

# Identifiability, Exchangeability, and Epidemiological Confounding

SANDER GREENLAND AND JAMES M ROBINS\*\*

Greenland S (Division of Epidemiology, UCLA School of Public Health, Los Angeles, California 90024, USA) and Robins J M. Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology* 1986, 15: 412–418.

Non-identifiability of parameters is a well-recognized problem in classical statistics, and Bayesian statisticians have long recognized the importance of exchangeability assumptions in making statistical inferences. A seemingly unrelated problem in epidemiology is that of confounding: bias in estimation of the effects of an exposure on disease risk, due to inherent differences in risk between exposed and unexposed individuals. Using a simple deterministic model for exposure effects, a logical connection is drawn between the concepts of identifiability, exchangeability, and confounding. This connection allows one to view the problem of confounding as arising from problems of identifiability, and reveals the exchangeability assumptions that are implicit in confounder control methods. It also provides further justification for confounder definitions based on comparability of exposure groups, as opposed to collapsibility-based definitions.

While confounding is widely recognized as one of the central problems in epidemiological research, a review of the literature will reveal little consistency among the definitions of confounding or confounder.<sup>1–10</sup> The definitions appearing over the past decade or so can, however, be roughly divided into two broad categories: the first, which we shall call 'comparability-based', considers confounding as arising from inherent differences in risk between exposed and unexposed populations (where 'inherent' means differences that would exist even if exposure was entirely absent from both populations); the second, which we shall call 'collapsibility-based', considers confounding as arising from differences between certain stratified (conditional) statistical measures of association, and the corresponding crude (unconditional or 'collapsed') measure. Miettinen's<sup>1</sup> and Rothman's<sup>2</sup> definitions fall in the former category, while those of Yanagawa<sup>3</sup> and Boivin and Wacholder<sup>10</sup> fall in the latter. (Some books offer definitions falling in the former category but then employ analysis methods based on the latter, eg, Kleinbaum *et al*<sup>7</sup>).

To our knowledge, no one has presented a theory of epidemiological confounding based on a model for individual effects—a somewhat surprising state of affairs, since a theory of synergy and antagonism has

been derived from such models.<sup>11</sup> Nevertheless, we shall here use a simple deterministic model of effects to derive such a theory. Using this model, it appears that confounding is a manifestation of a problem of non-identifiability of parameters—where non-identifiability is given the deterministic analogue of its statistical meaning (ie, distinct values for the unknown parameters of interest may determine the same data distribution).<sup>12</sup> The usual no-confounding assumptions that render the parameters identifiable may be viewed as exchangeability assumptions—where exchangeability is given the deterministic analogue of its subjective Bayesian<sup>13</sup> meaning (ie, the same data would be expected if the exposure states of the groups had been exchanged). The deterministic development has been chosen to limit the length and technical level of the presentation; nevertheless, analogous results are obtainable under a stochastic model.

## CAUSAL CONFOUNDING UNDER A DETERMINISTIC MODEL

### *Identifiability and Exchangeability in Individual Observations*

Consider a situation in which we wish to study the effect of a binary exposure factor on the risk of a disease over a specified period at risk. There are four possible types of individuals, according to their response to the presence and absence of exposure. Letting one indicate disease occurs and zero indicate disease does not occur over the period, we can tabulate these types as follows:

\* Division of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90024, USA.

\*\* Occupational Health Program, Harvard School of Public Health, Boston MA 02115, USA.

"Common" description of type	Exposed	Unexposed
Type 1. No effect (individual "doomed" <sup>11</sup> )	1	1
Type 2. Exposure causative (individual susceptible)	1	0
Type 3. Exposure preventive (individual susceptible)	0	1
Type 4. No effect (individual immune to disease <sup>11</sup> )	0	0

Suppose we observe a single exposed man over the risk period. If he gets the disease, without further information we cannot tell if he is Type 1 (would have got the disease regardless of exposure) or Type 2 (exposure caused the disease). If he does not get the disease, we cannot tell if he is Type 3 (exposure prevented the disease) or Type 4 (would not have got the disease). In other words, no matter what we observe, we cannot tell if exposure had an effect. This problem is one of *non-identifiability*: different possibilities for the effect predict identical data distributions, and so we cannot identify the effect from the data (here, the data comprise the observed outcome for one man).

