Commentary

Interaction and Exposure Modification: Are We Asking the Right Questions?

Clarice R. Weinberg*

* Correspondence to Dr. Clarice R. Weinberg, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709 (e-mail: weinber2@niehs.nih.gov).

Initially submitted August 8, 2011; accepted for publication August 25, 2011.

Most diseases arise not purely through genetic abnormalities nor purely through environmental causes, but as "complex" conditions brought about by the combined effects of genetic susceptibility factors, nongenetic experiences and exposures, and bad luck. Finding simple models capable of both characterizing such joint effects and providing new insight into pathogenesis remains an ongoing challenge in etiologic epidemiology. Additive null models can capture certain pure forms of independent etiologic effects in studies of rare conditions and can be useful for predicting possible effects of interventions. The concept of exposure modification is here proposed as useful, particularly in thinking about biologic interactions between exposures and genetic variants. Openness to parsimonious joint models and the insights they can provide is key to advancing our understanding of etiology.

attributable risk; effect modification; interaction; synergy

Abbreviation: PKU, phenylketonuria.

Editor's note: An invited commentary on this commentary appears on page 606.

Every epidemiologist knows that interactions among risk factors can be important to etiology, but arguments persist about how it should be detected, modeled, and interpreted. I am convinced that some of the confusion arises because we get needlessly stuck on the wrong questions. Here are some of the most stubborn wrong questions:

1. How much disease is due to environmental versus genetic factors?
2. Should we be using additive or multiplicative models when assessing gene-by-environment interaction?
3. Should our concern for public health relevance determine our choice of models?

I see these as ill-posed questions because they each rest on faulty premises.

Let's consider the first. Christiani recently wrote a piece about gene-by-environment interaction for the New England Journal of Medicine (NEJM). He stated, “Epidemiologists have long known that for most cancers, environmental factors have high attributable risks (as high as 85 to 95% in Western populations), …” (1, p. 792). The typical NEJM reader will assume that such a high attributable risk for environmental factors leaves little (perhaps 5%–15%) remaining for non-environmental causes. They will not appreciate the subtlety that risk attribution is not a 100% total that can be sliced up in this way. The notion that phenylketonuria (PKU), for example, is 100% attributable to the environment (phenylalanine in the diet) and 100% to genetics (one particular variant gene) is at first jarring, even to an epidemiologist. Although PKU is an extreme example, attributable risks for complex diseases typically do sum to more than 100%. If there are multiple overlapping sets of sufficient causes (2), the sum can even be more than 200%. Although useful in the context of risk reduction, attributable risk continues to be one of the most widely abused and misunderstood concepts in epidemiology.

What about the second issue? Should we be relying on additive models or multiplicative models for identifying factors that “interact”? What investigators (including toxicologists) typically want to mean by “interaction” is that the effect of 2 factors together is greater (synergism) or less (antagonism) when they occur jointly than what would have been expected on the basis of their separate effects, for example, when genotype has certain effects among the
unexposed and exposure has certain effects among those without the genetic variant. To assess interaction then, we need some method for combining the factors’ separate effects to enable us to form a “null” no-interaction model for what would be expected under independent joint action. Should our notion for “no interaction” be based on additivity? There are epidemiologists who vehemently advocate for additive approaches, to the point where they assert that the use of multiplicative models needs to be defended and even apologized for. It was this kind of rhetoric at the Third North American Congress of Epidemiology held from June 21 to 24, 2011, in Montreal that motivated this commentary.

I myself have argued for use of an additive no-interaction null model for a rare disease, and it is true that additivity sometimes captures a simple and strict meaning for independence of effect (3). Let me explain. If an outcome is rare and 2 causative factors act through completely separate causal pathways, then their joint effect would be expected to be additive. Suppose 2 hunters with very bad aim are both independently shooting at the same duck. Probabilistically, the risk the duck faces while flying over them is the sum of 2 (very small) risks. (This assumes, of course, that their aim is not so bad that the hunters might shoot each other).

