Ann. Rev. Public Health. 1982. 3:29–54

DESIGN AND ANALYSIS OF CASE-CONTROL STUDIES

Norman Breslow[1]

Department of Biostatistics SC-32, University of Washington, Seattle, Washington 98195

INTRODUCTION

A primary goal of epidemiological research is to describe how the exposure of members of a population to certain risk factors influences
quent disease incidence. Such a description is relatively straightforward for cohort studies, where exposed and unexposed members of the study group are followed forward in time so as to permit direct measurement of the relevant disease rates. Cohort studies of rare diseases are extremely costly and time consuming, however, since large groups must be assembled and followed up many years in order to collect enough cases for meaningful statistical analysis. The case-control study, wherein one compares cases and disease-free controls vis-à-vis exposure histories obtained by interview or other retrospective means, has been gaining favor in recent years as a valid and cost-effective alternative method (1).

Case-control studies have contributed to the solution of important public health problems. This is exemplified
ing endometrial cancer were more likely than controls to have received treatment with exogenous estrogens for menopausal symptoms (2, 3). However, the fierce debate that followed publication of these findings reminds us that such studies must be designed carefully in order to provide a true picture of the effect of exposure on incidence. Major issues that arise during the planning stages are: (*a*) the selection of subjects; (*b*) the choice of variables for measurement and analysis; (*c*) the extent to which comparability of cases and controls is to be assured through stratification or matching; and (*d*) the size of the case and control samples.

Recent developments in biostatistical theory and methodology have contributed to the resolution of some of these design issues and have also

[1]Research supported in part by grant KO7 CA00723 from the National Institutes of Health.

0163-7525/82/0510-0029$02.00

greatly extended the available methods of data analysis. Many of the basic ideas appear in the seminal 1959 paper of Mantel & Haenszel (4), which is now one of the most cited works in its field (5). Subsequent dissemination of electronic computers stimulated development of techniques for analysis of categorical and survival data in terms of multiplicative models (6–8). These are currently the methods of choice for the multivariate analysis of case-control data (9). More recently, several teams of statisticians have undertaken a concerted study of the role of matching in the design of case-control studies and of the extent to which it enhances or detracts from their efficiency (10–13).

The present review provides a brief summary of these statistical developments and a guide to the relevant literature. It starts with an account of the logical foundations of the case-control study in order to correct some common misconceptions about the types of risk factor/disease associations that are estimable therefrom. The discussion then considers each of the four design questions mentioned above, and finally turns to modern methods of data analysis.

WHAT CAN BE MEASURED?

For etiologic studies the most relevant measure of disease occurrence is the incidence rate: the number of new cases of disease that arise per person, per unit of time among those who are not yet affected. Denote by $\lambda_1 = \lambda_1(t)$ and $\lambda_0 = \lambda_0(t)$ the (instantaneous) rates of disease incidence among exposed and unexposed persons that occur at t time units into the study period in a particular substratum of the population, defined for example by age and sex. The effect of exposure on disease incidence is typically measured in terms of the difference $d = \lambda_1 - \lambda_0$ or the ratio $r = \lambda_1/\lambda_0$. While the difference has greater relevance as a measure of the impact of exposure on the total public health (14), the ratio is a better indicator of causal associations (15). For one thing, it is sensitive to the specificity of the association with a particular disease entity (16). For another, in order to explain an observed ratio r entirely on the basis of the confounding of exposure with some other causal factor, it is necessary that the other factor itself increase the incidence rates r-fold and also that it be r times more prevalent among the exposed. The ratio is usually more stable than the difference over population subgroups defined by age and other characteristics, which makes it more suitable as a summary measure (9). Of course no measure of association should be viewed in isolation, rather each should be interpreted in light of whatever information is available on the actual magnitude of the rates.

Rates vs Risks

When the disease is rare, or the study period short, the rate ratio will be approximately equal to the ratio of risks, or probabilities, of disease develop-

ment for exposed vs unexposed. More formally, the net disease probability for an exposed person is

$$P_1 = P_1(t) = 1 - \exp\left\{-\int_0^t \lambda_1(u)du\right\},$$

and similarly for the unexposed. Assuming that the ratio r of instantaneous rates remains reasonably constant during the course of the study, it follows that

$$\frac{P_1}{P_0} = \frac{1-(1-P_0)^r}{P_0} \approx r$$

provided both P_1 and P_0 are small (say less than 0.1). In other words, the ratio of risks and ratio of rates are nearly the same.

On the other hand, when the disease is common and the study period long, the rate ratio is substantially more extreme (further from unity) than the risk ratio. Then the question arises as to which is the more appropriate. Two considerations point to the rate ratio as the measure of choice. First, the risk ratio may depend heavily on the study duration, which is an arbitrary design feature rather than a basic property of the disease. For a common infectious disease one can imagine that over an extended period of time almost the entire population will be affected so that the risk ratio tends towards unity regardless of the relative magnitude of the underlying rates. Second, the domain of applicability of the risk ratio as a summary measure is limited by the requirement that the risks themselves cannot exceed one. If exposure is said to double the risk, this statement can only apply to subpopulations where the baseline risk is less than 50%.

In view of these restrictions it is surprising that the risk ratio continues to be utilized as a measure of association, especially in connection with cohort studies, even when the risks are substantially greater than zero. Of course there are situations where the concept of an (instantaneous) rate does not apply, for example when investigating the effect of prenatal exposures on pregnancy outcomes. Then the relevant measure of disease occurrence is in fact the probability of a discrete event like stillbirth. In such cases the logit (log odds ratio) or probit transformations suggest themselves as ways of obtaining a summary measure of the effect of exposure that has a wide domain of applicability (7).

Estimation of Rate Ratios from Case-Control Studies

Case-control investigations, in contrast to cohort studies, do not provide direct measures of disease incidence rates. Rather, one takes a sample of newly diagnosed cases of disease, together with a sample of disease-free controls, and compares the two in terms of prior exposure to the putative risk factor. Such studies are subject to well-known problems of bias arising

from the selection of subjects and from the ascertainment of exposure histories by interview or other retrospective means (17). In addition they have the apparent weakness of not providing information on the relevant question, namely the association between incidence rates.

Cornfield (18) was the first to demonstrate that separate samples of cases and controls did in fact contain information about incidence ratios, and in so doing successfully defended the case-control study against a major criticism. His demonstration was carried out in terms of risks rather than rates, and used the ratio of disease odds as an approximation to the ratio of risks. For the rare diseases that Cornfield had in mind, and for which the case-control study is particularly well suited, the subtle distinction between risks and rates was of little consequence. However, it spawned a myth that this study design was valid only under the so-called "rare disease assumption." Recently it has become clear that case-control studies as typically conducted provide information about the incidence rate ratio, whether or not the disease is rare (19, 20).

