



# American Journal of EPIDEMIOLOGY

Volume 135

Number 5

March 1, 1992

Copyright © 1992 by The Johns Hopkins University

School of Hygiene and Public Health

Sponsored by the Society for Epidemiologic Research

---

## REVIEWS AND COMMENTARY

---

### The "Over-exposed" Control Group

Ernst L. Wynder and Steven D. Stellman

---

Relative risk is determined as much by the level of exposure among controls as among cases. If cases and controls are drawn from a population in which the range of exposures is narrow, then a study may yield little information about potential health effects. This may be one reason why an association between dietary fat and cancer has not been consistently observed in Western populations. Since the fat intake as a percent of total calories in the US general population varies little, only very large relative risks can be detected in epidemiologic studies. Investigators of the dietary fat hypothesis need to select study groups from populations where the risk factor is not, essentially, narrowly distributed. *Am J Epidemiol* 1992;135:459-61.

diet; dietary fats; epidemiologic methods; relative risk

---

Epidemiologists commonly calculate risk for disease in persons exposed to a factor of interest, relative to an unexposed reference group. For instance, as we study the risk of lung cancer in a given population, we frequently use as a reference group persons who have never smoked. Since lung cancer in nonsmokers is quite rare, it is common to obtain odds ratios in smokers reaching 10, 20, or even 30 relative to nonsmokers, depending on the amount smoked and duration of smoking habit.

---

Received for publication July 2, 1991, and in final form July 25, 1991.

From the American Health Foundation, New York, NY.

Reprint requests to Dr. Steven D. Stellman, Department of Epidemiology, American Health Foundation, 320 East 43rd Street, New York, NY 10017.

This work was supported by US Public Health Service grant nos. CA-32617 and CA-17613 from the National Cancer Institute.

Our study of squamous cell lung cancer in women (1) shows a relative risk of 8.6 for lung cancer among current cigarette smokers relative to nonsmokers. The "exposed" group consists of persons who smoked one or more cigarettes daily, while "unexposed" means never having smoked.

Consider now what would happen if our reference population consisted of those who smoked up to 10 cigarettes per day, including nonsmokers, and the "exposed" were persons who smoked 11 or more: the relative risk of the "exposed" group would be far smaller than 10. By expanding the "unexposed" reference group to include women who had previously been considered "exposed," the relative risk is cut nearly in half, to 4.7. This halving results solely from altering the definition of "exposure," so that the reference group now includes both non-

smokers and light smokers. Obviously, no one's actual risk has changed.

While this exercise may seem simplistic, we believe it has relevance to a number of epidemiologic studies, notably those involving nutrition.

We have previously argued that a major problem with nutritional assessment in epidemiologic studies is that measurement errors can be so large that, in effect, the sought-for signal becomes obscured by noise (2). This is particularly true in homogenous populations whose consumption of nutrients of interest lacks variation.

As a specific example, in case-control studies of either breast or colon cancer—two kinds of cancer hypothesized to be causatively related to consumption of dietary fat (3)—the range of fat consumption in nearly any set of controls chosen according to the usual criteria of comparability to cases is almost certain to yield a “high” average consumption of dietary fat. Even if, as hypothesized, the average fat consumption by cases is higher still, the possibility of obtaining a significant relative risk is greatly diminished, because in effect the controls are “exposed,” like the cases, only their exposure level may not be so extreme.

The situation reveals an important limitation in the usefulness of nutritional epidemiologic studies in Western countries, at least as far as dietary fat is concerned. If, for instance, 90 percent of controls have a typical fat intake which ranges from 30–44 percent of total calories (2), one would have to undertake an extremely large case-control study (up to 2,500 cases) in order to detect a doubling of risk in the highest 5 percent relative to the lowest 5 percent, and even then the latter group will contain persons “exposed” to 29 percent fat. A smaller study could be justified only if many more subjects were “exposed” to diets in excess of 45 percent fat, a level found in Western populations as rarely as extremely low levels. The necessity for a large sample size is further compounded by severe problems of measurement error, as Freedman et al. (4) and Kushi et al. (5) have shown for cohort studies of dietary fat and colorectal cancer.

