Scand J Work Environ Health 1997;23(1):64-68

<https://doi.org/10.5271/sjweh.181>

Issue date: Feb 1997

On the epidemiologic notion of confounding and confounder identification

by [Nurminen M](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/9098915

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

On the epidemiologic notion of confounding and confounder identification

by Markku Nurminen, DrPH, PhD¹

Nurminen M. On the epidemiologic notion of confounding and confounder identification. *Scand J Work Environ Health* 1997;23(1):64–71.

In a recent commentary in this journal (1), Hernberg expressed an editor's stance against the use of statistical significance testing in deciding whether confounding is present in an epidemiologic study of an exposure-effect relationship or not. He concluded that the selection of confounders to be controlled in the analysis of the data should instead be made according to the following rule: "Formally one can compare the crude relative risk and the relative risk resulting after adjustment for the potential confounder. A difference indicates confounding, and in that case one should use the adjusted risk estimate. If there is no or a negligible difference, confounding is not an issue and the crude estimate is to be preferred. Personal judgment comes into play when what is 'negligible' is decided [p 316]". Both Hernberg's treatment of confounding and the prescription chosen to identify the problem invite an extended commentary for two main reasons. First, in causal inference the notion of confounding is more fundamental than any criterion of confounder control, despite the tacit assumption that the defining property of confounding is a difference between the crude and adjusted risk estimates resulting from the dissimilarities in the distribution of the covariate(s) between the exposed and unexposed groups. Second, the presence or absence of confounding depends on the parameter used to measure the effect of exposure on disease outcome. Because of their general interest, I would like to address both of these methodologic issues in some detail.

The distinction between the concepts of *confounding* and *confounder* is essential, but often ignored or obscured in epidemiologic literature. The definitions of confounding or confounder fall into two main classes. According to the first, based on the formal concept of *collapsibility*, confounding is a failure of the estimate for an adjusted (eg, stratified) effect parameter to equal the estimate for an unadjusted (crude) parameter obtained when a covariate is ignored as stratification is collapsed.

Thus confounding is equated with noncollapsibility (or change-in-parameter estimate). The second, *comparability*-based definition views the effect of exposure as a comparison of the group's outcome when exposed with what the *same* group's outcome would have been if exposure had been absent. Consequently, the observed value of the outcome measure in the exposed group is compared with the expected value of the outcome measure that would have been observed in the exposed group if it had, hypothetically, not been exposed. The unexposed group's actual outcome is used as a proxy for the exposed group's unobserved value. When the two groups' experiences differ, the groups are noncomparable, and confounding is said to result. Whittemore (2), Yanagawa (3), Boivin & Wacholder (4), and Becher (5) have defined confounding using the statistical concept of collapsibility, while Miettinen & Cook (6), Rothman (7), Greenland & Robins (8) and Checkoway et al (9) have adopted the epidemiologic comparability definition. Some books present confounding in accordance with the comparability definition but then espouse the change-in-parameter-estimate criterion for deciding whether confounding is a problem or not in the estimation of the strength of the exposure effect, for example, Kleinbaum et al (10) and Hernberg (11, see the preceding quotation from Hernberg's commentary).

As explicated by Greenland et al (12), the collapsibility approach requires that all the covariates assumed to be important predictors of the disease outcome be listed, and it specifies what is meant by their "control", before the problem of confounding can be defined. However, because residual confounding is always possible as a result of imperfect measurement of confounders with the use of imprecise covariates (see, eg, reference 13) or as a result of an unlimited number of undiscovered risk factors, a collapsibility-based definition cannot unambiguously estimate an unconfounded value for the causal

¹ Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FIN-002500 Helsinki, Finland.

Reprint requests to: Dr M Nurminen, Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FIN-002500 Helsinki, Finland.

parameter. Moreover, the conditions for nonconfounding depend on the true (unknown) parametric form of the model that relates the disease outcome to exposure and is conditional on extraneous covariates. A relevant statistical point is that the collapsibility definition implicitly assumes the constancy (homogeneity) of effect measures across strata. This assumption, which is made by procedures such as the Mantel-Haenszel estimates of risk ratios and the logistic regression estimates of odds ratios, is frequently not tenable in epidemiologic applications and can therefore be regarded as a weakness of the definition (14). Another cogent objection is that stratum-specific measures can themselves be confounded by further unknown risk factors (12) or by random imbalance caused by small sizes of the strata (8).