The basis for this claim can be understood by reviewing Cornfield's argument in detail. Table 1 introduces notation for the distribution of prior exposure and the disease status at the end of the study period for those who were disease-free at its start. The disease probabilities P_1 and P_0 are directly estimable from a cohort study. With case-control sampling, on the other hand, direct estimates are provided of the exposure probabilities

$$p_1 = pr(\text{exposed} \mid \text{case}) = \frac{pP_1}{pP_1 + qP_0}$$

$$p_0 = pr(\text{exposed} \mid \text{control}) = \frac{pQ_1}{pQ_1 + qQ_0}.$$

However, since the odds ratios of disease and exposure probabilities are identical, the exposure odds ratio as estimated from case-control studies provides a good approximation to the risk or rate ratio when P_0 and P_1 are small:

$$\frac{p_1 q_0}{p_0 q_1} = \frac{P_1 Q_0}{P_0 Q_1} \approx \frac{P_1}{P_0} \approx \frac{\lambda_1}{\lambda_0} = r.$$

Table 1 Distribution of exposure and disease in a particular stratum of the source population

	Exposed	Unexposed	Total
Diseased	pP_1	qP_0	$pP_1 + qP_0$
Disease-free	pQ_1	qQ_0	$pQ_1 + qQ_0$
Total	p	q	1

Stratification on Study Time

The only modification to this argument that is required when the disease is not rare is to divide the calendar period of the study into a number of intervals and to use these in conjunction with age and other variables as a basis for stratification of the sample. In this case P_1 and P_0 represent conditional probabilities of disease development over the relevant time interval among those who remain disease-free at its start. By making the interval sufficiently narrow about an instant of time t, the conditional probabilities become as small as we wish, so that the approximation

$$P_1/P_0 \approx \lambda_1(t)/\lambda_0(t) = r$$

holds to any desired degree of accuracy. The argument implicitly assumes that the controls are selected to be disease-free at the time of diagnosis of the corresponding cases rather than remaining disease-free throughout the entire study period. However, this requirement accords well with the actual conduct of most case-control studies. Thus, a summary estimate of the rate ratio r is obtained as the adjusted estimate of the exposure odds ratio after stratification on study time and other variables.

When the exposure probabilities for cases and controls are strongly correlated with time even as their odds ratios hold roughly constant, the requirement that the time matching of cases and controls be accounted for in the analysis has practical as well as theoretical relevance. For occupational cohort data collected over several decades the controls are chosen to be alive and under observation at the same age and calendar period that the corresponding case is diagnosed (21, 22). The variation in exposure probabilities from one matched set to another may be due to changes in exposure histories among new arrivals in the population, to differential elimination of exposed and unexposed persons through death or loss, or even to the attenuation of those at risk caused by the disease under study (if common). If one ignores the matching and calculates a single odds ratio from the 2 X 2 table of cases and controls pooled over all time intervals, the result is a conservatively biased estimate of the desired rate ratio. By preserving the

matching, a valid estimate results even if the population is not in "equilibrium" vis-à-vis exposure (20, 23, 24). However, as the numerical studies below make clear, the disequilibrium must be marked before the bias caused by ignoring the matching is of serious consequence.

In the sequel we assume that conditions hold so that the exposure odds ratio provides a valid estimate of the rate ratio, whether by virtue of the disease being rare or by having stratified the case-control sample on study time both in design and analysis. We continue to refer to the rate ratio as a relative risk, in accordance with standard practice.

SELECTION OF CASES AND CONTROLS

The preceding demonstration that rate ratios are theoretically estimable from case-control studies involves formal mathematical assumptions that can only be approximated in practice. Nevertheless, they are useful in defining an ideal standard against which the design of actual studies may be assessed. They suggest, for example, that the cases should consist of all new (incident) cases in a defined population, or at least of a random sample therefrom. Likewise the controls should form a random sample from the population "at risk" of developing the disease. Thus, controls for the endometrial cancer study should be limited to women with an intact uterus. Women with a prior hysterectomy may have exposure histories that differ substantially from those of the cases, but the comparison is not relevant to the question of endometrial cancer causation since such women are no longer at risk.

Population-Based Studies

The design that comes closest to meeting the ideal is the "population-based" case-control study. Cases are identified through a disease register covering a defined geographic area, while controls are obtained via well-established methods of area sampling. However, the register may not be complete, some identified cases may die or be too ill for interview, or controls may refuse to cooperate. Each of these possible selection factors needs be considered in terms of the degree to which it could bias the final result. Even with complete sampling, subtle problems of recall and other biases in the assessment of the exposure histories remain (17). For example, women with a recent diagnosis of endometrial cancer may be more aware of their medical histories, which suggests that evaluation of prior disease as a risk factor be based on the actual medical record instead of on patient interview.

Hospital-Based Studies

Hospital-based case-control studies offer obvious advantages in cost, convenience, and comparability of interview setting. They also have potential for

serious selection bias. Except in the case of a prepaid group practice, it is generally not possible to define with any degree of rigor the "population" from which the cases arose. One must assume that the exposure histories for patients with other diagnoses are representative of those "at risk" in the population, for which reason several control groups made up of different disease entities that may have different relations to the exposure are often used. Berkson (25) pointed out long ago the problem in studying the natural association between two or more diseases in a hospital setting, since each additional disease increased a patient's chance of being hospitalized. In spite of these apparent deficiencies, however, hospital-based studies have often provided the first indication of a major public health problem.

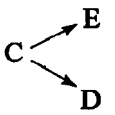
WHEN TO ADJUST

Case-control studies almost always require some sort of statistical control in their design or analysis in order to strengthen the validity of the conclusions that are reached. However, researchers are also aware that adjustment for the "wrong" factors may compromise the ability of the study to detect the association of interest (26). In this section we explore briefly the logical role of adjustment or control procedures and the question of when they should and should not be used.

Confounding

The main impediment to making valid causal inferences from observational studies in epidemiology is the confounding of the exposures of interest with other causal determinants of disease. One cannot logically exclude the possibility that some or all of the observed statistical association results from the fact that exposed persons are at higher risk of disease for reasons other than their exposure. An explanation suggested for the menopausal estrogen/endometrial cancer association, for example, is that women prone to the cancer because of hormonal imbalance would be more likely have a difficult menopause and thus receive symptomatic treatment.

True confounding may be diagrammed



where E represents exposure, D disease, C a confounder, and the arrows denote causal pathways. While it is impossible to completely rule out such effects, observational studies may nevertheless be designed and analyzed in such a way that known risk factors that could potentially confound the

association of interest are explicitly controlled, and so that large confounding effects from unknown or unmeasured variables are rendered implausible. Bradford Hill (27) lists as criteria for inferring causality the strength, specificity, and consistency of the association, the presence of a dose-response gradient, the temporal relationship between exposure and disease outcome and biological plausibility. On this basis, in fact, it is unlikely that the confounding proposed above by a constitutional hormonal factor is correct, since as Cole notes in his introduction to Ref. (9):

> 1. There is a dose-response relationship between oestrogen use and the relative risk of endometrial cancer; 2. the incidence rate of endometrial cancer has risen concurrently with the increase in oestrogen use; 3. the strength of the association between oestrogen use and endometrial cancer is similar in populations which are very dissimilar in their frequency of oestrogen use; and 4. cessation of oestrogen use is followed by a reduction in endometrial cancer incidence.

Cochran (28) gives a good general discussion of the role of matching, stratification, and covariance adjustments in controlling the effects of known confounding variables in observational studies. The basic idea is to compute separate measures of the effect of exposure within subgroups of the sample that are homogeneous for the confounding variables, and to aggregate these into a summary measure of association. Alternatively, the effects of quantitative variables may be controlled by modeling in an appropriate regression equation. Further discussion of the relevant techniques as they apply to case-control studies is given below in the section on methods of analysis.

