Epidemiology

Beyond the Basics

Moyses Szklo, MD, DrPH
Professor of Epidemiology
School of Hygiene and Public Health
Johns Hopkins University
Baltimore, Maryland

F. Javier Nieto, MD, PhD
Associate Professor of Epidemiology
School of Hygiene and Public Health
Johns Hopkins University
Baltimore, Maryland

AN ASPEN PUBLICATION®
Aspen Publishers, Inc.
Gaithersburg, Maryland
2000

PART III

Threats to Validity and Issues of Interpretation

nale for this procedure, and an important assumption justifying screening, is that, if left untreated, preclinical cases would necessarily become clinical cases; thus, there should not be a difference between the incidence of clinical and that of preclinical disease. However, when using available clinical disease incidence (eg, based on cancer registry data), it is important to adjust for differences in risk factor prevalence, expected to be higher in screenees than in the reference population from which clinical incidence is obtained. Thus, for example, a family history of breast cancer is more prevalent in individuals screened for breast cancer than in the female population at large.

Next, using the duration of the DPCP estimate, the estimation of the average lead time needs to take into account whether early diagnosis by screening is made at the first or in subsequent screening exams.

The estimation of the lead time of point prevalent preclinical cases detected at the first screening exam relies on the assumptions regarding the distribution of times of early diagnosis during the DPCP. For example, if the distribution of early diagnosis by screening can be assumed to be homogeneous throughout the DPCP—that is, if the sensitivity of the screening test is independent of time within the DPCP (Figure 4–10A)—the lead time of point prevalent preclinical cases can be simply estimated as

$$\text{Lead time} = \frac{\text{DPCP}}{2}$$

The latter assumption, however, may not be justified in many situations. For most diseases amenable to screening (eg, breast cancer), the sensitivity of the screening test, and thus the probability of early diagnosis, is likely to increase during the DPCP (Figure 4–10B) as a result of the progression of the disease as it gets closer to its symptomatic (clinical) phase. If this is the case, a more reasonable assumption would be that the average lead time is less than one half of the DPCP. Obviously, the longer the DPCP, the longer the lead time under any distributional assumption. Also, because the average lead time is a function of the validity of the screening exam, it becomes longer as more sensitive screening tests are developed.

The duration of the lead time for *incident* preclinical cases identified in a program where repeated screening exams are carried out is a function of how often the screenings are done—that is, the length of the interval between successive screenings. The closer in time the screening exams are, the greater the probability that early diagnosis will occur closer to the onset of the DPCP—and thus, the more the lead time will approximate the DPCP.

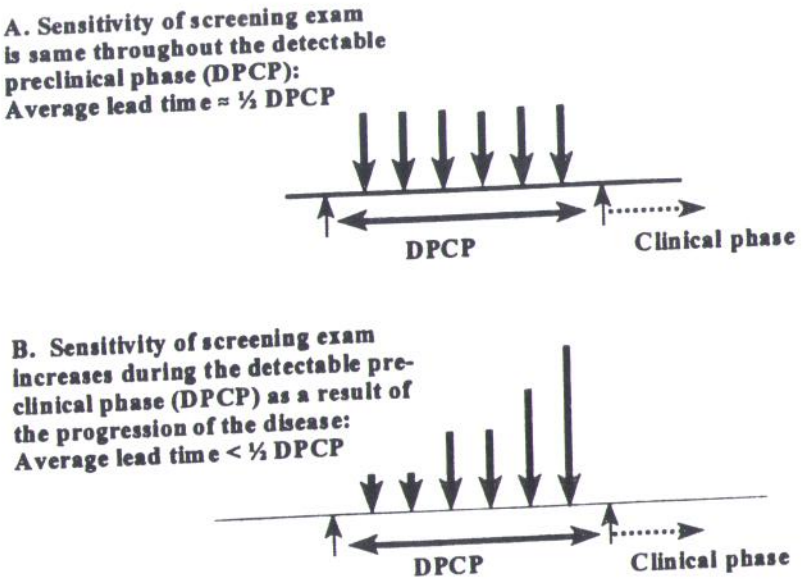

Figure 4–10 Estimation of lead time as a function of the variability of the sensitivity of the screening exam during the detectable preclinical phase.

