Statistics in Medicine

Special Issue Paper

Received 20 December 2010, Accepted 1 July 2011 Published online 4 October 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4354

Sensitivity analysis for interactions under unmeasured confounding

Tyler J. VanderWeele,[a,b*†] Bhramar Mukherjee[c] and Jinbo Chen[d]

We develop a sensitivity analysis technique to assess the sensitivity of interaction analyses to unmeasured confounding. We give bias formulas for sensitivity analysis for interaction under unmeasured confounding on both additive and multiplicative scales. We provide simplified formulas in the case in which either one of the two factors does not interact with the unmeasured confounder in its effects on the outcome. An interesting consequence of the results is that if the two exposures of interest are independent (e.g., gene–environment independence), even under unmeasured confounding, if the estimate of the interaction is nonzero, then either there is a true interaction between the two factors or there is an interaction between one of the factors and the unmeasured confounder; an interaction must be present in either scenario. We apply the results to two examples drawn from the literature. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: bias analysis; gene environment; independence; interaction; sensitivity analysis; unmeasured confounding

1. Introduction

Unmeasured confounding is a challenge in epidemiologic research and can bias effect measures. Various sensitivity analysis techniques have been developed and employed in the literature to assess the extent to which an unmeasured confounder would have to affect both the exposure and the outcome in order to change the qualitative conclusions drawn from an analysis [1–7]. Most of this literature has focused on sensitivity analysis for the overall effect. To the best of our knowledge, no sensitivity analysis technique is available at present to assess the impact of unmeasured confounding for interactions.

Biological and chemical exposures may interact with one another in producing their effects. The effects of such exposures may also be modified by various genetic factors, and as genetics research progresses, gene–gene and gene–environment interaction are gaining increasing prominence in the literature. Although genetic factors often are assumed effectively randomized, environmental factors and biological and chemical exposures in interaction analyses are subject to the same confounding as they would be in any observational study. In some studies, the effects of genetic factors may be confounded by population stratification if adequate control for this has not been made. Unmeasured confounding is clearly an issue in such gene–environment interaction analyses.

In this paper, we develop a sensitivity analysis approach using bias formulas to assess the sensitivity of interaction estimates to the presence of an unmeasured confounding. We consider a general setting in which one or more unmeasured confounders may affect both factors of interest. We give results for both additive and multiplicative scales. In addition to the general case, we also consider several more specific cases such that: (i) the unmeasured confounder affects only one of the two exposures; (ii) the unmeasured confounder does not interact with one of the exposures in its effects on the outcome; or (iii) the two exposures are independent of one another (e.g., gene–environment independence).

[a]*Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA*
[b]*Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA*
[c]*Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA*
[d]*Department of Biostatistics, University of Pennsylvania, Philadelphia, PA, USA*
**Correspondence to: Tyler J. VanderWeele, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA.*
[†]*E-mail: tvanderw@hsph.harvard.edu*

Copyright © 2011 John Wiley & Sons, Ltd.

2. Notation and definitions

We will let G and E denote our two factors or exposures of interest. These might well represent genetic and environmental factors respectively, but nothing in our development will restrict the application to only gene–environment interaction. The two factors might both be environmental or both genetic. We will let Y denote the outcome of interest.

The two exposures and the outcome may be binary or continuous. We let Y_{ge} denote the counterfactual or potential outcome [8, 9] for Y for each individual if possible, contrary to the fact that the first exposure had been set to g and the second to e. Thus, if the two exposures were both binary, then for each individual, there would be four counterfactual or potential outcomes, Y_{11}, Y_{10}, Y_{01}, and Y_{00}. We do not know the counterfactual outcomes for each individual, but we can hope to estimate them on average for the population. For example, if the two factors were both randomized, we could consistently estimate $\mathbb{E}[Y_{11}]$ by $\mathbb{E}[Y|G=1, E=1]$, $\mathbb{E}[Y_{10}]$ by $\mathbb{E}[Y|G=1, E=0]$, and so forth.

In an observational study, the exposures are not randomized and estimates are potentially subject to confounding. Thus, an investigator instead typically tries to collect data on a set of covariates C that suffices to control for this confounding. Essentially, within strata of C, the groups with different exposure status should be comparable. More formally, we use $A \coprod B|C$ to denote that A is independent of B conditional on C. We say that the effects of G and E on Y are unconfounded given C if for all g and e, $Y_{ge} \coprod \{G, E\}|C$. If the effects of G and E on Y are unconfounded given C, then we can consistently estimate $\mathbb{E}[Y_{ge}|c]$ by $\mathbb{E}[Y|G=g, E=e, C=c]$. Often, the set of measured covariates C will not suffice to control for confounding. Instead, we might hypothesize a set of unmeasured confounders U such that the effects of G and E on Y are unconfounded given $\{C, U\}$, that is, $Y_{ge} \coprod \{G, E\}|\{C, U\}$. Unfortunately, if we do not have data on U, we cannot stratify on or otherwise adjust for U.

If we only have data on C and we are interested in interaction on the additive scale, then we would typically use the following measure for additive interaction:

$$\mathbb{E}[Y|g_1, e_1, c] - \mathbb{E}[Y|g_0, e_1, c] - \mathbb{E}[Y|g_1, e_0, c] + \mathbb{E}[Y|g_0, e_0, c]. \tag{1}$$

An additive interaction measure of 0 corresponds to exact additivity (i.e., no additive interaction). If the effects of G and E on Y are unconfounded given C, then this will consistently estimate the true causal interaction on the additive scale:

$$\mathbb{E}[Y_{g_1e_1} - Y_{g_0e_1} - Y_{g_1e_0} + Y_{g_0e_0}|c]. \tag{2}$$

If however there are one or more unmeasured confounding variables U such that the effects of G and E on Y are unconfounded given $\{C, U\}$, but the effects are not unconfounded given only C, then the estimate in Equation (1) will not be consistent for the causal interaction in Equation (2). If the effects of G and E on Y are unconfounded given $\{C, U\}$, that is, $Y_{ge} \coprod \{G, E\}|\{C, U\}$, then the true causal interaction in Equation (2) is equal to the following:

$$\sum\nolimits_u \{\mathbb{E}[Y|g_1, e_1, c, u] - \mathbb{E}[Y|g_0, e_1, c, u] - \mathbb{E}[Y|g_1, e_0, c, u] + \mathbb{E}[Y|g_0, e_0, c, u]\} P(u|c) \tag{3}$$

However, without data on U, we cannot estimate the expression in Equation (3).

In the remainder of this paper, we will develop sensitivity analysis results to attempt to address this problem. We will express the difference between Equations (1) and (2) as the bias for the interaction on the additive scale, and we will derive expressions for the bias in terms of sensitivity analysis parameters that relate to the effects of the unmeasured confounder(s) U on the exposures and on the outcome. One can consider a variety of different sensitivity analysis parameters and assess how an unmeasured confounder with the properties indicated by the sensitivity analysis parameters would affect conclusions drawn about the causal interaction in Equation (2) from the estimated interaction in Equation (1). All inferences will be for interaction parameters for the overall population within strata of covariates C, though we also make comments in the following text about interaction parameters marginalized over C.

In the next section, we will give sensitivity analysis results for interaction on the additive scale. In the following section, we will then consider and give analogous results for the multiplicative interaction scale. The additive scale is generally considered the scale that is most relevant for assessing public health implications of interaction [10–13]; it is also most closely related to the notion of synergism within the sufficient cause framework, and these relations are described in detail elsewhere [13–16]. The sensitivity analysis technique for additive interaction will thus also be useful if investigators want to reason

Copyright © 2011 John Wiley & Sons, Ltd.

about synergism in the sufficient cause framework. Although additive interaction is most relevant for both public health purposes and for assessing mechanistic interaction, the multiplicative scale is what is most often used in practice (generally out of convenience), and thus we consider also sensitivity analysis techniques for interaction parameters on the multiplicative scale. Finally, sometimes case-control data are used and a multiplicative model is fit, but the risk ratios or odds ratio from the multiplicative model are used to estimate measures of additive interaction using Rothman's 'relative excess risk due to interaction' (RERI) [17–19], and we thus give sensitivity analysis techniques for this measure as well.

Before proceeding, one final point deserves attention. In this paper, we consider measures of causal interaction, that is to say, measures that examine the extent to which an outcome would change under interventions on both of the factors of interest. This is different than mere 'effect heterogeneity' in which the effect of one intervention varies across strata of another variable (which is not intervened upon) [20, 21]. The sensitivity analysis techniques in this paper apply to causal interaction. If effect heterogeneity is in view, then sensitivity analysis techniques for causal effects of a single exposure [7] could be applied to each stratum of the secondary factor separately. No additional theory or results are needed for sensitivity analysis in this case.

