results of a prospective analysis. *Cancer Res* 59:1225–1230.

- Giovannucci E, Stampfer MJ, Krithivas K, Brown M, Dahl DM, Brufsky A, Talcott J, Hennekens CH, Kantoff PW. 1997. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc Natl Acad Sci* 94:3320–3323.
- Godley PA, Schell MJ. 1999. Adjusted odds ratios under nondifferential misclassification: Application to prostate cancer. *J Clin Epidemiol* 52:129–136.
- Gohagan JK, Prorok PC, Hayes RB, Kramer BS. 2000. The prostate, lung, colorectal and ovarian (PLCO) cancer screening trial of the National Cancer Institute: History, organization, and status. *Control Clin Trials* 21:251S–272S.
- Hankey BF, Feuer EJ, Clegg LX, Hayes RB, Legler JM, Prorok PC, Ries LA, Merrill RM, Kaplan RS. 1999. Cancer surveillance series: Interpreting trends in prostate cancer—Part I: Evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. *J Natl Cancer Inst* 91:1017–1024.
- Kazemi-Esfarjani P, Trifiro MA, Pinsky L. 1995. Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: Possible pathogenic relevance for the (CAG)<sub>n</sub>-expanded neuropathies. *Hum Mol Genet* 4:523–527.
- Platz EA, Leitzmann MF, Rifai N, Kantoff PW, Pollak M, Stampfer MJ, Willett WC, Giovannucci E. 2003. Hormonal pathways and subsequent risk of prostate cancer in the PSA era. *Proc Am Asso Cancer Res* 44:720.
- Potosky AL, Miller BA, Albertsen PC, Kramer BS. 1995. The role of increasing detection in the rising incidence of prostate cancer. *J Am Med Assoc* 273:548–552.
- Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM. 2003. Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. *N Engl J Med* 349:335–342.
- Quaglia A, Parodi S, Grosclaude P, Martinez-Garcia C, Coebergh JW, Vercelli M. 2003. Differences in the epidemic rise and decrease of prostate cancer among geographical areas in southern Europe. An analysis of differential trends in incidence and mortality in France, Italy, and Spain. *Eur J Cancer* 39:654–665.
- Roehl KA, Antenor JA, Catalona WJ. 2002. Serial biopsy results in prostate cancer screening study. *J Urol* 167:2435–2439.
- Roehrborn CG, Boyle P, Gould AL, Waldstreicher J. 1999. Serum prostate-specific antigen as a predictor of prostate volume in men with benign prostatic hyperplasia. *Urology* 53:581–589.
- Scheck A. 2001. 'Extreme' prostate biopsies: Too many or not enough? *Urol Times* <http://ut.adv100.com/urologytimes/article/articleDetail.jsp?id=2500>
- Sies H, Stahl W. 1995. Vitamins E and C,  $\beta$ -carotene, and other carotenoids as antioxidants. *Am J Clin Nutr* 62(Suppl):1315S–1321S.
- Sirovich BE, Schwartz LM, Woloshin S. 2003. Screening men for prostate and colorectal cancer in the United States: Does practice reflect the evidence? *JAMA* 289:1414–1420.
- Stamey TA, Johnstone IM, McNeal JE, Lu AY, Yemoto CM. 2002. Preoperative serum prostate specific antigen levels between 2 and 22 ng/ml correlate poorly with post-radical prostatectomy cancer morphology: Prostate specific antigen cure rates appear constant between 2 and 9 ng/ml. *J Urol* 167:103–111.
- Stanford JL, Stephenson RA, Coyle LM, Cerhan J, Correa R, Eley JW, Gilliland F, Hankey B, Kolonel LN, Kosary C, Ross R, Severson R, West D. 1999. Prostate Cancer Trends 1973–1995, Bethesda, MD: SEER Program, National Cancer Institute NIH Pub. No. 99-4543.
- Stewart CS, Leibovich BC, Weaver AL, Lieber MM. 2001. Prostate cancer diagnosis using a saturation needle biopsy technique after previous negative sextant biopsies. *J Urol* 166:86–91.
- Thompson I, Goodman P, Tangen C, Lucia M, Miller G, Ford L, Lieber M, Cespedes R, Atkins J, Lippman S, Carlin S, Ryan A, Szczepanek C, Crowley J, Coltman C. 2003. The influence of finasteride on the development of prostate cancer. *N Engl J Med* 349:215–224.
- U.S. Preventive Services Task Force. 2002. Screening—Prostate cancer, U.S. Preventive Services Task Force Update.
- White E. 1986. The effect of misclassification of disease status in follow-up studies: Implications for selecting disease classification criteria. *Am J Epidemiol* 124:816–825.
- Whittemore AS, Kolonel LN, Wu AH, John EM, Gallagher RP, Howe GR, Burch JD, Hankin J, Dreon DM, West DW, Teh CZ, Paffenbarger RS. 1995. Prostate cancer in relation to diet, physical activity, and body size in blacks, whites, and Asians in the United States and Canada. *J Natl Cancer Inst* 87:652–661.
- Wu K, Erdman JW, Schwartz SJ, Platz EA, Leitzmann M, Clinton SK, DeGross V, Willett WC, Giovannucci EL. Plasma and dietary carotenoids and the risk of prostate cancer—A nested case-control study. *Cancer Epidemiol Biomarkers Prev* (in press).
- Yoshizawa K, Willett WC, Morris SJ, Stampfer MJ, Spiegelman D, Rimm EB, Giovannucci E. 1998. A study of prediagnostic selenium level in toenails and the risk of advanced prostate cancer. *J Natl Cancer Inst* 90:1219–1224.