Figure 4–11 illustrates schematically short and long between-screening intervals and their effect on the lead time. Assuming two persons with similar DPCPs whose disease starts soon after the previous screening, the person with the shorter between-screening interval (a to b, patient A) has his or her newly developed preclinical disease diagnosed nearer the beginning of the DPCP than the person with the longer between-screening interval (a to c, patient B). Thus, the duration of the lead time is closer to the duration of the DPCP for patient A than for patient B. The maximum lead time obviously cannot be longer than the DPCP.[43]

4.5 BIASES IN REPORTING STUDY RESULTS: PUBLICATION BIAS

Some common problems that may influence the quality of reporting of epidemiologic studies and that could potentially create report-

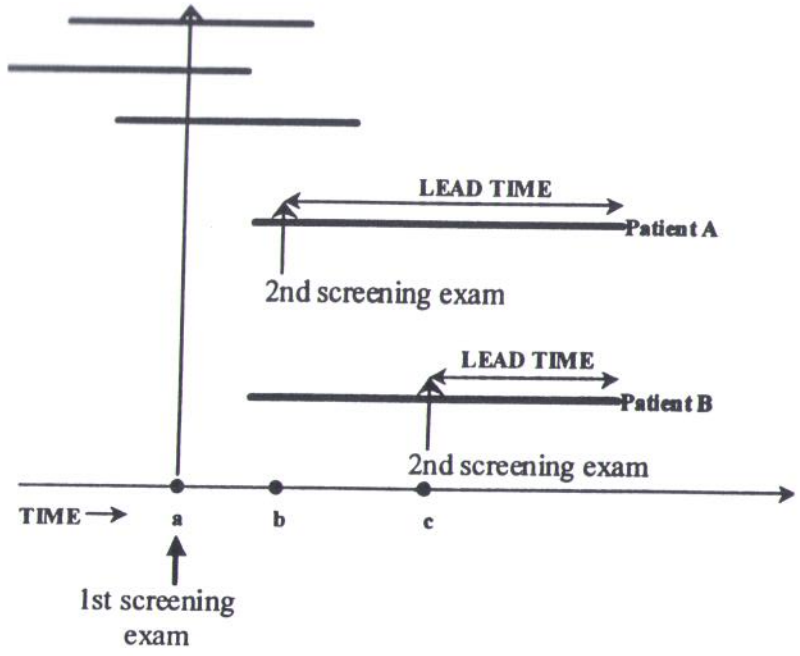

Figure 4–11 Relationship between frequency of screening and duration of lead time. Horizontal lines represent duration of detectable preclinical phase (DPCP). In patient A, the second screening exam is carried out soon after the first screening exam: lead time ≈ DPCP. In patient B, the between-screening interval is longer: lead time ≈ (.05) × DPCP.

ing biases are discussed in detailed in Chapter 9. In this section, we focus on a special type of reporting bias: *publication bias*.

Acceptance of the validity of published findings as applied to a given reference population is conditional on two important assumptions: (1) that each published study is unbiased and (2) that published studies constitute an unbiased sample of a theoretical "population" of unbiased studies (Figures 4–12 and 4–13). When these assumptions are not met, a literature review based on either meta-analytic or conventional narrative approaches will give a distorted view of the exposure-outcome association of interest. Whether assumption 1 above is met depends on several factors, including the soundness of the study designs and the quality of peer review. Publication bias, as conventionally defined, occurs when assumption 2 cannot be met because,

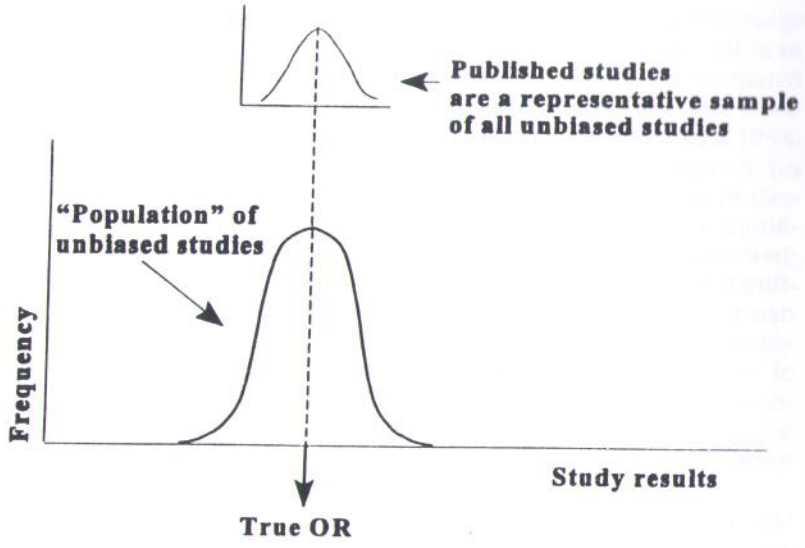

Figure 4–12 Published studies constitute a representative sample of a theoretical reference "population" of unbiased studies (no publication bias).

besides the quality of the report, other factors dictate acceptability for publication. Direction of findings is one such factor. For example, in a study carried out a few decades ago, Sterling[44] demonstrated that 97% of papers published in four psychology journals showed statistically significant results at the alpha level of 5%, thus strongly suggesting that results were more likely to be published if they had reached this conventional significance level.