3. Sensitivity analysis for interactions on the additive scale

The results that follow will relate the causal interaction contrast in Equation (2) to the interaction estimate with the data in Equation (1). We define the bias on the additive scale as

$$B_{\text{add}} = \mathbb{E}[Y|g_1, e_1, c] - \mathbb{E}[Y|g_0, e_1, c] - \mathbb{E}[Y|g_1, e_0, c] + \mathbb{E}[Y|g_0, e_0, c] - \mathbb{E}[Y_{g_1e_1} - Y_{g_0e_1} - Y_{g_1e_0} + Y_{g_0e_0}|c] \tag{4}$$

Theorem 1 gives a general formula for the bias for the interaction on the additive scale, B_{add}, in terms of various sensitivity analysis parameters. Appendix A presents the proofs of all results.

Theorem 1
Suppose that for all g and e, $Y_{ge} \coprod \{G, E\}|\{C, U\}$ and for any particular reference level u' of U define $\gamma_{ij}(u) = \mathbb{E}[Y|g_i, e_j, c, u] - \mathbb{E}[Y|g_i, e_j, c, u']$. We then have that

$$\begin{aligned} B_{\text{add}} = &\sum\nolimits_u \gamma_{11}(u)\{P(u|g_1, e_1, c) - P(u|c)\} - \sum\nolimits_u \gamma_{10}(u)\{P(u|g_1, e_0, c) - P(u|c)\} \\ &- \sum\nolimits_u \gamma_{01}(u)\{P(u|g_0, e_1, c) - P(u|c)\} + \sum\nolimits_u \gamma_{00}(u)\{P(u|g_0, e_0, c) - P(u|c)\}. \end{aligned}$$

The use of Theorem 1 requires specifying what might be interpreted as the effect of U in each strata of G and E, $\gamma_{ij}(u) = \mathbb{E}[Y|g_i, e_j, c, u] - \mathbb{E}[Y|g_i, e_j, c, u']$, and also the distribution of U in each strata of G and E, $P(u|g_i, e_j, c)$, along with the prevalence of U overall, $P(u|c)$. Each of these could be taken as sensitivity analysis parameters. Once the sensitivity analysis parameters have been specified and B_{add} has been calculated, a corrected estimated of additive interaction can be obtained by subtracting B_{add} from the estimated interaction measure in (1) using the observed data. Corrected confidence intervals could be obtained by bootstrapping. Theorem 1 applies to measures of additive interaction conditional on $C = c$. If the sensitivity analysis parameters are assumed to be constant over C, then the result also applies immediately to measures of additive interaction marginalized over C. Alternatively, an investigator could specify different sensitivity analysis parameters for each level of C and marginalize the stratum-specific corrected estimates over C. However, as can be seen, the use of Theorem 1 in its most general form requires the specification of a large number of sensitivity analysis parameters. Similar comments apply also to the results that follow.

Under the simplifying assumption that U does not interact with one of the two factors on the additive scale, the expression for the bias on the additive scale, B_{add}, simplifies considerably as stated in the next corollary.

Corollary 1A
Suppose that the effect of G and E on Y are unconfounded conditional on $\{C, U\}$. Suppose further that U is binary and that for fixed c, $\mathbb{E}[Y|g, e_1, c, U = 1] - \mathbb{E}[Y|g, e_1, c, U = 0] = \gamma_1$ and $\mathbb{E}[Y|g, e_0, c, U = 1] - \mathbb{E}[Y|g, e_0, c, U = 0] = \gamma_0$ are constant across strata of g so that G does not interact with U on the additive scale and let $\delta_1 = P(U = 1|g_1, e_1, c) - P(U = 1|g_0, e_1, c)$ and $\delta_0 = P(U = 1|g_1, e_0, c) - P(U = 1|g_0, e_0, c)$, then

$$B_{\text{add}} = \delta_1\gamma_1 - \delta_0\gamma_0.$$

Copyright © 2011 John Wiley & Sons, Ltd.

To use Corollary 1A, one needs to specify far fewer parameters than in Theorem 1. One simply needs to specify the effect of U, $\gamma_j = \mathbb{E}[Y|g, e_j, c, U = 1] - \mathbb{E}[Y|g, e_j, c, U = 0]$, for $E = e_1$ and $E = e_0$, along with the prevalence difference of U comparing $G = g_1$ and $G = g_0$, $\delta_j = P(U = 1|g_1, e_j, c) - P(U = 1|g_0, e_j, c)$, for $E = e_1$ and $E = e_0$. The use of Corollary 1A is far more straightforward than that of Theorem 1 but requires the stronger assumptions that U is binary and that G does not interact with U on the additive scale.

Note that by symmetry, if E does not interact with U on the additive scale, then $B_{\text{add}} = \delta_1^* \gamma_1^* - \delta_0^* \gamma_0^*$, where $\gamma_i^* = \mathbb{E}[Y|g_i, e, c, U = 1] - \mathbb{E}[Y|g_i, e, c, U = 0]$, $\delta_i^* = P(U = 1|g_i, e_1, c) - P(U = 1|g_i, e_0, c)$. With the bias formula from Corollary 1A, we can also obtain corrected confidence intervals simply by subtracting B_{add} from both limits of the confidence interval for the estimated interaction measure in (1) since for Corollary 1A, the bias formula is a deterministic function of the sensitivity analysis parameters.

Suppose now that U were only a confounder for E and that we had $G \times E$ independence in the sense that $\{E, U\} \coprod G|C$, we then have the following result.

Corollary 1B

Suppose that the effects of G and E on Y are unconfounded conditional on $\{C, U\}$ and we have $G \times E$ independence in the sense that $\{E, U\} \coprod G|C$, then if U does not interact with G on the additive scale in the sense that $\mathbb{E}[Y|g, e, c, u] - \mathbb{E}[Y|g, e, c, u']$ is constant across g, then $B_{\text{add}} = 0$.

By symmetry, if U were only a confounder for G and we had $G \times E$ independence in the sense that $\{G, U\} \coprod E|C$, then if U does not interact with E on the additive scale, then $B_{\text{add}} = 0$. Note that Corollary 1B does not assume that U is binary.

Remark 1

Suppose that U were an environmental factor that was a confounder only for E, not G. An interesting consequence of Corollary 1B is that if we have $G \times E$ independence and if we found that our estimated measure of interaction in Equation (1) were nonzero, then if there is no interaction between U and G on the additive scale, then $B_{\text{add}} = 0$. Thus, if we found that our estimated measure of interaction in Equation (1) were nonzero, then either there is an actual $G \times E$ interaction (because $B_{\text{add}} = 0$ and the estimated interaction is equal to the causal interaction) or there is a $G \times U$ interaction, another form of gene–environment interaction. Essentially, under gene–environment independence even with unmeasured confounding, we have some form of gene–environment interaction either with E or with U.

A result similar to Corollary 1B holds under $G \times E$ independence if there is an unmeasured genetic confounder U_1 for G and another unmeasured environmental confounder U_2 for E that are binary and independent of one another. In this case, if G does not interact with U_2, E does not interact with U_1, and U_1 does not interact with U_2 on the additive scale, then $B_{\text{add}} = 0$. Thus, if the estimated interaction measure in Equation (1) were nonzero, one could conclude either a true causal $G \times E$ interaction or a $G \times U_1$ interaction or a $E \times U_2$ interaction or a $U_1 \times U_2$ interaction, that is, some form of gene–environment interaction would be present. We give a formal statement of the result in Appendix A as Corollary 1C.