In contrast, the comparability-based definition of confounding does not suffer from such drawbacks because the definition makes no reference to confounders or their control and because, according to this view, nonconfounding is invariant across models underlying the effect measures. The problem of confounding stems from the *nonidentifiability* of the essential causal parameters of the assumed statistical model that determined the observations. To put it another way, we are not able to estimate the effect because the observed outcomes can be predicted by different possibilities (parameter values) for exposure impact. But these parameters can be made identifiable if we assume the exposed and unexposed groups are comparable (6) or exchangeable (8). In other words, the causal contrasts would be unbiasedly estimable even if the exposure states of the two groups had been exchanged. For the derivation of the theory of epidemiologic confounding based on the assumption of *exchangeability*, see reference 8. With this approach, we can deduce that confounders must fulfill well-known necessary conditions, namely, that confounders (i) must be predictive of disease risk (even among the unexposed), (ii) must be associated with exposure in the population under study, and (iii) must not be a link in the causal chain from exposure to outcome or a consequence of the outcome. (Compare with reference 6.) Sometimes these properties are taken as defining a confounder and not just being derived properties (15, 16). However, these three conditions cannot be considered to identify a confounder because they are not sufficient to make a covariate a confounder (8). Nevertheless, they can be used in practice as operational criteria for screening out nonconfounders, since any confounder must meet all three criteria.

The comparability definition led Miettinen & Cook (6) inductively to conclude that the presence or absence of confounding should not be equated with the presence or absence of collapsibility and that the decision regarding the existence of confounding should not depend on the chosen parameter. Moreover, the degree of confound-

ing — and hence the strength of an effect — cannot be directly measured from the observed data; it should always be evaluated against background disease risk, knowledge of subject matter, logical argument, evidence from previous studies, and the particulars of the empirical setting in which the study is being conducted. The dilemma posed by the conceptual identification of every important confounder is that, in addition to the factors brought forward by the study, one that, together with the included factors, constitutes a confounder may be omitted. There is a further problem in that, even if all the plausible confounders have suitably similar distributions between the compared groups, their joint effect may still cause bias. Criteria based singularly on associations within the data can also be misleading because the predictiveness of the potential confounders of the outcome can be a result of effect modification rather than a result of confounding (17). Therefore, the decision of whether a covariate requires control is both judgmental and subject to error.

However, the relative roles of prior information, as against data from the study population, involve subtleties that reach beyond the previous inductive examination of the problem (6). For example, given prior information that a covariate is a risk factor in a cohort study, inferences should be made conditional on appropriate measures (ancillary statistics) of discrepancies between this prior knowledge and the observed data (18). Conversely, if a covariate in a case-referent study is known a priori to be unassociated with either the exposure or the disease in the source population, then the crude estimate does not require adjustment irrespective of associations observed in the data (19). Statisticians have previously failed to recognize that epidemiologists have implicitly conditioned their causal analyses of exposure-effect relations with such ancillary statistics. This failure has probably prevented a coherent statistical definition of the epidemiologic concept of confounding (20).

Which definition of confounding should we adopt and which operational criteria for its detection should we select amid this conceptual discrepancy? The confusion can be explained by the use of statistical criteria for confounder control that are not coherent with the intuitive notions of confounding. Wichramaratne & Holford (21) offer a resolution by suggesting that (i) noncollapsibility of the chosen measure of effect and (ii) noncomparability of the exposed and unexposed groups are distinct phenomena. Exacerbating the problem is the fact that the criteria for a covariate to be a confounder according to the two definitions are not mathematically equivalent. However, Greenland et al (12) made the following important point: "If one elects to use a measure of effect for which comparability and collapsibility do *not* coincide, such as the incidence-odds ratio, then defining one of the two phenomena as

'confounding' will *not* automatically free one from the need to consider the other phenomenon when making inferences [p 1088]."

All these points raise the question of which effect measure is "coherent" or proper. Hernberg (1) brought up the change in the value of the relative risk parameter as the measure of confounding bias with which to decide whether a covariate needs to be controlled or not. The generic term "relative risk" is used synonymously with risk ratio, rate ratio, and odds ratio (22). The change-in-parameter-estimate criterion is problematic for two reasons. First, the criterion can be systematically misleading. It can lead to unnecessary and even bias-producing control of the covariate when misclassification of exposure or disease is present (23). Second, it depends on the chosen measure of effect. Miettinen & Cook (6) showed how comparability did not insure odds ratio collapsibility. They rejected the odds ratio parameter for its mathematical peculiarity — that its unconfounded crude value can fall outside the range of its stratum-specific values.