Subsequent research[45–49] has confirmed the tendency to publish "positive" results. In a study by Dickersin[47] comparing published randomized clinical trials with completed yet unpublished randomized trials by the same investigators, 55% of the published, but only 15% of the unpublished studies favored the new therapy being assessed.

A possible reason for publication bias is the reluctance of journal editors and reviewers to accept negative ("uninteresting"?) results. For example, in a study by Mahoney,[50] a group of 75 reviewers were asked to review different versions (randomly assigned) of a fictitious manuscript. The "Introduction" and "Methods" sections in all versions were identical, whereas the "Results" and "Discussion" sections were different, ranging from "positive" to "ambiguous" to "negative" results. Reviewers were asked to evaluate the methods, the data presen-

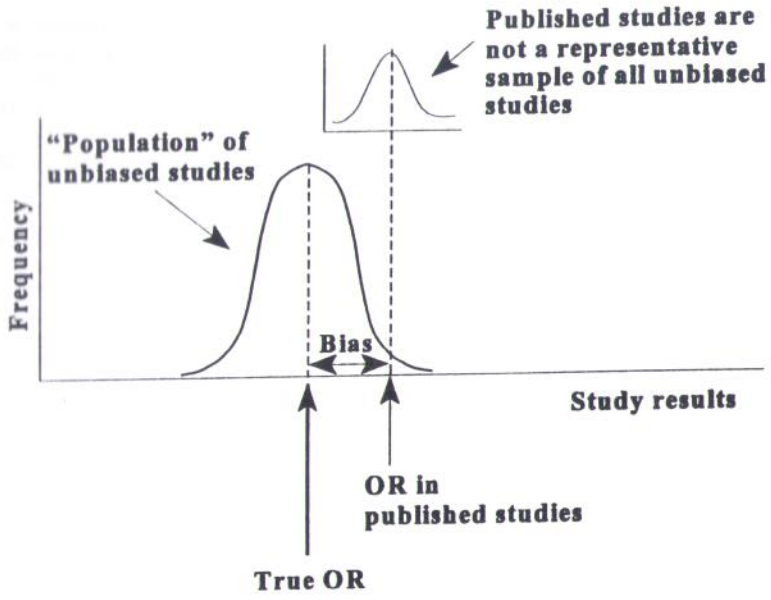

Figure 4–13 Publication bias: published studies constitute a biased sample of a theoretical reference "population" of unbiased studies (publication bias).

tation, the scientific contribution, and the publication merit. Compared with negative or ambiguous studies, the manuscripts with "positive" results systematically received higher average scores for all categories, including the category for evaluation of methods, even though the "Methods" section was identical in all sets.

Interestingly, however, editors and reviewers may not be the only source of publication bias. For example, in Dickersin et al's study,[47] for only 12% of unpublished yet completed studies, publication was intended by the authors. Reasons given by the authors why publication was not intended for the remaining 88% included "negative results" (28%), "lack of interest" (12%), and "sample size problems" (11%).

Even source of support seems to interfere with the likelihood of publication. Davidson,[51] for example, found that clinical trials favoring a new over a traditional therapy funded by the pharmaceutical industry had a publication odds 5.2 greater than that of trials supported by other sources, such as the National Institutes of Health.

Some forms of publication bias are rather subtle, such as "language bias" (ie, publication in an English language-journal vs publication in a journal in another language). For example, Egger et al[52] compared randomized clinical trials published by German investigators in either German journals or English journals from 1985 through 1994. When pairs of articles with the same first author were compared, no evidence of differences in quality between the papers written in German or English was found. In contrast, a strong, statistically significant difference with regard to the *significance* of findings was present: 63% of the articles published in English reported a statistically significant result, compared to only 35% of articles published in German (OR = 3.8; 95% confidence limits 1.3–11.3). These results strongly suggest that using language as one of the criteria to select studies to be included in a meta-analysis (a criterion that is frequently adopted because of practical reasons or reasons related to access) can seriously undermine the representativeness of published reviews, including those using meta-analysis methods.

The two most often cited systematic approaches to avoid publication bias are the creation of study registers[49] and advance publication of research designs.[53] Prevention of publication bias obviously requires efforts on the part of the scientific community as a whole, including researchers, peer reviewers, and journal editors. The latter should be aware that direction of findings and absence of "significant" results should not be used as criteria for rejection or acceptance of a paper.