Suppose now we find an unexposed man to observe over the risk period of interest. We now have four possible outcomes to our observations: (a) both men fall ill, (b) only the exposed man falls ill, (c) only the unexposed man falls ill, (d) neither man falls ill. For the same reason as before, we cannot tell if exposure had an effect: if we observe the exposed man to fall ill, it may be either because he was doomed or because of an exposure effect; if we do not observe him to fall ill, it may be because he was immune or because of an exposure effect. But, if we combine our observations with the *assumption* that both individuals are of the same type, we can deduce whether exposure had an effect: outcomes (a) and (d) with the assumption imply no effect, outcome (b) with the assumption implies a causative exposure effect occurred, and outcome (c) with the assumption implies a preventive exposure effect occurred (any exposure effect must, of course, occur only in the exposed individual).

Thus, the addition of both an observation and an unexposed individual and an assumption of equivalence of the individuals (in terms of response type) renders the effect identifiable: different possibilities for the effect now predict different data distributions (where the data now comprise the observed outcomes for two individuals).

Equivalence of response type may be thought of in terms of *exchangeability* of individuals: if the exposure states of the two individuals had been exchanged, the same data distribution would have resulted. Thus the

assumption employed to achieve identifiability was one of exchangeability of the individuals.

#### *Identifiability and Exchangeability in Populations*

Suppose now we observed a closed cohort of  $N_1$  initially disease-free exposed individuals observed over a specified risk period, and let  $p_j$ ,  $j=1$  to 4, be the proportion of cohort members of Type  $j$ ; suppose we also observe a closed cohort of  $N_0$  initially disease-free unexposed individuals to compare to the exposed, and let  $q_j$ ,  $j=1$  to 4, be the proportion of unexposed cohort members of Type  $j$ . Our observations may now be summarized in a  $2 \times 2$  table:

	Exposed	Unexposed
Cases	$A_1 = (p_1 + p_2)N_1$	$A_0 = (q_1 + q_3)N_0$
Non-cases	$B_1 = (p_3 + p_4)N_1$	$B_0 = (q_2 + q_4)N_0$
Totals	$N_1$	$N_0$
Incidence Proportions	$IP_1 = A_1/N_1$	$IP_0 = A_0/N_0$
Survival Proportions	$SP_1 = B_1/N_1$	$SP_0 = B_0/N_0$
Incidence Odds	$IO_1 = A_1/B_1$	$IO_0 = A_0/B_0$

There is no way to tell if there was any effect (in the exposed, of course) from the above data alone: neither  $IP_1 < IP_0$ ,  $IP_1 = IP_0$ , nor  $IP_1 > IP_0$  imply anything about  $p_2$  or  $p_3$ . But, if we combine our observations with the 'comparability' *assumption* that the proportion of each cohort that would fall ill if exposure is absent is the same, ie,  $q_1 + q_3 = p_1 + p_3$ , we can deduce that the incidence-proportion ('risk') difference IPD equals  $p_2 - p_3$ :

$$IPD = IP_1 - IP_0 = A_1/N_1 - A_0/N_0 = (p_1 + p_2) - (q_1 + q_3) \\ = (p_1 + p_2) - (p_1 + p_3) = p_2 - p_3,$$

so that  $IPD > 0$  implies there are some individuals in which the exposure caused disease (ie,  $p_2 > 0$ ), and  $IPD < 0$  implies there are some individuals in which the exposure prevented disease (ie,  $p_3 > 0$ ). If, however,  $IPD = 0$  we can only deduce that  $p_2 = p_3$ , *not* that there was no effect. Thus, the addition of both *observation* of an unexposed cohort and an *assumption* (of equality of the incidence proportions of the cohorts when exposure is absent) rendered our effect parameters ( $p_2$  and  $p_3$ ) only partially identifiable. To achieve full identifiability, we would have to employ another assumption: for example, additionally assuming exposure is never causative ( $p_2 = 0$ ) would force  $IPD = 0$  to imply  $p_3 = 0$ , while additionally assuming exposure is never preventive ( $p_3 = 0$ ) would force  $IPD = 0$  to imply  $p_2 = 0$ .

