(1985). *Evaluation of Short-Term Tests for Carcinogens: Report of the International Programme on Chemical Safety Collaborative Study on In Vitro Assays*. North Holland, Amsterdam.
COCHRAN, W. G. (1937). Catalogue of uniformity trial data. *J. Roy. Statist. Soc. Suppl.* **4** 233–253.
MARGOLIN, B. H. (1985). Statistical studies in genetic toxicology: A perspective from the U. S. National Toxicology Program. *Environ. Health Perspectives* **63** 187–194.
MARGOLIN, B. H., RESNICK, M. A., RIMPO, J. Y., ARCHER, P., GALLOWAY, S. M., BLOOM, A. D. and ZEIGER, E. (1986). Statistical analyses for in vitro cytogenetic assays using Chinese hamster ovary cells. *Environ. Mutagen.* **8** 183–204.

Comment

Richard G. Cornell

Rosenbaum's review provides a logical framework for thinking about the value of more than one control group in observational studies and contains useful discussion of the implications for the design and interpretation of investigations in which randomization to control and treatment groups is not feasible.

I found his discussion of case-control studies particularly interesting. Controls in these studies are "noncases" selected for comparison with a group of "cases" known to have a particular disease or other condition. Controls and cases are compared with respect to the extent of exposure to potential causative agents or with respect to other background variables. Rosenbaum emphasizes that a comparison of the histories of two or more control groups provides a check of the assumptions that underlie the estimation of the effect of exposure after covariance adjustment. The example he presents on the extent of exposure to sunlight of cataract cases and controls involves three control groups with other eye conditions. The use of multiple control groups enabled him to conclude that adjustment for age and sex is not sufficient for unbiased estimation of the effect of sunlight exposure on the prevalence of cataracts.

This example serves as a prototype for the interpretation of other case-control studies with more than one control group and a warning of possible undetected bias in case-control studies with only one control group. More importantly, it serves as a reminder that it is best to select more than one control group when the ideal control group cannot be formed through randomization. This allows a check on assumptions that cannot be attained through randomization and yet are crucial to conclusions on the effect of exposure after taking covariates into account.

Rosenbaum also refers to an example of a study in which the groups to be compared are exposure groups and the outcome is the prevalence of coronary thrombosis, with subclasses within each exposure group defined by covariates. Observational studies of this type are common in epidemiology and medicine. One example is a study, described by Cornell (1984), of cancer rates relative to exposure to the environment in steel plants, which produce stainless steel. One purpose of this study was to see if there is evidence of an increase in lung cancer rate attributable to the nickel and chromium used in stainless steel production. Exposure groups were formed by area worked within a plant. Comparisons were made after adjusting for age.

Another example is the comparison of survival rates for burn victims using registry data grouped by hospital, and subsequently by the speed of wound closure attained by the burn care practice in a hospital. Again, age is an important covariate for comparisons of burn survival. So is burn severity as measured by the extent of full thickness burn. A model that takes these variables as well as other demographic and severity variables into account for purposes of estimation, prediction and evaluation is discussed by Wolfe, Roi, Flora, Feller and Cornell (1983) and presented in detail by Cornell, Flora and Roi (1983).

These examples are typical of observational studies in epidemiology and medicine in that morbidity or mortality rates are compared between groups formed by exposure categories or type of treatment. Comparisons are made within categories defined by covariates or after adjustment for covariates. Common types of covariates are demographic variables, such as age and sex, and initial severity measures.

Rosenbaum gives guidance with respect to the design and analysis of such studies. He says that it is desirable to select two control groups in such a way that a possibly relevant, but unobserved, covariate has different distributions in the two control groups, and then to check to see if the responses in the two control groups are similar. If they are, then the unobserved

Richard G. Cornell is Professor of Biostastistics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109.

covariate can be ignored without concern for bias. If the responses in the two control groups are not similar, then care must be taken since there is evidence that the effect of exposure or treatment depends on variables not controlled for in the design and not measured for entry into the analysis.

How could this be done in the example of the study of stainless steel workers? Many times such studies contain comparisons with other workers in the same plant or with local, state or national population rates after adjusting for age. In comparisons within the same plant, cancer rates may be affected by substances in the environment other than nickel and chromium. The principle of control by systematic variation presented by Rosenbaum would lead to the formation of more than one control group distinguishable by different types and levels of exposure within the workplace environment, but not including nickel or chromium. For instance, one group could consist of office personnel. Another could consist of workers exposed to heat and dust in a plant that produces steel without nickel or chromium as ingredients.

In the example concerning the evaluation of burn care, the potential for survival may not be fully depicted by objective measures of the severity of the burn and by information on age and comorbidity that reflect the health and resilience of patients. Therefore it would be appropriate to include hospitals in the control, as well as treatment group, which treat a variety of types of patients. Possibilities include an inner city hospital that treats many indigent patients, a military hospital and a research hospital.

It may seen obvious that multiple control groups would be informative and desirable in these examples. Yet many observational studies use only one control group. Rosenbaum's review focuses attention on the benefit to be gained from multiple control groups and develops useful principles for their design and interpretation.

ADDITIONAL REFERENCES

CORNELL, R. G. (1984). Mortality patterns among stainless steel workers. In *Nickel in the Human Environment* (E. W. Sunderman, Jr., ed.) 65–71. International Agency for Research on Cancer, Oxford Univ. Press, N. Y.

CORNELL, R. G., FLORA, J. D. and ROI, L. D. (1983). The statistical evaluation of burn care. *Undergraduate Mathematics Applications Project J.* **4** 85–115.

WOLFE, R. A., ROI, L. D., FLORA, J. D., FELLER, I. and CORNELL, R. G. (1983). Mortality differences and the speed of wound closure among specialized burn care facilities. *J. Amer. Med. Assoc.* **250** 763–766.

Comment

Norman Breslow

The major conclusions of this paper should come as no surprise to biostatisticians and epidemiologists involved in the applications of statistical methods and concepts to clinical and observational studies in public health and medicine. For the most part, the arguments lend mathematical precision to principles of study design and interpretation that are well-known and widely used. Few epidemiologists involved with cohort studies, for example, would take issue with the statement that "... if, after adjustment for (the covariables), the two control groups differ with respect to the response ... then ... at least one of the control groups is not comparable to the treated group." The principles of "control by systematic variation" and "bracketing" seem intuitively clear and are evidently well-established in the social science literature. Although it is reassuring to note that these principles are confirmed by the randomization paradigm developed by Rubin and Rosenbaum, it would perhaps be of even greater interest to see examples where the paradigm led to novel insights and new methods of study design and analysis. Recent work of Robins (1986) on causal inference in occupational cohort studies represents a start in this direction.

Turning specifically to the field of case-control studies, the major conclusion is that inference regarding the validity of a covariable-adjusted relative risk estimated for a particular exposure is strengthened if one can demonstrate equality in the covariable-adjusted odds ratios that contrast the exposures in two or more control groups. This confirms the insights of Hill (1971): "If a whole series of control groups, e.g., of patients with different diseases, gives much the same answer and only the one affected group differs, the evidence is clearly much stronger than if the affected group differs from merely one other group." In their pioneering work on lung cancer and smoking, Hill and Doll in fact utilized two control groups. One consisted of patients with diagnoses other than cancer who were

Norman Breslow is Professor and Chairman, Department of Biostatistics, University of Washington, Seattle, Washington 98195.