REFERENCES

1. Sackett DL. Bias in analytical research. *J Chron Dis*. 1979;32:51–63.
2. Last JM. *A Dictionary of Epidemiology*. 3rd ed. New York, NY: Oxford University Press; 1995:15.
3. Lilienfeld DE, Stolley PD. *Foundations of Epidemiology*. New York, NY: Oxford University Press; 1994.
4. Schlesselman JJ. *Case-Control Studies: Design, Conduct, Analysis*. New York, NY: Oxford University Press; 1982.
5. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, Pa: Lippincott-Raven Publishers; 1998.
6. Byrne C, Brinton LA, Haile RW, et al. Heterogeneity of the effect of family history on breast cancer risk. *Epidemiology*. 1991;2:276–284.
7. MacMahon B, Yen S, Trichopoulus D, et al. Coffee and cancer of the pancreas. *N Engl J Med*. 1981;304:630–633.
8. Hsieh CC, MacMahon B, Yen S, et al. Coffee and pancreatic cancer. *N Engl J Med*. 1986;315:587–588. Letter.

9. Gordis L. *Epidemiology*. Philadelphia, Pa: WB Saunders; 1996.
10. Weinstock MA, Colditz GA, Willett WC, et al. Recall (report) bias and reliability in the retrospective assessment of melanoma risk. *Am J Epidemiol*. 1991;133:240–245.
11. Brinton LA, Hoover RN, Szklo M, et al. Menopausal estrogen use and risk of breast cancer. *Cancer*. 1981;47:2517–2522.
12. Chambless LE, Heiss G, Folsom AR, et al. Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: the Atherosclerosis Risk in Communities (ARIC) study, 1987-1993. *Am J Epidemiol*. 1997;146:483–494.
13. Mele A, Szklo M, Visani G, et al. Hair dye use and other risk factors for leukemia and pre-leukemia: a case-control study. Italian Leukemia Study Group. *Am J Epidemiol*. 1994;139:609–619.
14. Wei Q, Matanoski GM, Farmer ER, et al. DNA repair and susceptibility to basal cell carcinoma: a case-control study. *Am J Epidemiol*. 1994;140:598–607.
15. Commonwealth of Australia, National Health and Medical Research Council. *The Health Effects of Passive Smoking*. November 1997. Canberra: Australian Government Publishing Service.
16. Szklo M, Tonascia J, Gordis L. Psychosocial factors and the risk of myocardial infarctions in white women. *Am J Epidemiol*. 1976;103:312–320.
17. Doll R, Hill AB. Smoking and carcinoma of the lung: preliminary report. *B Med J*. 1950;2:739–748.
18. Perneger TV, Whelton PK, Klag MJ, Rossiter KA. Diagnosis of hypertensive end-stage renal disease: effect of patient's race. *Am J Epidemiol*. 1995;141:10–15.
19. Stewart WF, Linet MS, Celentano DD, et al. Age- and sex-specific incidence rates of migraine with and without visual aura. *Am J Epidemiol*. 1991;134:1111–1120.
20. Rose GA. Chest pain questionnaire. *Milbank Mem Fund Q*. 1965;43:32–39.
21. Bass EB, Follansbee WP, Orchard TJ. Comparison of a supplemented Rose questionnaire to exercise thalium testing in men and women. *J Clin Epidemiol*. 1989;42:385–393.
22. Garber CE, Carleton RA, Heller GV. Comparison of "Rose questionnaire angina" to exercise thalium scintigraphy: different findings in males and females. *J Clin Epidemiol*. 1992;45:715–720.
23. Sorlie PD, Cooper L, Schreiner PJ, et al. Repeatability and validity of the Rose questionnaire for angina pectoris in the Atherosclerosis Risk in Communities study. *J Clin Epidemiol*. 1996;49:719–725.
24. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol*. 1990;132:746–748.
25. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol*. 1980;112:564–569.
26. Wacholder S. When measurement errors correlate with truth: surprising effects of nondifferential misclassification. *Epidemiology*. 1995;6:157–161.
27. Flegal KM, Brownie C, Haas JD. The effects of misclassification on estimates of relative risk. *Am J Epidemiol*. 1986;123:736–751.
28. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol*. 1991;134:1233–1244.