4. Sensitivity analysis for interactions on the multiplicative scale

On the multiplicative scale, if we had data on covariates C, we would typically use the following measure for multiplicative interaction on the risk ratio scale:

$$\frac{\mathbb{E}[Y|g_1, e_1, c]}{\mathbb{E}[Y|g_0, e_1, c]} \Big/ \frac{\mathbb{E}[Y|g_1, e_0, c]}{\mathbb{E}[Y|g_0, e_0, c]}. \tag{5}$$

A multiplicative interaction measure of 1 corresponds to exact multiplicativity (i.e., no multiplicative interaction). If the effects of G and E on Y are unconfounded given C, then this will consistently estimate the true causal interaction on the multiplicative scale:

$$\frac{\mathbb{E}[Y_{g_1 e_1}|c]}{\mathbb{E}[Y_{g_0 e_1}|c]} \Big/ \frac{\mathbb{E}[Y_{g_1 e_0}|c]}{\mathbb{E}[Y_{g_0 e_0}|c]}. \tag{6}$$

If however there are one or more unmeasured confounding variables U such that the effects of G and E on Y are unconfounded given $\{C, U\}$, that is, $Y_{ge} \coprod \{G, E\}|\{C, U\}$, but the effects are not unconfounded given only C, then the estimate in Equation (5) will not be consistent for the causal interaction

Copyright © 2011 John Wiley & Sons, Ltd.

in Equation (6). We can then define the bias on the multiplicative scale as

$$B_{\text{mult}} = \left\{\frac{\mathbb{E}[Y|g_1,e_1,c]}{\mathbb{E}[Y|g_0,e_1,c]} \Big/ \frac{\mathbb{E}[Y|g_1,e_0,c]}{\mathbb{E}[Y|g_0,e_0,c]}\right\} \Big/ \left\{\frac{\mathbb{E}[Y_{g_1e_1}|c]}{\mathbb{E}[Y_{g_0e_1}|c]} \Big/ \frac{\mathbb{E}[Y_{g_1e_0}|c]}{\mathbb{E}[Y_{g_0e_0}|c]}\right\}.$$

The following theorem gives a general formula for the bias for the interaction on the multiplicative scale, B_{mult}, in terms of various sensitivity analysis parameters.

Theorem 2
Suppose that for all g and e, $Y_{ge} \coprod \{G,E\}|\{C,U\}$ and for any particular reference level u' of U define $\gamma_{ij}(u) = \frac{E(Y|g_i,e_j,c,u)}{E(Y|g_i,e_j,c,u')}$, then we have that

$$B_{\text{mult}} = \frac{\frac{\sum_u \gamma_{11}(u)P(u|g_1,e_1,c)}{\sum_u \gamma_{11}(u)P(u|c)} \Big/ \frac{\sum_u \gamma_{01}(u)P(u|g_0,e_1,c)}{\sum_u \gamma_{01}(u)P(u|c)}}{\frac{\sum_u \gamma_{10}(u)P(u|g_1,e_0,c)}{\sum_u \gamma_{10}(u)P(u|c)} \Big/ \frac{\sum_u \gamma_{00}(u)P(u|g_0,e_0,c)}{\sum_u \gamma_{00}(u)P(u|c)}}.$$

The use of Theorem 2 in its most general form requires the specification of a large number of sensitivity analysis parameters. Once the sensitivity analysis parameters have been specified and B_{mult} has been calculated, a corrected estimated of multiplicative interaction can be obtained by dividing the estimated interaction measure in (5) by the bias factor B_{mult}. Corrected confidence intervals could be obtained by bootstrapping. Under the simplifying assumption that U does not interact on the multiplicative scale with one of the two factors, the expression for the bias on the multiplicative scale, B_{mult}, simplifies considerably as stated in the next corollary. Note that it is not in general possible for U to not interact with a specific factor, say G, on both the additive and the multiplicative scales (unless it has no effect on the outcome or if, e.g., the baseline risk in the absence of U is the same in all strata of G). In general, at most, either no additive or no multiplicative interaction between G and U would hold.

Corollary 2A
Suppose that the effect of G and E on Y are unconfounded conditional on $\{C,U\}$. Suppose further that U is binary and that $\mathbb{E}[Y|g,e_1,c,U=1]/\mathbb{E}[Y|g,e_1,c,U=0] = \gamma_1$ and $\mathbb{E}[Y|g,e_0,c,U=1]/\mathbb{E}[Y|g,e_0,c,U=0] = \gamma_0$ are constant across strata of g so that G does not interact with U on the multiplicative scale, then

$$B_{\text{mult}} = \frac{1+(\gamma_1-1)P(U=1|g_1,e_1,c)}{1+(\gamma_1-1)P(U=1|g_0,e_1,c)} \Big/ \frac{1+(\gamma_0-1)P(U=1|g_1,e_0,c)}{1+(\gamma_0-1)P(U=1|g_0,e_0,c)}.$$

A similar result holds by symmetry if E does not interact with U on the multiplicative scale. With the bias formula from Corollary 2A, we can also obtain corrected confidence intervals simply by dividing both limits of the confidence interval for the estimated interaction measure in (5) by the bias factor B_{mult}, since for Corollary 2A, the bias formula is a deterministic function of the sensitivity analysis parameters. If U were only a confounder for E and that we also had $G \times E$ independence in the sense that $\{E,U\} \coprod G|C$, we then have the following result.

Corollary 2B
Suppose that the effect of G and E on Y are unconfounded conditional on $\{C,U\}$ and we have $G \times E$ independence in the sense that $\{E,U\} \coprod G|C$, then if U does not interact with G on the multiplicative scale in the sense that $\frac{E(Y|g,e,c,u)}{E(Y|g,e,c,u')}$ is constant across g, then $B_{\text{mult}} = 1$.

Remark 2
From Corollary 2B, it follows that if under $G \times E$ independence, we found that our estimated measure of interaction in Equation (5) were non-null, then either there is an actual $G \times E$ interaction or there is a $G \times U$ interaction. The result would apply to multiplicative interaction estimates from a case-only design or a case-control design under a rare outcome assumption or incidence density sampling. Note that by symmetry, if U were only a confounder for E and that we had $G \times E$ independence in the sense that $\{G,U\} \coprod E|C$, then if U does not interact with G on the multiplicative scale, then $B_{\text{mult}} = 1$. Note also that Corollary 1B does not assume that U is binary.

A result similar to Corollary 2B holds under $G \times E$ independence if there is an unmeasured genetic confounder U_1 for G and another unmeasured environmental confounder U_2 for E that are binary and independent of one another. In this case, if G does not interact with U_2, E does not interact with U_1, and U_1 does not interact with U_2 on the multiplicative scale, then $B_{\text{mult}} = 1$. Thus, if the estimated

Copyright © 2011 John Wiley & Sons, Ltd.

interaction measure in Equation (5) were non-null, one could conclude either a true causal $G \times E$ interaction or a $G \times U_1$ interaction or a $E \times U_2$ interaction or a $U_1 \times U_2$ interaction, that is, some form of gene–environment interaction would be present. We give a formal statement of the result in Appendix A as Corollary 2C.

5. Sensitivity analysis for the relative excess risk due to interaction

Often in case-control studies, logistic regression is used to accommodate the case-control design. In such studies, if investigators want to assess interaction on the additive scale for public health purposes [10–13] or mechanistic interaction [13–16], then a measure referred to as the RERI [17–19] is sometimes used. The measure is also sometimes used when a logistic regression model is fit to the data out of convenience rather than by necessity due to a case-control design. The RERI conditional on c would generally be estimated by

$$\frac{\mathbb{E}[Y|g_1,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]} - \frac{\mathbb{E}[Y|g_1,e_0,c]}{\mathbb{E}[Y|g_0,e_0,c]} - \frac{\mathbb{E}[Y|g_0,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]} + 1. \tag{7}$$

If the outcome is rare so that odds ratios approximate risk ratios, then each term $\mathbb{E}[Y|g,e,c]/\mathbb{E}[Y|g_0,e_0,c]$ can be approximated by the estimated odds ratio from the logistic regression.

Define the causal RERI conditional on $C = c$ by

$$\mathrm{RERI}_c = \frac{\mathbb{E}[Y_{g_1e_1}|c]}{\mathbb{E}[Y_{g_0e_0}|c]} - \frac{\mathbb{E}[Y_{g_1e_0}|c]}{\mathbb{E}[Y_{g_0e_0}|c]} - \frac{\mathbb{E}[Y_{g_0e_1}|c]}{\mathbb{E}[Y_{g_0e_0}|c]} + 1. \tag{8}$$

If the effects of G and E on Y were unconfounded conditional on (C, U) but data were only available on C, we might estimate the RERI by Equation (7), but this would be biased for the true quantity in Equation (8) because of the unmeasured confounding due to U. The following results can help reason about the causal RERI_c.