On the other hand, if the controls included many persons whose diet was more like that of the Japanese, who consume 10–20 percent of calories as fat, then an association between a 30–40 percent fat diet and breast or colon cancer could more easily and confidently be tested. This would be true for two reasons: 1) even a large measurement error would not necessarily obscure the signal, and 2) if the dietary hypothesis is correct, then the range of exposure among the controls would not be practically the same as that among the cases.

In the smoking example, the apparent shift in relative risk resulting from regrouping of exposure categories may seem an obvious mathematic result (as it is), but the decision process for grouping exposures in the first place often presents difficult choices. Wartenberg and Northridge (6) recently pointed out that epidemiologists frequently “choose to dichotomize their exposure data to make the analysis of the data easier and its presentation more straightforward” (6, p. 1058).

Textbooks of epidemiology, in illustrating analytic methods, sometimes present exposure as a binary choice. For instance, when explaining analytic procedures for case-control studies, both Hennekens and Buring (7) and Kelsey et al. (8) present classic  $2 \times 2$  tables with exposures classified simply as Yes or No. Likewise, Breslow and Day (9) introduce the concept of comparative rates by supposing “that the population has been divided into two such subgroups, one exposed to the risk factor in question and the other not exposed” (9, p. 55).

Other authors, however, discuss the intrinsic meaning of exposure at some length. Schlesselman (10, p. 45) notes, for instance, that since most American adults had consumed at least some saccharin in 1980, “in one sense, probably all persons had been ‘exposed,’” and suggests that as a refinement “one must define a minimal exposure that shall be taken as the reference level.” In the case of dietary fat, even a “minimal” exposure in Western populations may differ little from the norm.

While it is methodologically inappropriate

to deliberately select "unexposed" controls in a case-control study, it is equally defeating to conduct a study in a population that is essentially uniformly "exposed," because such a study could not detect even a modest increase in risk. It may well be that epidemiologic study of the dietary fat hypothesis in breast and colon cancer may best be pursued in populations with more diverse diets than are generally found in the United States.

---

#### REFERENCES

1. Wynder EL, Stellman SD. Impact of long-term filter usage on lung and larynx cancer risk: a case-control study. *J Natl Cancer Inst* 1979;62:471-7.
2. Wynder EL, Hebert JR. Homogeneity in nutritional exposure: an impediment in cancer epidemiology. *J Natl Cancer Inst* 1987;79:605-7.
3. National Academy of Sciences, National Research Council, Committee on Diet, Nutrition, and Cancer. Diet, nutrition, and cancer. Washington, DC: National Academy Press, 1982.
4. Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement error on planning sample size required in a cohort study. *Am J Epidemiol* 1990;132:1185-95.
5. Kushi LH, Zelterman D, Jacobs DR Jr, et al. Re: "The impact of dietary measurement error on planning sample size required in a cohort study." (Letter.) *Am J Epidemiol* 1991;134:1470-2.
6. Wartenberg D, Northridge M. Defining exposure in case-control studies: a new approach. *Am J Epidemiol* 1991;133:1058-71.
7. Hennekens CH, Buring JE. *Epidemiology in medicine*. Boston: Little, Brown & Company, 1987.
8. Kelsey JL, Thompson WD, Evans AS. *Methods in observational epidemiology*. New York: Oxford University Press, 1986.
9. Breslow NE, Day NE. *Statistical methods in cancer research. Vol I. The analysis of case-control studies*. IARC scientific publications no. 32. Lyon, France: International Agency for Research on Cancer, 1980.
10. Schlesselman JJ. *Case-control studies. Design, conduct, analysis*. New York: Oxford University Press, 1982.