The assumption that  $q_1 + q_3 = p_1 + p_3$  may also be

seen as a partial exchangeability assumption: it says that if the exposure states were exchanged, the value observed for the incidence in the absence of exposure would have been the same. (Complete exchangeability—the same incidence-exposure relation if exposure states were exchanged—would also require that  $q_1 + q_2 = p_1 + p_2$ ).

The preceding discussion may be cast in more familiar epidemiological terms by noting that  $A_1$  is the number of cases observed among the exposed, letting  $E_1 = (p_1 + p_3)N_1$  be the number of cases that would have occurred among the exposed had they not been exposed, and letting  $\hat{E}_1 = (q_1 + q_3)N_1 = IP_0N_1$  be the number of cases expected among the exposed had they the same incidence as the unexposed. Then  $SMR = A_1/E_1$  is the standardized morbidity ratio parameter, ie, the proportionate increase in incidence produced by the exposure (among the exposed), and  $A_1/\hat{E}_1$  is the classical ‘observed/expected’ standardized morbidity ratio estimate. The condition  $q_1 + q_3 = p_1 + p_3$  is then equivalent to  $\hat{E}_1 = E_1$ ; ie, the condition means the unadjusted expected value estimated from the unexposed group is equal to the number of cases that would have occurred in the exposed group had exposure been absent. This is one version of Miettinen and Cook’s criterion for no confounding in the absence of other biases.<sup>5</sup>

A problem arises in employing the no-confounding assumption ( $p_1 + p_3 = q_1 + q_3$ ) in small samples: it may be numerically impossible to satisfy, even approximately. For example, if we observed

	Exposed	Unexposed
Cases	1	1
Non-cases	1	2
Totals	2	3

we would have  $q_1 + q_3 = 1/3$ . But  $p_1 + p_3$  can only equal 1, 1/2, or 0, and so cannot equal  $q_1 + q_3$ . Note that this problem gradually disappears as either the number exposed ( $N_1$ ) or number unexposed ( $N_0$ ) grows larger: numerically,  $q_1 + q_3$  and  $p_1 + p_3$  can approximate each other to within the smaller of  $1/N_1$  and  $1/N_0$ .

*Randomization and Stratification*

What can be done to ensure no confounding, in the sense of  $p_1 + p_3 = q_1 + q_3$ ? The answer is: nothing can guarantee the condition holds, but some actions will make it more likely. In the absence of covariate information, the most effective one is *randomization* to exposure status. If we randomize, we can expect  $p_1 + p_3$  and  $q_1 + q_3$  to differ only in a random fashion, with

smaller differences being more likely than larger ones. In small samples (such as the preceding example) these random differences can easily be large, but as the sample sizes ( $N_1$  and  $N_0$ ) grow these random differences will tend to be smaller, so that when both samples are large, random differences will in probability be small.

If in the absence of randomization (or despite randomization) we believe there is confounding, ie, that  $p_1 + p_3 \neq q_1 + q_3$ , we might ask if we can identify subsets of our total study group ( $N_1 + N_0$ ) within which we believed confounding was nearly absent, ie, within which  $q_1 + q_3$  approximated  $p_1 + p_3$ . If so, we could conduct our analysis based on partitioning our data into such subsets (strata). To illustrate, suppose we have partitioned our data into  $K$  subsets; for subset  $k$  let  $p_{1k}, q_{1k}, p_{2k}$ , etc. represent the proportions of individual types within that subset; let  $N_{1k}, N_{0k}$  represent the number exposed and unexposed within that subset; and let  $IP_{0k} = q_{1k} + q_{3k}$  be the incidence proportion among the unexposed within that subset. If we assume that there is no confounding within subsets, ie, for all strata  $q_{1k} + q_{3k} = p_{1k} + p_{3k}$ , it follows that

$$\begin{aligned}
 p_1 + p_3 &= \sum (p_{1k} + p_{3k})N_{1k}/N_1 \\
 &= \sum (q_{1k} + q_{3k})N_{1k}/N_1 = \sum IP_{0k}N_{1k}/N_1, \\
 \text{or } E_1 &= (p_1 + p_3)N_1 \\
 &= \sum IP_{0k}N_{1k}.
 \end{aligned}$$

The content of the second equation should be familiar: it says that if there is no confounding within strata, the number of cases that would have occurred among the exposed if they had not been exposed,  $E_1$ , may be found by applying the stratum-specific incidences among the unexposed to the stratum-specific numbers of exposed subjects.