One usually does not know in practice all the causal determinants of disease or have accurate measurements for them. Nevertheless, some degree of adjustment is still possible based on correlates of the causal factors. The patient's age may indicate the degree of cumulative exposure to a number of nonspecific environmental agents or represent an increase in biological susceptibility, and is almost always considered a potential confounder. Social class is a crude indicator of a variety of shared environments. However, there are clear limitations on the ability of adjustment procedures to completely control confounding when imperfect measures are used.

Selection Bias

Another use for statistical adjustment procedures is to control the selection biases operating on cases or controls. We noted earlier that for the ideal case-control study, sampling of cases and controls was at random from the diseased and "at risk" populations, respectively. In some circumstances this requirement may be approximately met within defined subgroups of the population, although it does not hold overall. Then the same techniques of stratified analysis as used for confounding variables are usefully applied.

Multi-institutional studies are a case in point, especially when the case-control ratio varies from institution to institution.

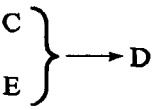
Overadjustment

Any set of case-control data may contain several apparent confounders, in the sense of variables that modify the exposure/disease association of primary interest when controlled for in the analysis. Unless he has a clear understanding of the subject matter, the data analyst may be tempted to select variables for the adjustment process on the basis of purely statistical relationships within the study data, so as to achieve an apparently bias-free estimate of relative risk. This approach can lead to distorted inferences for any of several reasons.

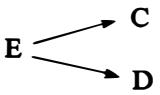
One such circumstance is when the apparent confounder represents an intermediate stage in the causal pathway between exposure and disease:

Endometrial hyperplasia acting as an intermediary between estrogen and endometrial cancer is a possible example. While adjustment for C will reduce the E/D association in such circumstances, this only confirms that the causal effect of E on D is mediated through C.

Another type of overadjustment occurs when both C and E represent correlated but imperfect measures of the same underlying causal factor. In the endometrial example, C and E may be two brands of hormone containing different amounts of the suspect conjugated estrogen. Rather than trying to disentangle their allegedly separate effects, adjusting one for the other, it makes more sense to create a single composite risk variable representing the factor of interest. The situation might be diagrammed:



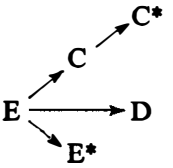
One must even beware of the situation where C is associated with D only through its correlation with an exposure E which is causally related to disease:



Adjustment for C should not affect the mean value of the estimate of the E/D association here, although it may inflate the variance. However, if one

chooses from among several such C variables, incorporating in the analysis those which result in the smallest (or largest) adjusted coefficient of the E/D association, chance fluctuations combine to produce a systematic bias. Day et al (29) have investigated these phenomena numerically.

Finally, if C and E cannot be observed directly, but are instead measured with error in variables C* and E*,



then C* will appear to confound the association with E* by carrying some of the effect that rightfully belongs to E (9, 30). Calendar year may play such a role in our example. Since the use of estrogen has varied rather sharply over the last two decades or so, the calendar year in which a woman has her menopause may carry some information about the likelihood she was treated over and above that contained in her own imperfect recollection of the fact. Both variables would appear to have some relationship to disease, but that involving calendar year would be noncausal.

The controversy surrounding the endometrial cancer studies provides an interesting illustration of a subtle interplay between selection bias and overadjustment. Horwitz & Feinstein (31) suggested that the observed association was the result of the increased likelihood for women treated with estrogens to have episodes of uterine bleeding. This presumably led to their being more closely examined for endometrial cancer than women who had not been so treated, so asymptomatic cases were diagnosed at a higher rate. Consequently, they "adjusted" for the possible selection bias in another study by restricting both cases and controls to women who had received dilation and curetage or hysterectomy because of uterine bleeding, and who therefore had equal chances of having the cancer diagnosed if present. This yielded a markedly reduced measure of estrogen effect. However, since controls having a history of bleeding would include many women with precursor lesions, especially hyperplasia of the endometrium, they are rightly regarded as being overmatched. Perhaps even more important, women with other conditions caused by estrogen exposure would be overrepresented in the control sample, which is therefore itself biased by selection (32).

WHEN TO MATCH

The object of stratification or covariance adjustment in case-control studies is thus to alleviate the distortion in the estimated effect of exposure caused

by confounding or selection bias. If another factor is causally related to disease, and if there is a chance it could be correlated with the exposure of interest, statistical adjustment is needed to produce a valid estimate. Since one rarely has good prior information about the degree of association between various risk factors in the population, a reasonable and prudent policy is to take account in the analysis of all known causal factors regardless of whether they may appear to be related to the exposure of interest in the data at hand. Significance testing of their relationship to either disease or exposure is irrelevant to the issue of whether adjustment for such effects modifies the association of interest (33).

Given that stratification on a potentially confounding variable will be used in analysis, the question then arises as to whether it is wise to account for this in the design by arranging some sort of balance between cases and controls. Extreme imbalance is obviously disadvantageous. If the disease under study occurs mostly at the older ages, for example, and if age is to be considered a potential confounder (as is almost always the case), then the selection of a control sample that contains large numbers of young people is wasteful. Few if any cases will be available in the younger age strata with which to compare such controls, and all the expense of interviewing and data collection will have gone for naught. What is not so immediately clear, however, is that case-control matching in less extreme situations rarely produces marked gains in efficiency and that under certain rather special conditions it may even be less advantageous to match than to take a simple random sample of cases and controls. Moreover, matching on a variable precludes one from evaluating its effects as a risk factor in the analysis.

Expected Frequency Distributions of Cases and Controls

In order to verify these points, and also to provide quantitative expression to some of the properties of bias made earlier, we calculate the mean values and variances of odds ratio estimates in both random and matched samples for the simplest possible situation, that of stratification on two levels of a potential confounder. The approach, as described in several recent articles (11, 12; D. C. Thomas and S. Greenland, unpublished manuscript), is to determine the expected joint distribution of exposure and confounder among N cases and M controls sampled for the study. The asymptotic means and variances of interest, obtained directly from the expected frequencies, are functions of five fundamental parameters describing the population at risk: p_C the proportion positive for the (potential) confounder; p_E the proportion exposed to the risk factor of interest; ψ the odds ratio associating exposure and confounder; R_C the relative risk associated with the confounder; and R_E the relative risk for exposure. The derivation assumes that R_E represents the effects of exposure whether the confounder

is present or absent, or equivalently that R_C determines the confounder's effect regardless of exposure. In other words, the presence of both risk factors is assumed to increase disease incidence $R_C R_E$ times over that when both are absent. If this assumption does not hold, i.e. in the presence of (multiplicative) interactions or effect modification (9, 34), a separate description of the exposure effect within the levels of C is required, and questions of bias and efficiency of the combined odds ratio estimate are moot.

Details on how to calculate the expected frequencies, means, and variances are available in Smith & Day (11) or from the author in an unpublished technical report.

