

## THEORETICAL EPIDEMIOLOGY

# Gene-Environment Interaction: Definitions and Study Designs

RUTH OTTMAN, PH.D.

*G. H. Sergievsky Center, Columbia University, and Epidemiology of Brain Disorders Research Department, New York State Psychiatric Institute, New York, New York 10032*

---

**Study of gene-environment interaction is important for improving accuracy and precision in the assessment of both genetic and environmental influences. This overview presents a simple definition of gene-environment interaction and suggests study designs for detecting it. Gene-environment interaction is defined as "a different effect of an environmental exposure on disease risk in persons with different genotypes," or, alternatively, "a different effect of a genotype on disease risk in persons with different environmental exposures." Under this strictly statistical definition, the presence or absence of interaction depends upon the scale of measurement (additive or multiplicative). The decision of which scale is appropriate will be governed by many factors, including the main objective of an investigation (discovery of etiology, public health prediction, etc.) and the hypothesized pathophysiologic model. Five biologically plausible models are described for the relations between genotypes and environmental exposures, in terms of their effects on disease risk. Each of these models leads to a different set of predictions about disease risk in individuals classified by presence or absence of a high-risk genotype and environmental exposure. Classification according to the exposure is relatively easy, using conventional epidemiologic methods. Classification according to the high-risk genotype is more difficult, but several alternative strategies are suggested.** © 1996 Academic Press, Inc.

**Key Words:** human genetics; epidemiology; interaction.

---

### INTRODUCTION

Rapid developments in molecular biology in recent years have led to a dramatic increase in our understanding of the genetic influences on many human diseases. At the same time, we have become increasingly aware that for many diseases, the genetic influences are exceedingly complex. These "complex" diseases

tend to aggregate in families, but their familial distributions are inconsistent with simple Mendelian modes of inheritance (autosomal or X-linked, dominant or recessive). Both genetic and environmental factors may contribute to susceptibility, and it is unclear how these factors interact in their influence on risk. The genetic or environmental mechanisms may differ among families or clinically defined subsets of the disease. Moreover, even within a single family, the effect of a susceptibility genotype might be variable because of the modifying effects of other genes or environmental factors.

Faced with this complexity, investigators have recognized the need to integrate the research tools of epidemiology and human genetics. Thus over the past 15-20 years, a new hybrid discipline, genetic epidemiology, has emerged [1]. This new field incorporates concepts and methods from epidemiology, biostatistics, clinical genetics, molecular genetics, and population genetics, as well as new methods developed to address problems specific to the study of the genetic contributions to "complex" diseases.

Study of gene-environment interaction is central to the emerging field of genetic epidemiology. In considering the joint effects of risk factors in disease causation, however, epidemiologists have debated intensely about what interaction is, where it comes from, and how to detect it [2-12]. These considerations apply no less to gene-environment interaction than to interaction between any two risk factors. The goals of this overview are to present a simple definition of gene-environment interaction and to suggest study designs for detecting it.

### GENE-ENVIRONMENT INTERACTION DEFINED

Consider an environmental risk factor, a "high-risk genotype," and a disease of interest. The environmental risk factor can be an exposure, either physical (e.g., radiation, temperature), chemical (e.g., polycyclic aromatic hydrocarbons), or biological (e.g., a virus); a behavior pattern (e.g., late age at first pregnancy); or a "life event" (e.g., job loss, injury). This is not intended

Supported by NIH RO1-NS20656.

as an exhaustive taxonomy of risk factors, but indicates as broad a definition as possible of environmental exposures. Similarly, the high-risk genotype is broadly defined, and can involve an autosomal or X-linked major gene, a polygenic model, or an epistatic model.

For simplicity, we assume that all three variables are dichotomous. This is obviously an oversimplification, to facilitate ease and clarity of discussion. Many exposures may be measured on a polychotomous or continuous scale (e.g., pack-years); several alternative genotypes may raise risk for some diseases (e.g., BRCA1 and BRCA2 for breast cancer, and different mutations in BRCA1) and some “disease” outcomes may be more appropriately measured on a continuous scale (e.g., body mass index, DNA adducts, cholesterol levels). The concepts of interaction described below could easily be extended to these polychotomous or continuous cases.