On the other hand, if the hunters are both pretty good shots, then the duck’s risk is instead multiplicative in the complement; that is, the probability that the duck escapes shots, then the duck’s risk is instead multiplicative in the pathways, then their joint effect would be expected to be about a 6-fold increase in risk for those subject to both factors. We see that such separate effects, which some would regard as biologically independent, do not lead to an additive formulation for the combined effect.

In my view, the question worth asking is the following: What can we learn about the etiology of a disease by carefully characterizing the joint effects of 2 (or more) causative factors?

Many of the most interesting gene-by-environment interactions that have been discovered to date involve what I think of as “exposure modification.” Good examples abound (7–11). These exposure modifiers are gene variants that act by influencing the function of some biologic pathway, for example, a detoxification pathway for products of cigarette smoking, pesticides, or a teratogen, or a metabolic pathway involving a dietary component such as folate or choline. Genetic variants often influence response to a given level of exposure by changing how (or how rapidly) that exposure (or some downstream product) is absorbed, processed, or metabolized, so that the dose of a toxic or beneficial product that is experienced by the fetus or a target organ (e.g., the bladder) is systematically altered in genetically susceptible individuals. Exposure modification is a clear instance where I think most of us would agree there is important biologic interaction.

We can formalize this algebraically. Perhaps the simplest example of exposure modification would be where the ultimate risk-conferring (but usually unmeasurable) dose experienced by the target organ or the fetus, say D, is a function of the (continuous) exposure E, but that function depends on the genotype, for example, \( D = f_G(E) \), where \( G \) is the genotype corresponding to one or more genetic exposure modifiers. PKU fits well into this framework, because in that syndrome a genetic abnormality influences clearance of what would normally be a beneficial amino acid, phenylalanine, resulting in neurotoxic levels. A simple example is where \( D = \beta \times E + \gamma \times G \times E = (\beta + \gamma \times G) \times E \). Here \( G \) is coded as 0/1 and serves as a marker for a risk-relevant genotype. This model reflects a scenario where \( G \) modifies the absorption or metabolic fate of exposure \( E \). If risk is logit-linear in the dose, \( D \), then this exposure modification by \( G \) will instead produce a departure from the multiplicative joint effects of \( E \) and \( G \). Notice that there may be no “main” effect of \( G \).

We can make this slightly more complicated by supposing that individuals are exposed to unmeasured xenobiotic chemicals that are not related to smoking but are handled by the same detoxification pathway. Now the dose, \( D \), might be \( D = \alpha + \beta_G \times E + \delta \times G \). In this example, \( G \) has an effect even in nonsmokers. Under a logit-linear model for risk as
a function of \( D \), we now have both a main effect of \( G \) and departure from a multiplicative model for the joint effect of \( E \) and \( G \). If, instead, the risk for disease is not logit-linear but a simple linear function of \( D \), then the joint effect would be linear in \( E \) and \( G \), but with a term for a product of the 2, indicating departure from additivity. It is easy to see that very complex relations can arise when risk depends on both a dose-response model and a linked exposure-modification model.

Exposure modifiers need not be genetic. For example, the use of filter facemasks during the spraying of pesticides and the use of gas scavenging equipment in dental offices that use nitrous oxide (laughing gas) can serve as important exposure modifiers. After the tsunami in Japan had produced a nuclear reactor emergency, the iodine distributed to children served as an exposure modifier by protecting thyroids from radiation by competitively docking onto receptors in their thyroids. Age can also be an exposure modifier. For example, the aging gastrointestinal tract can modify the effective dose due to dietary intake of vitamin \( B_{12} \), by inhibiting absorption of the nutrient in the intestine, with the unfortunate result being pernicious anemia.