Theorem 3

Suppose that for all g and e, $Y_{ge} \coprod \{G, E\}|\{C, U\}$ and for any particular reference level u' of U define $\gamma_{ij}(u) = \frac{E(Y|g_i,e_j,c,u)}{E(Y|g_i,e_j,c,u')}$, then we have that

$$\mathrm{RERI}_c = \frac{\sum_u \gamma_{00}(u)P(u|g_0,e_0,c)}{\sum_u \gamma_{00}(u)P(u|c)} \left[\frac{\frac{\mathbb{E}[Y|g_1,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_{11}(u)P(u|g_1,e_1,c)}{\sum_u \gamma_{11}(u)P(u|c)}} - \frac{\frac{\mathbb{E}[Y|g_1,e_0,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_{10}(u)P(u|g_1,e_0,c)}{\sum_u \gamma_{10}(u)P(u|c)}} - \frac{\frac{\mathbb{E}[Y|g_0,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_{01}(u)P(u|g_0,e_1,c)}{\sum_u \gamma_{01}(u)P(u|c)}} \right] + 1.$$

To apply the result, one would again specify the effect of U, $\gamma_{ij}(u) = \frac{E(Y|g_i,e_j,c,u)}{E(Y|g_i,e_j,c,u')}$, in each of the $G \times E$ strata along with the distribution of U, $P(u|g_1, e_0, c)$, for each of the $G \times E$ strata. The corrected causal RERI could then be computed using the expression in Theorem 3. Under simplifying assumptions that U is binary with a constant effect across $G \times E$, a more straightforward adjustment approach is possible as stated in the following corollary, which follows immediately from Theorem 3.

Corollary 3A

Suppose that for all g and e, $Y_{ge} \coprod \{G, E\}|\{C, U\}$ and suppose that U is binary and $\gamma = \frac{E(Y|g,e,c,U=1)}{E(Y|g,e,c,U=0)}$ is constant over g and e, then we have that

$$\mathrm{RERI}_c = \frac{\frac{\mathbb{E}[Y|g_1,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{1+(\gamma-1)P(U=1|g_1,e_1,c)}{1+(\gamma-1)P(U=1|g_0,e_0,c)}} - \frac{\frac{\mathbb{E}[Y|g_1,e_0,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{1+(\gamma-1)P(U=1|g_1,e_0,c)}{1+(\gamma-1)P(U=1|g_0,e_0,c)}} - \frac{\frac{\mathbb{E}[Y|g_0,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{1+(\gamma-1)P(U=1|g_0,e_1,c)}{1+(\gamma-1)P(U=1|g_0,e_0,c)}} + 1.$$

A simpler result is also possible under $G \times E$ independence as stated in the next corollary.

Corollary 3B

Suppose that for all g and e, $Y_{ge} \coprod \{G, E\}|\{C, U\}$ and we have $G \times E$ independence in the sense that $\{E, U\} \coprod G|C$. Suppose further that U is binary and that $\mathbb{E}[Y|g, e_1, c, U = 1]/\mathbb{E}[Y|g, e_1, c, U = 0] =$

Copyright © 2011 John Wiley & Sons, Ltd.

γ_1 and $\mathbb{E}[Y|g, e_0, c, U = 1]/\mathbb{E}[Y|g, e_0, c, U = 0] = \gamma_0$ are constant across strata of g so that G does not interact with U on the multiplicative scale, then

$$\text{RERI}_c = \frac{1}{\kappa}\frac{\mathbb{E}[Y|g_1, e_1, c]}{\mathbb{E}[Y|g_0, e_0, c]} - \frac{\mathbb{E}[Y|g_1, e_0, c]}{\mathbb{E}[Y|g_0, e_0, c]} - \frac{1}{\kappa}\frac{\mathbb{E}[Y|g_0, e_1, c]}{\mathbb{E}[Y|g_0, e_0, c]} + 1, \tag{9}$$

where

$$\kappa = \frac{1 + (\gamma_1 - 1)P(U = 1|e_1, c)}{1 + (\gamma_1 - 1)P(U = 1|c)} \Big/ \frac{1 + (\gamma_0 - 1)P(U = 1|e_0, c)}{1 + (\gamma_0 - 1)P(U = 1|c)}.$$

Corollary 3B assumes that the effect of U does not interact with G on the risk ratio scale. Under Corollary 3B, it is only necessary to specify γ_1 and γ_0 (the effect of U when $E = e_1$ and $E = e_0$, respectively) and the probability of $U = 1$, when $E = e_1$ and $E = e_0$, respectively; the value of $P(U = 1|c)$ could be calculated from these two probabilities. Once these are specified, one can calculate κ and use Equation (9) to obtain the corrected measure of the causal RERI. Note that even under the assumptions of Corollaries 3A or 3B, the CI for the corrected RERI cannot simply be obtained by applying some formula to the confidence limits of the uncorrected RERI. The confidence limits for the corrected RERI could be obtained either by using the delta method or by bootstrapping. Finally, note that if the estimated RERI in Equation (7) were found to be nonzero, then it would also follow that the quantity in Equation (1) was nonzero. If in addition we have $G \times E$ independence, we could still apply Corollary 1B to conclude that there was either a true causal $G \times E$ interaction or a $G \times U$ interaction.

As noted previously, the additive scale is most useful for assessing causal notions of interaction in the sufficient cause framework. Simple relations hold between the causal RERI and the presence of synergism in the sufficient cause framework. Specifically, if an investigator is interested in detecting 'sufficient cause interactions' [14, 15] corresponding to individuals with response patterns such that $D_{11} = 1$ but $D_{10} = D_{01} = 0$, then if both exposures are never preventive for any individual, then $\text{RERI}_c > 0$ implies the presence of this response pattern [14, 15]. Without this no-preventive-action assumption, $\text{RERI}_c > 1$ still implies the presence of this response pattern [14, 15]. If we are interested in detecting an even stronger notion of interaction that $D_{11} = 1$ but $D_{10} = D_{01} = D_{00} = 0$ (i.e., individuals for whom the outcome occurs if and only if both exposures are present, 'epistatic interactions' [16]), then $\text{RERI}_c > 2$ suffices without any assumptions about preventive action; $\text{RERI}_c > 1$ suffices if at least one of the exposures is never preventive for any individual; and $\text{RERI}_c > 0$ suffices if both are never preventive [16].

6. Applications

Using a case-only design [22, 23], Bennett *et al.* [24] studied the interaction between passive smoking and glutathione S-transferase M1 (GSTM1) on lung cancer risk among nonsmokers. Investigators genotyped 106 lung cancer cases and estimated a case-only measure of interaction. Let G denote GSTM1 (g_1 present, g_0 absent) and E passive smoking (e_1 present, e_0 absent). Using a case-only estimate, the investigators found

$$\frac{\mathbb{E}[Y|g_1, e_1, c]}{\mathbb{E}[Y|g_0, e_1, c]} \Big/ \frac{\mathbb{E}[Y|g_1, e_0, c]}{\mathbb{E}[Y|g_0, e_0, c]} = 2.6$$

The 95% CI for the estimate was (1.1, 6.1). The case-only study design assumes that the genetic factor is independent of passive smoking. The estimate itself and the CI suggest a gene–environment interaction between passive smoking and GSTM1 on lung cancer risk. The effect of smoking may, however, be confounded by air pollution. Poorer neighborhoods in which air pollution is, say, high may also have a higher prevalence of smoking or more extensive advertising for cigarettes.

Suppose that the genetic factor (GSTM1) is independent of both passive smoking and air pollution (note that the case-only design itself assumes that the genetic factor is independent of passive smoking). By Corollary 2B, it would then follow from the case-only estimate that either there is a true causal gene × passive smoking interaction or there is an interaction between the genetic factor and air pollution in the sense that there are some levels of air pollution, u and u', say such that the effect of air pollution when the genetic factor is present, $\frac{E(Y|g_1,e,c,u)}{E(Y|g_1,e,c,u')}$, differs from the effect when the genetic factor is absent, $\frac{E(Y|g_0,e,c,u)}{E(Y|g_0,e,c,u')}$.

Copyright © 2011 John Wiley & Sons, Ltd.

