Joint Effect of Genes and Environment Distorted by Selection Biases: Implications for Hospital-based Case-Control Studies

Sholom Wacholder,[1] Nilanjan Chatterjee, and Patricia Hartge

Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland 20892-7244

Abstract

The hospital-based case-control design enhances the response rates in studies that require the collection of biological samples from all of the participants. There are simple, established criteria for selecting controls so as to estimate the effect of a single factor without bias, but the analogous requirements for assessing an interaction are less clear. We derive these conditions by calculating the potential bias from selecting controls who were admitted for treatment of diseases related to either or both of the exposures of interest, designated as a gene variant (*G*) and an environmental agent (*E*). There is no bias in the estimate of the effect of *E* when *G* is associated with the control condition, whether causally or because of confounding. There is no bias in estimating multiplicative interaction between *G* and *E* for the disease of interest when there is no multiplicative *G-E* interaction for the control disease, even when the control condition is caused by *G* or *E*; if a mixture of several control diseases are used, however, the absence of *G-E* interaction in each individual disease does not ensure a lack of overall bias when controls are pooled. Hospital control designs are much less robust for assessing additive interaction. We conclude that the ideal control disease in a hospital-based study of gene-environment interaction is not caused by either *G* or *E* and that choosing controls from several conditions to act as a combined control group is a useful strategy. This formulation extends to the general problem of distortion of joint effects from selection biases or confounding.

Introduction

Epidemiologists are using an increasing variety of traditional, modified, and novel designs for epidemiological studies, designed to assess the effects of genes and environmental factors jointly. The hospital-based case-control study (1–5) has always been attractive for studies of disease like brain cancer with referral patterns that make it difficult to characterize the underlying study base (6). For molecular epidemiology studies, hospitalized controls facilitate the collection of blood and other biospecimens and may be more willing to participate than population controls. But who should be selected as a control in a study of gene-environment interaction? The selection of patients admitted to the hospital because of a disease that is either caused or prevented by an exposure of interest biases the estimate of the effect of the exposure (3). The analogous requirements for unbiased estimation of an interaction have not been established. For example, in a case-control study of lung cancer designed to assess how a gene modifies the effect of smoking on the risk of lung cancer, can controls be selected from patients admitted because of a smoking-related condition? By considering the components of the gene-environment interaction parameter, we trace the effects of retaining or dropping the exclusion rule. We demonstrate the impact of alternative exclusion rules and establish the criteria for determining which strategies are valid for assessing additive or multiplicative interaction. Finally, we show how the algebra we develop applies generally to the problems of assessing an interaction in the presence of selection bias or confounding in unmatched case-control studies.

Structure of diagnosis related selection bias

In general, valid inference from case-control studies dictates that the distribution of the study factors in controls must reflect the distribution of those factors in the study base from which the cases arise (6). In a hospital-based study, a person with the case-defining condition who would appear at the study hospital and be enrolled in the study as a case would also appear at the study hospital and be enrolled in the study as a control if he had the control-defining disease and *vice versa*; (3) throughout, we assume that this requirement is met. Second, when the goal is to estimate the effect of a single exposure on the risk of disease, the exposure of interest must be unrelated to the risk of developing the condition that brought the control to the hospital (3). But what are the requirements for valid hospital controls when the concern is the estimation of an interaction parameter or the effect of an exposure in a stratum defined by one or more genotypes?

This work is motivated by the planning of a case-control study of lung cancer. Our main interests lie in estimating the effects of genetic factors, such as alleles of a DNA repair gene, alone and jointly with smoking, on the risk of lung cancer; we are not particularly interested in estimating the effect of smoking on the risk of lung cancer. Controls for this study would be asked to respond to a questionnaire and provide blood and other biospecimens. Given the difficulties of choosing appropriate population-based controls with a high response rate, we considered using hospital controls instead, with the hope that confined patients would be more willing to participate. Before we could decide which conditions from which to choose hospital controls, we needed to determine the impact of choosing controls hospitalized for a disease related to smoking, the main

Received 7/27/01; revised 5/22/02; accepted 5/31/02.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
[1] To whom requests for reprints should be addressed, at Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, EPS 8046, 6120 Executive Boulevard, Bethesda, MD 20892-7244.

