MODERN EPIDEMIOLOGY

KENNETH J. ROTHMAN

Professor, Department of Family and Community Medicine,
University of Massachusetts Medical School,
Worcester, Massachusetts

LITTLE, BROWN AND COMPANY
Boston/Toronto

Fourth Printing

Library of Congress Catalog Card No. 85-82480

ISBN 0-316-75776-4

Printed in the United States of America

DON

Case-Control Studies

The sophisticated use and understanding of case-control studies is the most outstanding methodologic development of modern epidemiology. Conceptually there are clear links from experimental studies to nonexperimental follow-up studies to case-control studies, but case-control studies nevertheless differ enough from the scientific paradigm of experimentation that a casual approach to their conduct and interpretation invites misconception and possibly serious error.

RATIONALE FOR THE CASE-CONTROL STUDY DESIGN

Imagine a dynamic steady-state population of exposed and unexposed in-

dividuals. The relevant data on disease incidence for a time period of length t might be summarized as

$$I_1 = \frac{a}{P_1 t}$$

and

$$I_0 = \frac{b}{P_0 t}$$

where I_1 and I_0 are the incidence rates among exposed and unexposed, respectively, a and b are the respective numbers of individuals who developed disease during time interval t, and P_1 and P_0 are the respective population sizes. In a follow-up study the numerator and denominator of each rate are measured; doing so requires enumerating the entire population and keeping it under surveillance. A case-control study is an attempt to render the observations made on the population more efficient. The cases in a case-control study are the individuals who became ill during the time period, that is, a total of (a + b) individuals. The controls are a sample of the combined cohorts that gave rise to the cases. If a proportion, k, of the combined exposed and unexposed cohorts is taken as controls, and the number of such controls is c for exposed and d for unexposed, then the incidence rates among exposed and unexposed could be estimated as

$$I_1 = k\frac{a}{ct}$$

and

$$I_0 = k\frac{b}{dt}$$

If k, the sampling fraction for controls, is known, then estimates of disease incidence are obtainable for both exposed and unexposed groups, just as in a follow-up study. Even if k is unknown, however, which is usually the situation, the relative incidence, or rate ratio (RR, often referred to as *relative risk*), is obtained as

$$RR = \frac{I_1}{I_0} = \frac{ad}{bc} \qquad [6\text{-}1]$$

Since the sampling fraction, k, is identical for both exposed and unexposed, it divides out, as does t. The resulting quantity, ad/bc, is the exposure odds ratio (ratio of exposure odds among cases to exposure odds

among controls), often referred to simply as the *odds ratio*. This cancellation of the sampling fraction for controls in the odds ratio thus provides an unbiased estimate of the incidence rate ratio from case-control data [Sheehe, 1962; Miettinen, 1976]. The central condition for conducting valid case-control studies is that controls be selected independently of exposure status to guarantee that the sampling fraction can be removed from the odds ratio calculation.

The case-control design can be conceptualized as a follow-up design in which the person-time experience of the denominators of the incidence rates is sampled rather than measured outright. The sampling must be independent of exposure; by revealing the relative size of the person-time denominators for the exposed and unexposed incidence rates, the sampling process allows the calculation of the relative magnitude of incidence rates. Viewed in this way, the case-control study design can be considered a more efficient form of the follow-up study, in which the cases are the same as those that would be included in a follow-up study and the controls provide a fast and inexpensive means of inferring the distribution of person-time experience according to exposure in the population that gave rise to the cases.

Whereas follow-up studies are useful for evaluating the range of effects related to a single exposure, case-control studies can only provide information about the one effect that afflicts the cases selected. It is possible, of course, to select multiple series of cases with various diseases, but such an approach amounts to launching several simultaneous case-control studies (the different case series might or might not require separate control series depending on the conditions of ascertainment of the case series). On the other hand, a case-control study can conveniently provide information on a wide range of potentially etiologic exposures that might relate to a specific disease, whereas typically a follow-up study focuses on only one exposure (general population cohort studies are an exception, providing information on a wide range of exposures and diseases).

Some strengths and weaknesses of follow-up studies have symmetric weaknesses and strengths in case-control studies. Already mentioned is the ability of follow-up studies to evaluate a range of effects related to the exposure, whereas a case-control study can evaluate a range of exposures related to the disease. Whereas the evaluation of effects on rare diseases is problematic in follow-up studies, rare diseases are well suited to case-

control studies; on the other hand, case-control studies are inefficient for the evaluation of the effects of exposures that are rare in the source population for the cases. A concern in follow-up studies is the tracing of subjects to avoid loss to follow-up. In case-control studies the analogous concern is determining the correct exposure for all subjects to avoid the loss of subjects with unknown exposure. In follow-up studies, because exposure status is determined before the presence of disease is known either to the subject or to the investigator, there is no possibility of the disease outcome influencing the exposure classification; in case-control studies, if the exposure information comes from the subject after disease onset, knowledge of the disease could affect exposure data. This possibility for bias in case-control studies has no counterpart in follow-up studies. Of course, knowledge of exposure can affect the determination of disease status for subjects in follow-up studies, but this difficulty exists for any type of study. Offsetting the drawback of an additional potential source of bias in case-control studies is the important advantage that follow-up studies are usually large and expensive, whereas case-control studies are smaller and less expensive. The greater efficiency of case-control studies is a strength that may compensate for the greater possibility of bias.

Case-control studies have endured a poor reputation, sometimes serving as the "whipping boy" of epidemiologic research. Those who denigrate case-control research often refer disparagingly to "retrospective" studies without making the distinction made by epidemiologists between retrospective follow-up studies and case-control studies; the bulk of suspicion of so-called retrospective studies, however, is reserved for case-control studies. It is undeniable that case-control studies present more opportunities for bias and mistaken inference than other types of research. One reason is the possibility of recall bias in classifying exposure. The primary reason for error, however, has nothing to do with the validity of the information obtainable from case-control studies; it relates to the relative ease with which a case-control study can be mounted. Because it need not be extremely expensive nor time-consuming to conduct a case-control study, many studies have been conducted by would-be investigators who lack even a rudimentary appreciation for epidemiologic principles. A typical instance might be a study based on a series of patients seen by a single physician or a group of practitioners, with or without a control series. Occasionally such haphazard research can produce fruitful or even extremely important results, but often the results are wrong because basic research principles have been violated. The poor reputation suffered by case-control studies stems more from their inept conduct than from any inherent weakness in the conceptual approach. It is encouraging to observe the steady increase in case-control studies that have been designed, reported, and analyzed with respect for the principles of good study design, thereby minimizing the possibility for bias to distort the findings.