As a second example, we will consider a study of Ahsan *et al.* [25] examining the evidence for additive interaction between the effects of arsenic exposure in well water and body mass index (BMI) in producing premalignant skin lesions [25]. Data come from a large cohort study of 11,746 individuals in Bangladesh, many of whom had been exposed to various doses of arsenic through drinking well water. Following their analysis, let $G = 1$ for high versus low arsenic (<8 vs. >175 μg/L) and let $E = 1$ for low versus high BMI (<18.1 vs. >20.4) with $Y = 1$ denoting the presence of premalignant skin lesions. Ahsan *et al.* [25] adjusted for gender, age, education, cigarette smoking, hukka smoking, sun exposure, and land ownership. They used logistic regression to estimate the RERI in assessing potential additive interaction between BMI and arsenic exposure. Compared with the reference of $G = 0$, $E = 0$, the odds ratio for $G = 1$, $E = 1$ was 5.25 (95% CI: 3.07, 8.99); for $G = 1$, $E = 0$ was 2.96 (95% CI: 1.63, 5.37); and for $G = 0$, $E = 1$ was 0.71 (95% CI: 0.38, 1.32). The overall prevalence of skin lesions is 6.3%, which is generally considered sufficiently small so that odds ratio approximates risk ratios. The estimated RERI was thus $5.25 - 2.96 - 0.71 + 1 = 2.59$ with a 95% CI of (0.75, 4.24), suggesting evidence for positive additive interaction. Until the study was conducted, there was very little knowledge of which wells had high levels of arsenic; the correlation between arsenic exposure and other covariates is thus very weak. It is unlikely the effects of arsenic are subject to substantial confounding. The effects of BMI on skin lesions are, however, likely confounded by, say, nutritional intake. The conditional association between BMI and arsenic exposure is not statistically significant in the sample, and we could therefore potentially employ Corollaries 1B or 3B. By Corollary 1B, we would have that either there is an interaction between arsenic and BMI or between arsenic and the confounders of the effect of BMI, for example, nutritional intake. If we further wanted corrected estimates of the RERI between arsenic and BMI, we could use the sensitivity analysis technique in Corollary 3B. Let U denote a hypothetical binary unmeasured confounder with $U = 1$ indicating high versus low nutritional intake. Suppose high nutritional intake decreased the likelihood of skin lesions by threefold ($\gamma_1 = \gamma_0 = 1/3$) for all strata of arsenic and BMI, with prevalence of high nutritional intake of 0.6 in those with high BMI and a prevalence of 0.2 in those with low BMI, we then have that

$$\begin{aligned}\kappa &= \frac{1 + (\gamma_1 - 1)P(U = 1|e_1, c)}{1 + (\gamma_1 - 1)P(U = 1|c)} \Big/ \frac{1 + (\gamma_0 - 1)P(U = 1|e_0, c)}{1 + (\gamma_0 - 1)P(U = 1|c)} \\ &= \frac{1 + (1/3 - 1)(0.2)}{1 + (1/3 - 1)(0.6)} = 1.44.\end{aligned}$$

Under the rare outcome assumption so that odds ratios approximate risk ratios, we would have, by Corollary 3B, a corrected RERI of

$$\begin{aligned}\mathrm{RERI}_c &= \frac{1}{\kappa}\frac{\mathbb{E}[Y|g_1, e_1, c]}{\mathbb{E}[Y|g_0, e_0, c]} - \frac{\mathbb{E}[Y|g_1, e_0, c]}{\mathbb{E}[Y|g_0, e_0, c]} - \frac{1}{\kappa}\frac{\mathbb{E}[Y|g_0, e_1, c]}{\mathbb{E}[Y|g_0, e_0, c]} + 1 \\ &= \frac{1}{1.44}5.25 - 2.96 - \frac{1}{1.44}0.71 + 1 = 1.19.\end{aligned}$$

As an alternative scenario assuming weaker confounding, if high nutritional intake decreased the likelihood of skin lesions by twofold, with prevalence of 0.4 in those with high BMI and a prevalence of 0.2 in those with low BMI, we would have $\kappa = 1.13$ with a corrected $\mathrm{RERI}_c = 2.06$.

7. Discussion

In this paper, we have provided several methods for sensitivity analysis for unmeasured confounding in studies of interaction between biological or chemical exposures or between such exposures and genetic factors. We have considered both the additive and multiplicative scales along with additive interaction obtained from a multiplicative model (the RERI). We have given results in considerable generality and have also provided much simpler and easier to use techniques that can be employed under some simplifying assumptions such as that of a single binary unmeasured confounder. The techniques will likely be useful in a wide range of interaction studies and can be applied across numerous different study designs.

The results on additive interaction (Theorem 1 and Corollaries 1A–1C) are applicable to cohort designs. The multiplicative interaction results will also be applicable to cohort designs and will moreover be applicable case-control designs when the outcome is rare so that odds ratio approximates risk ratio. The multiplicative results will likewise be applicable to case-only designs, when the genetic and environmental factors are independent as under such an independence assumption, the case-only design allows

Copyright © 2011 John Wiley & Sons, Ltd.

one to estimate interaction on the multiplicative scale. The multiplicative results are also applicable to family-based study designs, which estimate the interaction on the log scale or in settings in which the outcome is rare. The results on the RERI will likewise also be applicable to cohort designs, case-control and case-only designs with a rare outcome, and family-based genetic designs.

The techniques may prove to be especially useful in assessing gene–environment interaction. Many such studies have not paid attention to potential confounding of the environmental factor. Ideally, better control for such confounding will be made. However, in settings in which the requisite data is not available, the techniques here will allow investigators to assess the extent to which unmeasured environmental confounding may affect unadjusted results. One issue that has not been examined here is the extent to which unmeasured confounding may affect the power of interaction analyses [26]. This is especially important because power to detect interaction is often quite low. Considerations of power in interaction analyses will be left to future work.

In many studies of gene–environment interaction, the genetic and environmental factors are assumed to be independent. We have seen previously that, under this assumption, interaction findings are particularly robust to unmeasured confounding insofar as if we are concerned about unmeasured confounding of the environmental factor by another unmeasured environmental exposure and with the observed data, we find interaction, then either there must be a true causal interaction between the genetic and environmental factor or there is interaction between the genetic factor and the unmeasured environmental confounding variable; in either case, we have gene–environment interaction. It is hoped that these various results will facilitate inference about interaction in genetic and epidemiologic practices and assist in assessing the robustness of findings that do not adequately control for confounding.

Appendix A. Proofs

Proof of Theorem 1
If for all g and e, $Y_{ge} \coprod \{G, E\}|\{C, U\}$, then

$$\begin{aligned}\mathbb{E}[Y_{ge}|c] &= \sum\nolimits_u \mathbb{E}[Y_{ge}|c,u]P(u|c)\\ &= \sum\nolimits_u \mathbb{E}[Y_{ge}|g,e,c,u]P(u|c)\\ &= \sum\nolimits_u \mathbb{E}[Y|g,e,c,u]P(u|c)\end{aligned}$$

where the first equality follows by the law of iterated expectations, the second by $Y_{ge} \coprod \{G, E\}|\{C, U\}$, and the third by consistency. Thus, for any fixed reference value of u' of U

$$\begin{aligned}&\mathbb{E}[Y|g,e,c] - \mathbb{E}[Y_{ge}|c]\\ &= \sum\nolimits_u \mathbb{E}[Y|g,e,c,u]P(u|g,e,c) - \sum\nolimits_u \mathbb{E}[Y|g,e,c,u]P(u|c)\\ &= \sum\nolimits_u \{\mathbb{E}[Y|g,e,c,u] - \mathbb{E}[Y|g,e,c,u']\}\{P(u|g,e,c) - P(u|c)\}.\end{aligned}$$

By applying this equality for (g_1, e_1), (g_1, e_0), (g_0, e_1), and (g_0, e_0), the result for B_{add} follows. □

Proof of Corollary 1A
If $\mathbb{E}[Y|g,e_1,c,U=1] - \mathbb{E}[Y|g,e_1,c,U=0] = \gamma_1$ and $\mathbb{E}[Y|g,e_0,c,U=1] - \mathbb{E}[Y|g,e_0,c,U=0] = \gamma_0$ are constant across strata of g, then

$$\begin{aligned}B_{\text{add}} &= \{\mathbb{E}[Y|g_1,e_1,c,U=1] - \mathbb{E}[Y|g_1,e_1,c,U=0]\}\{P(U=1|g_1,e_1,c) - P(U=1|c)\}\\ &\quad - \{\mathbb{E}[Y|g_0,e_1,c,U=1] - \mathbb{E}[Y|g_0,e_1,c,U=0]\}\{P(U=1|g_0,e_1,c) - P(U=1|c)\}\\ &\quad - \{\mathbb{E}[Y|g_1,e_0,c,U=1] - \mathbb{E}[Y|g_1,e_0,c,U=0]\}\{P(U=1|g_1,e_0,c) - P(U=1|c)\}\\ &\quad + \{\mathbb{E}[Y|g_0,e_0,c,U=1] - \mathbb{E}[Y|g_0,e_0,c,U=0]\}\{P(U=1|g_0,e_0,c) - P(U=1|c)\}\\ &= \gamma_1\{P(U=1|g_1,e_1,c) - P(u|c)\} - \gamma_1\{P(U=1|g_0,e_1,c) - P(u|c)\}\\ &\quad - \gamma_0\{P(U=1|g_1,e_0,c) - P(U=1|c)\} + \gamma_0\{P(U=1|g_0,e_0,c) - P(U=1|c)\}\\ &= \gamma_1\{P(U=1|g_1,e_1,c) - P(U=1|g_0,e_1,c)\} - \gamma_0\{P(U=1|g_1,e_0,c) - P(U=1|g_0,e_0,c)\}\\ &= \gamma_1\delta_1 - \gamma_0\delta_0.\end{aligned}$$

□

Copyright © 2011 John Wiley & Sons, Ltd.