Table 1 Odds ratio of hospitalization for CVD and lung cancer according to smoking (E) and DNA repair gene (G)[a]

	Odds ratios			
	$E+$ (smokers)		$E-$ (nonsmokers)	
	CVD	Lung cancer	CVD (source of improper controls)	Lung cancer
$G+$ (abnormal variant of genotype for DNA repair)	ABC	$\alpha\beta\gamma$	B	β
$G-$ (normal variant of genotype for DNA repair)	A	α	1	1

[a] E, exposure to an environmental agent such as smoking; G, exposure to a gene variant such as a DNA repair gene. These results apply directly to any environmental agent E and gene variant G.

Table 2 Interpretation of parameters[a,b]

Parameter (label)	Effect	Algebra	Estimate with proper controls	Estimate with improper (CVD) controls
RR[c] of E in $G-$	Effect of ***E*** specific to $G-$	R(**E+**,$G-$)/R(**E−**,$G-$)	α	α/A
RR of E in $G+$	Effect of ***E*** specific to $G+$	R(**E+**,$G+$)/R(**E−**,$G+$)	$\alpha\gamma$	$(\alpha\gamma)/(AC)$
RR of E adjusting for G[d]	Summary effect of E over G		α	α/A
RR of G in $E-$	Effect of **G** specific to $E-$	R($E-$,**G+**)/R($E-$,**G−**)	β	β/B
RR of G in $E+$	Effect of **G** specific to $E+$	R($E+$,**G+**)/R($E+$,**G−**)	$\beta\gamma$	$(\beta\gamma)/(BC)$
RR of G adjusting for E[d]	Summary effect of G over E		β	β/B
Additive interaction	Ratio of difference between stratum-specific risk differences and $R(E-,G-)$	{[R(**E+**,$G+$) − R(**E−**,$G+$)] − [R(**E+**,$G-$) − R(**E−**,$G-$)]}/R($E-$,$G-$) = {[R($E+$,**G+**) − R($E+$,**G−**)] − [R($E-$,**G+**) − R($E-$,**G−**)]}/R($E-$,$G-$)	$1-\alpha-\beta+\alpha\beta\gamma$	$1-\beta/B-\alpha/A-+\alpha\beta\gamma/ABC$
Multiplicative interaction	Ratio of stratum-specific RRs	[R(**E+**,$G+$)/R(**E−**,$G+$)]/[R(**E+**,$G-$)/R(**E−**,$G-$)] = [R($E+$,**G+**)/R($E+$,**G−**)]/[R($E-$,**G+**)/R($E-$,**G−**)]	$\alpha\gamma/\alpha=\beta\gamma/\beta=\gamma$	$[(\alpha\gamma)/(AC)]/[\alpha/A]=[(\beta\gamma)/(BC)]/[\beta/B]=\gamma/C$

[a] $R(E+,G+)$, the absolute risk of disease in those exposed to E and to G; $R(E+,G-)$, $R(E-,G+)$ and $R(E-,G-)$ are defined analogously.
[b] The arguments in bold-face type are used to emphasize the contrast being made and do not change the interpretation.
[c] RR, is the relative risk (odds ratio) for the specified level of environmental agent E and genetic variant G relative to those unexposed to E and G.
[d] If there is no G-E interaction for lung cancer or for CVD, *i.e.*, $\gamma = C = 1$.

risk factor for lung cancer but not the subject of this study, and possibly to the same genes that might be investigated for the study of lung cancer. However, we were unable to determine from the literature what the impact would be of choosing CVD[2] controls, given that smoking affects the risk of CVD, and that a DNA repair gene suspected of being related to lung cancer might be truly related to CVD.

We, therefore, decided to address this question ourselves. For simplicity, we categorize smoking and genotype into 2 levels, $E+$ or $E-$ for smokers and nonsmokers and $G+$ or $G-$ for those who do or do not carry the allele of interest. For our main points, we present a simple numerical example first to demonstrate our point. To generalize, we show the results abstractly as well. We refer to a hospital control group, such as those diagnosed with CVD, as "improper," if it does not meet the stringent requirement that E and G not be related to the risk of being hospitalized with the control condition.

Results and Examples

We take a two-track approach to make our points clear. We establish the general principles, and we show numerical examples for our most important points. We use the notation in Table 1 to present general models for the joint effects of an environmental exposure and a genetic variant on the odds ratio using Greek letters when the outcome is lung cancer and Roman letters when the outcome is hospitalization for CVD. That is, the odds ratios for lung cancer among those exposed to smoking alone or the gene alone are represented by α and β, respectively, and for the doubly exposed by $\alpha\beta\gamma$. Formally, α and β are the smoking and gene effects at the baseline level of the gene and smoking, respectively, and γ is the multiplicative interaction parameter for lung cancer. The variables A, B and C for CVD are analogous to α, β and γ for lung cancer.