Variance and Relative Efficiencies

Relative efficiency is measured by comparing the variances V_M and V_R of the stratified log odds ratio estimates for matched vs random samples. In order to gauge the loss in efficiency from matching when it is unnecessary to avoid bias, we also calculated the variance V_R^* of the pooled log odds ratio estimate for the random sample. All estimates and variances are based on the stratified data and take no account of any individual pair matching that may have been used. Of course for the situation we describe, such pair matching within levels of the confounder essentially will have been at random and a fully matched analysis therefore will be less efficient than the stratified analysis. However, the loss in using paired data is not great for situations of practical interest unless R_E is quite large (11).

Table 2 presents the main results as a function of the degree of the exposure/confounder association and the two relative risks. The calculations were carried out for fixed values of $p_C = 0.5$, $p_E = 0.3$ and $M = N$, but are at least qualitatively similar to those obtained with other parameter settings. For making inferences about the exposure/disease association under the null hypothesis ($R_E = 1$), the matched sample is more efficient than the random one. Its relative efficiency increases with the degree of confounding, especially as R_C increases. However, even in the most extreme situation shown ($\psi = R_C = 10$), no more than 26% of efficiency is lost by having failed to match at the design stage. Thus, a general conclusion is that matching does not produce major gains in efficiency except when the disease and confounder are very strongly related. This conclusion is reinforced by noting that when the exposure relative risk R_E and the degree of exposure/-confounder association ψ are both large, but R_C is small, matching may even lead to a loss in efficiency.

An explanation for this latter phenomenon is provided by Little & Rosenbaum (13), who have worked out the optimal within stratum case-control sampling ratios under the same no-interaction model used here. Suppose

Table 2 Relative efficiencies (variance ratios) and biases, in percent, for estimates of the relative risk in stratified case-control studies using frequency matched vs random sampling[a]

		$R_E = 1$				$R_E = 2$				$R_E = 5$				$R_E = 10$			
ψ	R_C	$\frac{V_M}{V_R}$	$\frac{V_M}{V_R^*}$	B_R	B_M	$\frac{V_M}{V_R}$	$\frac{V_M}{V_R^*}$	B_R	B_M	$\frac{V_M}{V_R}$	$\frac{V_M}{V_R^*}$	B_R	B_M	$\frac{V_M}{V_R}$	$\frac{V_M}{V_R^*}$	B_R	B_M
1	1	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0
	2	97	100	0	0	97	100	0	0	97	100	0	0	97	100	0	0
	5	88	100	0	0	87	100	0	0	87	100	0	0	88	100	0	0
	10	80	100	0	0	79	100	0	0	80	100	0	0	81	100	0	0
2	1	100	103	0	0	100	103	0	−4	100	103	0	−4	101	104	0	−6
	2	95	100	12	0	96	100	12	−4	97	101	12	−4	99	102	12	−5
	5	85	97	24	0	86	97	24	−1	88	98	24	−2	91	99	24	−3
	10	78	96	31	0	78	96	31	−1	81	97	31	−1	85	98	31	−2
5	1	100	113	0	0	99	114	0	−8	101	118	0	−18	105	122	0	−23
	2	93	106	27	0	93	106	27	−7	97	110	27	−14	103	114	27	−18
	5	82	99	58	0	83	99	58	−4	89	102	58	−8	95	105	58	−10
	10	76	95	75	0	77	96	75	−2	82	98	75	−5	88	100	75	−6
10	1	100	126	0	0	98	126	0	−14	100	134	0	−29	107	144	0	−37
	2	91	114	36	0	90	114	36	−11	96	120	36	−23	104	128	36	−28
	5	80	102	82	0	81	102	82	−6	88	107	82	−13	96	111	82	−16
	10	74	97	106	0	76	98	106	−4	82	101	106	−8	90	105	106	−9

[a] Key: ψ = Odds ratio relating exposure and confounder.
R_C = Relative risk (rate ratio) for the confounder.
R_E = Relative risk (rate ratio) for the exposure.
B_R = Bias of the pooled estimate of R_E from the random sample, expressed as a percentage of the true value.
B_M = Bias of the pooled estimate of R_E from the matched sample, expressed as a percentage of the true value.
V_M = Variance of the stratified estimate in the matched sample.
V_R = Variance of the stratified estimate in the random sample.
V_R^* = Variance of the pooled estimate in the random sample.

N_i cases fall in the i^{th} stratum and let δ_i denote the within stratum differences in exposure probabilities for cases and controls. Then the optimal allocation of controls is to distribute M_1 into the first stratum and M_2 into the second ($M_1 + M_2 = M$) so that the ratios $N_i/(N_i + M_i)$ are approximately proportional to $1/\sqrt{\delta_i}$. For the situations in Table 2 where V_M exceeds V_R^*, the random sample comes closer to meeting this ideal requirement than does the matched sample. Of course, as with most optimal design specifications, it is only possible to use this criterion as a rough guide to practice since the requisite data are not usually available until after the study has been completed.

Columns labeled $V_M \div V_R^*$ in Table 2 show the loss in efficiency imposed by matching when a stratified analysis is not needed to control bias. Since this occurs only when $\psi = 1$ or $R_C = 1$, entries for rows 1–5, 9, and 13 of the Table are of primary interest. They indicate no loss of efficiency when exposure and confounder are uncorrelated, but moderate losses for estima-

tion of large relative risks when the correlation is high. This provides further justification for the conclusion that matching in case-control studies is advantageous only when the confounder is strongly related to disease.

The Bias of Pooled Estimates

Table 2 also illustrates the earlier point that the confounding must be quite strong before the crude relative risk estimates are seriously affected. When $\psi = R_C = 10$, the unadjusted estimate is only about double the correct value, for instance. Note that the percentage bias does not increase with R_E since the effect of confounding on the crude estimate is simply to multiply R_E by a factor

$$\frac{R_C \psi p_C^* + q_C^*}{(R_C p_C^* + q_C^*)(\psi p_C^* + q_C^*)}$$

where $p_C^* = 1 - q_C^*$ is the probability that the confounder is present among the unexposed (35, 36). The conservative bias B_M associated with an unstratified analysis of the matched sample is minimal unless both the relative risk and the exposure/confounder correlation are large and R_C small. While the Table 2 data are for the special case $M = N$, other calculations show that the biases are unchanged as the M/N ratio increases, whereas the variance ratios are generally closer to 100 than those displayed here.

Further Considerations on the Role of Matching

The preceding calculations, carried out in a simple and rather idealized setting, suggest that matching in case-control studies is of limited value in increasing design efficiency except under rather extreme circumstances. If one considers that matching may sometimes reduce the size of the case series through failure to locate an appropriate match, or increase administrative costs, this conclusion holds with even greater force (10). When the confounding factors can be measured or categorized and thus controlled by modeling or stratification in the analysis, little or nothing is gained by individual pair matching over a design that assures rough comparability of cases and controls.

Further work is required to determine whether these results on the value of matching continue to hold for more complicated problems involving both continuous and discrete risk variables or the presence of interactions. Different efficiencies may result if the primary goal is to test for the homogeneity of the relative risk across several categories of a factor with which the exposure may interact.

Pair matching still has an important role to play in the control of confounding factors that cannot be defined or measured precisely. For example, siblings are often used to control approximately for genetic background, or a vaguely defined home environment. Identical twins control exactly for genetic factors whether or not they can be measured. Neighborhood controls are sometimes used when the confounding factors are social class, ambient air and water, or other ill-defined local residential factors. Of course overmatching results if these factors are not causally related to disease or do not even serve as surrogates for causal factors.