29. Willet W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. *Stat Med*. 1989;8:1031–1040.
30. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease. Relationships and methods of correction. *Annu Rev Public Health*. 1993;14:69–93.
31. Armstrong BG. The effects of measurement errors on relative risk regressions. *Am J Epidemiol*. 1990;132:1176–1184.
32. Kannel WB. CHD risk factors: a Framingham study update. *Hosp Prac*. 1990;25:93–104.
33. Giovannucci E, Ascherio A, Rimm EB, et al. A prospective cohort study of vasectomy and prostate cancer in US men. *JAMA*. 1993;269:873–877.
34. Goldberg RJ, Gorak EJ, Yarzebski J. A communitywide perspective of sex differences and temporal trends in the incidence and survival rates after acute myocardial infarction and out-of-hospital deaths caused by coronary heart disease. *Circulation*. 1993;87:1947–1953.
35. Perneger TV, Nieto FJ, Whelton PK. A prospective study of blood pressure and serum creatinine: results from the "CLUE" study and the ARIC study. *JAMA*. 1993;269:488–493.
36. Szklo M. Aplastic anemia. In: Lilienfeld AM, ed. *Reviews in Cancer Epidemiology*, vol 1. New York, NY: Elsevier North-Holland; 1980:115–119.
37. Antunes CM, Stolley PD, Rosenshein NB. Endometrial cancer and estrogen use: report of a large case-control study. *N Engl J Med*. 1979;300:9–13.
38. Horwitz RI, Feinstein AR. Estrogens and endometrial cancer: responses to arguments and current status of an epidemiologic controversy. *Am J Med*. 1986;81:503–507.
39. Brunekreef B, Groot B, Hoek G. Pets, allergy and respiratory symptoms in children. *Int J Epidemiol*. 1992;21:338–342.
40. Nieto FJ, Diez-Roux A, Szklo M, Comstock GW, Sharret AR. Short and long term prediction of clinical and subclinical atherosclerosis by traditional risk factors. *Int J Epidemiol*. 1999;52:559–567.
41. Shapiro S, Venet W, Strax P. Selection, follow-up, and analysis in the Health Insurance Plan study: a randomized trial with breast cancer screening. *Natl Cancer Inst Monogr*. 1985;67:65–74.
42. Hutchison GB, Shapiro S. Lead time gained by diagnostic time screening for breast cancer. *J Natl Cancer Inst*. 1968;41:665.
43. Morrison AS. *Screening in Chronic Disease*. New York, NY: Oxford University Press; 1985. Monographs in Epidemiology and Biostatistics, vol. 7.
44. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa. *J Am Stat Assoc*. 1959;54:30–34.
45. Dickersin K, Hewitt P, Mutch L, et al. Perusing the literature: comparison of MEDLINE searching with a Perinatal Trials database. *Control Clin Trials*. 1985;6:306–317.
46. Dickersin K, Chan S, Chalmers TC, et al. Publication bias and clinical trials. *Control Clin Trials*. 1987;8:343–353.
47. Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results. *JAMA*. 1992;267:374–378.

48. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. *Lancet*. 1991;337:867–872.

49. Dickersin K. Report from the Panel on the Case for Registers of Clinical Trials at the eighth annual meeting of the Society for Clinical Trials. *Control Clin Trials*. 1988;9:76–81.

50. Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cog Ther Res*. 1977;1:161–175.

51. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med*. 1986;1:155–158.

52. Egger M, Zellweger-Zahner T, Schneider M, et al. Language bias in randomised controlled trials publishd in English and German. *Lancet*. 1997;350:326–329.

53. Piantadosi S, Byar DP. A proposal for registering clinical trials. *Control Clin Trials*. 1988;9:82–84.

CHAPTER 5

Identifying Noncausal Associations: Confounding

5.1 INTRODUCTION

The term *confounding* refers to a situation in which a noncausal association between a given exposure and an outcome is observed as a result of the influence of a third variable (or group of variables), usually designated as a *confounding variable*, or merely a *confounder*. As discussed below, the confounding variable must be related to both the putative risk factor and the outcome under study. In an observational cohort study, for example, a confounding variable would differ between exposed and unexposed subjects and would also be associated with the outcome of interest. As shown in the examples that follow, this may result either in the appearance or strengthening of an association not due to a direct causal effect or in the apparent absence or weakening of a true causal association.

From the epidemiological standpoint, it is useful to conceptualize confounding as distinct from bias in that a confounded association, though not causal, is real (for further discussion of this concept, see Section 5.4.7). This distinction has obvious practical implications with respect to the relevance of exposures as *markers* of the presence or risk of disease for screening purposes (secondary prevention). If, on the other hand, the goal of the researcher is to carry out primary prevention, it becomes crucial to distinguish a causal from a noncausal association, the latter resulting from either bias or confounding. A number of statistical techniques are available to control for confounding; as described in detail in Chapter 7, the basic idea underlying adjustment is to use some *statistical model* to estimate what the as-