From these components, one can derive the expected values of the estimates of the effects of E and G and multiplicative or additive interaction parameters from a properly designed and conducted case-control study with lung cancer cases and CVD controls. For example, the odds ratio, equivalent for rare disease to risk ratio. RR for $E+,G-$ relative to the baseline $E-,G-$ will be distorted if $A > 1$ because smokers will be overrepresented in CVD controls compared with the study base. Table 2 uses the components of Table 1 to compare the expected values of the estimates from a study using proper controls with one using CVD controls.

Exposure and Gene Effects. The first two rows of Table 2 confirm the established principle that the exposure effect can be estimated without bias if the exposure is unrelated to the control disease. The estimated odds ratio of E or G using CVD controls will be correct when $A = 1$ or $B = 1$ in the absence of

[2] The abbreviation used is: CVD, cardiovascular disease.

Table 3 Hypothetical example in which the effect of *G* is estimated without bias even though the control disease is caused by *E*

In this hypothetical example, *E*+ triples the risk of CVD but *G* has no effect. If one used CVD controls with lung cancer cases, the odds ratio for *E* on lung cancer is biased by a factor of one-third; the odds ratio would be 5 instead of 15. But the effect of *G* is 3 both when proper controls and when CVD controls are used.

	E−,*G*−	*E*+,*G*−	*E*−,*G*+	*E*+,*G*+
Proper controls	100	100	100	100
Improper (CVD) controls	100	300	100	300
Cases (lung cancer)	100	1500	300	4500
Odds ratio: cases *vs.* proper controls	Reference	15	3	45
Odds ratio: cases *vs.* improper controls	Reference	5	3	15
Odds ratio: improper *vs.* proper controls	Reference	3	1	3

Table 4 Hypothetical example in which the effect of *G* is estimated without bias even though the control disease is caused by *E* and there is confounding of the effect of *G* by *E*

In this example, *G* and *E* are associated, *E* increases the risk by a factor of 2 in both *G*− and *G*+, and the effect of *G* is confounded by *E*. The odds ratio for *G* when *E* is ignored is 3 with proper controls and 3.2 with improper controls, instead of the true 2. Nevertheless, the effect of *G* is estimated correctly in subgroups defined by *E*+ and *E*− and, therefore, in adjusted analyses.

	E−,*G*−	*E*+,*G*−	*E*−,*G*+	*E*+,*G*+
Proper controls	100	200	200	100
Improper (CVD) controls	100	200	400	200
Cases (lung cancer)	100	400	800	800
Odds ratio: cases *vs.* proper controls	Reference	2	4	8
Odds ratio: cases *vs.* improper controls	Reference	2	2	4
Odds ratio: improper *vs.* proper controls	Reference	1	2	2

multiplicative interaction ($C = 1$). The asymmetry between *G*+ and *G*− and between *E*+ and *E*− when $C \neq 1$ are artifacts of the arbitrary definitions of + and −. Table 3 shows a hypothetical example in which the use of improper CVD controls for lung cancer cases produces bias in the estimate of the effect of smoking but not necessarily of the genetic variant.

Confounding. The estimates of the effects of *E* are unbiased at each level of *G* even when *G* is related to the risk of CVD ($B \neq 1$) when *E* is unrelated to the control disease ($A = 1$) and there is no multiplicative interaction for the control disease ($C = 1$). It follows that under these conditions the *G*-adjusted estimate is also unbiased because it is a weighted average of the unbiased estimates at the two levels of *G* (7). That is, if *G* and *E* are associated (in the controls or in the study base), then *G* confounds the crude estimate of the exposure effect when we choose a control group related to *G*, but the estimate of the effect of *E* can be deconfounded completely by adjusting for *G* (3). Of course, if there is a multiplicative *G*-*E* interaction, the weighted average will depend on γ.

It is noteworthy that if *E* and *G* are associated in the study base, one needs to adjust for *G* if it is a risk factor for either the case or the control diseases. For instance, one could consider using traffic-accident controls for a study of smoking and lung cancer under the assumption that alcohol, but not smoking, is an independent risk factor for accidents. Indeed, there will be more smokers among these controls than in a proper control group; standard adjustment in the analysis for those risk factors for accidents that are correlated with smoking completely eliminates the bias. As ever, unmeasured risk factors can induce bias. In particular, that an unmeasured factor X associated with exposure can confound the estimate of effect not only if X is a risk factor for the study disease but also if it is a risk factor for the control disease.