It is actually not necessary (or realistic) to require there be no confounding within strata in order to get no confounding overall. It is simply necessary that our stratified estimate of  $E_1$ ,  $\hat{E}_{1s} = \sum IP_{0k}N_{1k}$ , equal  $E_1$ . As the next section shows, this sufficient condition for no overall confounding may well hold even if confounding is complete or severe within most strata.

*Small-Stratum Confounding*

What if some or all of the strata are so small that substantial confounding must by numerical necessity be present in some strata? Consider the following example: a clinical trial using only twins, in which one of each twin is randomly allocated to treatment, and in which some twins are discordant on susceptibility type. Each pair forms its own stratum, so that there is exactly

one treated and one untreated person per stratum. This implies that each of the proportions  $p_{jk}$  and  $q_{jk}$  can take on only the values 0 or 1. Furthermore, within each stratum one and only one  $p_{jk}$  and one and only one  $q_{jk}$  will equal 1. Thus, the quantities  $p_{1k} + p_{3k}$  and  $q_{1k} + q_{3k}$  can equal only 0 or 1, and so to analyse within-stratum confounding we need consider only four possible combinations of values for these quantities:

$p_{1k} + p_{3k} = q_{1k} + q_{3k} = 1$ , in which case there is no confounding;

$p_{1k} + p_{3k} = 1$  and  $q_{1k} + q_{3k} = 0$ , in which case confounding is complete;

$p_{1k} + p_{3k} = 0$  and  $q_{1k} + q_{3k} = 1$ , in which case confounding is complete, and

$p_{1k} + p_{3k} = q_{1k} + q_{3k} = 0$ , in which case there is no confounding.

Suppose there are respectively T, U, V, and W pairs (strata) exhibiting the combinations of values for  $p_{1k} + p_{3k}$  and  $q_{1k} + q_{3k}$  just listed. Having randomized treatment, we can expect U and V to differ only in a random fashion, for U and V differ only because of randomization. Consequently, if the number of pairs is 'large', the difference between U and V will in probability be small. To see this, write  $E_1$  and its stratified estimate  $\hat{E}_{1s}$  in terms of the pair numbers:

$$\begin{aligned} E_1 &= \sum (p_{1k} + p_{3k})N_{1k} \\ &= T(1)I + U(1) + V(0)I + W(0)I \\ &= T + U; \\ \hat{E}_{1s} &= \sum (q_{1k} + q_{3k})N_{1k} \\ &= T(1)I + U(0)I + V(1)I + W(0)I \\ &= T + V. \end{aligned}$$

The difference  $E_1 - \hat{E}_{1s}$  is simply  $U - V$ , and the proportionate error due to confounding from using  $\hat{E}_{1s}$  in place of  $E_1$  in estimating (say) the  $SMR = A_1/E_1$  would be

$$\frac{A_1/\hat{E}_{1s} - A_1/E_1}{A_1/E_1} = \frac{U - V}{T + V}.$$

Thus the amount of confounding will be 'small' if  $U - V$  is 'small' relative to  $T + V$ . Since  $(U - V)/V$  converges to zero in probability as the number of pairs grows (if any U or V pairs exist),  $(U - V)/(T + V)$  must become 'small' as  $T + V$  becomes 'large'.

More precise meanings can be given to 'small' and 'large' above, but the point of the preceding argument is this: in a randomized trial, there is a certain positive

likelihood that a given degree of confounding remains in the adjusted (stratified summary) result because of confounding within strata; nevertheless, this likelihood becomes small as either the number of informative strata becomes large, or as one or more informative strata become large ('informative strata' here means strata within which the proportion of Type 2 or 3 individuals in the sampled population is non-zero). More simply put, 'random' confounding within strata will tend to cancel across strata, and will diminish in overall importance as the sample size grows, even if it remains large within strata.