Does the choice between use of a null model that is additive versus multiplicative actually matter much? One point to keep in mind in the choice between additive and multiplicative models is that if both factors have small effects, then the null additive model is very close to being the same as the null multiplicative model, implying that the choice is not really important. To see this, consider 2 risk factors that each confer a relative risk of 1.1. A multiplicative model predicts that their joint effect should be 1.1 squared, or 1.21. An additive model predicts instead that their joint effect should be \( 1.1 + 1.1 - 1 = 1.20 \). Thus, the null models for additivity and multiplicativity are almost identical. The alternative models that include a product term for the 2 factors (assuming they are binary) are de facto the same, one being a reparameterization of the other. It follows that testing for interaction by, for example, assessing the improvement in fit (or calculating a confidence interval for either interaction parameter) will deliver an inference that is virtually the same, regardless of whether the base model is additive or multiplicative. In practice, the choice of an additive versus a multiplicative null model matters only when at least one of the main effects is large, as in the case of lung cancer in relation to smoking with or without asbestos.

Nonetheless, presuming that we care about public health then, shouldn’t we be using additive models? Don’t risk differences do a better job of capturing the public health impact of a factor than would risk ratios? Yes, in the following sense: Ultimately we want to be able to calculate how many cases can be prevented by certain interventions, and those questions always involve risk differences. However, when developing etiologic models, the data should have a major voice. My view is that we can impact public health best by trying to understand the biologic basis for interaction; finding a parsimonious statistical model that fits the data well can sometimes (but not always) help inform that understanding, if we communicate well with the basic scientists who understand the relevant biologic pathways. Possible interventions to remove a causative factor may have larger impacts if we can target particularly susceptible populations, and a carefully constructed model for joint effects can be a useful step in building risk models for comparing those potential impacts.

What about protective factors? Wacholder et al. (12) described in a recent commentary how protective factors might display multiplicative effects when they act jointly with causative factors. A paper in the Journal of the National Cancer Institute (13) had reported that a protective genetic variant (rs3814113 (C allele single-nucleotide polymorphism) at 9p22.2) showed a nearly perfect multiplicative effect in women who carried deleterious variants in the genes \( BRCA1 \) and \( BRCA2 \), with the same hazard ratio for ovarian cancer of 0.80. Surprisingly, the hazard ratio in carriers of \( BRCA1 \) and \( BRCA2 \) was remarkably close to that seen in noncarriers. Because \( BRCA1 \) and \( BRCA2 \) mutations are strongly causative for ovarian cancer, the data were statistically incompatible with an additive model for joint effects. One implication of similar multiplicative effects in this context is that the absolute effect of the protective genetic variant is much greater (in terms of risk difference) in carriers than in noncarriers of \( BRCA1 \) or \( BRCA2 \), because the latter both confer such a high absolute risk compared with background. Of course, one could fit an additive model but make it fit the data just as well by including an interaction term. Is there, however, something that can be learned by noticing that for these factors the simple multiplicative formulation works so well?

In reflecting on possible implications of this multiplicative model, Wacholder et al. (12) proposed that independence for a protective and a causative factor could produce multiplicative joint effects if the protective factor acts through a separate pathway. Examples of such effects would be enhanced systems for repairing errors in DNA replication or mechanisms for stimulating the immune system (or the cells themselves) to recognize and eliminate cells that have undergone carcinogenic transformation.

In summary, most complex disease arises through the joint effects of exposures and life experiences—all acting together in a context of genetic susceptibility and bad luck, for example, where a random somatic mutation goes unrepaired in a key gene in a single cell and leads to cancer. As epidemiologists studying causation, we should stop arguing about additive versus multiplicative effects and biologic versus statistical interaction and instead set out to find illuminating ways to model and understand joint effects, to better understand the causal processes and to better identify individuals who should be targeted for screening to have maximal benefit. When this work is done with care, we may even be able to find points in the pathogenic process where preventive interventions could change an outcome.

ACKNOWLEDGMENTS

Author affiliation: National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina.

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences, under project number ZO1 ES040007.

Am J Epidemiol. 2012;175(7):602–605
Thanks are due to Drs. D. Umbach, S. London, and P. Howards for their comments on an earlier draft.
Conflict of interest: none declared.

REFERENCES