Table 4 illustrates the situation in which the adjusted estimates of effect of exposure are unbiased even though the crude effect would be biased when using improper controls.

Multiplicative or Additive Interaction. For unbiased estimation of multiplicative interaction, Table 2 shows that the requirement is no multiplicative interaction between *E* and *G* for the control disease ($C = 1$). Even if *E*, or *G*, or both cause or prevent CVD, the multiplicative interaction still can be estimated without bias. For additive interaction, often of greater interest, the requirement is more stringent, namely $A = B = C = 1$. Thus, because smoking increases the risk of CVD, it would be impossible to estimate the effect of an additive interaction between smoking and any gene on lung cancer risk unless *A* were known, even if there were no gene effect or interaction on CVD risk.

Table 5 shows the situation in which the overall and the stratum-specific *G* and *E* effects are both distorted, and yet the multiplicative interaction term is estimated correctly. In Table 6, the overall and stratum-specific *G* and *E* effects are both distorted, as is the multiplicative interaction. The additive interactions, which we define as the difference between the differences of the odds ratios for *E* in *G*+ and in *G*−, is also biased.

More Than One Disease Used to Define Patients Eligible to Be Controls. It is often noted that choosing hospital controls from more than one disease can provide protection against the possibility of a major bias in the estimation of the effects of *E* and *G*. A single control disease might indeed be related to the factor of interest. The bias would be mitigated if other diseases are used as controls as well, because the factor might be unrelated to disease or related in the opposite direction; however, with multiple diseases, there is more chance of at least some bias because each disease could be related to exposure, and it is unlikely that the net effect will entirely cancel. Of course, one or more sets of controls can be excluded from an analysis when their condition is found to be related to the exposure under study.

One unexpected consequence of the algebra of bias induction in the *G*-*E* interaction context is the effect of pooling

Table 5 Hypothetical example in which the control condition is caused independently by *G* and *E* in a multiplicative model, yet multiplicative interaction is estimated without bias

In this hypothetical example, both *E* and *G* increase the odds of the disease used for improper controls by 2-fold according to a multiplicative model. Nonetheless, the multiplicative interaction is estimated correctly as 1, even with improper controls. However, the additive interaction is estimated with bias. With improper controls, additive interaction parameter is estimated as $(2.25 - 1.5) - (1.5 - 1) = 0.25$, instead of $4 = (9 - 3) - (3 - 1)$. Because the control disease is related to both *E* and *G*, the effects of *E* and *G* are estimated with bias when using improper controls.

	E−,*G*−	*E*+,*G*−	*E*−,*G*+	*E*+,*G*+
Proper controls	100	100	100	100
Improper (CVD) controls	100	200	200	400
Cases	100	300	300	900
Odds ratio *vs.* proper controls	Reference	3	3	9
Odds ratio *vs.* improper controls	Reference	1.5	1.5	2.25
Odds ratio: improper *vs.* proper controls	Reference	2	2	4

Table 6 Hypothetical example in which the control condition is caused independently by *G* and *E* in an additive model, yet neither additive nor multiplicative interaction is estimated without bias

In this hypothetical example, both *E* and *G* increase the odds of the disease used for improper controls by a factor of 3 according to an additive model. Nonetheless, the additive interaction using proper controls is estimated as $-0.27 = (1.40 - 1.33) - (1.33 - 1)$. Not surprisingly, the multiplicative interaction is also estimated with bias either.

	E−,*G*−	*E*+,*G*−	*E*−,*G*+	*E*+,*G*+
Proper controls	100	100	100	100
Improper (CVD) controls	100	300	300	500
Cases (lung cancer)	100	400	400	700
Odds ratio: cases *vs.* proper controls	Reference	4	4	7
Odds ratio: cases *vs.* improper controls	Reference	1.3	1.3	1.4
Odds ratio: improper *vs.* proper controls	Reference	3	3	5

patients with a mixture of diseases into a single control group. Even if there is no multiplicative interaction for each of two control diseases, there is likely to be bias when estimating multiplicative interaction when the two are pooled in a single control group. Let us assume that the odds ratios for disease 1 in (*G*+,*E*−), (*G*−,*E*+) and (*G*+,*E*+) relative to (*G*−,*E*−) are 2, 4, and 8, and the odds ratios for disease 2 are 4, 2, and 8, respectively. With a combined control group consisting of equal numbers from disease 1 and disease 2, the odds ratios for the disease of interest will be 3, 3, and 8, respectively, no longer following a multiplicative pattern and, thereby, causing a violation of the requirement of no multiplicative interaction for the control series because *C* does not equal 1; most often the magnitude of the bias will be minor except when the magnitudes of the effects are large. Any average of the interaction estimates using each control series separately is unbiased, if the estimate from each individual series is unbiased; polytomous (8) logistic regression uses a weighting method that produces an efficient and unbiased estimator.