As an illustration of the latter problem, consider the cohort data in table 1, stratified by "balanced" covariate C that has equal distribution among the exposed and unexposed subjects. This procedure implies no confounding according to the comparability criterion. In addition there is no confounding of the crude comparisons with any of the three scales of risk measurement: the crude incidence difference, incidence ratio, and incidence-odds ratio all show exactly how much the exposure raised the crude incidence proportion. [Note that in epidemiologic terms "risk" pertains to an individual and "incidence proportions" are effect measures in populations, which are interpretable as "average risks" (14)]. Instead, the control of C changes the value of the incidence-odds ratio, which means negative confounding under the collapsibility definition. (Compare reference 4.) In this example the stratum-specific incidence-odds ratios are upward-biased, yet the crude incidence-odds ratio unbiasedly contrasts the odds of disease in the exposed and unexposed cohorts because the proportion of subjects with C is 50%. The value of the incidence-proportion

difference remains unchanged upon stratification. The situation is not quite so obvious for the incidence-proportion ratio, but the stratified estimate can be computed (24), and it equals the value of the pooled estimate (namely 2.3). This apparent anomaly has caused proponents of the collapsibility criterion for confounding to reject the incidence-odds ratio as a measure of exposure effect in favor of incidence-proportion difference and ratio measures (6, 14). A correct measure with which to assess the existence of confounding in the spirit of comparability requires the observed number of exposed cases (O) to be contrasted to the corresponding expected number (E), calculated on the assumption that the exposed subjects would incur the unexposed subjects' risk (6). In effect, the "Unexposed" columns of table 1 could be relabeled "Exposed cohort if not exposed". This application can be illustrated for the incidence-odds ratio by the data in table 1; we find from the "Pooled" column that O = 140 and from the "Stratum 1" and "Stratum 2" columns $E = (10/100)100 + (50/100)100 = 60$. Thus the adjusted estimate of the "observed/expected" incidence-odds ratio is $[O/(S - O)]/[E/(S - E)] = [140/(200 - 140)]/[60/(200 - 60)] = 5.4$, which is equal to the crude incidence-odds ratio estimate indicating no confounding. For a more-detailed criticism of the odds ratio as a measure of effect, see references 14 and 25. Greenland et al (12) provided an example of the reverse situation, in which the odds ratio was collapsible but there was extreme noncomparability (and, in their view, confounding) of causal comparison among the groups. We can also note that the equality of stratum-specific odds ratios or risk ratios is not a sufficient guarantee for the pooled estimator to be unconfounded or consistent (26, 27).

Robins & Morgenstern (28), and recently Greenland (29), have shown that the absence of confounding does not imply collapsibility of the person-time rate ratio (relative hazard); that is, the crude rate ratio need not equal the common stratum-specific rate ratio even if the exposed and unexposed cohorts (or stationary populations) have the same distribution of all risk factors. An analogous result holds for rate differences. Greenland (29)

Table 1. Hypothetical data stratified by a covariate (C) that has the same distribution in exposed and unexposed cohorts, illustrating the inadequacy of odds-ratio collapsibility as a criterion for nonconfounding.

	Stratum 1: C present		Stratum 2: C absent		Pooled (ignoring C)	
	Exposed	Unexposed	Exposed	Unexposed	Exposed	Unexposed
Diseased	50	10	90	50	140	60
Nondiseased	50	90	10	50	60	140
Size of subcohort (S)	100	100	100	100	200	200
Proportion with C					50%	50%
Incidence proportion	0.5	0.1	0.9	0.5	0.7	0.3
Incidence-proportion difference		0.4		0.4		0.4
Incidence-proportion ratio ^a		5.0		1.8		2.3
Incidence-odds ratio		9.0		9.0		5.4

^a The stratified estimate for the incidence-proportion ratio is 2.3.

explained that the reason for the "discrepancy between nonconfounding and collapsibility in rate comparisons arises when person-time is a post-exposure variable whose distribution can be altered by the effects of exposure and other risk factors [p 498]." However, no conflict would occur if incidence-proportion differences and ratios were used as the population measures of exposure effect on average risk and incidence-density rate comparisons were used only in situations in which they approximate survival-time (average hazard) comparisons.