In the absence of randomization, we have no probabilistic guarantee that within-stratum confounding will tend to disappear as the strata grow large, or cancel across strata as the number of strata grows large. Nevertheless, we must simply assume such cancellation occurs in order to proceed with the analysis. More precisely, we must assume that as the sample size grows the stratified version of the condition for no overall confounding, ie, that  $\hat{E}_{1s} = E_1$ , will tend to hold. Again, this may be viewed as a partial exchangeability assumption, to the effect that the unexposed incidence standardized to the first group ( $\hat{E}_{1s}/N_1$ ) would remain approximately unchanged if the exposure states of the two groups had been exchanged.

#### *Extension to Stochastic Models*

One can extend the above theory to situations involving stochastic elements, introduced either via random sampling, or stochastic individual mechanisms for disease occurrence. The chief change is that the fixed population quantities given above ( $A_1$ ,  $IP_1$ ,  $E_1$ , etc.) become expected values. Additional statistical considerations arise in construction of estimates of  $E_1$ .<sup>9</sup>

#### CONFOUNDERS

The last section provided a precise definition of confounding and a motivation for stratification in constructing unconfounded estimates of effect. Note that no mention of 'confounder' or 'covariate' was necessary. Thus we maintain that the notion of confounding in causal analysis is more fundamental than any notion of confounder or covariate control, despite a tendency of some writers to assume that confounding is solely the product of 'covariate imbalances'. In fact, the only 'covariate' that could serve as a foundation for the notion of confounding is the 'ultimate covariate', ie, the variable which takes on the value one if an individual is Type 1 or 3 and zero if an individual is Type 2 or 4. This covariate perfectly predicts the individual's outcome status if exposure is absent; if it could be measured, there would be no need for a

comparison group (since for every exposed individual one could tell whether the exposure had an effect merely by comparing the individual's actual outcome to the individual's value of this covariate). Since one cannot in general measure the 'ultimate covariate', one must make do with less perfect predictors of outcome in the absence of exposure.

One heuristic epidemiological definition defines a confounder as a variable that, when controlled, yields an estimate of  $E_1$  (and hence the SMR parameter) closer to the true value (ie, less biased) than when it is not controlled.<sup>5</sup> Implicit in this definition is the fact (not always made explicit) that a variable is a confounder only relative to everything else that is under control.<sup>1</sup> For example, upon control of various medical-care variables, socioeconomic status might cease to be a confounder in a perinatal-mortality study. It is perhaps less widely realized that a variable may become a confounder upon control of certain other variables (the example given below illustrates this if one takes the 'ultimate covariate' as the variable in question).

When starting from the preceding definition, many

ing the parameter of interest as a measure of the effect of exposure.

Occasionally, one sees writings in which the above properties are taken as defining a confounder (as opposed to merely being derived properties).<sup>6</sup> This is a mistake, because the above properties are not sufficient to guarantee a variable is a confounder; in other words, there can be variables that satisfy all three properties but for which control is not helpful.<sup>14,15</sup>

There is also a logical conflict between the definition based on whether control 'helps' (ie, reduces the net bias in estimating  $E_1$ ) and the definition based on whether the above three properties are satisfied: it may be that a variable satisfies the above properties, but that the residual confounding left upon its control is greater than the confounding present when it is not controlled (indeed, in the absence of randomization such a situation would not necessarily be unusual).

*Example.* Suppose (unknown to us) exposure has no effect, ie, there are no Type 2 or 3 individuals, and that we measure a risk factor having the following joint distribution with exposure and type:

Type	Factor present		Factor absent		Crude	
	Exposed	Unexposed	Exposed	Unexposed	Exposed	Unexposed
1 ('doomed')	60	70	40	180	100	250
4 ('immune')	40	30	60	220	100	250
Total	100	100	100	400	200	500
Incidence	0.60	0.70	0.40	0.45	0.50	0.50

authors deduce the following as necessary properties of a confounder (assuming there is no selection bias or misclassification) (cf ref. 5):

- (1) It must be predictive of risk among the unexposed (in the present context, this means it must be predictive of being 'Type 1 or 3').
- (2) It must be associated with exposure in the population under study (ie, the cohort in a cohort study, and the source population of cases and controls in a case-control study).