Disease risk for each of the four combinations of genotype and environmental risk factor is denoted by  $r_{11}$  (for exposed persons with the high-risk genotype),  $r_{10}$  (for those with the exposure alone),  $r_{01}$  (for those with the genotype alone), and  $r_{00}$  (for those with neither) (Table 1).

To define gene-environment interaction, it is useful to review the concept of *conditional independence*, which stipulates that the relationship between two factors is the same across strata defined by a third factor. In the context of the effects of two risk factors, A and B, on disease risk, conditional independence implies that the effect of factor A on disease risk is the same across strata defined by factor B. If this condition is not met, then an interaction between A and B can be said to exist. Based on this concept, gene-environment interaction can be defined as “a *different effect* of an environmental exposure on disease risk in persons with different genotypes,” or, equivalently, “a *different effect* of a genotype on disease risk in persons with different environmental exposures.”

Under this strictly statistical definition, whether or

not interaction is said to exist will depend on the scale of measurement. If risks are measured on an additive scale, the effect of an environmental exposure differs among persons with different genotypes (i.e., interaction on an additive scale) when  $r_{11} - r_{01} \neq r_{10} - r_{00}$ . If risks are measured on a multiplicative scale, the effect of an environmental exposure differs among persons with different genotypes (i.e., interaction on a multiplicative scale) when  $r_{11}/r_{01} \neq r_{10}/r_{00}$ .

Table 1 shows the data layout for a cohort study (part A) or case-control study (part B) in which the effects of an environmental exposure and a genotype on disease risk are assessed. As shown in the table, the data from a cohort study can be used to compute four relative risks (RRs), using persons with neither the exposure nor the high-risk genotype as the reference group:  $RR_{11}$  for persons with both genotype and exposure,  $RR_{10}$  for persons with the exposure alone,  $RR_{01}$  for persons with the genotype alone, and  $RR_{00}$  for persons with neither genotype nor exposure. The data from a case-control study can be used to compute four analogous ORs. Under the definition of interaction given above, Table 2 shows the expected relations among these RRs or ORs, under the conditions of no interaction, synergistic interaction, and antagonistic interaction, for the additive and multiplicative scales, respectively. If the effects of two variables meet the condition of “no interaction on a multiplicative scale,” the data can be said to fit a “multiplicative model,” and if their effects meet the condition of “no interaction on an additive scale,” the data can be said to fit an “additive model” (Table 2).

A given dataset can be tested to see whether it conforms to an additive or multiplicative model, assuming appropriate data can be collected for this purpose. However, the question of which scale should be used to define interaction has been intensely debated [2-5]. Rothman has advocated use of a fixed reference point to define interaction in epidemiologic studies, and argues that the additive scale of measurement is the only

**TABLE 1**  
Epidemiologic Measures of Effects of High-Risk Genotype and Environmental Exposure

| Disease status        | High-risk genotype present |                           | High-risk genotype absent |                          |
|-----------------------|----------------------------|---------------------------|---------------------------|--------------------------|
|                       | Exposed                    | Unexposed                 | Exposed                   | Unexposed                |
| A. Cohort study       |                            |                           |                           |                          |
| Affected              | $a$                        | $b$                       | $e$                       | $f$                      |
| Unaffected            | $c$                        | $d$                       | $g$                       | $h$                      |
| Risk ( $r_{ij}$ )     | $r_{11} = \frac{a}{a+c}$   | $r_{01} = \frac{b}{b+d}$  | $r_{10} = \frac{e}{e+g}$  | $r_{00} = \frac{f}{f+h}$ |
| Relative risk         | $RR_{11} = r_{11}/r_{00}$  | $RR_{01} = r_{01}/r_{00}$ | $RR_{10} = r_{10}/r_{00}$ | $RR_{00} = 1.0$ (ref)    |
| B. Case-control study |                            |                           |                           |                          |
| Case                  | $a$                        | $b$                       | $e$                       | $f$                      |
| Control               | $c$                        | $d$                       | $g$                       | $h$                      |
| Odds ratio            | $OR_{11} = ah/cf$          | $OR_{01} = bh/df$         | $OR_{10} = eh/gf$         | $OR_{00} = 1.0$ (ref)    |