Regarding Hernberg's (1) chief point that "... confounding is bound to time and place and cannot be generalized. Therefore statistical significance testing is not the way to scrutinize potential confounding [p 315]", not all statistician-epidemiologists see the matter as categorically, even though the procedure has been strongly criticized for a long time. (See, for example, reference 30.) It is clear that the problems in selecting potential confounding factors are primarily substantive and not statistical, although there are some statistical guidelines that may prove useful. In the absence of prior knowledge about confounding in the exposure-effect relationship, investigators frequently rely on the data at hand to guide them in their decision of whether to adjust for a covariate or not. The data-based criteria for confounders always pertains to the particular population sample experience yielding the exposure-effect relation of the study. If the use of significance testing of potential confounders is rejected solely because it is a sample-to-population inference criterion (in frequentist statistics), then one is logically compelled to reject *all* similar criteria that can be used for such a purpose. There are several other statistical testing strategies besides the significance-test-of-the-covariate criterion, including the equivalence-test-of-the-difference and collapsibility test criteria (31). The equivalence test is preferable to the significance test on logical grounds, because it tests whether the amount of confounding present is worth worrying about, rather than testing a certainly false null hypothesis of nonconfounding (32). The impact of various confounder selection criteria on effect estimation was investigated in simulation studies for both case-referent data (33) and cohort data (34). The change-in-parameter-estimate criterion (with a specified percentage of difference between the crude and adjusted estimate) tended to perform best for both case-referent and cohort studies, even though significance testing methods can perform acceptably if their significance levels are set much higher than conventional levels (to values of 0.20 or more). Thus using statistical significance tests of potential confounders can be defended in some cases, but, as recommended by Robins & Greenland (35), if doing a frequentist test for confounding is insisted upon, such testing (i) should only be made in a *preliminary* style for deciding which of the strongest confounders to include in a multivariate adjustment and

(ii) it should be reserved for situations in which there is little or no prior information about the investigated associations. In most cases, however, prior information exists that implies that the disease and the exposure are associated conditional on the covariate, or else that they have no such association. In such cases a preliminary test can be seriously misleading (31).

More advanced statistical methods that combine the favorable characteristics of stratification and multivariate modeling have been proposed for the multiconfounder situation (36, 37). Although criticized for exaggerating statistical significance (38), simulation findings (39) support the use of the multivariate confounder score and the propensity score analyses as valid methods for detecting and controlling confounders in most of the practical situations encountered in epidemiologic research.

Hernberg (1) also remarked that "The general rule is that confounders should be controlled, but that the nonconfounders should be left without control to enhance the sensitivity of the study [p 316]". The control of covariates, either by stratification or regression modeling, actually involves two goals (40). Primarily, bias (or statistical inconsistency) should be prevented. Secondly, imprecision should be avoided. Miettinen (41, p 229) specified, according to the comparability definition, that, if a covariate is only a determinant of the outcome and has no association with the exposure, its control, though irrelevant for validity, nevertheless reduces the residual variance of the outcome variate and thus enhances the statistical efficiency of the analysis (ie, improves the precision of the estimates of exposure effect). But, if a nonconfounder is merely a correlate of exposure, its control reduces efficiency. However, proponents of the collapsibility approach maintain that ignoring a balanced covariate may have different implications for the validity and precision of both the estimates and the significance tests of exposure effect, depending on the scale of the outcome measurement (42).

In conclusion, the problem of confounding is a central, yet highly complicated issue in epidemiology, as in all nonexperimental causal research. Any attempt to clarify "confounding" in simple conceptual or statistical terms is destined to omit some important aspect of the topic. I hope that the commentary by Hernberg and my rejoinder will stimulate further discussion on the correct use of statistics in epidemiologic studies among the readers of this journal.

Acknowledgments

I thank Tuula Nurminen for her valuable comments.

References

1. Hernberg S. Significance testing of potential confounders and other properties of study groups — misuse of statistics [commentary]. *Scand J Work Environ Health* 1996;22:315—6.