Proof of Corollary 1B

If there is no interaction between G and U on the additive scale, then we have $\gamma_{1j}(u) = \gamma_{0j}(u)$. By Theorem 1 and $\{E, U\} \coprod G|C$, we have

$$\begin{aligned}
B_{\text{add}} &= \sum\nolimits_u \gamma_{11}(u)\{P(u|g_1, e_1, c) - P(u|c)\} - \sum\nolimits_u \gamma_{10}(u)\{P(u|g_1, e_0, c) - P(u|c)\} \\
&\quad - \sum\nolimits_u \gamma_{01}(u)\{P(u|g_0, e_1, c) - P(u|c)\} + \sum\nolimits_u \gamma_{00}(u)\{P(u|g_0, e_0, c) - P(u|c)\} \\
&= \sum\nolimits_u \gamma_{11}(u)\{P(u|e_1, c) - P(u|c)\} - \sum\nolimits_u \gamma_{10}(u)\{P(u|e_0, c) - P(u|c)\} \\
&\quad - \sum\nolimits_u \gamma_{11}(u)\{P(u|e_1, c) - P(u|c)\} + \sum\nolimits_u \gamma_{10}(u)\{P(u|e_0, c) - P(u|c)\} \\
&= 0.
\end{aligned}$$

This completes the proof. □

Corollary 1C

Suppose for all g and e and for binary U_1, U_2 we have $Y_{ge} \coprod \{G, E\}|\{C, U_1, U_2\}$ and we have $G \times E$ independence in the sense that $\{G, U_1\} \coprod E|C$ and $\{E, U_2\} \coprod G|C$, then if G does not interact with U_2 on the additive scale in the sense that $\mathbb{E}[Y|g, e, c, u_1, u_2] - \mathbb{E}[Y|g, e, c, u_1, u_2']$ does not vary with g, if E does not interact with U_1 on the additive scale in the sense that $\mathbb{E}[Y|g, e, c, u_1, u_2] - \mathbb{E}[Y|g, e, c, u_1', u_2]$ does not vary with e, and if U_1 does not interact with U_2 on the additive scale in the sense that $\mathbb{E}[Y|g, e, c, u_1, u_2] - \mathbb{E}[Y|g, e, c, u_1, u_2']$ does not vary with u_1, then $B_{\text{add}} = 0$.

Proof of Corollary 1C

If we let $U = (U_1, U_2)$ and $u' = (0, 0)$, then by Theorem 1, we have that

$$\begin{aligned}
B_{\text{add}} &= \sum\nolimits_u \gamma_{11}(u)\{P(u|g_1, e_1, c) - P(u|c)\} - \sum\nolimits_u \gamma_{01}(u)\{P(u|g_0, e_1, c) - P(u|c)\} \\
&\quad - \sum\nolimits_u \gamma_{10}(u)\{P(u|g_1, e_0, c) - P(u|c)\} + \sum\nolimits_u \gamma_{00}(u)\{P(u|g_0, e_0, c) - P(u|c)\}.
\end{aligned}$$

Now, $\sum_u \gamma_{ij}(u)\{P(u|g_i, e_j, c) - P(u|c)\}$

$$\begin{aligned}
&= \gamma_{ij}(1,1)\{P(U = (1,1)|g_i, e_j, c) - P(U = (1,1)|c)\} \\
&\quad + \gamma_{ij}(1,0)\{P(U = (1,0)|g_i, e_j, c) - P(U = (1,0)|c)\} \\
&\quad + \gamma_{ij}(0,1)\{P(U = (0,1)|g_i, e_j, c) - P(U = (0,1)|c)\} \\
&= \gamma_{ij}(1,1)\{P(U_1 = 1|g_i, c)P(U_2 = 1|e_j, c) - P(U_1 = 1|c)P(U_2 = 1|c)\} \\
&\quad + \gamma_{ij}(1,0)\{P(U_1 = 1|g_i, c)P(U_2 = 0|e_j, c) - P(U_1 = 1|c)P(U_2 = 0|c)\} \\
&\quad + \gamma_{ij}(0,1)\{P(U_1 = 0|g_i, c)P(U_2 = 1|e_j, c) - P(U_1 = 0|c)P(U_2 = 1|c)\}.
\end{aligned} \tag{A1}$$

Let $\tau_{ij} = \mathbb{E}[Y|g_i, e_j, c, U_1 = 1, U_2 = 1] - \mathbb{E}[Y|g_i, e_j, c, U_1 = 1, U_2 = 0]$ so that

$$\begin{aligned}
\gamma_{ij}(1,1) &= \mathbb{E}[Y|g_i, e_j, c, U_1 = 1, U_2 = 1] - \mathbb{E}[Y|g_i, e_j, c, U_1 = 0, U_2 = 0] \\
&= \tau_{ij} + \gamma_{ij}(1,0)
\end{aligned}$$

Summing Equation (A1) over $i = 0, 1$ and $j = 0, 1$ and noting that because G and U_2 do not interact on the additive scale, $\tau_{1j} = \tau_{0j}$ and $\gamma_{1j}(0,1) = \gamma_{0j}(0,1)$, and because G and U_2 do not interact on the additive scale, $\gamma_{i1}(1,0) = \gamma_{i0}(1,0)$, we then have that B_{add}

$$\begin{aligned}
&= \tau_{11}\{P(U_1 = 1|g_1, c)P(U_2 = 1|e_1, c) - P(U_1 = 1|g_0, c)P(U_2 = 1|e_1, c)\} \\
&\quad - \tau_{10}\{P(U_1 = 1|g_1, c)P(U_2 = 1|e_1, c) - P(U_1 = 1|g_0, c)P(U_2 = 1|e_1, c)\} \\
&\quad + \gamma_{11}(1,0)\{P(U_1 = 1|g_1, c)P(U_2 = 1|e_1, c) - P(U_1 = 1|g_1, c)P(U_2 = 1|e_0, c)\} \\
&\quad - \gamma_{01}(1,0)\{P(U_1 = 1|g_0, c)P(U_2 = 1|e_1, c) - P(U_1 = 1|g_0, c)P(U_2 = 1|e_0, c)\} \\
&\quad + \gamma_{11}(1,0)\{P(U_1 = 1|g_1, c)P(U_2 = 0|e_1, c) - P(U_1 = 1|g_1, c)P(U_2 = 0|e_0, c)\} \\
&\quad - \gamma_{01}(1,0)\{P(U_1 = 1|g_0, c)P(U_2 = 0|e_1, c) - P(U_1 = 1|g_0, c)P(U_2 = 0|e_0, c)\} \\
&\quad + \gamma_{11}(0,1)\{P(U_1 = 0|g_1, c)P(U_2 = 1|e_1, c) - P(U_1 = 0|g_0, c)P(U_2 = 1|e_1, c)\} \\
&\quad - \gamma_{10}(0,1)\{P(U_1 = 0|g_1, c)P(U_2 = 1|e_0, c) - P(U_1 = 0|g_0, c)P(U_2 = 1|e_0, c)\}.
\end{aligned}$$

Copyright © 2011 John Wiley & Sons, Ltd.