**TABLE 2**  
Definition of Interaction on Additive and Multiplicative Scales

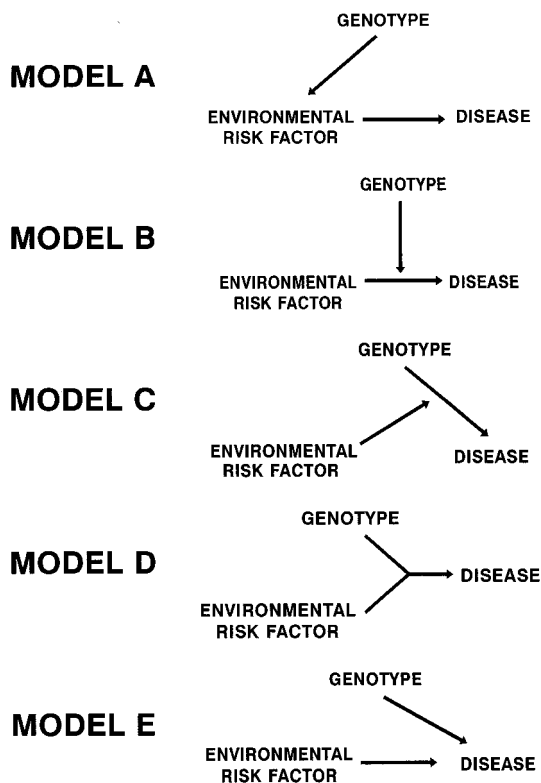
|                          | Scale of measurement   |  |
|--------------------------|--|--|
|                          | Additive   | Multiplicative   |
| No interaction           | $RR_{11} = RR_{01} + RR_{10} - 1$<br>$OR_{11} = OR_{01} + OR_{10} - 1$<br>("additive model") | $RR_{11} = RR_{01} \times RR_{10}$<br>$OR_{11} = OR_{01} \times OR_{10}$<br>("multiplicative model") |
| Synergistic interaction  | $RR_{11} > RR_{01} + RR_{10} - 1$<br>$OR_{11} > OR_{01} + OR_{10} - 1$                       | $RR_{11} > RR_{01} \times RR_{10}$<br>$OR_{11} > OR_{01} \times OR_{10}$                             |
| Antagonistic interaction | $RR_{11} < RR_{01} + RR_{10} - 1$<br>$OR_{11} < OR_{01} + OR_{10} - 1$                       | $RR_{11} < RR_{01} \times RR_{10}$<br>$OR_{11} < OR_{01} \times OR_{10}$                             |

meaningful reference point [2]. This argument is based on a conceptual, rather than purely statistical, definition of interaction: coparticipation of two factors in a single causal mechanism. Rothman asserts that two factors that are part of different causal mechanisms will generally have an additive relationship, whereas those that are part of the same mechanism will be more than additive (i.e., they will interact on an additive scale). Others have argued that the choice of measurement scale is arbitrary in the absence of a specific pathophysiologic model, and both the additive and the multiplicative scales can be appropriate under different conditions [3–5]. For example, if the disease etiology involves a multistage process (such as with the initiation and promotion stages in cancer), two factors that act at the same stage will generally fit an additive model, whereas those that act at different stages will generally fit a multiplicative model [7,8]. Rothman et al. also pointed out that the choice of the measurement scale depends in part on the goal of the investigation [7]. If the primary goal is to unravel disease etiology, it may be more appropriate to use a multiplicative scale, whereas if it is to predict the number of cases in the population, it may be more appropriate to use an additive scale [7].