2. Whittemore AS. Collapsing multidimensional tables. *J R Stat Soc B* 1978;40:328—40.
3. Yanagawa T. Designing case-control studies. *Environ Health Perspect* 1979;32:143—56.
4. Boivin J-F, Wacholder S. Conditions for confounding of the risk ratio and the odds ratio. *Am J Epidemiol* 1985;121:152—8.
5. Becher H. The concept of residual confounding in regression models and some applications. *Stat Med* 1992;11:1747—58.
6. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981;114:593—603.
7. Rothman KJ. *Modern epidemiology*. Boston (MA): Little, Brown and Company, 1986.
8. Greenland S, Robins JA. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;15:413—9.
9. Checkoway H, Pearce N, Crawford-Brown DJ. *Research methods in occupational epidemiology*. New York (NY): Oxford, 1989.
10. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. Belmont (CA): Lifetime Learning Publications, 1982.
11. Hernberg S. *Introduction to occupational epidemiology*. Chelsea (MI): Lewis Publishers, Inc, 1992.
12. Greenland S, Morgenstern H, Poole C, Robins JM. Re: "Confounding confounding" [letter]. *Am J Epidemiol* 1989;129:1086—9.
13. Nurminen M. Some remarks on the operation of biases. *Scand J Work Environ Health* 1983;9:377—83.
14. Greenland S. Interpretation of and choice of effect measures in epidemiologic analysis. *Am J Epidemiol* 1987;125:761—8.
15. Schlesselman JJ. *Case-control studies: design, conduct, analysis*. New York (NY): Oxford University Press, 1982.
16. Kelsey JL, Thompson WD, Evans AS. *Methods in observational epidemiology*. New York (NY): Oxford University Press, 1986.
17. Miettinen OS. Confounding and effect-modification. *Am J Epidemiol* 1974;100:350—3.
18. Robins JM. *The statistical foundations of confounding in epidemiology*. Boston (MA): Harvard School of Public Health, 1983. Occupational health program, technical report no 2.
19. Robins JM, Morgenstern H. *Confounding and prior knowledge*. Boston (MA): Harvard School of Public Health, 1983. Occupational health program, technical report no 1.
20. Robins J, Morgenstern H. Confounding and prior knowledge [abstract]. *Am J Epidemiol* 1983;118:426.
21. Wickramaratne PJ, Holford TR. Confounding in epidemiologic studies: the adequacy of the control group as a measure of confounding. *Biometrics* 1987;43:751—65.
22. Last JM. *A dictionary of epidemiology*. 2nd ed. New York (NY): Oxford University Press, 1988.
23. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985;122:495—506.
24. Miettinen O, Nurminen M. Comparative analysis of two rates. *Stat Med* 1985;4:213—26.
25. Nurminen M. To use or not to use the odds ratio in epidemiologic analyses? *Eur J Epidemiol* 1995;11:365—71.
26. Gart JJ. On the combination of relative risks. *Biometrics* 1962;18:601—10.
27. Nurminen M. On pooled and stratified analysis of epidemiologic rates. *Public Health Rev* 1988;16:163—87.
28. Robins JM, Morgenstern H. The foundations of confounding in epidemiology. *Comput Math Applic* 1987;14:869—916.
29. Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology* 1996;7:498—501.
30. Dales LG, Ury HK. An improper use of statistical testing in studying covariables. *Int J Epidemiol* 1978;7:373—5.
31. Greenland S. Cautions in the use of preliminary-test estimators [comment]. *Stat Med* 1989;8:669—73.
32. Greenland S. Confounding in epidemiologic studies [reader reactions]. *Biometrics* 1989;45:1309—22.
33. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol* 1989;129:125—37.
34. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol* 1993;138:923—5.
35. Robins JM, Greenland S. The role of model selection in causal inference from nonexperimental data. *Am J Epidemiol* 1986;123:392—40.
36. Miettinen OS. Stratification by a multivariate confounder score. *Am J Epidemiol* 1976;104:609—20.
37. Rosenbaum PR, Rubin BD. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41—55.
38. Pike MC, Anderson J, Day N. Some insights into Miettinen's confounder score approach to case-control analysis. *J Epidemiol Community Health* 1979;33:104—6.
39. Cook EF, Goldman L. Performance of tests of significance based on stratification by a multivariate confounder score or by a propensity score. *J Clin Epidemiol* 1989;42:317—24.
40. Cochran WG. *The planning and analysis of observational studies*. New York (NY): Wiley, 1983.
41. Miettinen OS. *Theoretical epidemiology: principles of occurrence research in medicine*. Albany (NY): Delmar Publishers Inc, 1985.
42. Gail MH. Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In: Moolgavkar SH, Prentice RL, editors. *Modern statistical methods in chronic disease epidemiology*. New York (NY): Wiley, 1986:3—18.