Furthermore, because U_1 and U_2 do not interact on the additive scale, $\tau_{1j} = \gamma_{1j}(0,1)$, we can group the first and the seventh, the second and the eighth, the third and the fifth, and the fourth and the sixth terms to get B_{add}

$$\begin{aligned}
&= \tau_{11}\{P(U_2 = 1|e_1, c) - P(U_2 = 1|e_1, c)\} - \tau_{10}\{P(U_2 = 1|e_1, c) - P(U_2 = 1|e_1, c)\} \\
&\quad + \gamma_{11}(1,0)\{P(U_1 = 1|g_1, c) - P(U_1 = 1|g_1, c)\} - \gamma_{01}(1,0)\{P(U_1 = 1|g_0, c) - P(U_1 = 1|g_0, c)\} \\
&= 0.
\end{aligned}$$

This completes the proof. □

Proof of Theorem 2
If for all g and e, $Y_{ge} \coprod \{G, E\}|\{C, U\}$, then as in the proof of Theorem 1, $\mathbb{E}[Y_{ge}|c] = \sum_u \mathbb{E}[Y|g, e, c, u]P(u|c)$, and thus for any fixed value of u' of U, we have that

$$\begin{aligned}
&\mathbb{E}[Y|g, e, c]/\mathbb{E}[Y_{ge}|c] \\
&= \sum\nolimits_u \mathbb{E}[Y|g, e, c, u]P(u|g, e, c)/\sum\nolimits_u \mathbb{E}[Y|g, e, c, u]P(u|c) \\
&= \sum\nolimits_u \{\mathbb{E}[Y|g, e, c, u]/\mathbb{E}[Y|g, e, c, u']\}P(u|g, e, c)/\sum\nolimits_u \{\mathbb{E}[Y|g, e, c, u]/\mathbb{E}[Y|g, e, c, u']\}P(u|c).
\end{aligned}$$

By applying this equality for (g_1, e_1), (g_1, e_0), (g_0, e_1), and (g_0, e_0) and taking ratios, the result for B_{mult} follows. □

Proof of Corollary 2A
If U is binary and $\mathbb{E}[Y|g, e_1, c, U = 1]/\mathbb{E}[Y|g, e_1, c, U = 0] = \gamma_1$and $\mathbb{E}[Y|g, e_0, c, U = 1]/\mathbb{E}[Y|g, e_0, c, U = 0] = \gamma_0$ are constant across strata of g, then B_{mult}

$$\begin{aligned}
&= \frac{\frac{\sum_u \gamma_1(u)P(u|g_1,e_1,c)}{\sum_u \gamma_1(u)P(u|c)} \Big/ \frac{\sum_u \gamma_0(u)P(u|g_0,e_1,c)}{\sum_u \gamma_0(u)P(u|c)}}{\frac{\sum_u \gamma_1(u)P(u|g_1,e_0,c)}{\sum_u \gamma_1(u)P(u|c)} \Big/ \frac{\sum_u \gamma_0(u)P(u|g_0,e_0,c)}{\sum_u \gamma_0(u)P(u|c)}} \\
&= \frac{\sum_u \gamma_1(u)P(u|g_1, e_1, c)/\sum_u \gamma_0(u)P(u|g_0, e_1, c)}{\sum_u \gamma_1(u)P(u|g_1, e_0, c)/\sum_u \gamma_0(u)P(u|g_0, e_0, c)} \\
&= \frac{1 + (\gamma_1 - 1)P(U = 1|g_1, e_1, c)}{1 + (\gamma_1 - 1)P(U = 1|g_0, e_1, c)} \Big/ \frac{1 + (\gamma_0 - 1)P(U = 1|g_1, e_0, c)}{1 + (\gamma_0 - 1)P(U = 1|g_0, e_0, c)}.
\end{aligned}$$

□

Proof of Corollary 2B
If there is no interaction between G and U on the multiplicative scale, then we have $\gamma_{1j}(u) = \gamma_{0j}(u)$. By Theorem 2 and $\{E, U\} \coprod G|C$, we have

$$\begin{aligned}
B_{\text{mult}} &= \frac{\frac{\sum_u \gamma_{11}(u)P(u|g_1,e_1,c)}{\sum_u \gamma_{11}(u)P(u|c)} \Big/ \frac{\sum_u \gamma_{01}(u)P(u|g_0,e_1,c)}{\sum_u \gamma_{01}(u)P(u|c)}}{\frac{\sum_u \gamma_{10}(u)P(u|g_1,e_0,c)}{\sum_u \gamma_{10}(u)P(u|c)} \Big/ \frac{\sum_u \gamma_{00}(u)P(u|g_0,e_0,c)}{\sum_u \gamma_{00}(u)P(u|c)}} \\
&= \frac{\frac{\sum_u \gamma_{11}(u)P(u|e_1,c)}{\sum_u \gamma_{11}(u)P(u|c)} \Big/ \frac{\sum_u \gamma_{11}(u)P(u|e_1,c)}{\sum_u \gamma_{11}(u)P(u|c)}}{\frac{\sum_u \gamma_{10}(u)P(u|e_0,c)}{\sum_u \gamma_{10}(u)P(u|c)} \Big/ \frac{\sum_u \gamma_{10}(u)P(u|e_0,c)}{\sum_u \gamma_{10}(u)P(u|c)}} = 1.
\end{aligned}$$

This completes the proof. □

Corollary 2C
Suppose for all g and e and for binary U_1, U_2 we have $Y_{ge} \coprod \{G, E\}|\{C, U_1, U_2\}$ and we have $G \times E$ independence in the sense that $\{G, U_1\} \coprod E|C$ and $\{E, U_2\} \coprod G|C$, then if G does not interact with U_2 on the multiplicative scale in the sense that $\mathbb{E}[Y|g, e, c, u_1, u_2]/\mathbb{E}[Y|g, e, c, u_1, u_2']$ does not vary with g, if E does not interact with U_1 on the multiplicative scale in the sense that $\mathbb{E}[Y|g, e, c, u_1, u_2]/\mathbb{E}[Y|g, e, c, u_1', u_2]$ does not vary with e, and if U_1 does not interact with U_2 on the multiplicative scale in the sense that $\mathbb{E}[Y|g, e, c, u_1, u_2]/\mathbb{E}[Y|g, e, c, u_1, u_2']$ does not vary with u_1, then $B_{\text{mult}} = 1$.

Copyright © 2011 John Wiley & Sons, Ltd.

Proof of Corollary 2C
If we let $U = (U_1, U_2)$ and $u' = (0, 0)$ then by Theorem 2, we have

$$B_{\text{mult}} = \frac{\frac{\sum_u \gamma_{11}(u)P(u|g_1,e_1,c)}{\sum_u \gamma_{11}(u)P(u|c)} / \frac{\sum_u \gamma_{01}(u)P(u|g_0,e_1,c)}{\sum_u \gamma_{01}(u)P(u|c)}}{\frac{\sum_u \gamma_{10}(u)P(u|g_1,e_0,c)}{\sum_u \gamma_{10}(u)P(u|c)} / \frac{\sum_u \gamma_{00}(u)P(u|g_0,e_0,c)}{\sum_u \gamma_{00}(u)P(u|c)}}.$$

Let $\tau_{ij} = \mathbb{E}[Y|g_i, e_j, c, U_1 = 1, U_2 = 1]/\mathbb{E}[Y|g_i, e_j, c, U_1 = 1, U_2 = 0]$ so that $\gamma_{ij}(1, 1) = \tau_{ij} \times \gamma_{ij}(1, 0)$. We then have that

$$\begin{aligned}
&\sum\nolimits_u \gamma_{ij}(u)P(u|g_i, e_j, c) / \left\{\sum\nolimits_u \gamma_{ij}(u)P(u|c)\right\} \\
&= [\gamma_{ij}(1,1)P(U = (1,1)|g_i, e_j, c) + \gamma_{ij}(1,0)P(U = (1,0)|g_i, e_j, c) \\
&\quad + \gamma_{ij}(0,1)P(U = (0,1)|g_i, e_j, c) + P(U = (0,0)|g_i, e_j, c)] \,/ \\
&\quad [\gamma_{ij}(1,1)P(U = (1,1)|c) + \gamma_{ij}(1,0)P(U = (1,0)|c) \\
&\quad + \gamma_{ij}(0,1)P(U = (0,1)|c) + P(U = (0,0)|c)] \\
&= [\gamma_{ij}(1,1)P(U_1 = 1|g_i, c)P(U_2 = 1|e_j, c) + \gamma_{ij}(1,0)P(U_1 = 1|g_i, c)P(U_2 = 0|e_j, c) \\
&\quad + \gamma_{ij}(0,1)P(U_1 = 0|g_i, c)P(U_2 = 1|e_j, c) + P(U_1 = 0|g_i, c)P(U_2 = 0|e_j, c)] / \\
&\quad [\gamma_{ij}(1,1)P(U_1 = 1|c)P(U_2 = 1|c) + \gamma_{ij}(1,0)P(U_1 = 1|c)P(U_2 = 0|c) \\
&\quad + \gamma_{ij}(0,1)P(U_1 = 0|c)P(U_2 = 1|c) + P(U_1 = 0|c)P(U_2 = 0|c)]
\end{aligned}$$

Using this in Theorem 2 for B_{mult} for $i = 0, 1$ and $j = 0, 1$ and noting that because G and U_2 do not interact on the multiplicative scale, $\tau_{1j} = \tau_{0j}$ and $\gamma_{1j}(0, 1) = \gamma_{0j}(0, 1)$, G and U_2 do not interact on the multiplicative scale, $\gamma_{i1}(1, 0) = \gamma_{i0}(1, 0)$, and U_1 and U_2 do not interact on the multiplicative scale, we have $\tau_{1j} = \gamma_{1j}(0, 1)$ and thus we then have that B_{mult}.