#### MODELING RELATIONS BETWEEN GENOTYPE AND EXPOSURE

I recently described five biologically plausible models of relations between a genotype and an environmental exposure in terms of their effects on disease risk [13] (Fig. 1). Each of the five models leads to a different set of predictions about disease risk in individuals classified by presence or absence of the high-risk genotype and environmental exposure (Table 3). In the table, ">1" denotes a relative risk exceeding 1.0, and "≥1" denotes a greater increase in risk.

In Model A, the effect of the genotype is to produce, or increase expression of, a "risk factor" that can also be produced nongenetically. An example is the relation of the autosomal recessive disorder, phenylketonuria (PKU), to high blood phenylalanine and mental retar-



**FIG. 1.** Five models of the relation between a high-risk genotype and an environmental exposure, in terms of their effect on disease risk.

dation. Individuals who are homozygous for the PKU gene have a deficiency in the enzyme required to convert phenylalanine to tyrosine. If left untreated, a buildup of blood phenylalanine occurs after birth (before birth, the mother's enzymes are used), and the high blood phenylalanine levels cause mental retardation. The retardation can be prevented if blood phenylalanine levels are kept low by dietary restriction. Mental retardation can also result from exposure to high blood phenylalanine in persons who are not homozygous for PKU: offspring of PKU mothers, who have intrauterine exposure to high blood levels because of their *mother's* enzyme deficiency.

In epidemiologic terms, the high blood phenylalanine level in Model A is an *intervening* (or *intermedi-*

**TABLE 3**  
Expected Relative Risks Under Different Models

| Model | High-risk genotype present |                     | High-risk genotype absent |                           |
|-------|----------------------------|---------------------|---------------------------|---------------------------|
|       | Exposed $RR_{11}$          | Unexposed $RR_{01}$ | Exposed $RR_{10}$         | Unexposed $RR_{00}$ (ref) |
| A     | >1                         | 1                   | >1                        | 1                         |
| B     | ≥1                         | 1                   | >1                        | 1                         |
| C     | ≥1                         | >1                  | 1                         | 1                         |
| D     | >1                         | 1                   | 1                         | 1                         |
| E     | ?                          | >1                  | >1                        | 1                         |

ate) variable [14], and the effect of the exposure is the same in persons with and without the high risk genotype. This is explicitly *not* interaction, as defined above. It is an important model, however, because discovery of the mechanisms by which susceptibility genes influence disease is a central goal of genetic epidemiology. The same biologic mechanisms may apply to both genetic and nongenetic causal pathways.

In Model B, the genotype exacerbates the effect of the risk factor, but there is no effect of the genotype in unexposed persons. One example is the relation of xeroderma pigmentosum, an autosomal recessive disorder, to ultraviolet (UV) radiation and skin cancer. Excessive exposure to UV radiation increases risk for skin cancer in the general population, but individuals with xeroderma pigmentosum are deficient in an enzyme required for repair of DNA damage induced by UV radiation, and hence have even higher risk. If sun exposure could be prevented completely in these persons, they would not have increased risk for skin cancer.

In Model C, the exposure exacerbates the effect of the genotype, but there is no effect of the exposure in persons with the low-risk genotype. Individuals with porphyria variegata, an autosomal dominant disorder, have skin problems of varying severity, including unusual sun sensitivity and a tendency to blister easily. If they are exposed to barbiturates, an innocuous exposure in the general population, they experience acute attacks that may involve paralysis or even death.

In Model D, both the exposure and the genotype are required to increase risk. Most individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency, an X-linked recessive disorder, are asymptomatic. However, some persons with this genotype develop severe hemolytic anemia if they eat fava beans. Dietary exposure to fava beans does not produce this reaction in individuals without G6PD deficiency.

In Model E, the exposure and the genotype each have some effect on disease risk, and when they occur together risk is higher or lower than when they occur alone. An example is the relation between  $\alpha$ -1-antitrypsin deficiency, smoking, and chronic obstructive pulmonary disease (COPD). Risk of COPD is increased both in nonsmokers with  $\alpha$ -1-antitrypsin deficiency and in smokers without  $\alpha$ -1-antitrypsin deficiency. Risk is increased to a greater extent in smokers with  $\alpha$ -1-antitrypsin deficiency.