HOW MANY SUBJECTS?

A critical question faced by the designers of any scientific investigation is whether the available resources are adequate to answer the questions of primary interest within a reasonable period of time. Sample sizes determined from simple models are useful as a rough guide to feasibility. Since rather arbitrary prescriptions for statistical significance and power enter into their calculation, however, and since accurate prior information is rarely available regarding the joint distribution of risk factors in the population, refinements based on attempts to model more accurately the true situation are rarely worth the extra effort entailed.

Graphical Determination of Sample Size

Chase & Klauber (37) present a useful graph for determining the size of the sample needed to detect a specified relative risk R with prescribed statistical power $1-\beta$ and significance level α. This is reproduced in Figure 1. The derivation assumes that equal numbers of cases and controls are drawn at random from the source population, and that the relative risk is to be estimated for a single binary risk factor (exposed vs unexposed) without consideration of the adjustments for confounding variables that will almost certainly be used in the analysis. Specifically, if the proportion of exposed in the population at risk of disease development is p_0 and exposure increases incidence by a factor R, then the expected proportion of cases exposed is $p_1 = Rp_0/(1 + Rp_0 - p_0)$. The requisite number of cases for testing the equality of p_0 and p_1 is approximately

$$N = \frac{(Z^{\alpha} + Z^{\beta})^2}{2(\arcsin\sqrt{p_0} - \arcsin\sqrt{p_1})^2},$$

where Z^{α} and Z^{β} denote the percentiles of the normal distribution associated with the specified error probabilities α and β (38).

Figure 1 presents the values of N obtained from this formula as a function of p_0 for detecting relative risks of $R = 2$, 3, 4, and 5 using a one-sided significance level of $\alpha = 0.05$ and power $1 - \beta = 0.95$. These may be converted into sample sizes for other specifications of power and significance level by using the multipliers

$$(Z^{\alpha} + Z^{\beta})/(2 \times 1.645)^{2}.$$

The Advantage of Sampling More Controls than Cases

For studies of rare chronic diseases, the number of available cases may be severely limited, whereas potential controls are abundant. Also, the relative costs of locating and interviewing cases and controls may differ sharply. In these and similar situations it makes sense to try to increase the power of the statistical design by varying the case-control ratio, usually by increasing

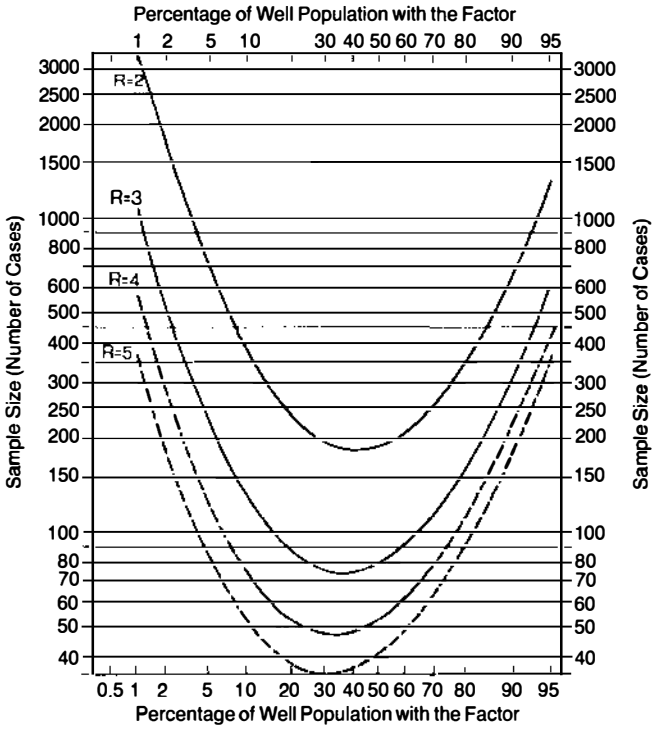


Figure 1 Sample sizes for case-control studies. Graph shows the approximate size of the case series needed to detect a relative risk of $R = 2$, 3, 4, or 5 with 95% power using a 5% one-sided significance level, assuming a control series of equal size is also sampled. Redrawn with permission from Chase & Klauber (37), who used dotted lines to indicate regions for which the approximation was not as accurate.

the number of controls taken per case. Under the random sampling model, the requisite number of cases needed when $S = M/N$ controls are sampled per case is obtained by multiplying the number previously derived for a 1:1 ratio by the factor $(S + 1)/2S$. Thus with two cases per control, only 3/4 the earlier number of cases (but 50% more controls) are needed to achieve the same α and β error probabilities.

These results imply that the efficiency of a design with N cases and a $S = M/N$ ratio is $2S/(S + 1)$ relative to a design with $M = N$, at least for making tests of the null hypothesis. An identical conclusion holds for studies that utilize pair matching of cases and controls both in design and analysis (39). Accordingly, the power of a study with a fixed number of cases can never be increased more than two-fold over the equal sample size design, and studies with as few as four controls per case achieve 80% of the theoretical maximum. Moreover, if it costs C times as much to locate and interview a case as it does a control, the optimum ratio in a study with fixed total cost rather than a fixed number of cases is $M/N = \sqrt{C}$ (40, 41).

A limitation of these calculations, however, is that they refer only to tests of the null hypothesis against "nearby" alternatives. When the relative risk is large, increasing the M/N ratio may in fact increase both the power of the study and the efficiency of estimation substantially more than they would suggest. Figure 2 shows the relative efficiencies (ratios of asymptotic variances) of various M/N ratios for relative risk estimation in matched studies with specified N, derived for the special case when the control exposure probabilities in the matched sets range over the three values 0.1, 0.2, and 0.3 with equal frequency (22). When $R = 1$ the efficiencies relative to $M = N$ rapidly approach their theoretical maximum of 2, in accordance with the previous formula. However, when R exceeds 10, the maximum efficiency increases well beyond that limit and the M/N ratio must be considerably larger than 4 if 80% of the maximum is to be achieved. Practical experience with testing and estimation of large relative risks, as well as with more complicated regression analyses of continuous and interacting risk factors, also suggests that in situations where controls are cheap it may be advantageous to increase the M/N ratio beyond the values of 2, 3, or 4 usually suggested. This is especially relevant for case-control analyses of large cohort data files, since the information on potential controls will already have been collected and the costs are simply those of analysis.

