American Journal of Epidemiology
Copyright © 1995 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved

Vol. 141, No. 12
Printed in U.S.A.

HISTORICAL PAPER

AMERICAN JOURNAL OF EPIDEMIOLOGY
Copyright © 1974 by The Johns Hopkins University

Vol. 100, No. 5
Printed in U.S.A.

CONFOUNDING AND EFFECT-MODIFICATION

OLLI MIETTINEN[1]

The preceding commentary by Fisher and Patil (1) deals with criteria for a confounding factor and with the distinction between confounding and effect-modification. I take the authors and Editor up on their kind invitation to comment, and I expand on the subtleties of both confounding and effect-modification.

CONFOUNDING: DETAILS FOR CRITERIA

Conditionality on other controlled factors

Fisher and Patil note that two or more factors might jointly constitute a confounder even though each of them singly is devoid of the confounding property. They conclude that the basic criteria for a confounding factor—"relatedness" to both the exposure and the illness at issue (2)—must refer to relationships *conditional* upon all other factors that are considered for control, and they advise efforts to evaluate these conditional associations *instead* of unconditional ones to enhance the detection of confounding.

The formal conclusion reached by Fisher and Patil is, I believe, familiar to epidemiologists: We are aware that the need to control a given factor depends ultimately on its characteristics (relatedness to the exposure and the illness), conditional on whatever other factors are controlled (by restriction, matching, stratification or modelling). This is not only widely recognized but also quite routinely applied in decisions about the need to control.

On the other hand, the practical advice given by Fisher and Patil differs from that already implicit in epidemiologic decision-making. According to these authors one should not exclude a potential confounding factor from further consideration if it fails to meet the simple (unconditional) criteria. However, epidemiologists in general seem rather content to do just that inasmuch as it seems customary to confine the consideration of conditional relationships to factors for which the simple ones do indicate confounding. In other words, whereas Fisher and Patil advocate *routine* consideration of the conditional relationships with the aim of *detecting* confounding that might otherwise be missed, the prevailing tendency is for *selective* use of the conditional criteria with the aim of *excluding* from control a factor which superficially would appear to be a confounder.

Is there, then, a need for a change in epidemiologic research practice as to the use of simple and conditional criteria for confounding? I believe not. Fisher and Patil are, or course, formally correct even in their procedural advocacy. However, I believe that to pursue routinely the conditional relationships is a policy whose productiveness is much too low to justify the added efforts and complexities relative to the prevailing approach of first "screening" on the basis of unconditional criteria. Rarely is adequate information about the conditional relationships even available, as Fisher and Patil point out; and when it is, the conditional criteria only very exceptionally bring out the confounding property where it was not apparent in terms of the simple relationships. On the other hand, the conditional view is often helpful, even without data, in disposing of a potential confounder. As an example of the latter, when controlling "family" or "neighborhood" through matching, the control of, say, income would be irrelevant even if it were unconditionally associated with both

[1] Department of Epidemiology, School of Public Health, Harvard University, Boston, Mass. 02115.

Supported by grants 5 P01 CA 06373 HE 10436 from the National Institutes of Health.

the exposure and the illness.

Meaning of "relatedness"

Fisher and Patil imply that "relatedness" in each of the two criteria for a confounding factor means statistical dependence, but they give no details. Several specifications are worthy of note:

1) The relatedness to the exposure should not be simply a reflection of the exposure influencing the potential confounder. In such a case the factor at issue could be a link in one causal path from the exposure to the illness, and its control would serve only to block the manifestation of this path.

2) The relatedness to the illness must be strictly predictive, i.e., a confounding factor is necessarily a risk indicator for the illness. In fact, in the strictest sense, a confounding factor has a causal connection to the illness. However, since it is generally infeasible to control all causal risk indicators with the confounding property, it is necessary to control them indirectly through their correlates—noncausal risk indicators that are associated with the exposure; by the same token, it is reasonable to regard the latter as actual confounders rather than simply as correlates thereof.