Note that in Models B, C, and D, interaction is always present, regardless of whether risks are measured on an additive or multiplicative scale. However, Model E encompasses situations both with and without interaction, and moreover, the choice of the scale of measurement will determine whether or not interaction is said to exist. In the  $\alpha$ -1-antitrypsin example, Khoury et al. [15] reported relative risks of COPD of  $RR_{10} = 3.8$  in smokers with the low-risk genotype (PiM),  $RR_{01} = 1.6$  in nonsmokers with the high-risk

genotype (PiMZ), and  $RR_{11} = 4.7$  in smokers with the high-risk genotype, using nonsmokers with the low-risk genotype as the reference group. These relative risks are consistent with an additive model (i.e., no interaction on an additive scale), since  $3.8 + 1.6 - 1 = 4.4 \approx 4.7$ .

In a limited sense, Models B, C, D, and E form an exhaustive set of possible models of gene-environment interaction. There are four possible combinations of genotype and exposure, in terms of their individual effects on disease risk: (a) an effect of the exposure but not the genotype, (b) an effect of the genotype but not the exposure, (c) an effect of neither the genotype nor the exposure, and (d) an effect of both the genotype and the exposure. If we add interaction to each of these four possibilities, the result is Models B through E, respectively.

Model E can encompass antagonism, or a joint effect lower than expected from a multiplicative or additive model, as shown by the  $RR_{11} = ?$  in Table 3. In Models B, C, and D, however, the interactions are synergistic. If antagonistic interactions are included, there are three additional models: the genotype *suppresses* the effect of the environmental exposure, but has no effect when acting alone (Model B'), the exposure *suppresses* the effect of the genotype, but has no effect when acting alone (Model C'), and risk is *reduced* in persons with both exposure and genotype, but neither has an effect when acting alone (Model D'). With increasing emphasis on research on chemopreventive agents, the importance of these models is likely to increase.

As an illustration of the application of these models to disorders with complex genetic influences, consider the joint effects of genetic susceptibility and heavy alcohol drinking on risk for epilepsy (recurrent unprovoked seizures). There is strong evidence for a genetic component in some forms of epilepsy [16]. However, the familial distribution does not follow a simple Mendelian pattern, suggesting that if susceptibility genes are important in some families, their phenotypic expression depends on unidentified environmental factors. Ng et al. [17] reported an association between heavy drinking and risk of a first unprovoked seizure (i.e., a seizure not associated with an acute structural or metabolic insult to the central nervous system): the odds ratio increased from 3-fold in persons who drank 51–100 g of ethanol per day to almost 20-fold in those who drank more than 200 g per day, compared with nondrinkers.

The relations between heavy drinking and a genetic susceptibility to epilepsy are unclear, and the models described above provide a framework for considering several alternatives. A mechanism consistent with Model A, for example, might involve a genotype that increased risk of alcoholism. In this case, the genotype would not affect risk of epilepsy directly, but would lead to increased levels of exposure to alcohol, and thus

to an indirect effect on epilepsy risk. With Model B, the effect of the susceptibility genotype would be to increase the brain's sensitivity to the effects of alcohol; the genotype would then have no effect on risk for epilepsy in nondrinkers. With Model C, the susceptibility genotype would raise risk for epilepsy regardless of drinking behavior. Heavy drinking would raise risk further in those with the genotype, but would have no effect in those without the genotype. With Model D, the effect of the susceptibility genotype would be restricted to heavy drinkers, and at the same time, drinking would have no effect on epilepsy risk in those without the genotype. With Model E, the genotype and heavy drinking would each affect risk in the absence of the other, and the combined effects might fit a multiplicative model, an additive model, or neither.