The Effects of Selection Bias and Confounding on the Estimates of Joint Effect of Two Factors in Case-Control Studies. Hospital controls are just one example of controls selected with potential bias. Our results extend not only to studies with hospital controls but to any situation in which cases or controls are selected in a biased manner with respect to a factor of interest. That is, Table 2 applies when the ratio in controls:cases of the odds of selection of a control with (*E*+,*G*−), (*E*−,*G*+), or (*E*−,*G*−) relative to (*E*−,−*G*) are *A*, *B*, and *ABC*, respectively, just as for the improper controls in Table 1. In fact, these results apply to the possible distortion of a two-way interaction estimate attributable to the confounding effects of an unmeasured third factor that is differential in cases and controls, possibly caused by a three-way interaction.

Discussion

This work merges the quantitative evaluation of selection bias with the measurement of the joint effects of two factors. We have specifically discussed the joint effects of a hereditary and an environmental or behavioral factor on disease, which motivated this work. Our results apply to the general problems of selection bias and confounding, beyond the focus on hospital-based studies in this paper and, equally, to interactions between two genes or between two environmental factors. The parameter estimates in Table 2 hold, for example, for a population-based case-control study in which the case ascertainment is complete and "improper controls" are a consequence of nonresponse in controls differential by a level of one or more factors. In principle, extensions to multiple categories or even to continuous forms of *E* and to third and higher-order interactions follow the same logic.

In considering the effects of selection bias, we have assumed the fulfillment of each of the standard case-control requirements, including those related to case and control ascertainment and selection, common catchment area, and equivalent exposure assessment. In addition, we assume, when appropriate, that special problems peculiar to hospital controls, such as Berkson's bias, do not have an important impact (3).

For concreteness, we discuss several important lessons from this work in the context of the use of hospital controls in studying the joint effects of *G* and *E*. First, there is no bias in the estimate of the multiplicative *G*-*E* interaction when there is no *G*-*E* interaction for the disease used for controls, even when the control condition is caused by either *G* or *E*. Similarly, even the effect of *E*, stratified on *G*, can be estimated without bias even when *G* causes the control diseases. The analogous statement holds when *E* and *G* are switched.

Second, bias can arise when assessing multiplicative in-

teraction using two or more control diseases, even if there is no multiplicative interaction in each control disease. Even if each of two control conditions, perhaps CVD and accidents, individually lead to no bias, a pooled control group can produce bias. Thus, the protection provided by the use of multiple diseases instead of only one for estimating the effects of one factor does not extend completely to studying multiplicative interaction in the situation when both factors are related to disease; the use of polytomous logistic regression (8) might alleviate this problem.

Third, the additive interaction effect is less robust against bias introduced when *E* or *G* is related to the control condition. To measure additive interaction of *G* and *E*, when *E* is smoking, in a study of lung cancer, one must use only those diseases that *G* neither causes nor prevents among either smokers or nonsmokers. For example, using bladder cancer controls can produce bias in assessing a gene-smoking additive interaction when studying lung cancer, even if *G* is unrelated to the control diseases. On the other hand, there would be no bias in assessing multiplicative interaction if there was a *G* effect on the risk of bladder cancer but no multiplicative interaction with smoking.

In hospital-based case-control studies, a strategy of using several control conditions, each of which is thought to be causally unrelated to any factor to be studied, seems advisable. If new evidence suggests that one of the control conditions is related to one of the factors of interest, it is an easy exercise to exclude those controls from analyses aimed at estimating the effect of an individual factor or interaction involving that condition. We cannot rely on our knowledge completely; *e.g.*, it might seem reasonable to use CVD controls for a study of Alzheimer's disease until one realizes that variants in the *APO-E* gene causes both (9, 10). Second, a sensitivity analysis, examining the effect of excluding each control set, also seems appropriate. An objective method could be devised to determine which disease groups to include for each hypothesis, with decisions-to-exclude based on how much of an outlier the exposure distribution is for the one group among all of the others. Third, sometimes the cases can be used in a case-only analysis to provide additional, although not fully independent, evidence.