3) The relatedness to the illness must obtain even without mediation by the exposure, i.e., it should not simply reflect association with exposure and the "effect" of the latter. More specifically, the risk indicator status should obtain even "on the null hypothesis" or, equivalently, among the nonexposed. (Risk indicator status among the exposed could result from effect-modification by the potential confounder.)

4) The relationship may be spurious—results of differential biases of selection and/or differential errors of observation over the factor at issue. Thus, a true confounding factor (social class, say) may simply *appear* to be associated with the exposure (intrauterine x-ray exposure, say) as a result of variations according to this factor in the accessibility to subjects or their records or in errors of information about the exposure; and the factor may be simply an *ostensible* risk indicator because of differential errors in case detection, depending on the category of the factor.

5) Both relationships must obtain in the data of the study at issue. When deciding upon control in the selection of subjects (restriction of range, matching in selecting control subjects), one must, of course, rely on *a priori* information about the outlook —in the absence of control—for the appearance of relationships in the data, and statistical inference might be involved in this (1). On the other hand, in the data-analysis stage the decision about possible control (exclusion of categories, stratification, model fitting) is *not* to be based on statistical inference about (significance testing of) the pertinent associations. In both stages the decisions are to be based on quantitation of the distortive impact of the factor in the absence of control. Formal estimation of this may be feasible (3).

EFFECT-MODIFICATION

First principles

Fisher and Patil point out that the magnitude of the "effect"—the parameter for association conditional on controlled confounders—may vary among categories of some factor, and that this factor need not be a confounder of the association. They go no further into the principles of this phenomenon. For placing their note in its proper perspective, it is necessary to appreciate some basic principles of effect-modification:

1) All causally/preventively insufficient exposures have extremely powerful effect-modifiers, which render the effects either total or nonexistent, i.e., which make the exposure causally/preventively either sufficient or inoperative. This may be deduced from the facts that a given illness always results from a sufficient cause, and that the

presence/absence of a given insufficient cause/preventive is a necessary part of some of these sufficient ones; where it is, it is (necessary and) sufficient, whereas otherwise it has no effect at all.

2) All intermediate effect-modification results from differential (latent) weightings of the total and zero effects among the categories of the factor at issue. Since the weights are nothing but relative frequencies of the regions of causal/preventive (sufficiency and) necessity of the exposure at issue, an intermediate effect-modifier need not be a risk indicator conditional on exposure or nonexposure. By the same token, it need not be a confounder (cf. specification no. 3 for "relatedness").

3) All risk indicators, conditional on exposure or nonexposure, modify at least one of the two common epidemiologic measures of effect—risk difference or risk ratio (minus one). Thus, the common assumption of uniformity of risk ratio over a confounder is tantamount to the assumption that, over the range of the confounder, (the absolute value of) risk difference is proportional to the risk of the nonexposed.

Appreciation in data analysis

Fisher and Patil point out, without elaboration, that data analysis without regard for effect-modification can be of suboptimal sensitivity in the detection of the very existence of the association and incomplete in the estimation of its magnitude.

At the same time, routine approaches in epidemiologic data analysis are generally oblivious to possible effect-modification: Common significance tests—the McNemar test (4) and its extension (5) for matched series as well as the more general Mantel-Haenszel test (6) for (matched and) unmatched series—do not provide for nonuniformity of effect. They are, in fact, optimal in the very case where the favorite parameter, risk ratio (or, more accurately, the "odds ratio"), is uniform over the matching categories or the strata of analysis (7). Similarly, common procedures for estimating this parameter either presuppose uniformity over the matching categories (8) or strata of analysis (9), or they suppress the nonuniformity through "summarization" (6) or standardization (10).