### TESTING THE MODELS

In order to test the models, individuals must be classified by presence or absence of both the exposure and the high-risk genotype. Classification according to exposure histories is relatively easy, using conventional epidemiologic methods such as physical measurements, interviews, medical record reviews, etc. However, one complication relevant to family studies is that measurement of exposures is easier for *proband*s (i.e., the affected or unaffected persons who lead to ascertainment of the families under study) than for relatives: some relatives will always be deceased or otherwise unavailable, and probands may be unable to provide accurate information about them [18].

Classification according to the high-risk genotype is

**TABLE 4**

Alternative Strategies for Measuring or Approximating the Genotype

| Measure  | Limitations   |
|--|---|
| 1. Identified susceptibility gene.   | May be used only when mutation has been identified; typing may be expensive; carriers may be rare.  |
| 2. Ecogenetics or candidate genes.   | Requires knowledge of disease pathogenesis.   |
| 3. Associated genetic marker.  | Associated markers seldom available and may be rare.  |
| 4. Linked genetic marker.  | Useful only when linkage is established and marker data are available for entire families.  |
| 5. Positive vs negative family history or relatives of cases vs relatives of controls. | Surrogate for genotype subject to misclassification; exposure history of cases and controls is associated with probability that they have the high-risk genotype. |
| 6. Monozygous vs dizygous cotwins of affected twins.                                   | Difficult to ascertain sufficient numbers; limited statistical power.   |

more problematic. Table 4 lists six possible strategies for measuring or approximating the genotype. The first strategy, testing for an identified susceptibility gene (e.g., BRCA1 in breast cancer or the APP gene in familial Alzheimer's disease) can be used only in those relatively rare, "ideal" instances when the disease-causing mutation has been identified. Further, even if the mutation has been identified, testing may be expensive, and carriers may be so rare that assembling sufficient numbers for a test of gene-environment interaction is impractical.

The second strategy involves measurement of candidate genes or ecogenetic markers (e.g., the cytochrome P450 CYP1A1 or glutathione transferase  $\mu$ , which play a role in activating/detoxifying carcinogens) [19]. Use of this approach requires some knowledge of disease pathogenesis, because the genes selected must have a plausible role in disease causation.

In the third strategy, a genetic marker that is associated with the disease is used as a surrogate for the high-risk genotype. This design can be used only in special situations where a population association between a disease and a marker allele has been demonstrated. For example, risk for Alzheimer's disease (AD) has consistently been found to be increased in individuals either heterozygous or homozygous for the  $\epsilon$ 4 allele of the apolipoprotein E (APOE) gene on chromosome 19 [20]. In a recent study, Mayeux et al. studied the combined effects of head injury and genetic susceptibility on risk for AD, using APOE genotypes as a measure of genetic susceptibility [21]. Using individuals with neither APOE- $\epsilon$ 4 nor head injury as the reference group, risk was increased 2-fold in individuals with APOE- $\epsilon$ 4 without head injury, and 10-fold in individuals with both APOE- $\epsilon$ 4 and head injury; risk was not increased in head-injured individuals without APOE- $\epsilon$ 4. These results are consistent with Model C.

The fourth strategy applies to situations where evidence has been obtained for linkage of a genetic marker to a disease susceptibility gene, but the actual disease-causing mutation has not yet been identified. In this case, generally no association will be found at the population level between the disease susceptibility allele and a specific allele at the marker locus. Hence a different allele at the marker locus will be found to be segregating with the disease susceptibility allele in different families. In families with confirmed evidence for linkage, however, the marker data for the family can be used to assign to each family member a probability that he or she carries the disease susceptibility allele. These probabilities (based on marker sharing *within* families) can then be grouped across families to explore the effects of environmental risk factors in persons with high vs low probabilities of carrying the susceptibility gene.