Refinements

The rather simple procedures outlined here should be adequate for the typical study in which rough estimates of sample size are desired as a guide to feasibility. More refined methods have been developed for particular

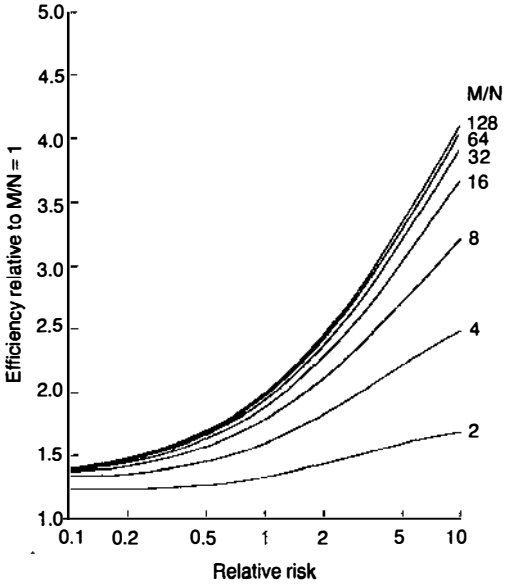


Figure 2 Relative efficiencies of variable ratio matched case-control designs. Graph shows the inverse ratio of asymptotic variance of the relative risk estimates for a study with N cases and $M > N$ controls vs a study with N cases and N controls, as a function of the relative risk R. Calculated according to a formula in Ref. (22) under the assumption that the control exposure probabilities take on the values 0.1, 0.2, and 0.3 with equal frequency in the N matched sets.

problems. For example, conditional test statistics including the Fisher exact test are often used in the final data analysis. Because of their inherent conservatism, when such tests are used the sample sizes from Figure 1 are a little too small to achieve the nominal α and β specifications. Casagrande et al (42) present more accurate approximations for this situation.

No account is taken above of stratification or covariance adjustment procedures used to control for confounding. Approximate power and sample size formulas are available for analyses based either on stratification of the sample into a series of 2 X 2 tables (43) or on a logistic regression model with concomitant variables (44). However, their utilization in practice is severely limited by the requirement that the distribution of the control exposure probabilities across strata, or the joint distribution of the exposure and concomitant variables, be specified in advance. Similarly, sample size formulas for matched pair designs require prior knowledge of the probability that a pair is discordant for exposure (45), which depends in a rather complicated way on the joint distribution of matching and exposure variables. Of course when such information is available from pilot studies or from the early results, it may be used to adjust the initial sample size requirements as derived from the simple model.

METHODS OF ANALYSIS

Before the advent of high speed computers, methods of design and analysis of case-control studies were limited to two basic choices: either matched sets of case and control(s) were constructed on the basis of confounding factors and preserved intact during analysis; or else a random sample of cases and controls was stratified during analysis into a series of 2 X 2 tables and a simple noniterative summary odds ratio was computed as a measure of relative risk. The inflexibility of the available analytic machinery was a particular problem for matched studies. In order to account for matched sets that contained fewer than the target number of controls, or to account in the analysis for a confounding variable not incorporated in the design, the investigator may have felt compelled to discard those sets with too few controls or where case and controls were not also matched for the additional confounding variable. This wasted valuable data. If he utilized all the data but ignored the matching, the result was a conservatively biased estimate of the relative risk. He may have chosen the matched design in the first place because it seemed to represent the only way to control adequately the joint effects of several confounding variables.

Such practices are no longer necessary or defensible now that flexible analytic tools have been developed. Foremost among these is maximum likelihood fitting of the linear logistic regression model or, for matched sets, its conditional likelihood analog. Identical models are used to estimate relative risks from cohort studies, which confirms the essential point that the two types of study share a common conceptual foundation that largely extends to methods of analysis.

After reviewing new developments related to stratified analyses, this section outlines in rather general terms the major contributions made by such models. A more detailed presentation, which includes many illustrative datasets and worked examples, is available in a recent monograph (9).

A New Look at Some Old Odds Ratio Estimates

Two widely used noniterative estimates of the common odds ratio in a series of 2 X 2 tables are the weighted least squares (WLS) or empirical logit estimate (7) and the Mantel-Haenszel (MH) estimate (4). The former is obtained from a weighted average of the observed log odds ratios for each table, usually after each cell has been augmented by 1/2 so as to avoid infinities. The latter is a weighted average of the empirical odds ratios themselves. The WLS estimate is known to be optimal when the sample sizes within each 2 X 2 table are large (46) and compares favorably with the MH estimate in terms of mean square error (47). However, when there are a rather large number of strata with few observations each, it is subject to unacceptable conservative bias and can therefore lead to seriously distorted

inferences (48). The MH estimate, on the other hand, is unbiased in such situations. Moreover, estimates of its variance that are suitable for use with moderately sized samples have recently been proposed to fill a long-standing gap (49, 50).

Two maximum likelihood estimates are available for this problem also: an asymptotic maximum likelihood (AML) estimate, which involves explicit estimation of the nuisance parameters associated with each 2 X 2 table; and a conditional (CML) estimate in which the nuisance parameters are eliminated by consideration of an appropriate conditional likelihood (51). Both require iterative calculation, but the latter is impractical even on a computer unless the numbers of cases and controls within each stratum are reasonably small. When these numbers are large, both share the optimal theoretical properties of the WLS estimate. More important, for moderately sized tables, the AML estimate compares favorably in terms of mean square error with the MH estimate, especially if cell entries are augmented by 1/4 (47).

However, for matched or finely stratified data consisting of a large number of rather small tables, the AML estimate is seriously biased. If the total number of cases plus controls in each table equals a constant T, for example, it tends to inflate the true log relative risk by the factor $T/(T-1)$ (48). Of the two estimates that remain unbiased in this situation, MH has an asymptotic variance almost identical to that of CML if the relative risk is not too large. Hence the robust MH estimate, long a favorite among practicing statisticians and epidemiologists, remains the method of choice in terms of lack of bias, efficiency, and ease of computation over a wide range of data configurations.

Prospective vs Retrospective Logistic Regression

Methods of analysis based on stratification of the sample start to break down when a large number of confounding factors need to be controlled simultaneously. Typically, the range of each continuous variable is partitioned into a small number of levels, rarely over five. With only one variable, this permits a reasonable analysis with adequate control of confounding (28). However, when there are several such variables the strata formed by the cross-classification may number several hundred, many of which will be effectively lost from analysis because they contain only cases or only controls. Moreover, the residual confounding bias that results from the partition of a continuous variable tends to increase with each additional variable (9).

An obvious remedy is to model the effects of the confounding in a regression analysis. Here the logistic model (7) is the method of choice because its parameters may be interpreted as log odds ratios or relative

risks. Let us denote by y a binary indicator of disease status at the end of the specified study period or interval ($y = 1$ for cases and $y = 0$ for controls), by x a variable representing the exposures of interest, and by z the confounding variables. Both x and z may be continuous or vector-valued. Then the model equation for the probability of disease given exposure and confounders is

$$\text{logit } pr(y \mid x,z) = \alpha + \beta x + \gamma z + \delta xz,$$

where

$$\text{logit } (p) = \log \left\{ p/(1-p) \right\}.$$

This may be generalized for a stratified analysis by permitting separate constant terms α_i within each stratum.

An important consequence is that the relative risk of disease for individuals with two different exposure variables x^* and x, but a common set z of confounders, is simply

$$\exp \left\{ \beta(x^* - x) \right\} \text{ or } \exp \left\{ \beta(x^* - x) + \delta(x^* - x)z \right\},$$

depending upon the presence or absence (δ=0) of interactions between exposure and concomitant variables. In their absence, $\exp(\beta)$ represents the fractional change in incidence associated with a unit increase in the corresponding exposure variable, which applies regardless of z. In their presence, the multiplicative effect is modified by z.

