

## CONCEPTS OF INTERACTION

KENNETH J. ROTHMAN,<sup>1</sup> SANDER GREENLAND<sup>2</sup> AND ALEXANDER M. WALKER<sup>1</sup>

Readers of the *American Journal of Epidemiology* have seen a lively discourse on the topic of synergy, a major conceptual area in epidemiology for which there exists fundamental controversy as to definitions. In 1974, one of us (KJR) (1) proposed that synergy (or its negative counterpart, antagonism) between two or more causes of disease ought to be evaluated in reference to a specific yardstick. The reference point was one that equated independence of the causes with the situation in which the joint effect was equal to the sum of the separate effects, with effect defined as excess risk. Rothman emphasized the distinction between biologic and statistical interaction, arguing that biologic interaction, unlike statistical interaction, could not be defined with arbitrariness in the choice of scale of measurement.

The arguments given were not completely persuasive. Walter and Holford (2) contended that the choice among statistical models of independence depended on specification of the causal models under evaluation, which differed for various etiologic situations. They argued also that choice of a statistical model of independence should at times reflect statistically convenient properties of the model. Kupper and Hogan (3) also expounded the view that the existence of interaction is "model-dependent." Blot and Day (4), re-

sponding to the arguments of Kupper and Hogan, emphasized the distinction between the terms synergy and interaction, and proposed that public-health objectives should be considered separately from (and in preference to) statistical objectives in judging whether two factors are synergistic.

We believe that the controversy surrounding the concept of interaction can be laid to rest with specification of the context in which the interaction is being evaluated. Four broad contexts can be distinguished: statistical, biological, public health, and individual decision-making. Each has different implications for the evaluation of interaction.

### STATISTICAL INTERACTION

The term *statistical interaction* is intended to denote the interdependence between the effects of two or more factors within the confines of a given model of risk. The evaluation of interaction depends on the model chosen. The general form of models of disease risk for two dichotomous risk factors *A* and *B* may be abbreviated as follows:

$$R(A,B) = F(K_0, K_aA, K_bB, K_{ab}AB)$$

where  $R(A,B)$  is the risk or rate associated with specified combinations of factors *A* and *B*,  $F$  is the risk or rate function, and the coefficients  $K_0$ ,  $K_a$ ,  $K_b$ , and  $K_{ab}$  are estimated from the data and are chosen in such a way as to maximize the compatibility of the estimates with the observed data according to some criterion such as least squares, maximum likelihood, etc. For dichotomous factors *A* and *B* and a data set with complete observations, the set of estimates  $K_0$ ,  $K_a$ ,  $K_b$ ,  $K_{ab}$  will give a complete description of the observed risks

<sup>1</sup> Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115 (reprint requests to Dr. Rothman).

<sup>2</sup> Division of Epidemiology, University of California, Los Angeles, CA.

This work was supported in part by a contract (NO1-NS-8-2399) with the National Institute for Neurological and Communicative Disorders and Stroke and a grant from the Millbank Memorial Foundation.

or rates. Two commonly used forms of  $F$  are the logistic model for risk

$$R(A,B) = \frac{e^{(K_0 + K_aA + K_bB + K_{ab}AB)}}{1 + e^{(K_0 + K_aA + K_bB + K_{ab}AB)}}$$

and the additive model for rates

$$R(A,B) = K_0 + K_aA + K_bB + K_{ab}AB.$$

For the purposes of describing risk variation, the coefficient  $K_{ab}$  may be taken as a parameter of interaction; the magnitude of  $K_{ab}$  depends on the choice of the model. If the choice is dictated solely by concern for statistical convenience, without reference to biologic mechanism, no scientific inferences about biologic interaction are possible, as has been pointed out by Walter and Holford (2) as well as Kupper and Hogan (3).

A major goal for statistical analysis has been simple description of observed phenomena, often with prediction as a related goal. Description and prediction need not be linked to biologic inference. For example, the early identification of persons at high risk for cancer might proceed by intensive screening of those for whom a high probability of developing the cancer is predicted from a multivariate risk function. Such a function need not shed any light on etiology to be effective. In fact, all the risk factors in the function could be noncausal risk factors—correlates of the actual causes—and the function might still perform well.

For predictive functions, simplicity is desirable. Models and scales of measurement which lead to a simple function without sacrificing accuracy are reasonable goals. The evaluation of statistical interaction, as a step to finding a simple, accurate predictive function, need not be tied to specific biologic models. For the sole purpose of prediction, the rules for model building and evaluation of interaction are not necessarily restrained by considerations of biologic mechanism or interpretation of individual terms in the

model. On the other hand, prediction, especially extrapolation, may be more accurate if the mathematical model corresponds well to the biologic process; for this reason, the biologic accuracy of a model should be considered along with statistical convenience and simplicity, even for models used solely for prediction.