The fifth strategy involves use of family history data as a surrogate for the genotype. In a case-control

study, both cases and controls can be classified by presence or absence of a family history (defined as, say,  $\geq 1$  affected first-degree relative), and those with a positive family history can be assumed to be more likely to have a genetic susceptibility than those without. This strategy should be used with caution. First, it is subject to great misclassification. The family history may be positive in persons without a high-risk genotype (e.g., because of phenocopies) and negative in those with a high-risk genotype (e.g., because of competing causes of death, small family size, chance, etc.). Second, the probability of a “positive family history” varies with the number of relatives per family, the genetic distance of included relatives to the proband, and the years-at-risk of the relatives. If these factors vary between cases and controls, serious confounding may result.

An alternative method of using family history data involves comparison of *relatives* of cases (assumed more likely to have the high-risk genotype) and *relatives* of controls (assumed less likely to have the high-risk genotype). The disease experience of the relatives is “reconstructed” from birth until current age at the time of study or age at death. With this method, called the “reconstructed cohort” approach by Susser and Susser [18], differences between cases and controls in family size are controlled for, because the denominator is the number of relatives (rather than the number of cases and controls). Also, survival analysis can be used to control for the effects of variability in years-at-risk of the relatives [21], and the analysis can easily be stratified on important attributes of the relatives, such as gender and relationship to the proband.

There are two important problems with this approach, however. First, since the populations under study are the *relatives* of the cases and controls, exposure histories of the relatives must be obtained, and this, as noted above, is often difficult. Second, the probability that an individual carries a high-risk genotype is related to his or her disease and exposure status, and this relation depends on the underlying model of gene-environment interaction. Thus the degree of difference between relatives of cases and relatives of controls, in terms of the probability that they carry a high-risk genotype, may depend on the exposure status of the cases and controls. To illustrate this point, we consider the expectations from four hypothetical examples, derived from simple algebra:

**Example 1:** The genotype does not affect disease risk in unexposed persons (e.g., Models B and D). In this case, unexposed cases and unexposed controls will not differ in terms of the probability of having the high-risk genotype.

**Example 2:** The *risk difference* between exposed and unexposed persons is the same regardless of genotype (i.e., no interaction on an additive scale). In this case, the probability that an affected person has the high-risk genotype is greater if he or she was unexposed

than if he or she was exposed. (This might seem intuitively obvious—we might expect, for example, that a nonsmoker who develops lung cancer is more likely to be genetically susceptible than a smoker who develops lung cancer. However, it turns out that this expectation is restricted to the additive model.)

**Example 3:** The *relative risk* of disease in exposed vs unexposed persons is the same regardless of genotype (i.e., no interaction on a *multiplicative* scale). In this case, the probability that an affected person has the high-risk genotype is the same, regardless of exposure status.

**Example 4:** The relative risk of disease in exposed vs unexposed persons is greater in the presence than in the absence of the high-risk genotype ( $r_{11}/r_{01} > r_{10}/r_{00}$ ) (i.e., synergism on a multiplicative scale). In this case, an affected person will be *more* likely to have the genotype if he or she was exposed than if unexposed.

The expectations summarized above can be used as a test of the additive and multiplicative models in family studies. For example, we found that the degree of increased risk of epilepsy was greater in relatives of probands with idiopathic/cryptogenic epilepsy (i.e., epilepsy occurring in the absence of an identified environmental insult) than in relatives of those with remote symptomatic epilepsy (associated with an identified environmental insult to the central nervous system such as severe head trauma or stroke) [23]. The greater risk in relatives of probands with idiopathic/cryptogenic epilepsy suggests that these “unexposed” probands were more likely to be genetically susceptible than were the “exposed” probands. This result is inconsistent with the multiplicative model, which predicts that exposed and unexposed probands would have the same probability of carrying the high-risk genotype.

Finally, twin data can be used to approximate the genotype. Consider a sample of monozygotic (MZ) and dizygotic (DZ) twins in which one member of each pair is affected (i.e., a proband), and exposure data are available for both cotwins. Assuming there is no difference between MZ and DZ twins in terms of prevalence of the exposure, prevalence of the genotype, or disease risk conditional on genotype and exposure, no difference is expected between MZ and DZ twin *proband*s in the probability of having the high-risk genotype. However, prevalence of the genotype is expected to be higher in *cotwins* of MZ than DZ probands. Thus, associations between disease and exposure can be investigated in MZ and DZ cotwins, in order to evaluate gene-environment interaction. However, this approach has limited statistical power [24].