Estimates of the parameters in the logistic model are obtained by maximization of the joint likelihood function, obtained as the product of each individual's disease probability. This is clearly appropriate for cohort studies, for then y is a random outcome variable. A somewhat remarkable property of the logistic model is that the same "prospective" likelihood function yields valid statistical inferences regarding the relative risk parameters β and δ of interest even when used with separate samples of cases and controls (52, 53). Thus, widely available computer programs for logistic regression analysis may be applied, treating case-control status as the "dependent" outcome variable.

Alternative formulations of the logistic model for case-control studies have been proposed that seem to correspond more closely to the actual sampling design in that they treat exposure rather than disease status as the

random outcome (54). With a dichotomous variable denoting exposed ($x = 1$) vs unexposed ($x = 0$), the retrospective equation becomes

$$\text{logit } pr(x \mid y,z) = \alpha^* + \beta^* y + \gamma^* z + \delta^* yz.$$

The relative risk parameters β^* and δ^* associating y and x here have precisely the same interpretation as for the prospective model. It is therefore fortunate that the estimates and standard errors for (β,δ) or (β^*,δ^*) will tend to agree when both models are fitted to the same set of data, provided each accounts for the effects of z by inclusion of a sufficiently rich parameterization in γ or γ^* (53, 55). While the choice between the two models is evidently of little consequence for this simple problem, the prospective model is generally to be preferred. First, it has the logical advantage of focusing attention on the correlates of disease and thus reinforces the notion that variables are selected as confounders primarily because of their effect on incidence rates. Second, the prospective model easily generalizes to allow the risk variables x of principal interest to be multivariate or continuous. Extensions of the retrospective model, on the other hand, may require the fitting of exceedingly large numbers of parameters.

Conditional Logistic Regression for Matched Sets

The preceding development noted that the logistic regression model could be adapted to stratified data by utilizing a separate α_i parameter for each stratum. This is a perfectly satisfactory approach so long as the number of strata is small relative to the total number of observations. However, with matched sets or otherwise finely stratified samples, the number of nuisance parameters may increase to the point where it breaks down. The bias of the AML estimate of the odds ratio when there are a large number of 2×2 tables exemplifies the basic problem.

The usual solution to this dilemma is via conditional inference. Suppose that in a particular stratum of N cases and M controls the exposure vectors are $x_1, x_2, \ldots, x_{N+M}$, but it is not specified which of these go with the cases and which with the controls. (Any concomitant confounding variables z are ignored for the moment.) Then the conditional probability that the first N x's go with the cases, as observed, and the remainder with the controls, may be written

$$\frac{\prod_{j=1}^{N} pr(x_j \mid y = 1) \prod_{j=N+1}^{N+M} pr(x_j \mid y = 0)}{\sum_{l \in R(N,M)} \prod_{j=1}^{N} pr(x_{l_j} \mid y = 1) \prod_{j=N+1}^{N+M} pr(x_{l_j} \mid y = 0)}$$

where the denominator sum ranges over the $\binom{N+M}{N}$ possible ways of dividing the numbers from 1 to $N + M$ into one group $[l_1, \ldots, l_N]$ of size N and its complement $[l_{N+1}, \ldots, l_{N+M}]$. Writing

$$pr(x \mid y) = pr(y \mid x)pr(x)/pr(y)$$

and substituting for $pr(y|x)$ the probabilities specified by the logistic model, one arrives at the conditional likelihood contribution

$$\frac{\prod_{j=1}^{N} \exp(\beta x_j)}{\sum_{l \in R(N,M)} \prod_{j=1}^{N} \exp(\beta x_{l_j})}$$

which depends only on the β parameters of interest (20, 21). The full conditional likelihood consists of a product of such terms, one for each matched set or stratum. It has the same form as the likelihood used to analyze data from prospective incidence studies, although for these the "risk-sets" $R(N,M)$ consist of all those at risk of disease development during the corresponding time period rather than just cases plus a control sample (8, 56). This illustrates once again the fundamental point that statistical inference procedures for data collected in cohort or case-control studies from the same population are essentially identical.

In practice, the conditional likelihood approach has been used most often for matched sets consisting of a single case and one or several controls (57, 58). Statistical inferences resulting from its use in this situation generalize many of the analytic procedures previously suggested for matched studies (59, 60). By including selected confounding variables z among those that are modeled in the logistic equation, it is now possible to control in the analysis for variables that were not included in the matching process without discarding vast quantities of data (9). Computer programs are available to perform the requisite calculations, which, through clever use of recursive formulas, permit as many as 20 or 30 cases or controls to be included in each stratum (61, 62).

SUMMARY

Recent developments in biostatistical theory have clarified the role of the case-control study in providing information on incidence rate ratios, have reevaluated the place of matching in case-control design, and have provided flexible new methods for the multivariate analysis of case-control data.

Incidence rate ratios may be estimated directly, without any "rare disease" or "equilibrium" assumptions, provided only that the relevant time aspects are accounted for in design and analysis. Now that alternative methods based on statistical modeling are available for control of confounding, matching appears less attractive than heretofore. It cannot be expected to increase design efficiency unless confounder and disease are strongly related and may reduce efficiency in other situations. The Mantel-Haenszel estimate continues to offer a simple, robust, and efficient method of combining data from a series of 2 X 2 tables, even if these are numerous in relation to the total sample size, and thus is preferred to its more complex competitors. Conditional and unconditional logistic regression models provide a suitable framework for multiple relative risk function estimation in matched or random samples.

Literature Cited

1. Cole, P. 1979. The evolving case-control study. *J. Chron. Dis.* 32:15–27
2. Smith, D. C., Prentice, R. L., Thompson, D. J., Herrmann, W. L. 1975. Association of exogenous estrogens and endometrial cancer. *N. Engl. J. Med.* 293:1164–67
3. Ziel, H. K., Finkle, W. D. 1975. Increased risk of endometrial cancer among users of conjugated estrogens. *N. Eng. J. Med.* 293:1167–70
4. Mantel, N., Haenszel, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* 22:719–48
5. Bailar, J. C., Anthony, G. G. 1977. Most cited papers of the Journal of the National Cancer Institute 1962–75. *J. Natl. Cancer Inst.* 59:709–14
6. Bishop, Y. M. M., Fienberg, S. E., Holland, P. W. 1976. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge: MIT Press. 557 pp.
7. Cox, D. R. 1970. *The Analysis of Binary Data.* London: Methuen. 142 pp.
8. Cox, D. R. 1972. Regression models and life tables (with discussion). *J. R. Stat. Soc. B* 34:187–202
9. Breslow, N. E., Day, N. E. 1980. *Statistical Methods in Cancer Research.* I. *The Analysis of Case-Control Studies.* Lyon: Intl. Agency Res. Cancer. 338 pp.
10. Kupper, L. L., Karon, J. M., Kleinbaum, D. G., Morgenstern, H., Lewis, D. K. 1981. Matching in epidemiologic studies: Validity and efficiency considerations. *Biometrics* 37:271–91
11. Smith, P. G., Day, N. E. 1982. Matching and confounding in the design and analysis of epidemiological case-control studies. *Proc. European Symp. Medical Statistics.* London: Academic. In press
12. Samuels, M. L. 1981. Matching and design efficiency in epidemiological studies. *Biometrika* In press
13. Little, R. J. A., Rosenbaum, P. R. 1981. Confounding and selection bias in case-control studies. *EPA Tech. Rep.* 520/8–81–004. Washington DC: GPO
14. Berkson, J. 1958. Smoking and lung cancer: Some observations on two recent reports. *J. Am. Stat. Assoc.* 53: 28–38
15. Cornfield, J., Haenszel, W. 1960. Some aspects of retrospective studies. *J. Chron. Dis.* 11:523–34
16. Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., Wynder, E. L. 1959. Smoking and lung cancer: Recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.* 22:173–203
17. Sackett, D. L. 1979. Bias in analytic research. *J. Chron. Dis.* 32:51–63
18. Cornfield, J. 1951. A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix. *J. Natl. Cancer Inst.* 11:1269–75
19. Miettinen, O. S. 1976. Estimability and estimation in case-referent studies. *Am. J. Epidemiol.* 103:226–35
20. Prentice, R. L., Breslow, N. E. 1978. Retrospective studies and failure time models. *Biometrika* 65:153–58
21. Liddel, F. D. K., McDonald, J. C., Thomas, D. C. 1977. Methods of cohort analysis: Appraisal by application to as-