Some authors add:

- (3) It must not be an intermediate in the causal sequence from exposure to outcome, or a consequence of the outcome.

The first two properties are implicitly taken as conditional on other controlled factors, and follow immediately from algebraic arguments based on collapsibility of the SMR parameter.<sup>9</sup> The third property, however, is a logical consequence of identify-

Note that only the last two lines of the table would be observable in a real study. From them we can see that the factor is associated with exposure (100 of 200 exposed have the factor, as opposed to 100 of 500 unexposed), and is predictive of outcome among the unexposed (among the unexposed, 0.45 of those without the factor will get the disease, as opposed to 0.70 of those with the factor). Nevertheless, if we fail to control the factor, we will observe disease in an equal proportion (1/2) of the exposed and unexposed, correctly indicating no exposure effect, while if we control the factor, exposure will incorrectly appear to be preventive of disease in both strata. More precisely, we may note that  $p_3 = 0$  and thus  $p_1 + p_3 = p_1 = 0.50 = q_1 = q_1 + q_3$ , so that the crude estimate is unconfounded, while  $\hat{E}_{1,3} = 0.70(100) + 0.45(100) = 115 \neq 0.50(200) = E_1$ , so that the adjusted estimate is confounded. One might explain this phenomenon by saying that the 'downward' bias left after control of the factor is cancelled by the 'upward' bias produced by not controlling the factor, leaving an unbiased crude

estimate of effect; Robins, however, gives an example in which risk factor control increases bias but this explanation fails.<sup>16</sup>

In practice, one may have no idea whether the bias produced by failing to control some factor (or factors) is added to or opposed by any other bias. In most situations one is much more likely to be certain that the derived properties 1–3 above are satisfied for a given factor than to be certain that the control of the factor actually reduces the net bias in one's estimate. Thus although properties 1–3 are not sufficient to define a confounder, they remain practical criteria for screening out non-confounders (since any confounder must meet all three criteria).

One point deserves special emphasis, if only because it was often overlooked in early epidemiological literature: the property of being a confounder is not directly verifiable from data.<sup>5</sup> As the preceding example illustrates, what we observe (incidence and factor distributions) is not sufficient to judge with certainty whether a factor is a confounder, or how much confounding it is responsible for. Thus, the decision of whether to control a factor is subject to error (even if the factor is known with certainty to satisfy properties 1–3 given earlier).

#### COLLAPSIBILITY

Certain authors<sup>3,10</sup> employ a definition of confounder that is based on the statistical notion of 'collapsibility'.<sup>14,15</sup> In this approach, no reference is made to individual 'effects', and notions of 'cause' enter only in the *a priori* identification of which factors should be considered potential confounders. (Usually, in the absence of misclassification or selection bias, a variable possessing properties (1) and (3) given earlier, ie, a non-intermediate risk predictor among the unexposed, will be considered a potential confounder.) One then considers whether the crude estimator of effect is a 'biased' (or statistically inconsistent) estimator for a stratified parameter in a statistical model. For example, a model commonly used in case-control studies asserts that upon stratification of the underlying source population on all potential confounders, each stratum-specific exposure-disease association in the population would be manifested by an odds ratio (cross-product ratio) that is common to all strata (ie, homogeneous across strata). If this common population odds ratio equals the crude population odds ratio, the odds ratio is said to be collapsible,<sup>14</sup> and the crude odds ratio is said to be unconfounded.<sup>3,10,11</sup>

As noted by Miettinen and Cook,<sup>5</sup> the identification of confounding with non-collapsibility (which they reject) makes the presence or absence of confounding

depend on which parameter is chosen to measure exposure effect. For example, in a cohort study one may find the risk difference to be collapsible but the odds ratio not, so that whether one judged confounding present would depend on which measure one chose.