How do hospital-based studies compare with available alternatives? Family-based designs are particularly useful for identifying and characterizing genetic effects but seldom are ideal for studying joint effects. Studies using siblings as controls are often infeasible for studies of diseases of old-age and often suffer from overmatching on genetic and environmental effects that aggregate within families, but they are very efficient for studies of interactions between rare alleles and environmental factors (11). Population-based case-control studies can be ideal in theory but can suffer from poor response rates and attendant biases. Case-only designs prohibit the estimation of the individual (12) effects of *E* or *G* or of additive interactions (1) without outside information. They permit a powerful test of multiplicative interaction between an environmental and a genetic factor but only under the assumption, impossible to verify directly, that the factors are independent in the study base (1, 13).

As explorations of gene-environment effect continue, investigators will develop new designs and modify old ones to increase efficiency. Subtle opportunities for bias can arise, as illustrated by the potentially biased estimation of the effects of genes among those exposed to the environmental variable, depending on the precise choice of eligible diagnoses. Nevertheless, the fundamental logic of hospital-based case-control design and the attendant control selection requirements hold in the setting of research on gene-environment interactions. The serious problems with hospital controls, particularly for additive interactions, must be considered against a background of other problematic control selection strategies. Alternatives, including sibling controls, population controls, and case-only designs, face their own challenges to validity or efficiency, including poor response rates, overmatching, and important assumptions of independence that are difficult to verify.

References

1. Lilienfeld, A. M., Pedersen, E., and Dowd, J. E. Cancer Epidemiology: Methods of Study. Baltimore, MD: The Johns Hopkins Press, 1967, pp. 69–79.

2. Hennekens, C. H., and Buring, J. E. Epidemiology in Medicine. Boston: Little, Brown and Company, 1987, pp. 137–139.

3. Wacholder, S., Silverman, D. T., McLaughlin, J. K., and Mandel, J. S. Selection of controls in case-control studies. II. Types of controls. Am. J. Epidemiol., *135:* 1029–1041, 1992.

4. Kelsey, J. L., Whittemore, A. S., Evans, A. S., and Thompson, W. D. Methods in Observational Epidemiology. New York: Oxford University Press, 1996, pp. 198–199; 201–205.

5. Rothman, K. J., and Greenland, S. Modern Epidemiology. Philadelphia: Lippincott-Raven Publishers, 1998, pp. 100.2; 107.

6. Wacholder, S., McLaughlin, J. K., Silverman, D. T., and Mandel, J. S. Selection of controls in case-control studies. I. Principles. Am. J. Epidemiol., *135:* 1019–1028, 1992.

7. Flanders, W. D., Boyle, C. A., and Boring, J. R. Bias associated with differential hospitalization rates in incident case-control studies. J. Clin. Epidemiol., *42:* 395–401, 1989.

8. Dubin, N., and Pasternack, B. S. Risk assessment for case-control subgroups by polychotomous logistic regression. Am. J. Epidemiol. *123:* 1101–1117, 1986.

9. Menzel, H. J., Kladetzky, R. G., and Assmann, G. Apolipoprotein *E* polymorphism and coronary artery disease. Arteriosclerosis, *3:* 310–315, 1983.

10. Small, G. W., Mazziotta, J. C., Collins, M. T., Baxter, L. R., Phelps, M. E., Mandelkern, M. A., Kaplan, A., La Rue, A., Adamson, C. F., Chang, L., *et al.* Apolipoprotein *E* type 4 allele and cerebral glucose metabolism in relatives at risk for familial Alzheimer disease. J. Am. Med. Assoc., *273:* 942–947, 1995.

11. Witte, J. S., Gauderman, W. J., and Thomas, D. C. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. Am. J. Epidemiol., *149:* 693–705, 1999.

12. Piegorsch, W. W., Weinberg, C. R., and Taylor, J. A. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. Stat. Med., *13:* 153–162, 1994.

13. Albert, P. S., Ratnasinghe, D., Tangrea, J., and Wacholder, S. Limitations of the case-only design for identifying gene-environment interactions. Am. J. Epidemiol., *154:* 687–693, 2001.

Joint Effect of Genes and Environment Distorted by Selection Biases: Implications for Hospital-based Case-Control Studies

Sholom Wacholder, Nilanjan Chatterjee and Patricia Hartge

Cancer Epidemiol Biomarkers Prev 2002;11:885-889.