bestos mining. *J. R. Stat. Soc. A* 140: 469–91
22. Breslow, N. E., Patton, J. 1979. Case-control analysis of cohort studies. In *Energy and Health,* ed. N. Breslow, A. Whittemore, pp. 226–42. Philadelphia: Soc. Indust. Appl. Math.
23. Halperin, M. 1977. Re: "Estimability and estimation in case-referent studies." *Am. J. Epidemiol.* 105:496–98
24. Miettinen, O. 1977. The author replies. *Am. J. Epidemiol.* 105:498–502
25. Berkson, J. 1946. Limitations of the application of fourfold table analysis to hospital data. *Biometr. Bull.* 2:47–53
26. Miettinen, O. S. 1970. Matching and design efficiency in retrospective studies. *Am. J. Epidemiol.* 91:111–18
27. Hill, A. B. 1971. *Principles of Medical Statistics.* New York: Oxford Univ. Press. 390 pp. 9th ed.
28. Cochran, W. G. 1965. The planning of observational studies of human populations (with discussion). *J. R. Stat. Soc. A* 128:234–65
29. Day, N. E., Byar, D. P., Green, S. B. 1980. Overadjustment in case-control studies. *Am. J. Epidemiol.* 112:696–706
30. Mosteller, F., Tukey, J. W. 1977. *Data Analysis and Regression.* Reading: Addison-Wesley. 588 pp.
31. Horwitz, R. I., Feinstein, A. R. 1978. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *N. Engl. J. Med.* 299:1089–94
32. Hulka, B. S., Brimson, R. C., Greenberg, B. G., Kaufman, D. G., Fowler, W. C. Jr., Hogue, C. J. R., Berger, G. S., Pulliam, C. D. 1981. "Alternative" controls in a case-control study of endometrial cancer and exogenous estrogen. *Am. J. Epidemiol.* 112:376–87
33. Dales, L. G., Ury, H. K. 1978. An improper use of significance testing in studying covariables. *Intl. J. Epidemiol.* 7:373–77
34. Miettinen, O. 1974. Confounding and effect modification. *Am. J. Epidemiol.* 100:350–53
35. Schlesselman, J. J. 1978. Assessing effects of confounding variables. *Am. J. Epidemiol.* 108:3–8
36. Ejigou, A., McHugh, R. 1977. On the factorization of the crude relative risk. *Am. J. Epidemiol.* 106:188–93
37. Chase, M. S., Klauber, M. R. 1965. A graph of sample size for retrospective studies. *Am. J. Publ. Health* 55: 1993–96
38. Feigl, P. 1978. A graphical aid for determining sample size when comparing two independent proportions. *Biometrics* 34:111–22
39. Ury, H. K. 1975. Efficiency of case-control studies with multiple controls per case: Continuous or dichotomous data. *Biometrics* 31:643–49
40. Pike, M. C., Casagrande, J. T. 1979. Re: "Cost considerations and sample size requirements in cohort and case-control studies." *Am. J. Epidemiol.* 110:100–2
41. Walter, S. D. 1980. Matched case-control studies with a variable number of controls per case. *Appl. Stat.* 29:172–79
42. Casagrande, J. T., Pike, M. C., Smith, P. G. 1978. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* 34:483–86
43. Gail, M. 1973. The determination of sample sizes for trials involving several independent 2 X 2 tables. *J. Chron. Dis.* 26:669–73
44. Whittemore, A. 1981. Sample size for logistic regression with small response probability. *J. Am. Stat. Assoc.* 76: 27–32
45. Walter, S. D. 1977. Determination of significant relative risks and optimal sampling procedures in prospective and retrospective comparative studies of various sizes. *Am. J. Epidemiol.* 105: 387–97
46. Gart, J. J. 1962. On the combination of relative risks. *Biometrics* 18:601–10
47. Hauck, W. W., Anderson, S., Leahy, F. J. 1981. Finite-sample properties of some old and some new estimators of a common odds ratio from multiple 2 x 2 tables. *J. Am. Stat. Assoc.* In press
48. Breslow, N. 1981. Odds ratio estimators when the data are sparse. *Biometrika* 68:73–84
49. Hauck, W. W. 1979. The large sample variance of the Mantel-Haenszel estimator of a common odds ratio. *Biometrics* 35:817–20
50. Breslow, N. E., Liang, K. Y. 1982. The variance of the Mantel-Haenszel estimator. *Biometrics.* In press
51. Gart, J. J. 1971. The comparison of proportions: A review of significance tests, confidence intervals and adjustments for stratification. *Int. Stat. Rev.* 39: 148–69
52. Anderson, J. A. 1972. Separate sample logistic discrimination. *Biometrika* 59: 19–35
53. Prentice, R. L., Pyke, R. 1979. Logistic disease incidence models and case-control studies. *Biometrika* 66:403–11
54. Prentice, R. L. 1976. Use of the logistic

model in retrospective studies. *Biometrics* 32:599–606
55. Breslow, N. E., Powers, W. 1978. Are there two logistic regressions for retrospective studies? *Biometrics* 34:100–5
56. Breslow, N. E. 1978. The proportional hazards model: Applications in epidemiology. *Commun. Stat. Theory Methods.* A7:315–32
57. Holford, T. R., White, C., Kelsey, J. L. 1978. Multivariate analysis for matched case-control studies. *Am. J. Epidemiol.* 107:246–56
58. Breslow, N. E., Day, N. E., Halvorsen, K. T., Prentice, R. L., Sabai, C. 1978. Estimation of multiple relative risk functions in matched case-control studies. *Am. J. Epidemiol.* 108:299–307
59. Miettinen, O. S. 1970. Estimation of relative risk from individually matched series. *Biometrics* 26:75–86
60. Pike, M. C., Casagrande, J., Smith, P. G. 1975. Statistical analysis of individually matched case-control studies in epidemiology: Factor under study a discrete variable taking multiple values. *Br. J. Prev. Soc. Med.* 29:196–201
61. Smith, P. G., Pike, M. C., Hill, A. P., Breslow, N. E., Day, N. E. 1981. Multivariate conditional logistic analysis of stratum matched case-control studies. *Appl. Stat.* 30:190–97
62. Gail, M. H., Lubin, J. H., Rubinstein, L. V. 1981. Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika* 68:703–7