If, however, one's objective is to compare the exposed population's actual incidence with its 'null' incidence (the incidence it would have experienced had exposure been absent), confounding is present (absent) if the unexposed incidence does not equal (at least in expectation) this 'null' incidence. This is the 'comparability-based' definition. Using this, one is led, as Miettinen and Cook were, to reject the odds ratio as a measure of intrinsic interest and instead employ the risk ratio and risk difference, for unlike the odds ratio both the incidence ratio and difference will be collapsible whenever confounding (in the 'comparability-based' sense) is absent.<sup>5</sup>

Even if one defines a confounder as a 'non-collapsible potential confounder', one should note that this property is not directly verifiable from the data: first, the collapsibility referred to is in the statistical source population,<sup>5,14,15</sup> which the observed data only reflect with error; second, the defining properties of a potential confounder (properties (1) and (3)) are not wholly verifiable from the data.

#### CONCLUSION

We have attempted to trace back to the most simple level (observation of deterministic outcomes in individuals) the origin of the epidemiological notion of confounding. In so doing, we have concluded that the problem of confounding arises from our inability to identify (or estimate) from data along the fundamental causal parameters that determine our observations. The 'no-confounding' assumptions that render these parameters at least partially identifiable may be recognized as assumptions about exchangeability of individuals or groups. This view leads to a definition of confounding that is equivalent to the 'comparability-based' definitions given by certain epidemiologists, and thus conflicts with 'collapsibility-based' definitions. In particular, it reinforces the conclusions of Miettinen and Cook that presence or absence of confounding should not be equated with absence or presence of collapsibility, and that confounding should not be regarded as a parameter-dependent phenomenon. It also reinforces the notion that the degree of confounding of results is an unknown quantity, not directly measurable from observed data.

#### ACKNOWLEDGEMENTS

The authors would like to thank Hal Morgenstern

and Ken Rothman for their comments, and Virginia Hansen for preparation of the manuscript.

## REFERENCES

- <sup>1</sup> Miettinen O S. Stratification by a multivariate confounder score. *Am J Epidemiol* 1976; **104**: 609–20.
- <sup>2</sup> Rothman K J. Epidemiological methods in clinical trials. *Cancer* 1977; **39**: 1771–5.
- <sup>3</sup> Yanagawa T. Designing case-control studies. *Env Health Perspect* 1979; **32**: 143–56.
- <sup>4</sup> Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol* 1980; **9**: 361–7.
- <sup>5</sup> Miettinen O S, Cook E F. Confounding: essence and detection. *Am J Epidemiol* 1981; **114**: 593–603.
- <sup>6</sup> Schlesselman J J. *Case-Control Studies: Design, Conduct, Analysis*. New York, Oxford University Press, 1982.
- <sup>7</sup> Kleinbaum D G, Kupper L L, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. Belmont, CA, Lifetime Learning Publications, 1982.
- <sup>8</sup> Last J M. *A Dictionary of Epidemiology*. New York, Oxford University Press, 1983.
- <sup>9</sup> Robins J M, Morgenstern H. Confounding and prior knowledge. Technical report No. 1. Boston, Occupational Medicine Program, Harvard School of Public Health, 1983.
- <sup>10</sup> Boivin J-F, Wacholder S. Conditions for confounding of the risk ratio and of the odds ratio. *Am J Epidemiol* 1985; **121**: 152–8.
- <sup>11</sup> Miettinen O S. Causal and preventive interdependence. Elementary principles. *Scand J Work Environ Health* 1982; **8**: 159–68.
- <sup>12</sup> Graybill F A. Theory and application of the linear model. North Scituate, Duxbury, 1976.
- <sup>13</sup> de Finetti B. *Probability, Induction, and Statistics*. New York, Wiley, 1972.
- <sup>14</sup> Whittemore A S. Collapsibility of multidimensional contingency tables. *J Roy Statist Soc B* 1978; **40**: 328–40.
- <sup>15</sup> Shapiro S H. Collapsing contingency tables—a geometric approach. *Am Statist* 1982; **36**: 43–6.
- <sup>16</sup> Robins J M. A new approach to causal inference in mortality studies with a sustained exposure period. Technical report No. 3. Boston, Occupational Medicine Program, Harvard School of Public Health, 1985.

(Revised version received January 1986)