This completes the proof. □

Proof of Theorem 3
Using the result in the proof of Theorem 1, we have that

$$\begin{aligned}
\frac{\mathbb{E}[Y_{g_i e_j}|c]}{\mathbb{E}[Y_{g_0 e_0}|c]} &= \frac{\mathbb{E}[Y|g_i, e_j, c]}{\mathbb{E}[Y|g_0, e_0, c]} \Big/ \left[\frac{\sum_u \gamma_{ij}(u)P(u|g_i, e_j, c)}{\sum_u \gamma_{ij}(u)P(u|c)} \Big/ \frac{\sum_u \gamma_{00}(u)P(u|g_0, e_0, c)}{\sum_u \gamma_{00}(u)P(u|c)}\right] \\
&= \frac{\frac{\mathbb{E}[Y|g_i,e_j,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_{ij}(u)P(u|g_1,e_1,c)}{\sum_u \gamma_{ij}(u)P(u|c)}} \times \frac{\sum_u \gamma_{00}(u)P(u|g_0, e_1, c)}{\sum_u \gamma_{00}(u)P(u|c)}.
\end{aligned}$$

Thus, RERI_c

$$\begin{aligned}
&= \frac{\mathbb{E}[Y_{g_1 e_1}|c]}{\mathbb{E}[Y_{g_0 e_0}|c]} - \frac{\mathbb{E}[Y_{g_1 e_0}|c]}{\mathbb{E}[Y_{g_0 e_0}|c]} - \frac{\mathbb{E}[Y_{g_0 e_1}|c]}{\mathbb{E}[Y_{g_0 e_0}|c]} + 1 \\
&= \frac{\sum_u \gamma_{00}(u)P(u|g_0, e_0, c)}{\sum_u \gamma_{00}(u)P(u|c)} \left[\frac{\frac{\mathbb{E}[Y|g_1,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_{11}(u)P(u|g_1,e_1,c)}{\sum_u \gamma_{11}(u)P(u|c)}} - \frac{\frac{\mathbb{E}[Y|g_1,e_0,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_{10}(u)P(u|g_1,e_0,c)}{\sum_u \gamma_{10}(u)P(u|c)}} - \frac{\frac{\mathbb{E}[Y|g_0,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_{01}(u)P(u|g_0,e_1,c)}{\sum_u \gamma_{01}(u)P(u|c)}}\right] + 1.
\end{aligned}$$

□

Proof of Corollary 3B
If U does not interact with G on the multiplicative scale, then $\gamma_{10}(u) = \gamma_{00}(u) = \gamma_0(u)$ and $\gamma_{11}(u) = \gamma_{01}(u) = \gamma_1(u)$, and if we have $G \times E$ independence, then by Theorem 3, RERI_c

$$\begin{aligned}
&= \frac{\sum_u \gamma_0(u)P(u|e_0, c)}{\sum_u \gamma_0(u)P(u|c)} \left[\frac{\frac{\mathbb{E}[Y|g_1,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_1(u)P(u|e_1,c)}{\sum_u \gamma_1(u)P(u|c)}} - \frac{\frac{\mathbb{E}[Y|g_1,e_0,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_0(u)P(u|e_0,c)}{\sum_u \gamma_0(u)P(u|c)}} - \frac{\frac{\mathbb{E}[Y|g_0,e_1,c]}{\mathbb{E}[Y|g_0,e_0,c]}}{\frac{\sum_u \gamma_1(u)P(u|e_1,c)}{\sum_u \gamma_1(u)P(u|c)}}\right] + 1 \\
&= \frac{1}{\kappa}\frac{\mathbb{E}[Y|g_1, e_1, c]}{\mathbb{E}[Y|g_0, e_0, c]} - \frac{\mathbb{E}[Y|g_1, e_0, c]}{\mathbb{E}[Y|g_0, e_0, c]} - \frac{1}{\kappa}\frac{\mathbb{E}[Y|g_0, e_1, c]}{\mathbb{E}[Y|g_0, e_0, c]} + 1,
\end{aligned}$$

Copyright © 2011 John Wiley & Sons, Ltd.

where

$$\kappa = \frac{\sum_u \gamma_1(u)P(u|e_1,c)}{\sum_u \gamma_1(u)P(u|c)} \Big/ \frac{\sum_u \gamma_0(u)P(u|e_0,c)}{\sum_u \gamma_0(u)P(u|c)}$$
$$= \frac{1+(\gamma_1-1)P(U=1|e_1,c)}{1+(\gamma_1-1)P(U=1|c)} \Big/ \frac{1+(\gamma_0-1)P(U=1|e_0,c)}{1+(\gamma_0-1)P(U=1|c)}.$$

□

Acknowledgements

The authors thank two anonymous referees for their helpful comments and the editors of the special issue, Enrique Schistermann and Paul Albert, for facilitating this research. The Long-Range Research Initiative of the American Chemistry Council, the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institutes of Health, and the National Institute of Environmental Health Sciences partially supported this research.

References

1. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder LL. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* 1959; **22**:173–203.
2. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society Series B* 1983; **45**:212–218.
3. Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology* 1990; **1**:239–246.
4. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998; **54**:948–963.
5. Robins JM, Scharfstein D, Rotnitzky A. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models for Epidemiology, the Environment, and Clinical Trials*, Halloran E, Berry D (eds). Springer-Verlag: New York, 2000; 1–95.
6. Rosenbaum PR. *Observational Studies*, 2nd edn. Springer-Verlag: New York, 2002.
7. VanderWeele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments and confounders. *Epidemiology* 2011; **22**:42–52.
8. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974; **66**:688–701.
9. Neyman J. Sur les applications de la thar des probabilities aux experiences Agaricales: essay des principle. (1923). Excerpts reprinted in English (D. Dabrowska and T. Speed, Trans.) *Statistical Science* 1990; **5**:463–472.
10. Blot WJ, Day NE. Synergism and interaction: are they equivalent? *American Journal of Epidemiology* 1979; **110**:99–100.
11. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *American Journal of Epidemiology* 1980; **112**:467–470.
12. Saracci R. Interaction and synergism. *American Journal of Epidemiology* 1980; **112**:465–466.
13. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, 3rd edn. Lippincott Williams & Wilkins: Philadelphia, 2008.
14. VanderWeele TJ, Robins JM. The identification of synergism in the suffcient-component cause framework. *Epidemiology* 2007; **18**:329–339.
15. VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology* 2009; **20**:6–13.
16. VanderWeele TJ. Empirical tests for compositional epistasis. *Nature Reviews Genetics* 2010; **11**:166.
17. Rothman KJ. *Modern Epidemiology*, 1st edn. Little, Brown and Company: Boston, MA, 1986.
18. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology* 1992; **3**:452–456.
19. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. *Epidemiology* 1996; **7**:286–290.
20. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology* 2009; **20**:863–871.
21. VanderWeele TJ, Knol MJ. The interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Annals of Internal Medicine* 2011; **154**:680–683.
22. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! *American Journal of Epidemiology* 1996; **144**:207–213.
23. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine* 1994; **13**:153–162.
24. Bennett WP, Alavanja MCR, Blomeke B, Vähäkangas KH, Castrén K, Welsh JA, Bowman ED, Khan MA, Flieder DB, Harris CC. Environmental tobacco smoke, genetic susceptibility, and risk of lung cancer in never-smoking women. *Journal of the National Cancer Institute* 1999; **91**:2009-2014.
25. Ahsan H, Chen Y, Parvez F, Zablotska L, Argos M, Hussain I, Momotaj H, Levy D, Cheng Z, Slavkovich V, van Geen A, Howe GR, Graziano JH. Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the Health Effects of Arsenic Longitudinal Study. *American Journal of Epidemiology* 2006; **163**:1138–1148.
26. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. *Statistics in Medicine* 1983; **2**:243–251.

Copyright © 2011 John Wiley & Sons, Ltd.