## IMPORTANCE

Gene-environment interaction can obscure both environmental effects (which may be evident only in genetically susceptible persons) and genetic effects

(which may be evident only in those with appropriate exposure histories). Thus, study of gene–environment interaction is important for improving accuracy and precision in the assessment of both genetic and environmental influences. An understanding of gene–environment interaction also has important implications for public health. It aids in predicting disease rates and provides a basis for well-informed recommendations for disease prevention.

#### ACKNOWLEDGMENT

I am grateful to Giancarlo Logroscino, M.D., for many helpful discussions of these ideas.

#### REFERENCES

1. Khoury MJ, Beaty TH, Cohen BH. Fundamentals of genetic epidemiology. New York: Oxford Univ. Press, 1993.
2. Rothman KJ. Interactions between causes. In: Modern epidemiology. Boston: Little, Brown, 1986:311–26.
3. Koopman JS. Causal models and sources of interaction. *Am J Epidemiol* 1977;106:439–44.
4. Kupper LL, Hogan MD. Interaction in epidemiologic studies. *Am J Epidemiol* 1978;108:447–53.
5. Walter SD, Holford TR. Additive, multiplicative, and other models for disease risks. *Am J Epidemiol* 1978;108:341–6.
6. Blot WJ, Day NE. Synergism and interaction: are they equivalent? [letter] *Am J Epidemiol* 1979;110:99–100.
7. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol* 1980;112:467–70.
8. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int J Epidemiol* 1981;10:383–7.
9. Miettinen OS. Causal and preventive interdependence: elementary principles. *Scand J Work Environ Health* 1982;8:159–68.
10. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. *Stat Med* 1983;2:243–51.
11. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scand J Work Environ Health* 1988;14:125–9.
12. Pierce N. Analytical implications of epidemiological concepts of interaction. *Int J Epidemiol* 1989;18:976–80.
13. Ottman R. An epidemiologic approach to gene–environment interaction. *Genet Epidemiol* 1990;7:177–85.
14. Susser M. Causal thinking in the health sciences: concepts and strategies in epidemiology. New York: Oxford Univ. Press, 1973.
15. Khoury MJ, Beaty TH, Newill CA, Bryant S, Cohen BH. Genetic–environmental interactions in chronic airways obstruction. *Int J Epidemiol* 1986;15:65–72.
16. Ryan SG. Partial epilepsy: chinks in the armour. *Nat Genet* 1995;10:4–6.
17. Ng SKC, Hauser WA, Brust JCM, Susser M. Alcohol consumption and withdrawal in new-onset seizures. *N Engl J Med* 1988;319:666–73.
18. Susser E, Susser M. Familial aggregation studies: a note on their epidemiologic properties. *Am J Epidemiol* 1989;129:23–30.
19. Perera FP, Whyatt R. Biomarkers and molecular epidemiology in mutation/cancer research. *Mutat Res* 1994;313:117–29.
20. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele  $\epsilon 4$  with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993;43:1467–72.
21. Mayeux R, Ottman R, Maestre G, et al. Synergistic effects of traumatic head injury and apolipoprotein- $\epsilon 4$  in patients with Alzheimer's disease. *Neurology* 1995;45:555–7.
22. Chase GA, Folstein M, Breitner J, Beaty T, Self S. The use of lifetables and survival analysis in testing genetic models, with an application to Alzheimer's disease. *Am J Epidemiol* 1983;117:590–7.
23. Ottman R, Annegers JF, Risch N, Hauser WA, Susser M. Relations of genetic and environmental factors in the etiology of epilepsy. *Ann Neurol* 1996;39:442–9.
24. Ottman R. Epidemiologic analysis of gene–environment interaction in twins. *Genet Epidemiol* 1994;11:75–86.