The prevailing habits in significance-testing (for the main null hypothesis of no association conditional on the controlled confounders) cannot be blamed on epidemiologists. For, seemingly no test is known which, by virtue of allowing for nonuniformity of the "odds ratio", would in general tend to be more powerful than the Mantel-Haenszel test (6) or its derivatives (4, 5).

On the other hand, the common failure to look for, and quantify, modification of risk ratio ("odds ratio") is not readily justifiable. As long as the numbers of subjects are sufficient in the categories of the potential modifier, uniformity of this parameter may be tested—against an unspecific alternative—by the use of the statistic $\chi^2(J-1) = \Sigma_j[\chi^2(1)]_j - \chi^2(1)$, where $\chi^2(1)$ is the usual overall chi-square computed without regard for nonuniformity (4-6), and where $[\chi^2(1)]_j$ is the same statistic computed for the j^{th} one of the J categories of the potential modifier (cf. Zelen (11)). More sensitive testing may also be feasible using an appropriate model for the risk ratio as a function of the modifying factor.

If might seem that the ideal framework for the evaluation of possible modifiers of the "odds ratio" is the "log-linear" model (12), with parameters for interaction between the exposure and other factors interpreted as measures of modification. Thus, if the exposure, coded as 0 for absent and 1 for present, were to have the estimated "main effect" of b_e and "interaction effect" b_{ea} with age (A) but none with other factors in the model, then the age-function of the "odds ratio" might be thought to be $\exp(b_e + b_{ea}A)$. I believe that this would be wrong in the usual situation where the model involves other age-related factors (as confounders). After all, when deprived of

its correlates (by "keeping them fixed" through the model), "age" becomes a hollow concept, a biologically meaningless number. This pitfall of model interpretation tends to apply, though less strikingly, to sex and many other factors as well, and the problem is not specific to the "log-linear" model but characteristic of multivariate analysis in general.

The solution to this problem of model interpretation is to be sought, I believe, from an appropriate choice of the effect parameter. It seems to me that, in ratio terms, the appropriate parameter is the "standardized morbidity/mortality ratio" for the exposed relative to the nonexposed, evaluated separately within each category of the potential modifier. The standard would vary among the categories (10), but the ratios (minus one) would express the varying "effects", without confounding by the "standardized" factors, for the varying kinds of exposed people in the different categories of the potential modifier. The confounders would thus be controlled as confounders but not as determinants of effect-modification attributed to another factor.

Summary

Confounding and effect-modification —both very central to epidemiologic thinking and research on causality—are closely related but distinctly separate concepts and phenomena. Both of them involve considerable subleties, with implications for problem conceptualization, study design, data analysis and inference. Some of these subtleties and their implications may warrant greater appreciation in the practice of epidemiologic research, while others require further conceptual development.

References

1. Fisher L, Patil K: Matching and unrelatedness. Am J Epidemiol 100:347-349, 1974
2. Miettinen OS: Matching and design efficiency in retrospective studies. Am J Epidemiol 91:111-118, 1970
3. Miettinen OS: Components of the crude risk ratio. Am J Epidemiol 96:168-172, 1972
4. McNemar Q: Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12:153-157, 1947
5. Miettinen OS: Individual matching with multiple controls in the case of all-or-none responses. Biometrics 25:339-355, 1969
6. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719-748, 1959
7. Birch MW: The detection of partial association. I: The 2 × 2 case. J Roy Statist Soc B 26:313-324, 1964
8. Miettinen OS: Estimation of relative risk from individually matched series. Biometrics 26:75-86, 1970
9. Gart J: The comparison of proportions: a review of significance tests, confidence intervals and adjustments for stratification. Rev Internatl Statist Inst 39:148-169, 1971
10. Miettinen OS: Standardization of risk ratios. Am J Epidemiol 96:383-388, 1973
11. Zelen M: The analysis of several 2 × 2 contingency tables. Biometrika 58:129-137, 1971
12. Worcester J: The relative odds in the 2^3 contingency table. Am J Epidemiol 93:145-149, 1971