#### BIOLOGIC INTERACTION

*Biologic interaction* may be defined as the interdependent operation of two or more causes to produce disease. To achieve agreement on a more precise definition of biologic interaction would probably be difficult, but fortunately it is not necessary. Consider a simple example—the initiation-promotion model for carcinogenesis: agent  $A$  increases the number of cells susceptible to carcinogen  $B$ , which can transform a susceptible cell into a malignant cell.  $A$  and  $B$  are both causes, but some might say they act independently whereas others might call them synergistic. As long as the model is explicit and its implications can be deduced, however, the question of whether the causes are said to act synergistically or independently is moot. Specification of the biologic model replaces the purely abstract and vague notion of interdependence of effects with a specific form of interdependence which is subject to evaluation. Little information about the biologic process can be gained from classifying the hypothesized biologic mechanism as one of synergy or independence. Such classification could be a useful shorthand to describe categories of mechanisms, but only if such categories were widely and explicitly agreed upon in the scientific community. Two categories of general interest are those in which etiologic factors act interchangeably in the same step in a multistep process, or alternatively act at different steps in the process. These broad categories correspond generally to mathematical models in which the effects of factors are additive or multiplicative,

respectively. A data analysis which evaluates the relative tenability of such mathematical models can, therefore, cast light on the nature of the interrelation between effects of causal factors; likelihood ratios based on the alternative models would provide a reasonable means for comparison. Such an analysis provides insight into etiologic mechanisms without the need to specify a definition of synergy, much less achieve consensus on the definition.

#### PUBLIC-HEALTH INTERACTION

The primary concerns when evaluating interaction in public-health contexts are the number of cases of disease occurring in a population and the proportional contribution of each risk factor to this case burden. For public-health purposes, the effects of two factors, *A* and *B*, may be considered independent if the number of cases of disease that would occur in the population does not depend on the extent to which *A* and *B* occur together in the same individuals. If the number of cases does depend on the extent to which risk factors *A* and *B* occur together in the same individuals, then the effects of *A* and *B* interact in a manner which bears on any public concern relating to the disease in question. The preceding considerations imply, as Blot and Day (4) pointed out, that for public-health purposes interaction between risk factors is equivalent to a departure from additivity of incidence rate differences ("attributable risks").

Consider, for example, a population in which the risk of lung cancer in a given time period is one per thousand for those not exposed to either of two risk factors, say smoking and asbestos. Suppose that the risk for smokers who are not exposed to asbestos is 10 per thousand in the same time period. For those with asbestos exposure, the risks are three and 30 per thousand for nonsmokers and smokers, respectively. The risk ratio is identical for

smoking (and for asbestos) at each level of the other risk factor. To some data analysts, this might indicate statistical or biologic independence, but for public-health applications it is not appropriate to consider the effects of smoking and asbestos as independent. The number of cases of lung cancer caused by cigarette smoking depends on how many of the smokers are also exposed to asbestos (or symmetrically, the number caused by asbestos depends on how many of those exposed to it are also smokers). Consequently, the public-health implications of the effects of smoking and asbestos depend on the proportion of the population in which these factors occur jointly. No alternative exists to an additive criterion for evaluating independence of effects for the public-health viewpoint, as long as the public-health burden is directly proportional to the number of cases.

#### INTERACTION IN INDIVIDUAL DECISION-MAKING

Decision-making in evaluating personal risk entails considerations parallel to decision-making for public-health issues. Thus, a physician advising a woman on whether to use oral contraceptives would reasonably inquire about the presence of hypertension, despite the fact that the risk ratio for cerebrovascular complication from oral contraceptives is about the same among normotensive and hypertensive women (5). The increase in absolute risk, however, is considerably greater for hypertensive women, thus making hypertension interactive with oral-contraceptive use for the purposes of evaluating individual risk. On this basis, hypertension becomes a contraindication for oral-contraceptive use. It is noteworthy that no reference to a biologic model is pertinent to this type of evaluation of interaction: departures from additivity of risk differences are the focus whatever the underlying biologic mechanisms might be, provided that cost to the indi-

vidual is taken to be a linear function of risk, as it usually is, especially when risks are small.

#### CONCLUSION

We believe that the disagreement about the methodologic principles appropriate to the evaluation of interaction stems from a failure to separate the four contexts discussed here. In statistical contexts, independence and interaction may well be defined in an arbitrary manner. In biologic contexts addressing specific causal mechanisms, defining interaction or synergy between factors is unnecessary, since these terms do not enhance the intelligibility of a mechanism which is already specified in detail. We believe that in public-health contexts synergy and antagonism should ordinarily be interpreted

as departures from additivity of incidence rate differences, and analogously, in the context of individual decision-making, synergy and antagonism should ordinarily be interpreted as departures from additivity of risk differences.

#### REFERENCES

1. Rothman KJ: Synergy and antagonism in cause-effect relationships. *Am J Epidemiol* 103:506-511, 1976
2. Walter SD, Holford TR: Additive, multiplicative and other models for disease risks. *Am J Epidemiol* 108:341-346, 1978
3. Kupper LL, Hogan MD: Interaction in epidemiologic studies. *Am J Epidemiol* 108:447-453, 1978
4. Blot WJ, Day NE: Synergism and interaction: are they equivalent? (Letter to the Editor.) *Am J Epidemiol* 110:99-100, 1979
5. Collaborative group for the study of stroke in young women: Oral contraceptives and stroke in young women. Associated risk factors. *JAMA* 231:718-722, 1975